

***TIPO, BRAZIL'S 'LIKE': SYNCHRONIC FUNCTIONAL AND
PHONETIC ANALYSES OF NOMINAL, GRAMMATICAL, AND
DISCOURSE FUNCTIONS***

A Dissertation Submitted to the
College of Graduate and Postdoctoral Studies
In Partial Fulfillment of the Requirements
For the Degree of Doctor of Philosophy
In the Department of Linguistics
University of Saskatchewan
Saskatoon

By

CHRISTIANI PINHEIRO THOMPSON

© Copyright Christiani P. Thompson, August 2021. All rights reserved.
Unless otherwise noted, copyright of the material in this thesis belongs to the author

Permission to Use

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Linguistics
University of Saskatchewan
9 Campus Drive
Saskatoon, Saskatchewan S7N 5A5 Canada

Dean
College of Graduate and Postdoctoral Studies
University of Saskatchewan
116 Thorvaldson Building, 110 Science Place
Saskatoon, Saskatchewan S7N 5C9 Canada

Abstract

Previous research in Brazilian Portuguese has indicated that the noun *tipo* 'type,' 'kind' is undergoing grammaticalization (Bittencourt, 1999; Lima-Hernandes, 2005). Review of the literature, however, reveals a limited number of studies that provide an account of its current state in conversational speech. Moreover, research on the grammaticalization of *tipo* has been mostly limited to the examination of its multifunctionality (Bittencourt, 1999; Laurentino, 2016; Lima-Hernandes, 2005), resulting in a gap as to how the processes of grammaticalization may be reflected on its use and production.

Using data from the *Projeto Sociolinguístico Contemporâneo Brasileiro* corpus (Thompson & Onosson, 2016), comprised of sociolinguistic interviews conducted with teenage public-school students in Rio de Janeiro, this dissertation presents the findings of a study that examined the current state of *tipo* in conversational discourse. An innovative multimethodological approach was employed aiming to address *tipo*'s functional diversity. Distributional, functional, acoustic, and perceptual investigations were conducted with the goal to gain insight into some of the processes *tipo* is undergoing as it sheds its nominal properties and acquires new grammatical, discourse, and pragmatic functions.

Results reveal a functional expansion of *tipo*, which was found to be performing roles such as a preposition, a conjunction, and a discourse marker among others. More notably, results from a subsequent acoustic analysis reveal consistent differences in pronunciation between nominal and non-nominal forms, suggesting that as *tipo* expands to perform new functions, speakers are encoding such changes at the segment level. A discrimination task conducted with 98 teenage students also confirmed that speakers are able to discriminate nominal from non-nominal forms, suggesting that other processes beyond durational differences may be playing a role in the grammaticalization of that noun.

Acknowledgements

There are many people that have inspired me and supported me along the way who I wish to thank. I know I will not remember everyone, but know that I am extremely grateful to every single one of you who have helped make this degree possible.

First and foremost, I want to thank my family who has been absolutely incredible throughout this degree. Michael, Anthony, and Spencer, thank you for handling everything as I was away and for making it possible for me to pursue my dreams. Without your support, I would have never started this program. Aos meus pais Linda and Elias Thompson e irmãos Cláudia e Luciano, muito obrigada do fundo do meu coração. Seu amor, apoio, e paixão por estudos e conhecimento foram e ainda são minha maior fonte de inspiração.

The Department of Linguistics at the University of Saskatchewan offered me one of the most positive educational and professional experiences I have ever had. I want to thank my advisor, Dr. Jesse Stewart, for believing in me and in my potential when I said that I wanted him to be my supervisor. My admiration for his research, his ongoing curiosity about things, and the knowledge he shared with me made pursuing this degree a truly amazing experience. From U of S, I would also like to express my sincere gratitude to Dr. Veronika Makarova, who has been so supportive since the beginning of this journey and who, together with Dr. Bettina Spreng, offered me incredible opportunities to work with the department. Dr. Makarova and Dr. Spreng, thank you for the encouragement and advice you offered me the past few years. They have meant a lot to me and have shaped me both personally and professionally. Dr. Lovick, thank you for all your words of encouragement and your support. Brianna Groot, your knowledge about EPrime and discrimination tasks made the last part of this study come to life. Thank you for the hours you spent with me and for your patience as I asked you all the million questions I had. Finally, I want to thank the librarians at Murray Library, in special Lisa Carpenter, who, due to my daily trips to the library, eventually memorized my name, number, etc. Thank you all for helping me access the material I needed so many times. You are awesome.

Finishing this degree means I had the support of many to get me started. From the University of Victoria, I would like to express my sincere gratitude to Dr. Alex D'Arcy, who was the first to believe that the PSCB project was interesting enough to be pursued. Without her support, this would have probably been a very different degree. I also want to thank Dr. Dave McKercher, Dr. Martha McGinnis-Archibald, Dr. Sonya Bird, and Jenny Jessa, who believed in me and my dream of pursuing a doctoral program in

linguistics. I am very grateful for all you have done for me throughout the years. From Brazil, I want to thank Dr. Maria Marta Pereira Scherre, who was the first to listen to the initial idea for my doctoral study and who not only offered me uncountable resources but also mentored me throughout the way.

On a personal note, there are many I have to thank. Anne Gilbody, the one who said back in 2012 that I liked studying too much and should go back to university - you started it all! Thank you for believing in me and for being there for me along the way. I also want to express my gratitude for the support I have received from the Hoenisch family, my dear friends, who have been cheering me up since I became a Canadian citizen all those years back, and to Susan Simpson, the best neighbour one could ever have. Sky Onosson, I don't even know where to start - thank you for absolutely everything you have done and still do for me (including listening to me all the times I was frustrated while you patiently taught me how to do things in R). Nick Bartzis, thank you for doing what you do best - being there for me whether things are good or bad and being bossy when I was being a bit (ok, maybe more than a bit) difficult. Patricia Cathers, Pam Rubidge, and Robynne Mills, thank you for always saying the kindest words and for your amazing friendship.

Ah, and I cannot forget my USask family. Sandra Rickard, Ian Moy, Marie-Eve Presber, Melissa Je, Abdullah Olewi, Chrysa Morfi, Amanda Prudente de Moraes, Tamer Marzook, Adam McInnes, Elie Abillamaa, and Sam Rzd - thank you for the days of laughter, weekends of studying, late night dinners, tears of frustration as well as for the hugs, music, movie sessions, leprechauns, Oliver, and D&D. Gabriela Prado, thank you for being super awesome and for all you did for me when I got to Saskatoon. I know I am forgetting some (I apologize for that), but I have not forgotten the impact you have made in my life. Thank you all for making Saskatoon feel like home.

Finally, I want to express my gratitude to the committee members who have offered so many insights and who were so inspirational during my degree and the defense, in special to Dr. Carvalho. Your, questions, insights, and suggestions were invaluable. I sincerely appreciate all the knowledge you shared with me.

Thank you all for making this the best and most fulfilling educational experience ever.

The study detailed in this dissertation was supported by funding from the Social Sciences and Humanities Research Council of Canada (Joseph-Armand Bombardier CGS Doctoral Award, number 767-2016-1395).

Dedication

To my parents, Linda and Elias, and my sons, Anthony and Spencer. You are my inspiration.

Table of Contents

Permission to Use.....	i
Abstract.....	ii
Acknowledgements	iii
Dedication	v
Table of Contents	vi
Introduction	1
Chapter 1 The Portuguese Language	5
1.1 Background	5
1.2 The Origin of Brazilian Portuguese	7
1.3 A Description of the Carioca Dialect of Rio de Janeiro.....	13
1.4 Linguistic Variation and Social Stratification in Rio de Janeiro	21
1.5 Linguistic Innovation and Language Change.....	23
1.6 Projeto Sociolinguístico Contemporâneo Brasileiro.....	29
1.6.1 Data Collection - Methods	32
1.6.2 Transcriptions.....	34
Chapter 2 Linguistic Innovation	36
2.1 Preliminary Analysis	36
2.1.1 <i>Aí</i>	38
2.1.2 <i>Então</i>	40
2.1.3 <i>Sei lá</i>	43
2.2 Summary.....	47
Chapter 3 <i>Tipo</i>	48

3.1	<i>Tipo</i>	48
3.2	Grammaticalization.....	60
3.3	Frequency Analysis.....	66
3.4	Methods	67
3.4.1	Participants.....	67
3.4.2	Results.....	68
3.5	Patterns of Use of <i>Tipo</i>	75
3.5.1	Results.....	78
3.6	Functional Analysis.....	80
3.6.1	Functional Analysis - Results	82
3.6.1.1	Noun.....	83
3.6.1.2	Preposition.....	84
3.6.1.3	Conjunction.	85
3.6.1.4	Adverb.....	86
3.6.1.5	Sentence-final adverb.....	86
3.6.1.6	Quotative complementizer.....	87
3.6.1.7	Discourse Marker.....	93
3.6.1.8	Discourse Particle.	104
3.7	Functions by Gender.....	107
3.8	Additional Findings.....	108
3.8.1	Co-variants.....	108
3.8.2	<i>Tipo</i> and Subject Doubling	111
3.9	Summary.....	114
Chapter 4	Production.....	116
4.1	Phonological reduction/erosion	117
4.2	Participants.....	123
4.3	Data Collection	123

4.4	Data.....	124
4.5	Exploratory Analysis	126
4.6	Results.....	128
4.6.1	Full and Elided Vowels	128
4.6.2	Affricate	137
4.6.3	Stop.....	139
4.6.4	Vowels.....	141
4.7	Summary.....	151
Chapter 5 Perception.....		154
5.1	Sound perception and discrimination.....	154
5.2	Methods	161
5.2.1	Data Collection	161
5.2.2	Participants.....	162
5.2.3	Equipment	162
5.2.4	Stimuli.....	163
5.2.5	Experiment Design.....	166
5.2.6	Participants.....	169
5.2.7	Experiment	170
5.3	Results.....	173
5.4	Summary.....	181
Chapter 6 Discussion		183
6.1	The Current State of Tipo in the Carioca Dialect.....	184
6.2	Limitations.....	197
6.3	Future Research	198
References		201

Figures

Figure 1.1. <i>The Emperor and his ‘conversation guide.’</i>	11
Figure 1.2. <i>Map of Brazil divided by states with Brasília (the capital) and the state of Rio de Janeiro highlighted.</i>	14
Figure 1.3. <i>The vowels of Brazilian Portuguese.</i>	16
Figure 1.4. <i>Você vai sair hoje? (‘Are you going out today?’)</i>	18
Figure 1.5. <i>Tu vai sair hoje? (‘Are you going out today?’)</i>	19
Figure 1.6. <i>Map of Rio de Janeiro showing Ilha do Governador (in red).</i>	31
Figure 3.1. <i>Frequency of nominal and non-nominal functions of tipo found in the data from the PSCB corpus.</i>	69
Figure 3.2. <i>Frequency of tokens of nominal and non-nominal functions of tipo by corpus.</i>	72
Figure 3.3. <i>Comparison of the percentage of nominal vs. non-nominal functions of tipo per corpus (1970 - 2018).</i>	72
Figure 3.4. <i>Number of nominal vs. non-nominal functions of tipo produced by teenagers in the PEUL and PSCB corpora.</i>	74
Figure 3.5. <i>Percentage of nominal vs. non-nominal functions of tipo by teenagers identified in the PEUL and PSCB corpora.</i>	74
Figure 3.6. <i>Frequency of grammatical categories following tipo.</i>	79
Figure 3.7. <i>Distribution of tipo in reported speech constructions.</i>	91
Figure 3.8. <i>‘The terrible acute tiponitis’.</i>	105

Figure 3.9. <i>'Acute tiptonitis'</i>	106
Figure 3.10. <i>'So?' 'What's up?' 'Like!' 'Cool!' (De Souza, 2004)</i>	106
Figure 4.1. <i>The vowel system of Brazilian Portuguese in stressed syllables and word-final position (vowels are circled).</i>	119
Figure 4.2. <i>Example of some general notes indicating possible elements of interest.</i>	125
Figure 4.3. <i>Elided /i/ and /o/ vowels in tipo.</i>	129
Figure 4.4. <i>Visualization of predictions from the /i/ generalized mixed effects model.</i>	131
Figure 4.5. <i>Visualization of predictions from the /o/ generalized mixed effects model.</i>	133
Figure 4.6. <i>Boxplot of token duration by function and gender.</i>	135
Figure 4.7. <i>Overall duration of the voiceless affricate [tʃ] by function by gender.</i>	138
Figure 4.8. <i>Overall duration of the voiceless stop /p/ by function and gender.</i>	140
Figure 4.9. <i>Overall duration of /i/ and /o/ (which is realized as [u] in BP) by function.</i>	142
Figure 4.10. <i>Duration of the /i/ vowel by gender and function.</i>	143
Figure 4.11. <i>Duration of the /o/ vowel by gender and function.</i>	143
Figure 4.12. <i>Plotting of F1 and F2 measurements (in Hz) of /i/ and /o/ vowels by function.</i>	147
Figure 4.13. <i>Plotting of F1 and F2 measurements of the /i/ vowel (left) and /o/ vowel (produced as [u]) (right) by function.</i>	148
Figure 5.1. <i>Textgrid showing the nominal token of tipo used in the experiment.</i>	163
Figure 5.2. <i>Textgrid showing a token of tipo from the PSCB corpus with an elided /i/ vowel.</i>	165

Figure 5.3. <i>Textgrid showing a token of tipo from the PSCB corpus with an elided /o/ vowel.</i>	165
Figure 5.4. <i>Textgrid showing a token of tipo from the PSCB corpus with elided /i/ and /o/ vowels.</i>	165
Figure 5.5. <i>Layout of the different screens that comprised the experiment.</i>	172
Figure 5.6. <i>Plotting of the average reaction time (ms) for accurate (TRUE) and inaccurate (FALSE) responses.</i>	178
Figure 6.1. <i>Design of the current study.</i>	184
Figure 6.2. <i>Tweet containing tipoassim from February 22, 2021.</i>	188
Figure 6.3. <i>YouTube channel Tipassim Né.</i>	188
Figure 6.4. <i>Tweet containing tipassim from February 18, 2021.</i>	188
Figure 6.5. <i>Tipassim in tweets from February 21 and February 26, 2021.</i>	193
Figure 6.6. <i>Hashtag #tipassim showing the elision of the /o/ vowel and coalescence of forms.</i>	193
Figure 6.7. <i>The functions of tipo in the PSCB corpus including transcriptions, definitions, and examples.</i>	197

Tables

Table 1.1. <i>Phoneme inventory of Brazilian Portuguese with common allophones in brackets. ..</i>	16
Table 1.2. <i>Students who participated in the PSCB project by age, gender, and school.....</i>	30
Table 3.1. <i>Functions of tipo identified in previous studies on BP.</i>	57
Table 4.1. <i>Elided and non-elided /i/ and /o/ vowels by function.....</i>	130
Table 4.2. <i>Generalized linear mixed effects model results for the vowel /i/ presence in tipo.</i>	132
Table 4.3. <i>General linear mixed effects model results for the vowel /o/ presence in tipo.....</i>	134
Table 4.4. <i>Total duration of tipo by function and gender.....</i>	136
Table 4.5. <i>Duration of the /i/ vowel by function and gender.....</i>	145
Table 4.6. <i>Duration of the /o/ vowel by function and gender.....</i>	146
Table 4.7. <i>F1 measurements of /o/ by function and gender.....</i>	149
Table 4.8. <i>F2 measurements of /o/ by function.....</i>	150
Table 5.1. <i>Composition of stimulus pairings used in the discrimination task including distractors.</i>	168
Table 5.2. <i>Total number of accurate and inaccurate responses.</i>	174
Table 5.3. <i>Frequency of accuracy by pair type.....</i>	175
Table 5.4. <i>Glmer model results of accuracy in discriminating the paired stimuli.</i>	177
Table 5.5. <i>Reaction time (in ms) across different paired stimuli.....</i>	179
Table 6.1. <i>Summary of the results.....</i>	185

Introduction

“Custom alone can determine what is right and wrong, not the dictum of grammarians, however eminent.” (Sayce, 1900, p. xxiv)

Review of the literature reveals that extensive research has been conducted on various aspects of Brazilian Portuguese including syntax, morphology, bilingualism, and language contact (Carvalho, 2014; Carvalho & Lucchesi, 2020; Munn & Schmitt, 2002; Pilati, Naves, & Salles, 2017; Thomas, 1969). This dissertation aims to contribute to this literature by describing an exploratory study aimed to investigate empirical synchronic evidence of language innovation and change in the use of linguistic elements in Contemporary Brazilian Portuguese.

Initial findings from an analysis of data retrieved from the *Projeto Sociolingüístico Contemporâneo Brasileiro* (PSCB) corpus (Thompson & Onosson, 2016) suggested that *tipo* (masculine noun meaning 'type,' 'kind') was grammaticalizing and performing functions beyond those of a noun in the *Carioca* dialect of Rio de Janeiro. Review of the literature shows that the possible acquisition of grammatical functions by *tipo*, first postulated by Bittencourt (1999), still remains largely underinvestigated, with studies often limited to the examination of its multifunctionality (Bittencourt, 1999; Laurentino, 2016; Lima-Hernandes, 2005). Using data from the PSCB corpus, the exploratory study detailed in this dissertation examined the current state of grammaticalization of *tipo* in the *Carioca* dialect of Brazilian Portuguese and explored

some of the processes it is undergoing as it expands from its original nominal usage to grammatical and discourse-pragmatic functions.

To address its possible ongoing grammaticalization, a multilayered methodological approach was employed aiming to address its functional diversity in the *Carioca* dialect of Brazilian Portuguese. More specifically, this study aimed to address the following questions:

1. What is the current state of *tipo* in the *Carioca* dialect of Brazilian Portuguese?
2. What are the patterns of use of *tipo*? Where is it often found in utterances?
3. What functions is *tipo* currently performing in the speech of teenagers?
4. Are different functions of *tipo* acoustically distinct? If so:
 - a. Are these durational differences?
 - b. Are these quality differences at the segmental level?
5. If acoustic differences are identified, are speakers able to discriminate nominal from non-nominal functions of *tipo*?

To answer these questions, the following sequence of analyses was conducted:

1. an initial distributional analysis, which aimed to provide details about the frequency of nominal and non-nominal functions of *tipo* in the speech of *Carioca* teenagers who, at the time of data collection, attended one of the municipal public schools in Rio de Janeiro;

2. a functional analysis that describes the current functions *tipo* is performing in the oral discourse of that group;
3. an acoustic analysis, which focused on the examination of the segments of *tipo* as it performs different grammatical and discourse-pragmatic functions;
4. a perceptual investigation, which investigated speakers' ability to discriminate different acoustic realizations of *tipo* (more specifically, of nominal vs. non-nominal tokens).

This dissertation aims to contribute to the literature by offering insights gained from this exploratory study, designed to investigate empirical synchronic evidence of language innovation and change in the use of *tipo* in Contemporary Brazilian Portuguese. By examining frequency, function, production, and perception, this study offers a valuable insight into important aspects of the grammaticalization of *tipo*. The value of the unique research design employed in this study goes beyond the identification of how speakers are using *tipo* as it also examines the structure of the word itself and how it is produced, perceived, and interpreted through objective and quantifiable measurements. Additionally, this systematic and accountable examination paints a more nuanced picture of what some of the processes involved in the grammaticalization of *tipo* entail.

This dissertation is organized as follows: Chapter 1 describes the Portuguese language and the genesis of Brazilian Portuguese. Chapter 2 presents a discussion on linguistic innovation and preliminary findings of an analysis of innovation in the *Projeto Sociolinguístico Contemporâneo Brasileiro* corpus (Thompson and Onosson, 2016). Chapter 3 discusses the lexical item *tipo* and provides the description and results of frequency, distributional, and functional analyses. An acoustic analysis of tokens of *tipo* as it performs different functions is described in Chapter 4 while Chapter 5 presents the description and results of a perception experiment. Chapter 6 summarizes the results of this study and draws conclusions¹.

¹ This study was supported in part by funding from the Social Sciences and Humanities Research Council.

Chapter 1 The Portuguese Language

1.1 Background

Portuguese, a Romance language (i.e., of Latin origin), is the sixth most spoken language in the world, with over 244 million speakers (Tinsley & Board, 2014). It is the national language in nine countries around the world: Cape Verde, Mozambique, Guinea-Bissau, East Timor, Equatorial Guinea, São Tomé and Príncipe, Angola, Portugal, and Brazil. Research has shown an increase in its relevance worldwide (both in number of translations and number of speakers) (Sanchez, 2014). Brazil, the fifth largest country in the world, holds the most speakers of Portuguese (204 million) (IBGE, 2018). Linguistic research has shown that the variety of Portuguese spoken in Brazil differs from European Portuguese (de Mello, 1996; Galves, Moraes, & Ribeiro, 2005; Scherre & Naro, 1993; Varejão, 2009). One of the differences found in vernacular Brazilian Portuguese (BP)² is the loss of flectional morphology, such as plural agreement (1) and variation in person-verb agreement (2).

(1)	<i>Eles</i>	<i>são</i>	<i>estudante</i>	vs.	<i>Eles</i>	<i>são</i>	<i>estudante-s</i>
	3P	be.3P	student.S		3P	be.3P	student-PL
	'They	are	student'		'They	are	students'

² In contrast with standard Brazilian Portuguese, which follows the rules of prescriptive grammar (Parkvall & Álvarez López, 2003).

(2)	<i>Nós</i>	<i>somos</i>	<i>estudante-s</i>	vs.	<i>Nós</i>	<i>é</i>	<i>estudante</i>
	1P	be.1P	student-PL		1P	be.3S	student.S
	'We	are	students'		'We	is	student'

Differences can also be found in the phonology of the two variants. For instance, in BP phonology, when the consonants /t/ and /d/ are followed by a high front vowel (/i/), they undergo affrication to [tʃ] and [dʒ], respectively (de Brito, 1986), as in the case of the word *tigre* 'tiger,' which is realized as /'tʃigri/ in BP and not /'tigri/ as commonly heard in European Portuguese (henceforth EP), or the preposition *de* 'of', realized as [dʒi] in BP rather than [di]. Another phonological process of BP not present in EP is diphthongization³, the process of a monophthong, a vowel that has a single identifying vowel target value, becoming a diphthong, in which two targets are found within a syllable nucleus (Clark, Yallop, & Fletcher, 2008). More specifically, this process is often found in stressed vowels that are followed by a voiceless alveolar fricative (/s/) or voiceless postalveolar fricative /ʃ/, such as in *três* 'three' (/treʃ/ → /treiʃ/) (Cagliari, 2002; Vasilévski, 2012a). In addition to the aforementioned processes, previous research has shown that in EP, unstressed vowels may be deleted, resulting in consonant clusters (e.g.: *fonologia* /funulu'ʒiɐ/ → /fnulu'ʒiɐ/) 'phonology' (Mateus & d'Andrade, 2000) while in BP epenthesis (the addition of a sound between two consonants) is common

³ According to Leiria (2000), "diphthongization is strongly correlated to geographic dialectal variation" in BP (p. 134).

process used to simplify syllables by breaking consonant clusters, such as in *pneu* 'tire' /pneu/ → /pi'neu/ or in *advogado* 'lawyer' /advo'gadu/ → /adʒivo'gadu/ (Câmara, 1972; Frota & Vigário, 2001).

The undeniable differences between the two dialects have given rise to questions regarding the evolution of BP and the development of the idiosyncratic features that sets it apart from the other dialects of Portuguese spoken around the world.

1.2 The Origin of Brazilian Portuguese

In the 16th century, during the expansion of the Portuguese empire, the Portuguese language arrived in Brazil. In the land, Portuguese explorers found indigenous groups who spoke over 1,300 different languages (Gaspar, 2012). This linguistic diversity was expanded with the arrival of between four and five million slaves from Africa, who brought with them their own languages, such as Yoruba, Ewe, and Fon, and a number of Bantu languages, along with their cultural practices (Rodrigues, 2010).

The importance of the African languages was soon recognized, with books such as *A arte da língua de Angola* ('The art of the language of Angola') (Dias, 1697) and *A obra nova de língua geral de mina* ('The new work on the general mine language'), published in Portugal, which focused on an Ewe-Fon pidgin spoken in the gold mines in the 18th century (Peixoto, 1944), or more noticeably, the acknowledgment that Portuguese speakers were borrowing words of African origin and their subsequent inclusion in the

Dicionário da Língua Portuguesa (Morais e Silva, 1789). A review of the literature shows that researchers have attributed several features of the BP vernacular⁴ to the influence of African languages spoken in Brazil, such as pronunciation (as in the case of the deletion of /d/ in the gerund form -ndo → -no as in *fazendo* /fa'zẽdu/ → *fazeno* [fa'zẽnu]) 'doing' (Cristino, 2004) and apheresis (the loss of one or more sounds in the beginning of a word, as in *estou* → *tô* '[I] am' or *você* → *cê* 'you') (Mendonça, 2012) (see Avelar & Galves, 2014 and Mendonça, 2012 for more on the topic).

Some authors have recently posited that BP was also influenced by the languages spoken by the indigenous groups who lived in the region upon the arrival of the Portuguese (Lobato, 2006; Mattos e Silva, 1993). Some of the arguments regard the possible introduction of suffixes in BP, such as *-rana*, *-guara*, and *-oara* (Robl, 1985) and the undeniable lexical contribution Tupi has made to BP (Mattos e Silva, 1993). Albeit limited to the lexical repertoire of indigenous languages, Góis and Martins (2019) claim that this contribution has had a direct impact on how the linguistic identity of the country was shaped. To support that argument, Góis and Martins (2019) discuss three clear cases in BP that, according to the authors, have helped shape the Brazilian

⁴ The vernacular refers to the style in which speakers minimally monitor their speech (Labov, 1991).

linguistic identity. These cases, which show the inclusion of Tupi elements, can be divided into⁵:

1. Compound nouns – in some compound nouns, at least one of the lexical items is of Tupi origin while the other(s) item(s) may be of Brazilian or foreign origin (e.g., the name of the bird *sabiá-branco* (*sabiá* from the Tupi word *sawi'a* 'sabiá bird' and *branco*, from the Germanic word *blank* 'polished, shiny');
2. Lexical items – in some instances, Portuguese bound morphemes are used with some lexical items of Tupi origin, such as in the case of the verb *empipocar* (*em-pipoc-ar*) 'break out in hives', in which the word *pipoca* (from the Tupi *pi'poka*, 'popcorn' is preceded by the Portuguese verbal prefix *em-* and followed by the Portuguese infinitival marker *-ar*);
3. Lexical items (2) – in other instances, Tupi morphemes are used with some lexical items of Portuguese origin, such as in the case of *canarana* (*cana* + *rana* 'a type of grass typical of the Amazon and the Guianas): *cana* comes from the Latin *canna* ('sugar cane') and *-rana* is a Tupi suffix meaning 'similar to'.

⁵ The examples presented below are from Góis and Martins (2019, p. 435).

Researchers have long debated about the emergence of the BP vernacular (Guy, 1989; Scherre & Naro, 2010). Arguments are hinged on the hypotheses (among others) of: creolization and the influence of West African languages, most notably Yoruba, which would have given rise to the Brazilian Portuguese vernacular⁶ (Guy, 1989); semi-creolization, a process in which a language would undergo partial restructuring of its grammar but would not undergo full creolization (Ferreira, 2009; Holm, 1991, 1992; Parkvall & López, 2003); and natural drift, which posits that the differences encountered between the European and Brazilian varieties are simply the result of a natural drift that occurred in Indo-European languages, and, more specifically, in Romance languages (Naro & Scherre, 2010). Naro and Scherre (2007) argue that elements that researchers have considered to be proof of the creole origin of BP (such as the loss of the plural morpheme -s) were, in fact, tendencies that were present in EP that developed in BP.

The changes that occurred in Brazilian Portuguese have been a topic of interest since at least the 19th century, with its idiosyncratic features being the object of interest and discussion by European speakers. Examples include discussions about the divergence between the European and Brazilian variants (de Alencastro, 1997). To some, the linguistic elements characteristic of BP were considered almost incomprehensible as exemplified in the political cartoon by Pinheiro (1872), which

⁶ In contrast to the standard BP, which would have retained more features of EP.

shows the Emperor D. Pedro II in possession of a 'Brazilian-Portuguese / Portuguese-Brazilian conversation guide' (p.2) (see Figure 1.1).



Figure 1.1. The Emperor and his 'conversation guide.'

In the image, a Brazilian boy shows the emperor a supposed conversation guide consisting of Brazilian-Portuguese/Portuguese-Brazilian translations while the other boys hold a bottle of a "popularity elixir" and another a book of "brief notes on civility."

Retrieved from Pinheiro, 1872, p.2.

Historical documents show scholars' and politicians' awareness of the growing differences between the European and the Brazilian varieties of Portuguese. One such example is that of Domingos Borges de Barros (Viscount of Pedra Branca), who made a direct request to geographer Adriano Balbi to include in his *Introduction à l'Atlas Ethnographique du globe* (Balbi, 1826) information that he had collected on the differences

between the two varieties of the language (Balbi, 1826, p. 172). In his contribution, the Viscount discusses not only how Brazilian Portuguese “breathes the gentleness of the climate and of the character of its inhabitants” (qtd. in Goldberg, 2018, p. 17) but also the lexical differences found between the two languages (Alkmim, 2012).

Another example of this interest in the BP variant and its unique features is illustrated in Vasconcelos’ (1901) *Esquisse d’une dialectologie portugaise*. The author claims that

La langue nationale du Brésil est le portugais, qui transporté dans un milieu si différent de celui de son origine, y a éprouvé beaucoup de modifications. Les écrivains brésiliens ont beaucoup discuté, au point de vue patriotique, si le portugais du Brésil est ou non un dialecte. Si j’appelle dialecte, par exemple, le portugais de Tras-os-Montes, à plus forte raison je dois donner ce nom au portugais du Brésil, ou brésilien. (p.159)

‘The national language of Brazil is Portuguese, which, transported to an environment so different from that of its origin, has experienced a great deal of patriotic discussion as to whether the Portuguese of Brazil is a dialect or not. If I call, for example, the Portuguese of Tras-os-Montes a dialect, the more reason I have to give this name to the Portuguese of Brazil, or Brazilian.’

Some of the questions raised by grammarians and anthropologists in the 19th century regarded the possibility of the ‘differences’ found in BP being the result of language contact between the European colonizers, and, mainly, African slaves (Rodrigues, 2010). One of the first to clearly postulate this possible influence was Nina Rodrigues in his work *Os africanos no Brasil* (‘The Africans in Brazil’), written between 1890 and 1905 (but only published in 1932). Nina Rodrigues’ arguments were built on

previous work, mainly that by Ribeiro (1897), who expressed what he considered to be an undeniable influence of African languages on the BP vernacular. As Ribeiro states,

Under the denomination Black Element we define all the types of changes that have taken place in the Brazilian language due to the influence of African languages by the slaves brought to Brazil. These changes are not as superficial as some researchers argue; on the contrary, they are quite profound, not only regarding the vocabulary but even the grammatical system of the language (qtd. in Bonvini, 2009, p. 15).

Researchers have yet to reach a consensus: while some postulate that African languages have had considerable influence on the BP vernacular (see Lucchesi, Baxter, & Ribeiro, 2009, and Mattos e Silva, 1993), others have categorically refuted such influence (see Naro & Scherre, 2007).

1.3 A Description of the Carioca Dialect of Rio de Janeiro

Brazil is divided into 26 states and the federal district of Brasília, its capital. The vast expansion of landmass, socioeconomic differences, and distribution of the population amidst the states and the federal district has resulted in the emergence of different dialects. One such example is the dialect that emerged in the city of Rio de Janeiro, the second largest city in Brazil (pop. 6.3 million in 2010) (IBGE, 2018) and one of the cultural centres of the country.



Figure 1.2. Map of Brazil divided by states with Brasília (the capital) and the state of Rio de Janeiro highlighted. This map is freely licensed under the Open Data Commons Open Database License (ODbL) by Pngwing.com - <https://www.pngwing.com/en/free-png-yqkhs> .

Between 1807 and 1810, Rio de Janeiro witnessed an explosion in population growth, which quadrupled the population from 60,000 to nearly 250,000 (Callou, 2009; Cavalcanti, 2004). At a time when the majority of the population consisted of African slaves, this growth, caused by the arrival of the Royal Family and members of the court to the city, led to a stronger influence of European Portuguese in Rio de Janeiro, and language contact most likely had a strong impact in the Portuguese spoken in the city. This, in addition to its geographic isolation from inland areas, resulted in innovative linguistic features that are unlike those of other dialects spoken in Brazil. Among the dialects spoken in the country, the influence and distinctiveness of the dialect spoken in

Rio de Janeiro has been and still is of interest to researchers, with the history of the city providing relevant information regarding its rise to a “prestigious dialect.

This dissertation will focus on the *Carioca* (a demonym used to refer to elements related to the city of Rio de Janeiro) dialect, which is found in and around the city of Rio de Janeiro. *Carioca* is considered a dialect of prestige as well as the standard dialect of Brazil (de Brito, 2016). *Cariocas*, individuals born and raised in the city, are described as possessing a peculiar dialect that is distinguishable from any other in the country.

Websites such as *Como Aprender o Sotaque Carioca* (‘How to learn the *Carioca* accent’), *Rio para Paulistas* (‘Rio for those who are from São Paulo’) and YouTube channels such as *Como Ter o Sotaque Carioca* (‘How to have the *Carioca* accent’), *Como ser Carioca* (‘How to be *Carioca*’), or *Sotaque Carioca: Ensinando Cariquês* (‘The *Carioca* accent: Teaching how to speak the *Carioca* dialect’) are some of the numerous examples that illustrate the importance of the dialect in Brazil.

To better situate the *Carioca* among other Brazilian dialects, it is important that some basic information about Contemporary Brazilian Portuguese is presented.

Contemporary Brazilian Portuguese has the following phonemes: seven oral vowels (see Figure 1.3); five nasal vowels /ã/, /ê/, /ĩ/, /õ/, /ũ/; and the consonants and allophones detailed in Table 1.1 (Câmara Jr., 1977, 1986; Cardoso, 2009; Seara, Nunes, & Lazzarotto-Volcão, 2011; Vasilévski, 2012a).

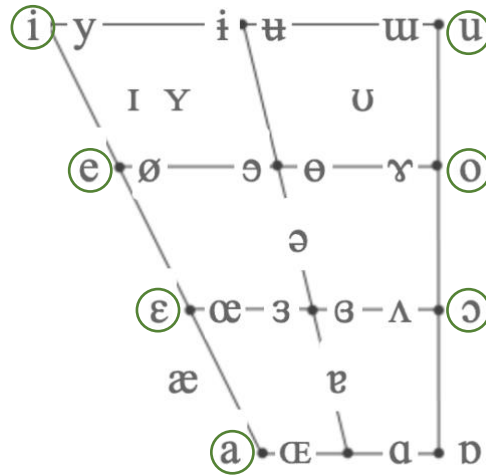


Figure 1.3. The vowels of Brazilian Portuguese.

Table 1.1. Phoneme inventory of Brazilian Portuguese with common allophones in brackets.

	labial		coronal				dorsal			radical	laryngeal
	Bilabial	Labio-dental	Dental	Alveolar	Palato-alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Nasal	m		n					ɲ			
Plosive	p b		t d					k g			
Fricative		f v		s z	ʃ ʒ			[x] [χ]	[χ] [ʁ]		[h] [ɦ]
Affricate											
Approximant				[ɹ]		[ɻ]		[ɰ]			
Tap			r								
Trill			[r]								
Lateral Fricative											
Lateral Approximant								ʎ			
Lateral Flap				l							

The *Carioca* dialect is said to be unique among Brazilian dialects in the way that some phonological processes happen in tandem to create its distinctive ‘sound’. Some of the most notable phenomena⁷ are:

1. palatalization of the syllable final -s [s] → [ʃ] (also called *chiado Carioca*) (e.g.: *gasto* /'gastu/ → /'gaʃtu/ ‘(I) spend’, *lápiz* /'lapis/ → /'lapiʃ/ ‘pencil’) (Callou, 2009; Lipski, 1973);
2. realization of post-vocalic -s as [h] or [ø], which is often stigmatized as it is commonly associated with lower social class and/or lower educational level (e.g., *os carro-ø* /uʃ'kahu/ ‘the cars’) (de Brito, 2016; de Lima, 2005);
3. raising of the unstressed final vowels [e] and [o] to [i] and [u], respectively (although this is also found in other dialects) (e.g., /'gato/ → /'gatu/ ‘cat’, /'pare/ → /'pari/ ‘(you) stop’) (Houaiss, 1958; Lipski, 1973);
4. weakening (and oftentimes suppression) of unstressed vowels (e.g., *vinte* /'vĩtʃi/ → /vĩʃ/ (Lipski, 1973);
5. raising of [e] and [o] in atonic positions (Houaiss, 1958; Lipski, 1973) (e.g., *tomar* /to'ma/ → /tu'ma/ ‘to take,’ ‘to drink’;

⁷ It is important to note that some of these phenomena are not restricted to the Carioca dialect.

6. diphthongization (addition of a glide, either [i] or [u], which is environment-dependent) (e.g., *paz* /paʃ/ → /paiʃ/ ‘peace’ and lengthening of syllables with primary stress (Major, 1985);
7. prosody (i.e., the intonation, rhythm, stress, and tone of spoken language).

Figures 1.4 and 1.5 illustrate intonational differences in production of the same question by a *Carioca* speaker and by a speaker of the *Florianopolitano* dialect of the south of Brazil. In Figure 1.4, the pitch accent (PA) is found on the penultimate syllable of the final prosodic word (PW) while in Figure 1.5 a high boundary tone (BT) is found in the final syllable of the final PW:

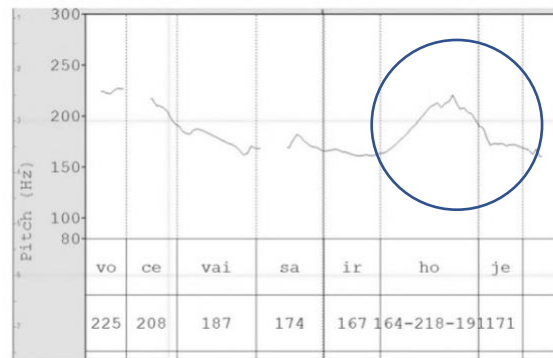


Figure 1.4. *Você vai sair hoje?* (‘Are you going out today?’)
 Female speaker from Rio de Janeiro.
 Based on Silva & Cunha, 2011.

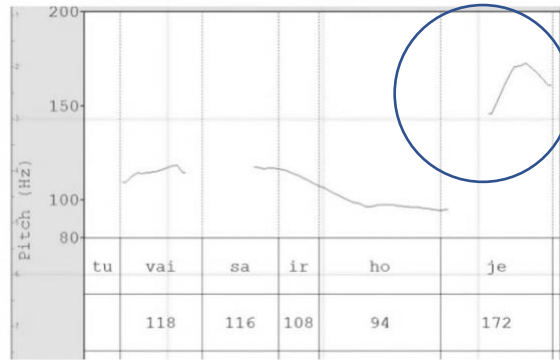


Figure 1.5. *Tu vai sair hoje?* ('Are you going out today?')
 Female speaker from Florianópolis,
 south of Brazil. Based on Silva & Cunha, 2011.

8. occasional insertion of a schwa offglide immediately after a vowel with primary stress (found in many environments but not when followed by [s]) (e.g., *alô* /alo/ → /aloə/ 'hello' (Lipski, 1973);
9. affrication of *t* and *d* when followed by [i] (realized as [tʃ] and [dʒ]) (e.g., /'tipu/ → /'tʃipu/ 'type', /di'jẽru/ → /dʒi'jẽru/ 'money' (de Brito, 2016, analyzed in detail herein);
10. realization of [r] in syllable final position as [x], [ʀ], or [h] (environment-dependent) (e.g., *poder* /'po'der/ → /po'dex/, /po'der/, or /po'deh/, 'power' (Azevedo, 1981);
11. lateral vocalization, in which [l] is realized as [u] in coda (syllable final) position (e.g.: *mel* /mɛl/ → /mɛu/) 'honey', *alto* /'altu/ → /'autu/ 'tall') (Câmara, 1986; Cristófaros-Silva & de Oliveira, 2001).

Lexical differences have also been shown to distinguish *Cariocas* from other BP speakers. For instance, Nascentes (1922) presents an introductory analysis of some of the most distinguishable features of that dialect, including a list of words that were considered dialect-specific. These neologisms/innovations would later spread to other areas of the country due to the city's role as a national cultural reference (see also de Alencastro, 1997). As Nascentes states, "everything is born in Rio first to later be exported to all other states" (qtd. in Flan, 1953, p. 38). *Cariocas* are described by Nascentes as "the greatest creators of the words [that make up] our (Brazilian) slang", such as the word *balzaqueana*, popularized after first being used in Rio to refer to a woman in her 30s (a reference to Balzac's *La femme de trente ans*) (Flan, 1953, p. 38). These innovations, he argues, are at times new terms coined by speakers, at other times words derived from existing words, but, often, just regular words that acquire new meaning due to the widespread innovative use by the population of the city (Nascentes, 1922). Nascentes (1922) also shows interest in the process of linguistic change, raising important questions about the emergence, incorporation, and dissemination of new terms in the *Carioca* dialect and their role in speakers' discourse; he describes these changes as being at times sudden and at other times slow, so slow that he postulates that an analysis of such progression might be an impossible feat.

Cross-linguistically, innovation is not a phenomenon that is uniform across dialects. Elements such as gender, age, and social class have been shown to influence

how individuals speak (Eckert, 2014; Labov, 1966, 2006). One important and often mentioned aspect of the *Carioca* dialect is the different sociolects spoken in Rio de Janeiro. Social stratification, a sad reality for which the city is well known, has been shown to not only influence how *Cariocas* speak but also how they may negatively perceive speakers who belong to social groups other than the dominant group (Roth-Gordon, 2009).

1.4 Linguistic Variation and Social Stratification in Rio de Janeiro

Linguistic variation has been shown to be correlated with several factors including speakers' age, network, and social class (Drager, 2015; Labov, 1972; Rickford, 1986). Rio de Janeiro is a city of social contrasts in which the wealthiest and the poorest commonly live side-by-side. This gap between the upper and the lower classes is reflected in speakers' sociolect. While the standard dialect is associated with the dominant class, the social exclusion and the negative stigma associated with *favelados*⁸ (those who live in the *favelas*, 'slums,' 'shanty towns') and individuals of low socioeconomic status is reflected in the noticeable prejudice shown by some towards their social dialect (Patel, 2015; Roth-Gordon, 2009).

⁸ The word *favelado* is often used pejoratively by those who do not live in the *favelas* to refer to someone who acts rude, looks poor, or has no manners.

Previous research has shown that variants used by speakers of a social dialect (or sociolect) come to index an individual's social identity (Guy, 2013; Labov, 2006).

Findings have also shown a correlation between social class membership and the use of linguistic variables in a given dialect (see Chambers & Trudgill, 1998). For instance, in the 1960s, William Labov conducted a ground-breaking study on the differential use of /r/ by individuals working in three department stores of different social statuses in New York City. His results showed a correlation between the production of /r/ and the social status of the store, indicating that the production of the phoneme was socially stratified, with individuals working at the store that caters to customers of higher socioeconomic class using the /r/ more frequently than those working at the store whose customers are mainly those of lower socioeconomic status (Labov, 1966).

The social meaning of the linguistic forms used by speakers (in conjunction with other practices) show that the language choices they make index their social identity (Labov, 2006). These linguistic forms have also been shown to result in social assumptions and judgments towards speakers (Campbell-Kibler, 2005). This view, shared by those who live in the favelas, is clearly depicted by Alexandre Lucena (qtd. in Patel, 2015), a resident of one of the *favelas* in Rio, who states that "If they don't accept us as people, how will they accept our language as valid? That is where it starts. In

Zona Sul⁹ they do not accept me because of how I speak, I use slang as a word. But this is my dialect, it can't be changed. That is exactly why they don't accept me."

In addition to the social stigma, prescriptivists have also described the dialect of the *favelas* as an example of "regression", a corruption of the standard language (Patel, 2015). It is not surprising, therefore, that this sociolect has a direct impact on individuals' lives, from education inequality to their exclusion from Rio's social fabric (Patel, 2015). Although their sociolects (i.e., social dialects) convey social information, elements that have been shown to be 'detrimental' to language are also examples of linguistic innovation and reflect the communicative needs of those speakers.

1.5 Linguistic Innovation and Language Change

Language change is a by-product of language use. Natural living languages are in a continuous process of change, and discourse pragmatics is no exception. Speakers create structures that reflect their ideas, feelings, and inner thoughts. As their communicative needs change, speakers spontaneously recruit linguistic elements that make communication possible and, if needed, will create new forms or use linguistic elements in innovative ways. Meanwhile, other forms will disappear. These changes are, at times, idiosyncratic and may not result in language change beyond the use of an

⁹ Zona Sul refers to the south area of Rio, where most of the inhabitants are from the upper class. Zona Sul includes Copacabana, Ipanema, Leblon, and Leme among other neighbourhoods.

individual speaker (Weinreich, Labov, & Herzog, 1968). At other times, these changes or innovative uses of linguistic elements become accepted into a community's speech repertoire, spreading to other speakers and becoming part of their linguistic inventory.

Historically, linguists have primarily focused on traditionally core areas such as syntax and phonology (among others) as well as on the structure of languages. One example is etymology, in which scholars investigate the origin and development of words in a language (e.g., the Portuguese word *espada* 'sword,' which has been shown to come from the Greek *spáthe* > Latin *spatha* > Modern Portuguese) (see seminal works such as Campbell, 1999; Chomsky, 1959, 2002; Ohala, 1974; Pierrehumbert, 1980; Sapir, 1949).

Research has shown that innovation can be both externally and internally motivated. In studies of language use and internally motivated innovations, one topic of great interest is that of *productivity*, defined as the ability speakers have to create new forms (Bauer, 1994). Innovation results from speakers' choice of using forms for expressive purposes and for the fulfillment of their communicative needs. Therefore, innovations are seen as idiosyncratic in nature, i.e., they would not necessarily promote language change beyond an individual (Weinreich, Labov, & Herzog, 1996). Take, for instance, the pronunciation of the word *ônibus* 'bus' /'onibuʃ/ in BP. If a speaker

produces that as **ônibos* /'onibof/¹⁰, that does not necessarily imply this variation will spread to other speakers. On the other hand, the spread of innovative forms among a community and their acceptance as new items in their linguistic repertoire would reflect the concept of external language change (Brinton & Traugott, 2005). An often-cited example of this is found in dialects of North American English in the so-called 'cot-caught' merger, in which the phonemic distinction between the back vowels /ɑ/ and /ɔ/ is lost, resulting in several homophonous words that were traditionally minimal pairs (Herold, 1990, 1997; Labov, 1994, 2001; Labov, Ash, & Boberg, 2006).

To investigate possible linguistic innovation, researchers mostly focus on informal conversational speech as speakers tend to be less concerned about what prescriptivists would call 'proper' or 'correct' language. In informal speech, researchers have also identified that speakers make more active use of new forms in systematic ways (Chambers & Trudgill, 1998). Despite their presence in all living languages, innovative forms have been perceived by prescriptivists as negative and undesirable as well as a sign of inarticulateness and degradation of the standard language (see e.g., Bittencourt, 1999; D'Arcy, 2005; Tagliamonte, 2016 for discussions on the topic).

¹⁰ Examples of this pronunciation have been found in spoken BP and are considered examples of *hypercorrection*, an attempt to correct an apparent incorrect form.

Innovations have been seen by prescriptivists as threats to 'proper' language as they could cause irreversible damage.

When examining linguistic variables, those used by the higher class are often seen as more prestigious. Assigning a higher value to these variables means that speakers who make use of other variables may be discriminated against and the innovative forms used by them possibly stigmatized. Consider, for example, how *Dicionário Informal* ('Informal Dictionary') refers to the innovative ways in which *tipo* (masculine noun, 'type, kind') is used in vernacular Rio-chat and how Marleth Silva, a writer for the newspaper *A Gazeta do Povo*, sees its usage as highly pejorative in spontaneous speech:

- *Expressão idiomática, indicando miséria vernacular absoluta* 'Idiomatic expression which indicates absolute vernacular poverty' (Dicionário Informal, 2019)
- *"Meu filho adolescente anda falando muito "tipo assim" ou apenas "tipo." Eu me pego falando também. Dou-lhe um tapinha cada vez que ele deixa escapar um e ele faz o mesmo comigo. Achei que era um problema nosso, um vício familiar. Até me surpreender com pessoas, tipo assim cultas, intercalando suas frases com a lamentável muleta."* 'My teenage son is saying "tipo assim" or just "tipo" a lot. I catch myself saying it, too. I give him a light slap every time he lets one slip and he does the same to me. I thought it was our problem; a family vice. Then I was surprised to see, *tipo*

(‘like’), educated people, interspersing their sentences with the deplorable crutch’ (Silva, 2013, para. 2).

It also seems that other words and expressions in Brazilian Portuguese are not immune to offending prescriptivists as a similar stance can be found when the topic under discussion is the phrase *sei lá* (‘I don’t know’):

- *Expressão que significa o mesmo que "não sei", simplesmente, mas que dependendo da "entonação" usada, seja na fala ou na escrita, pode querer dizer simplesmente "não sei" ou ter algo de reticências, de evasiva, de preguiça de pensar ou então de desaforo.*
‘Expression that simply means ‘I don’t know,’ but which, depending on the intonation used, whether in speech or in writing, may simply mean ‘I don’t know’ or work as an ellipsis, evasion, laziness in thinking, or even effrontery’ (Dicionário Informal, 2019).

Prior research suggests that teenagers play an important role in linguistic innovation (D’Arcy, 2005; Tagliamonte 2005, 2016), most notably female speakers, who have been shown to be in the forefront of linguistic innovation (Tagliamonte & D’Arcy, 2009). As they spend their youth under the care of their parents or caregivers, children are influenced by the vernacular of those individuals when acquiring language. The teenage years have been shown to be one of the first stages in an individual’s life in which the influence of groups outside the family becomes quite important (Chambers,

2003; Eckert, 1997; Tagliamonte, 2005). The acquisition and the increase in the use of variants during such an important developmental stage (in addition to variants that are acquired from the exposure to the language of their caretakers) offers valuable information on how innovation may lead to language change (Tagliamonte, 2016).

The participation of adolescents in groups has been shown to promote the spread of language variation and linguistic change as they become more exposed to linguistic innovation (Eckert, 1988; Tagliamonte, 2016) at a stage when they are developing their social identity (Eckert, 1988). At such an important transitional period, the natural tendency to search for distinctness may also be reflected in the development of linguistic innovation. As Eckert (2018) states, adolescence, "... provides greater motivation than at any other time in life to adapt linguistic patterns to community structure" (p. 64).

Adolescents have been shown to promote linguistic change, including (but not limited to) structural change, lexical variation and innovation, and phonological variation (Andersen, 2000). As Eckert (1988) argues, this age group may give researchers access to "an important key to the study of the mechanisms of such adaptation" (qtd. in Tagliamonte, 2005, p. 1912). Teenagers are, therefore, a valuable resource to studies of linguistic innovation as language variation identified in this group may reflect unique social realities and may provide insights into linguistic change at a particular moment in time. Although the language of teenagers may

provide an insight into real-time linguistic change, research focusing on the speech of this age group remains scant in Brazilian Portuguese.

1.6 Projeto Sociolinguístico Contemporâneo Brasileiro

The importance of the *Carioca* dialect and its unique elements resulted in the development of a project that aimed to register the vernacular and gain access to possible linguistic innovations in oral discourse of teenagers in the city of Rio. In 2015, Sky Onosson and the author developed the *Projeto Sociolingüístico Contemporâneo Brasileiro* ('Contemporary Brazilian Sociolinguistic Project,' henceforth PSCB) (Thompson & Onosson, 2016)¹¹. The project involved fieldwork conducted in municipal public schools in that city, which are mostly attended by students of low to middle-low socioeconomic status who, oftentimes, live in one of the many *favelas* 'slums' in the city. This was the first project of its kind to receive the support from the Secretary of Education of Rio de Janeiro, giving full access to the 560,000 students that comprised the school district (currently the largest one in Latin America)¹² at the time of data collection. 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.

¹¹ This project received ethics approval from the University of Victoria, the University of Saskatchewan, and the Secretaria Municipal de Educação da Cidade do Rio de Janeiro (SME).

¹² We would like to thank the Municipal Secretary of Education of the city of Rio de Janeiro for allowing us access to the schools in that school district.

In an attempt to glean information on innovation and variation in the speech of the students, a modified version of Labov's (1984) sociolinguistic interview script, which encourages speakers to use informal language and speak as naturally as possible (D'Arcy, 2011), was used. Being a native speaker of the language and a member of the community of the area where the data collection took place, the author conducted the interviews. Between 2015 and 2018, 58 male and 120 female (N = 178) preteen and teenage middle school students (ages 11 to 16) were interviewed (see Table 1.2).

Table 1.2. *Students who participated in the PSCB project by age, gender, and school.*

Age	School A		School B		Total
	Female	Male	Female	Male	
11	1	1	4	0	6
12	5	1	27	15	48
13	3	0	43	22	68
14	4	0	25	11	40
15	5	5	3	2	15
16	0	1	0	0	1
Total	18	8	102	50	N = 178

Data collection sessions took place in Ilha do Governador, an island in the Guanabara Bay (see Figure 1.6). Ilha do Governador, which is connected to the mainland by a freeway, has a population of over 212,000; of those, 32% live in one of its 29 slums which correspond to over 42% of the total area of the island (Cavallieri &

Lopes, 2006; Cavallieri & Vial, 2010). Ilha do Governador was chosen as the site for the data collection for the following reasons:

1. The author is a member of the community and also attended municipal public schools on the island;
2. The administration of the schools on the island showed immediate interest in the study;
3. The easiness of mobility within the island made it easier for data collection to take place as data collection sessions were arranged according to school schedule and student availability.

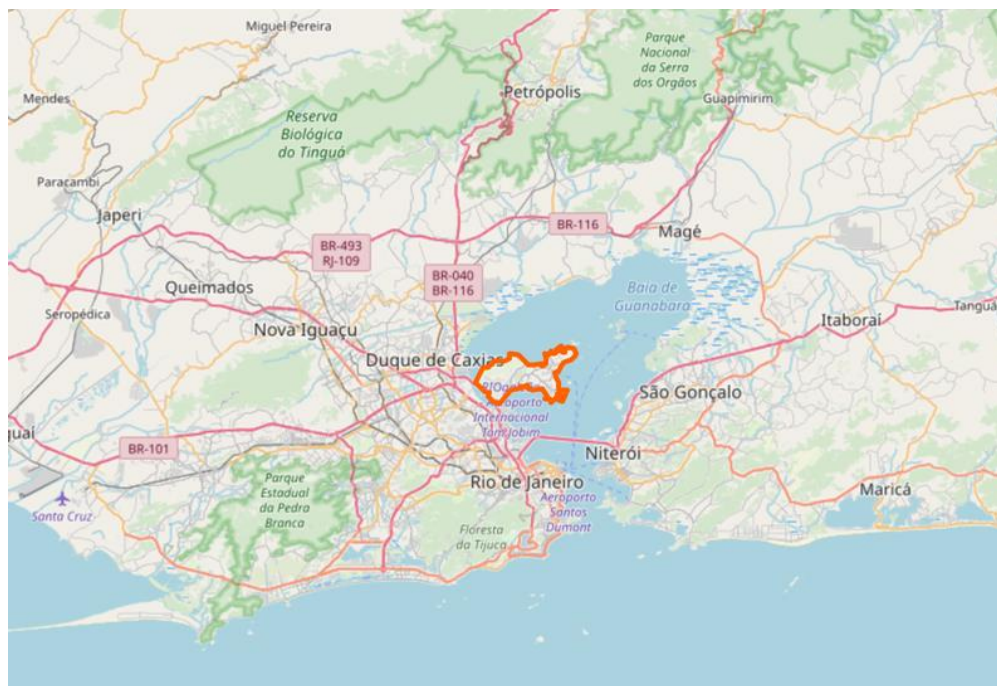


Figure 1.6. Map of Rio de Janeiro showing Ilha do Governador (in red). This map is freely licensed under the Open Data Commons Open Database License (ODbL) by the OpenStreetMap Foundation (OSMF).

Before data collection started, ethics approval was obtained through the University of Victoria, University of Saskatchewan¹³, and the Secretary of Education of the city of Rio de Janeiro. Data were collected in two municipal schools on the island¹⁴. During 5 field trips to Rio, 84 interviews were conducted, each ranging between 30 to 70 minutes¹⁵ in length, for a total of approximately 57 hours of data. To date, this is the largest sociolinguistic project to be held in the schools of Rio de Janeiro.

1.6.1 Data Collection - Methods

Prior to the commencement of the data collection sessions, students and school administration received information letters about the study¹⁶. Middle-school students interested in participating received a consent form to be signed by their caregivers and by the students themselves. Once the signed forms were returned, the researchers and the school staff discussed optimal times for data collection sessions to take place.

Interviews were held in rooms assigned by each school and were conducted in pairs or groups of three. To increase the likelihood of natural conversational speech being used, students chose their interview partner(s). To minimize the effects of possible formality

¹³ University of Victoria – protocol number 15-109; University of Saskatchewan – BEH-16-343.

¹⁴ We would like to thank the city of Rio de Janeiro and both EM Jornalista Orlando Dantas and GEO Nelson Prudêncio for the support provided and for the interest in our study (the names of the schools are included with their permission and that of the city of Rio de Janeiro).

¹⁵ The duration of the interviews was based on students' availability and their daily schedules.

¹⁶ While the first school visited caters to elementary and middle-school students (the latter being of interest to this study), the second school only caters to middle-school students.

and to try to mitigate the “Observer’s Paradox”¹⁷ (Labov, 1972), the researcher, who was also raised on the island, engaged with the students when appropriate by answering questions and discussing events that were pertinent to the community¹⁸. Quite often, students would lead the conversation, allowing them to pursue topics of their interest with minimal interruption from the interviewer. The end goal was to make students feel comfortable enough to speak as if they were not being observed (Labov, 1972).

Video and audio data were captured to provide access to both verbal and nonverbal data of all interviews. During the first two field trips in 2015 and 2016, a ZOOM Q4 Handy Video Recorder with built-in mic XY stereo format (120°), maximum sound pressure input of 130 dB SPL, input gain of +7db – +47dB was used to record audio and video data. As background noise was found to have affected the quality of the audio data collected, a ZOOM H4n Pro Handy Recorder and compatible SHURE WH20 cardioid headset microphones¹⁹ were used to collect audio data during the three data collection sessions that took place between 2017 and 2018 in addition to the visual data collected by the ZOOM Q4 Handy Video Recorder. The interviews were recorded in 16-bit Waveform Audio File Format (.wav) with a sample rate of 44.1 kHz. It is

¹⁷ However, as Schilling (2013) states, “it may be better [...] to admit that there is no such thing as non-observed language data” (p.112).

¹⁸ The positive experience participants reported after the interview resulted in a cascading effect and a large number of students being interested in taking part in the study. In some cases, over 90% of the students in a classroom were interviewed.

¹⁹ The choice of headset mics was based on the fact that they are adjustable and non-intrusive, allowing participants to feel more comfortable during the interviews.

important to clarify that, although the data from 2017 and 2018 were not collected under desirable laboratory conditions, all the recordings were of high quality.

1.6.2 Transcriptions

Upon conclusion of data collection sessions, the audio recordings of 64 interviews²⁰ were annotated²¹ and transcribed in Praat (Boersma & Weenink, 2017) by the author, a native speaker of the dialect. Separate tiers were used for the transcription of each participant's speech, which followed orthographic conventions. To maintain consistency (despite the acknowledgment of the idiosyncratic nature of speech), colloquialisms were kept to a minimum in the transcripts but were included in detail in the annotation tier, and inaudible parts of the audio were marked as such. Interjections, filled pauses (e.g., 'uh', 'ah'), and restarts were transcribed. The transcriptions were later imported into Elan (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006), allowing for the addition of video recordings to the audio recordings and the transcriptions. No audio or video recording was discarded. To protect participants' anonymity, files were saved under individual labels that include the participant's number, the number of the interview session, gender, and age (e.g., RDJ_150_M_14, in

²⁰ This number reflects the interviews that had been transcribed prior to the beginning of data analysis; it includes all of the most recent interviews and most of the interviews conducted in 2015.

²¹ Annotations included information on linguistic form, including phonological processes and unexpected features and structures that will be revisited at a later time.

which RDJ stands for Rio de Janeiro, 150 for student 1 of interview session 50, M for male, and 14 for the student's age).

Chapter 2 Linguistic Innovation

This chapter focuses on the results of an exploratory study conducted to examine salient elements in the speech of teenagers in the PSCB corpus. Information on data collection sessions and a preliminary analysis are presented followed by a summary of some of the features found in the data. Next, the chapter focuses on one of the most prominent and least investigated elements found during the initial analysis, the noun *tipo* ('type,' 'kind').

2.1 Preliminary Analysis

To gain insight into whether words and expressions are being used in innovative ways in the vernacular, the corpus data retrieved from the PSCB were examined from a synchronic (at a specific point in time) perspective. The goal was to identify possible trends of features performing functions other than their referential (original) function (e.g., a noun being used as a conjunction). Although historical analyses have been proven essential to determine the progression of change and the dissemination of innovative elements, investigations from a single point in time provide information about the current state of a language. In addition, the unavailability of audio recordings means that much of the historic data is comprised of written documents, which often reflect the normative use of language rather than the vernacular. Labov (1994) also argues that such documents may include scribal errors and hypercorrections (p.11).

Most importantly, Labov (1994) says that, “[h]istorical documents can only provide positive evidence. Negative evidence about what is ungrammatical can only be inferred from obvious gaps in distribution (p.11). Researchers have suggested that the application of a synchronic-diachronic interface to studies of linguistic change is valuable (Ramat, Mauri, & Molinelli, 2013). However, due to the still limited number of diachronic studies on language variation in Brazilian Portuguese (especially on the speech of *Carioca* teenagers), this study focused on salient linguistic elements and their current state in spoken language to investigate whether they are used in new, innovative ways.

An initial cursory analysis of the data was conducted once transcriptions were finished. The goal of this analysis was to identify salient elements²² in the speech of participants. Some of the elements such as *aí* (adverb of place ‘there’) and *então* (‘therefore’) have been the focus of several studies in Brazilian Portuguese (see Braga, 2003; Martelotta & da Silva, 1996; Pezatti, 2001; Tavares, 1999); of the features initially identified, *tipo* (‘type,’ ‘kind’) is unarguably one of the most prominent and least investigated. The next subsections highlight some of the findings from the initial analysis.

²² In this context, *salient* refers to either high frequency use of a linguistic element or one used frequently in unexpected ways.

2.1.1 *Aí*

Researchers have proposed that *aí*, an adverb of place that refers to a location that is far from the speaker and close to the person being talked to (Dicionário Priberam, 2019), has developed pragmatic functions other than its referential meaning as illustrated in the example below (Souza, 2009; Tavares, 1999, 2003).

(3) *E tudo que tem, algum tipo de festa, de eu num tô lá, elas ligam cá pra casa da minha mãe:*

*“Ah, Leila tá **aí**?” LEI, 04 (Braga, 2003)*

*‘And everything there is, some kind of party, that I am not there, they call me here at my mom’s house: “Ah, is Leila **there**?”*

For instance, Braga (2003) has argued that *aí* has been shown to perform both anaphoric and cataphoric functions as illustrated in (4):

(4) Anaphoric: *Uns seis meses depois que a gente chegou, ele resolveu ir lá pedir uma vaguinha pra deixar jogar. **Aí** desde **aí** ele não largou mais.* GU, 62 (Braga, 2003)

*‘About six months after we arrived, he decided to go there to ask for chance to play. **Then** since **then** he has not dropped it.’*

Aí has also been shown to be used to connect cause and effect as exemplified in (5).

(5) *Eles botaram ela, assim, num monte de aparelhos, sabe? **Aí** ela deu uma melhorazinha.*

FLP 03 L 1222 (Tavares, 1999)

*‘They connected her to a whole bunch of machines, you know? **Then** she improved a bit.’*

Findings also show that *ái* may be used to indicate textual (rather than temporal) sequencing, thus functioning as a topic organizer (Braga, 2003; Tavares, 1999) as illustrated in examples 6 and 7:

- (6) *É, ali tinha o Rox, Cine Rox, e tinha o Cine Ritz também. Mas só o Cine Ritz também. (inint) hoje, né? Existia naquela época também. Aí o Cine Ritz só [<ti->]- tinha cinema pra criança, mas era só durante a tarde, e à noite não podia ir, né? É porque naquela época a censura não era dezoito anos, era vinte e um anos.* FLP 18. L 1109 (Tavares, 1999)

‘Yeah, there was Rox, Cine Rox, and there was Cine Ritz there, too. But only Cine Ritz, too. (unint) today, right? It existed then too. **Then**, Cine Ritz only [<ha->] - had movies for children, but it was only in the afternoon, and at night children could not go, right? It’s because at that time the movie rating was not eighteen years, it was twenty-one years.’

- (7) F: *Ele começou a se... a gostar de uma garota, mas a garota só dava fora nele. Aí um dia ela resolveu dizer que tinha um tal de um teste lá. Aí ele foi lá concorrer. Aí nesse teste ele consegue, ele vai trabalhar na Broadway.* GU, 62 (Braga, 2003)

‘He started to ... liking a girl, but the girl kept turning him down. **Then** one day she decided to say that there was a test there. **Then** he went there to take part in it. **Then** on this test he manages, he gets to work on Broadway.’

Braga and Paiva (2003) observe that the functions of *ái* include the conclusion or revisiting of a subtopic. In the case of the *Carioca* dialect, Braga (2003) proposes that *Cariocas* are seemingly using *ái* at the beginning of utterances to refer back to a topic that

was previously mentioned. This function was also identified in the PSCB corpus as example 8 illustrates:

(8) *Tipo, ela ficou quieta na dela, só que eu sou assim, sempre que entra alguém nova, sempre vou, chego perto da pessoa, vou falar. Ela, não, ela fica no canto dela. Aí eu fui assim, "Bom, vou chamar pra minha amiga."* (RDJ_101_F_14)²³

'Like, she was quietly by herself, but I am like this, anytime anyone new gets in (the class), I always go, get closer to this person, talk (to him/her). She won't, she is always on her own. Then I was like, "Well, I am going to ask her to be my friend."

2.1.2 *Então*

Então 'at that time/occasion' (Dicionário Priberam, 2019) is an adverb of time that has caught the attention of researchers due to the innovative ways (in addition to its referential meaning) in which it is being used in the speech of Brazilians. Grammarians Cunha and Cintra (2000) argue that *então* (among other adverbs) could belong to a class of "denotative words" (p.372), a group of adverbs that, in certain constructions, are deemed not only unclassifiable as such but not easily classifiable at all (Cunha & Cintra, 2000).

²³ To ascertain anonymity, codes were created to replace students' names. RDJ refers to data collected in Rio de Janeiro; the three following numbers refer to participant number in individual interviews and the interview number (e.g., 204 refers to participant number 2 in interview number 04); F stands for female participants and M male participants; the last two numbers refer to the participant's age at the time of data collection.

Research has shown that *então* is performing pragmatic functions, an evolution that does not seem to be limited to Brazilian Portuguese, with similar findings being shown regarding the Spanish *entonces* 'so' (Flores-Ferrán, 2014) and the English *so* (Bolden, 2009; Byron & Heeman, 1997). In BP, Braga (2003) observes that *então* can not only provide cohesion by introducing sequences or subtopics, but it can also introduce utterances that express conclusion, effect, or consequence as example 9 illustrates.

(9) “...ao saltar do carro Marcelo tentou conversar com Neuza, para que entrássemos em casa sem acordar os pais de Márcia. Mas a nojenta da garota em vez de ajudar a irmã ficou reclamando e falando besteiras, **então** Marcelo deu-lhe um fora curto e grosso...” (Braga, 2003)

‘When he jumped out of the car, Marcelo tried to talk to Neuza, so that we could enter the house without waking up Marcia's parents. But the disgusting girl instead of helping her sister kept complaining and talking bullshit, **so** Marcelo gave her the brush-off.’

Pezatti (2001) observes that *então* may perform a semantic connection suggesting conclusion between statements. Such changes seem to be following the loss of temporal reference as in example 10 below:

(10) *hoje em dia se você depois passou uma época que você ia ao cinema tinha que ficar de pé numa fila e NORme... não é? Então não era divertimento aquilo... era::eu acho que era nem divertimento (ruídos) passava porque a pessoa ficava cansada de ficar em fila adquirir ingresso ficava na fila de ingresso.* DID-SP-234:582 (Pezatti, 2001)

'Nowadays if you have lived in a time when you went to the movies and had to stand in a huge line... isn't it? **So** that was not entertainment... it was::I think that it wasn't even entertaining (noise) would be over because the person would get tired from standing in line to purchase a ticket would stand in the ticket line.'

Martelotta and Rodrigues (1996) claim that *então* also orients listeners by providing information about the linearity of events, such as in example 11:

- (11) *"fui/ao banheiro, para escovar os dentes. Entretanto, já no banheiro,/ saíu debaixo do cesto de roupas, uma grande aranha, marrom, horrorosa!... Eu dei um berro (pois morro de medo de aranhas), e ela continuou lá. Dei outro berro, e então ouvi o meu pai dizer: "corre, filhinha"; ele naturalmente percebeu que se tratava de uma aranha, mesmo não estando lá. (Martelotta & Rodrigues, 1996)*

'I went/to the bathroom to brush my teeth. However, already in the bathroom,/a huge brown ugly spider came from under the hamper!... I screamed (because I am afraid of spiders) and it remained there. I screamed again, and **then** I heard my dad say: "run, honey"; he naturally assumed it was a spider even though he was not there.'

In addition, *então* seems to be used to indicate the relationship between a state and an inference. In such cases, Pezotti (2001) argues that the argument is often presented in a circular way, with the presentation of the argument first, followed by the cause, and finally returning to the argument, which is then introduced by *então* as in example 12:

(12) *se eu começo a a a a pensar em estatística se verá que o lugar mais perigoso do mundo é a cama... porque noventa por cento das pessoas morrem na cama... então é o lugar mais perigoso... não vá pra cama que você não more... bem* D2-RE-05:105 (Pezatti, 2001)

'if I start thinking in terms of statistics you will see that the most dangerous place in the world is the bed... because ninety percent of the people die in their beds... so it is the most dangerous place... do not go to bed and you will not die... well'

2.1.3 *Sei lá*

The periphrastic construction *sei lá* ('I don't know') is constituted by the verb *saber* ('know') inflected in the first person singular in the simple present tense ('[I] know') followed by the adverb of place *lá*, a reference to a place far from the speaker (Dicionário Priberam, 2019). These two words have been shown to be used as a fixed expression in oral discourse. Previous research has investigated the desemanticization (loss of referential meaning) of *sei lá* (de Araújo, 2014; de Oliveira & dos Santos, 2011; Martelotta & Rêgo, 1996) and the acquisition of new textual and interactional functions.

Researchers have proposed that *sei lá* may help re-establish discourse structure, be used as a filler (i.e., filled pauses), be indicative of avoidance of commitment, or be a sign of hesitation (de Araújo, 2014; de Oliveira & dos Santos, 2011; Martelotta & Rêgo, 1996) as in the examples 13 and 14 below, retrieved from de Araújo (2014) and Martelotta and Rêgo (1996):

Hesitation:

- (13) *Acho que já tem três ou é quatro, sei lá. Já perdia a conta desse nego-ço já, faz muito tempo, óh. [estrala s dedos]. Já perdi da conta e não quero nem lembrar da conta mais.*

(Araújo, 2014)

'I think it has been two or three, **I don't know**. I have lost count of this stuff, it has been a long time. [snaps fingers]. I have lost count and don't even want to remember the bill anymore.'

- (14) *eu pensei que eles fossem ficar ali... quando eu virei pra trás que eu fui procurar eles... para... me entrosar no assunto... eles tinham sumido... conclusão... a música já estava acabando... e eu tive que...que... sei lá... tive que... ver sozinho como é que fazia as coisas...*

(Martelotta & Rêgo, 1996)

'I thought that they were going to stay there... when I turned around to look for them...to... get more info on the topic...they were gone...in summary... the song was almost over... I had to... to... **I don't know**.... I had to.... learn how to do things on my own...'

In the PSCB corpus, the use of *sei lá* is not usually an indication that the teenagers either lack the knowledge about the fact under discussion or are unable to answer a question but rather to indicate an upcoming dispreferred response or to attenuate/downplay the previous/following message (for instance, when voicing their disagreement or giving negative feedback as in example 15). As example 16 shows, speakers also use the expression when they are uncertain about how a statement will be received, seemingly trying to save face (both theirs and the listeners'), in accord with findings on the English expression 'I don't know' (Tagliamonte, 2016).

- (15) *Às vezes é até o contrário, não sei, dependendo de do colega, sei lá.*
'Sometimes it is even the opposite, I don't know, depending on on the friend, I **don't know.**' (RDJ_147_M_13)
- (16) *Num gosto, sei lá, acho estranho, pô.*
'I don't like it, I **don't know**, I find it weird.' (RDJ_147_M_13)

As previously mentioned, results of the analysis of innovative use of linguistic forms indicate that the word *tipo* ('type', 'kind') is one of the most ubiquitous elements used by the speakers; nonetheless, *tipo* remains largely under-investigated. As a result, an analysis of the current state of grammaticalization of *tipo* was conducted. The question underpinning the analyses conducted regard the possibility of identification of discourse-pragmatic changes and consequent classification of innovation in its use in conversational speech in the city of Rio de Janeiro.

To establish the current state of the use of *tipo* in the *Carioca* dialect, an initial documentation of linguistic predictions of its use in oral discourse was conducted. Excerpts from the PSCB corpus (Thompson & Onosson, 2016) were examined (see Section 3.3). Results will not only present the first account of possible systematic syntactic constraints in the use of *tipo* in the *Carioca* dialect but will also provide the foundation for an analysis of the functions it is currently performing in the dialect. Findings will offer an insight into its possible grammaticalization (see Section 3.2).

Using synchronic dialect data, this study initially aims to:

1. Uncover synchronic patterns of variation in the function of *tipo* in the corpus of interview data of which PSCB is comprised;
2. Disambiguate the different grammatic and discourse-pragmatic functions of *tipo* in oral discourse by conducting a careful functional analysis.

Upon attesting the multifunctionality of *tipo*, the following analyses will concentrate on the most frequent functions it is performing in discourse. The investigation will consist of the analyses of:

1. The distribution of tokens by frequency, function, and speakers' sex (Chapter 4)
2. Examination of possible phonological reduction (erosion) of *tipo* in oral discourse (Chapter 4);
3. (if phonological reduction/erosion is identified) Examination of possible correlation between function and phonological reduction (Chapter 4);
4. (if erosion is identified) Investigation of speakers' capacity to discriminate phonological differences among the different functions (Chapter 5).

2.2 Summary

This chapter described the results of a preliminary analysis of the PSCB corpus which aimed to explore possible innovations in the speech of teenagers in Rio de Janeiro.

Focusing on salient linguistic elements, the analysis revealed that the elements *áí*, *então*, *sei lá*, and *tipo* were often used by speakers, with *tipo* being the least investigated of the four terms.

Chapter 3 *Tipo*

This chapter consists of the initial analysis of the uses of *tipo* in the *Carioca* dialect of Rio de Janeiro. To examine *tipo*'s possible grammaticalization, an initial analysis of linguistic predictions was conducted. Upon the completion of this first stage, a functional analysis was performed to identify the grammatical and pragmatic functions of *tipo* in the *Projeto Sociolinguístico Contemporâneo Brasileiro* corpus.

3.1 *Tipo*

In 1999, Brazilian researcher Bittencourt proposed that the noun *tipo* 'type,' 'kind' was undergoing a process of semantic change. Bittencourt (1999) posited that possible innovation in the use of *tipo*, at times alongside the adverb *assim* 'this way/manner', had resulted in an expression that means 'something of this/that type'.

Tipo, a taxonomic masculine noun (canonical meaning 'type', 'kind'), comes from the Greek noun *typos* 'mold.' In Brazilian Portuguese, it is defined as 'type', 'kind', or 'example' but may also refer to someone who is 'unlike others' or 'eccentric' (Bittencourt, 1999). As a noun, *tipo* is often used in binominal constructions²⁴ (N1 of N2 – *tipo + de* ('of') + noun) in which *tipo* is the head of N1 (Mihatsch, 2016). Bittencourt (1999) argued that through a process of semantic bleaching and semantic change

²⁴ In this dissertation, I do not provide a detailed account of binominal taxonomic nouns; for an in-depth investigation discussion of these class of nouns, see Mihatsch (2016).

resulting from speakers' innovative use, *tipo* might be acquiring multiple functions beyond that of its original nominal meaning.

In its original nominal form, *tipo* can be inflected for singular or plural and is commonly preceded by an article, a numeral, or a quantifier (e.g., *alguns tipos* 'some types') and may be followed by a prepositional phrase. The examples below illustrate constructions that include the referential (nominal) form that were retrieved from the Programa de Estudos sobre o Uso da Língua (PEUL) and Projeto Sociolingüístico Contemporâneo Brasileiro (PSCB) corpora.

(17) *Todo mundo tem esse tipo de problema e de violência*
every world has this **type of** problem and of violence
'Everyone has this type/kind of problem and of violence' (11 MIR - PEUL²⁵)

(18) *Eu gosto de todo tipo de música*
I like of all **type of** music
'I like all types of music.' (RDJ_266_F_12)

Work on the innovation in the use of taxonomic nouns such as *tipo* can be found in in languages such as English (Traugott, 2008), Italian (Voghera, 2013), Russian

²⁵ 11 - MIR - PEUL refers to the age of the participant, the participant's initials, and the corpus.

(Kolyaseva & Davidse, 2018), and Spanish (Fernández, 2017; Mihatsch, 2016).

Conversely, to date, few studies have examined the possible multifunctionality of *tipo* in oral discourse in Brazilian Portuguese and, more specifically, in the *Carioca* dialect. For instance, de Castilho (1991) discusses functional changes of *tipo* in their analysis of the speech of educated *Carioca* participants. Using data from the *Projeto da Norma Urbana Oral Culta do Rio de Janeiro* corpus (NURC-RJ), de Castilho (1991) argues that *tipo* is performing different functions in her analysis of semantic constraints of different structures, including that of an adverb indicating an approximation. The author also discusses in length the two main structures in which *tipo* is found in the data: preceding a prepositional phrase (PP) (e.g., *O Mangalarga é um tipo de cavalo... não muito grande...* 'The Mangalarga is a type of horse that is not very big.') and preceding a noun phrase (NP) (e.g.: *Aquelas locomotivas de... maria-fumaça...tipo maria-fumaça* 'Those locomotives of ... steam train... like steam train') (de Castilho, 1991). Using oral and written data, Lima-Hernandes (2005) conducted both synchronic and diachronic analyses of four words, including *tipo*. The data for her synchronic analysis of *tipo*, *feito* 'as, like', *igual* 'equal to', and *como* 'as' were retrieved from *the Projeto Norma Linguística Urbana Culta* (NURC, Rio de Janeiro) and *Programa de Estudos sobre o Uso da Língua* (PEUL, Rio de Janeiro) corpora while the diachronic analysis focused on printed documents between the 13th and the 20th century. Lima-Hernandes (2005) argues that *tipo* has five different functions: reference noun, classification noun, preposition, comparative element, and discourse

marker as well as an indicator of an approximation. The example below from Lima-Hernandes (2005) illustrates *tipo* functioning as a comparative element:

- (19) *Vai morar para lá, ("então") carrega sua família. Tem muitos capixabas, tem muitos cariocas, gaúchos, mineiros. Mineiro dá muito! Então, aí é tipo uma colônia, não é?*
(E23-PEUL-Amostra 80)

"Go live there, ("then") take your family. There are many capixabas, there are many cariocas, gaúchos, mineiros. Lots of mineiros! So, then, it's **like** a colony, right?"

Tipo has also been investigated in other dialects of BP. For instance, Bertozzo (2014) examined the functions of *tipo* in the speech of twelve individuals between the ages of 7 and 49 divided into two groups: one consisting of 7- to 14-year-olds and another of 25- to 49-year-olds. The data were collected in the city of Chapecó, south of Brazil. Findings indicate that *tipo* is undergoing the process of grammaticalization (see Section 2.2) and is performing the functions of a conjunction, an adverb, a pronoun, a noun, and as a comparative element. Laurentino (2016), on the other hand, investigated the functions of *tipo* in *Natalense* dialect of Natal, Rio Grande do Norte. The data, retrieved from the corpus *Banco de Dados FALA-Natal*, consisted of oral discourse of 15- to 21-year-olds. Based on the results of his analysis, Laurentino (2016) argues that *tipo* is performing four pragmatic and eight discursive functions. He describes the functions as the following:

1. *Exemplification*, in which *tipo* is used to illustrate or provide examples of an idea previously shared in the discourse, such as in example (20).

(20) *Sobre assassinato também, suspense. Tipo, uma pessoa andando num*
about murder too suspense **like** one person walking in
parque ou numa floresta... escura... aí chega um assassino
park or in a forest dark then arrives a murderer
e faz o homicídio
and makes the homicide

'About murders as well, suspense. Like, one person walking in a park or in a dark forest, and then a murderer arrives and commits the homicide.' (Natal, m.f.) (Laurentino, 2016)

2. *Comparative* - in this function, *tipo* would establish proximity or similarity between elements as (21) illustrates.

(21) *Naruto... tipo uma história tipo quase mangás japoneses*
Naruto like a story **like** almost manga japanese
'Naruto, like a story, like almost a Japanese manga.' (Natal, m.f.) (Laurentino, 2016)

3. *Explanation* - introducing a clarification or explanation about what has been previously said as in example 22 from Laurentino (2016).

(22) *Achei legal também. Tipo, eu não esperava não ganhar...*
 thought cool too **like** I not expected not to win
 'I thought it was cool, too. Like, I did not expect not to win...' (Natal, m.f.)

4. *Conclusion* - used to conclude a previously mentioned idea (meaning *therefore*) as in (23) (Laurentino, 2016).

(23) *Do respeito entre os professores os amigos... respeito em*
 of respect between the teachers the friends respect in
casa, na rua, assim... tipo foi bem legal essa oficina
 home in street this way **like** was very cool this workshop
 'The respect among the teachers, friends... Respect at home, on the streets like, this workshop was very cool.' (Natal, m.f.)

Regarding the discursive functions of *tipo*, Laurentino classifies them as follows:

1. *Phatic introduction* - when performing this function, *tipo* marks the beginning of a speaker's turn, with examples being found when the speaker starts answering a question (see example 24).

(24) Interviewer: *Como é que faz? Me diz como é que é?*
 how is that do me tell how is that is
 'How is it done? Tell me, how is it done?'

Participant: *Tipo você pega cinco ou seis palitos de coqueiro*
like you get five or six sticks of coconut
 'Like, you get five or six coconut sticks.' (Natal, m.f.)

2. *Information imprecision marker* - when performing this function, *tipo* would

indicate uncertainty or imprecision about what is being said, such as in (25):

(25) *Tipo acho que era umas quarenta e poucas pessoas numa sala*
like think that was some forty and few people in a room
'Like, I think there were around forty-something people in one room.' (Natal, m.f.)

3. *Elaboration marker* - when performing this function, *tipo* is used as a way to

manage discourse, more specifically a tool to buy time while organizing the utterance as in (26) (Laurentino, 2016).

(26) *Ficar... tipo... quando meu avô morreu... tipo... conhecer mais ele...*
stay **like** when my grandpa died **like** know more he
'Stay... like... when my grandpa died... like, get to know him better...' (Natal, m.f.)

4. *Introducing internal dialogue* - in this function, *tipo* introduces direct speech that

reflects the speaker's inner thoughts (see example 27).

(27) *Mas depois, acho que vai ficar se roendo, tipo:*
but after think that will stay self regret **like**
"Meu Deus, por quê? Isso tá acontecendo mesmo?"
my God why this is happening really
'But after, I think he will regret it, like: "My God, why? Is this really happening?"' (Natal, f.m.)

5. *Rectification marker* - in this function, *tipo* introduces a rectification or correction about what has been previously mentioned as example 28 illustrates.

(28) *Tentaria... mudar o Brasil, tipo corrupção que tá muito...*
 would try change the Brazil like corruption that is very
tipo, não tentar mudar tudo...
like not try to change everything
 'I would try... to change Brazil, like, the corruption is too... Like, not try to change everything...' (Natal, m.f.)

6. *Emphasis marker* - as this marker, *tipo* highlights what is to be said, with a emphatic pronunciation (Laurentino, 2016) (see example 29).

(29) *Só se fosse tipo crimes cruéis mesmo, assim*
 only if were **like** crimes cruel really this way
 'Only if they were like really cruel crimes.' (Natal, m.f.)

7. *Sequential marker* - as a sequential marker, *tipo* indicates that the utterance has not yet come to an end, thus suggesting the continuity of the speaker's turn (see example 30).

- (30) *Mas nós teve que levar ele pro canil. Tipo*
 but we had that take him to pound **like**
se ele morreu, nós num sabe
 if he died we not know
 'But we had to take him to the pound. Like, we don't know if he died.' (Natal, m.f.)

8. *Approximate delimitation* - when performing this function, *tipo* would indicate the non-precise nature of what is being said (meaning *approximately*) as in (31).

- (31) *Eu acho que na minha cabeça passou horas, mas*
 I think that in my head passed hours but
passou tipo cinco minutos.
 passed **like** five minutes
 'I think that, in my mind, hours had gone by, but it was like five minutes.'
 (Natal, f.m.)

Laurentino's study also presents a comparative analysis of the results from its morphosyntactic analysis to the results of the syntactic analysis of *like* conducted by D'Arcy (2005). The author states that only two positions previously observed by D'Arcy were not found in his data: *internal and to the left of the NP* and *internal and to the left of an AP* (Laurentino, 2016, p. 85).

Finally, Thompson (2013) argues that in her analysis of written and audio data, *tipo* is shown to be performing nine functions: (1) *noun [- generic]* (meaning

'individual'); (2) *noun [+ generic]* (meaning 'model'); (3) *delimiting noun* (introducing a noun phrase); (4) *delimiting articulator* (connecting the verb to its complement); (5) *modifying articulator* (connecting two noun phrases and indicating their similarities); (6) *approximative articulator* (meaning 'approximately'); (7) *comparative articulator* (sentence-initial, preceding a comparison to what was previously said); (8) *addendum articulator* (introducing a phrase that expands on the meaning of the main clause); and (9) *marker* (discursive function, in which most of its meaning has been bleached) (Thompson, 2013). A summary of the functions of *tipo* identified in recent studies is presented in Table 3.1.

Table 3.1. *Functions of tipo identified in previous studies on BP.*

<i>Author</i>	<i>Year</i>	<i>Function</i>
Lima-Hernandes	2005	reference noun
		classification noun
		preposition
		comparative element
		discourse marker
		indicator of an approximation
Thompson	2013	noun [-generic]
		noun [+generic]
		delimiting noun
		delimiting articulator

		modifying articulator approximative articulator comparative articulator addendum articulator marker
Bertozzo	2014	conjunction adverb pronoun noun comparative element
Laurentino	2016	exemplification comparative explanation conclusion phatic introduction information imprecision marker elaboration marker introducing internal dialogue rectification marker emphasis marker sequential marker approximate delimitation

Despite the aforementioned research, Brazilian scholars have discussed the limited number of studies that have examined *tipo*'s (multi)functionality in oral discourse. The reason, they argue, may be the limited number of tokens of *tipo* found in both oral discourse and written texts (Lima-Hernandes, 2005, p. 60)²⁶. In addition, studies that investigated *tipo* in the *Carioca* dialect have predominantly used already existing corpora (mostly collected between 1970 and 2000), resulting in limited ability to account for its current patterns of use. This study will shed light on the current state of *tipo* in oral discourse using recently collected data. This fresh look at its present-day state in the speech of *Cariocas* will provide a valuable contribution to the field of sociolinguistics in Brazil. The analysis will also give an insight into how teenagers, who have been shown to be in the forefront of linguistic innovation (Tagliamonte, 2005, 2016), are using this form in conversational speech. The analyses discussed in this chapter aim to address the following research questions:

1. What is the current state of *tipo* in the *Carioca* dialect of Brazilian Portuguese?
2. What are the patterns of use of *tipo*? Where is it often found in utterances?
3. What functions is *tipo* currently performing in the speech of teenagers?

²⁶ The argument raised by Lima-Hernandes (2005) was based on data retrieved from corpora that consist of oral data collected between the 1970s and the year 2000. I intend to question this claim in later chapters as, impressionistically, *tipo* seems to be ubiquitous in the PSCB corpus.

3.2 Grammaticalization

As previously discussed, Bittencourt (1999) postulated that the noun *tipo* 'type, kind' was possibly undergoing a process of grammaticalization, in which a linguistic element comes to acquire grammatical functions over time (Hopper & Traugott, 2003, p. 2). The term, coined by Meillet (1912), refers to changes from lexical to grammatical forms, which have been of interest to researchers since the 19th century (see, for example, Bopp, 1816, Gabelentz, 1901, and Humboldt, 1825).

Grammaticalization has been defined as the process by which lexical items (content words such as nouns, verbs, adjectives, and adverbs) come to acquire grammatical function or how grammatical items (function words such as prepositions, auxiliaries, determiners, connectives, etc.) acquire new grammatical functions (Hopper & Traugott, 2003). It has been defined by Kuryłowicz (1965) as "the increase of the range of a morpheme advancing from a lexical to a grammatical or from a less grammatical to a more grammatical status . . ." (p. 69) and by Meillet (1912) as "the passage of an autonomous word into the role of grammatical element" (p.131). This gradual shift from content to functional category would be reflected in both semantic and syntactic reanalyses (Brinton, 1996; Langacker, 1977). Researchers have posited that, to undergo grammaticalization, lexical items must at first be "semantically general"; during grammaticalization, the item would become progressively contextually constrained and would be reanalyzed. This reanalysis would allow the

item to perform new morphological and syntactic functions (Hopper & Traugott, 2003, p. 100).

Researchers have postulated that syntactic and semantic reanalysis may take place during grammaticalization (Brinton, 1996; Langacker, 1977). Items undergoing grammaticalization have been shown to undergo gradual semantic bleaching (or desemanticization), with loss of the element's propositional meaning and the acquisition and increase of pragmatic meaning by the grammaticalized word (Bybee & Pagliuca, 1985; Traugott, 1995). Hopper and Traugott (2003) observe that although weakening of the item's lexical meaning may take place, an expansion of the grammatical functions it performs results in an increase in its polysemy. Research on grammaticalization has indicated that this change, rather than being a binary switch, is a gradual progress from *more lexical* to *more grammatical* or *more concrete* to *more abstract*, typically following a unidirectional cline (Hopper & Traugott, 2003).

The hypothesis of unidirectionality is one of the core elements of the theory of grammaticalization. As items follow the cline, they would undergo processes that would allow them to acquire grammatical and pragmatic functions. The gradual path down the cline from lexical to grammatical as proposed by Hopper and Traugott (2003) is represented as follows:

- **content word > grammatical word > clitic > inflectional affix (Hopper & Traugott, 1995, p.7)**

while the verbal cline, showing the progression of acquisition of functional roles can be represented as:

- **lexical verb > vector verb > auxiliary > clitic > affix > zero (Fischer, 2007, p. 182)**

The clines as described above provide a visual representation of the unidirectional steps grammaticalizing elements would undergo from their canonical form. Although the clines represent the unidirectional steps that words would go through during grammaticalization, it is important to stress that not all words that undergo grammaticalization will reach the end of the cline.

Different taxonomies of processes pertaining to grammaticalization have been posited. For instance, Heine (2003) describes *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) as the four main mechanisms of grammaticalization (Heine, 2003). Hopper (1996) argues that five principles are involved in the ongoing process of grammaticalization:

- *layering* (i.e., different forms expressing the same meaning, such as the various ways one can express the future form in English, e.g.: 'I am going to travel,' 'I will travel')
- *divergence* (the coexistence of grammaticalized and lexical items such as in the case of English verb 'have,' which is the main verb in 'I have two cars' and an auxiliary verb in 'He has travelled')
- *specialization* (the process through which one form becomes the single option to perform a grammatical function, such as in the case of the negative construction in French, which saw *pas* becoming the most frequent complement to *ne* in negation despite the existence of other nouns such as *point*, *mie*, and *goutte* which also performed the function of reinforcing negation) (Hopper & Traugott, 2003, p.117)
- *persistence* (i.e., some of the referential meaning still associated to the item, which restricts its distribution, a characteristic of the early stages of grammaticalization, such as in the case of the verb *will* which likely developed its predictive future meaning from its Old English use indicating 'intention,' 'promise') (Hopper, 1991, p. 29)
- *decategorialization* (change in the category, with minor categories deriving from major categories and not vice-versa, such as the main verb 'like' as in 'He likes

chocolate' functioning as an adverb indicating approximation in 'She is like 6 feet tall') (Hopper, 1996).

In addition to these processes, Traugott (1995) posits that *subjectivity* (i.e., changes associated with the attitudes of the speaker) increases in the early stages of grammaticalization. Subjectification can be better explained as the subjective perspective that the speaker shares when making choices about the linguistic elements used express their thoughts and ideas. By negotiating meaning, speakers encode a subjective view which may, at one point, lead to the reanalysis of the word/expression (Traugott, 2010).

An increase in frequency has also 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 oftentimes 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, 2003b, 2003c). Compare, for example, the sentences 'I am going to visit my family' and 'I am travelling to see my family'; while movement verbs such as *go* and *travel* can be used in the construction *progressive + to + infinitive*, only *go* has come to accept non-mobile subjects (Bybee, 2003b). Therefore, increase in frequency would be

one of the main mechanisms not only of language change but also of grammaticalization (Bybee, 2011). As a result, Cacoullos and Walker (2011), among others, 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.

The complexity involved in the investigation of all the processes associated grammaticalization would make the completion of this study within the expected timeframe unlikely. Therefore, this study will focus on processes that are of special interest to the author, namely *phonological reduction/phonetic erosion; divergence; generalization; and acquisition of pragmatic functions*. Despite limiting the number of examined processes, this study is still innovative in nature, having been designed to offer an opportunity for different processes pertaining to grammaticalization to be examined within the scope of one single investigation.

To examine the aforementioned processes within the context of the use of *tipo* in the *Carioca* dialect, the analyses conducted in this study are synchronic in nature. Although grammaticalization is commonly associated with diachronic studies, which focus on the investigation of the evolution of forms over time, synchronic studies have been shown to provide valuable information regarding the state of elements that may be undergoing grammaticalization at a specific point in time (Lichtenberk, 1991; Pichler & Levey, 2011). Persistence of older forms alongside new forms and constructions (i.e.,

layering) may give researchers fresh information about the state of grammaticalization of elements and an insight into its mechanism.

3.3 Frequency Analysis

To explore the grammaticalization of *tipo* in the speech of participants, an investigation of the frequency of nominal vs. non-nominal use was conducted. Extracts containing the word *tipo* were retrieved from the 64 interviews (those that were fully transcribed by the time data analysis was conducted) in the PSCB corpus. In total, 1,496 examples were retrieved, and, of those, 63 examples were excluded from the analysis due to the uncertainty of their function. For example, the utterance in (32), as *tipo* was being uttered by the speaker, speaker RDJ_104_M_15 interrupted her and took the floor, preventing speaker RDJ_204_F_15 from concluding her thought. The lack of context led to uncertainty about the possible function *tipo* would have performed if the speaker had concluded her thought.

(32) *Tipo...* (RDJ_204_F_15)
'Like...' or 'Type...'

In total, 1,433 tokens were analyzed. Tokens were classified according to whether *tipo* was performing nominal or non-nominal functions. Non-nominal functions were identified based on the following features²⁷:

- loss of nominal properties (e.g., impossibility to be inflected for number), such as in *A minha vó é tipo essa* ('My grandma is like that') but not **Minhas avós são tipos essas*;
- difficulty in classifying *tipo* according to lexical categories, such as in *É tipo fazer bobagem* ('It's like doing something silly'); and, in addition,
- use in constraints other than those of the nominal form (e.g., introducing direct speech), such as in *Mas eu fiquei, tipo, "Eu não sou a mesma pessoa que a minha irmã"* ('But I was, like, "I'm not the same person as my sister"').

3.4 Methods

3.4.1 Participants

Data for this analysis were collected from the interviews of 106 participants (male, N = 30, female, N = 76) between the ages of 11 and 16 who attended one of two public schools in the city of Rio de Janeiro at the time of data collection. For more information on participants, see Section 1.6.

²⁷ The examples were retrieved from the PSCB corpus and were produced by speakers RDJ_239_F_14, RDJ_112_F_13, and RDJ_226_F_13, respectively.

3.4.2 Results

Results from a descriptive statistical analysis involving percentages of 1,433 excerpts show that the occurrences of the nominal function of *tipo* is slightly over 2% (N = 29) of the total number of occurrences in the data analyzed while the non-nominal forms were found in almost all the excerpts (N = 1,404, 97.98%). Example 21 illustrates the nominal use of *tipo* in a binominal construction (N1 of N2) while examples 22 and 23 illustrate its non-nominal use:

(21) *A gente gosta praticamente do mesmo tipo de música*
we like practically of the same **type of** music
'We basically like the same **type** of music.' (RDJ_126_F_13)

(22) *Aí tem dinheiro no final de semana pra sair tipo esse*
then have money on end of week to go out **like** this
final de semana
end of week
'Then you have money on the weekend to go out **like** this weekend.'
(RDJ_101_F_14)

(23) *Tipo ela tem vários amigos*
like she has several friends
'**Like**, she has several friends.' (RDJ_201_F_14)

Figure 3.1 clearly depicts the difference in frequency of the nominal and non-nominal forms of *tipo*, confirmed by a *Chi-squared test* that reveals that the non-nominal

functions of *tipo* are significantly more common than its nominal counterpart (χ^2 -squared = 1383, $df = 1$, $p\text{-value} < 2.2e-16$).

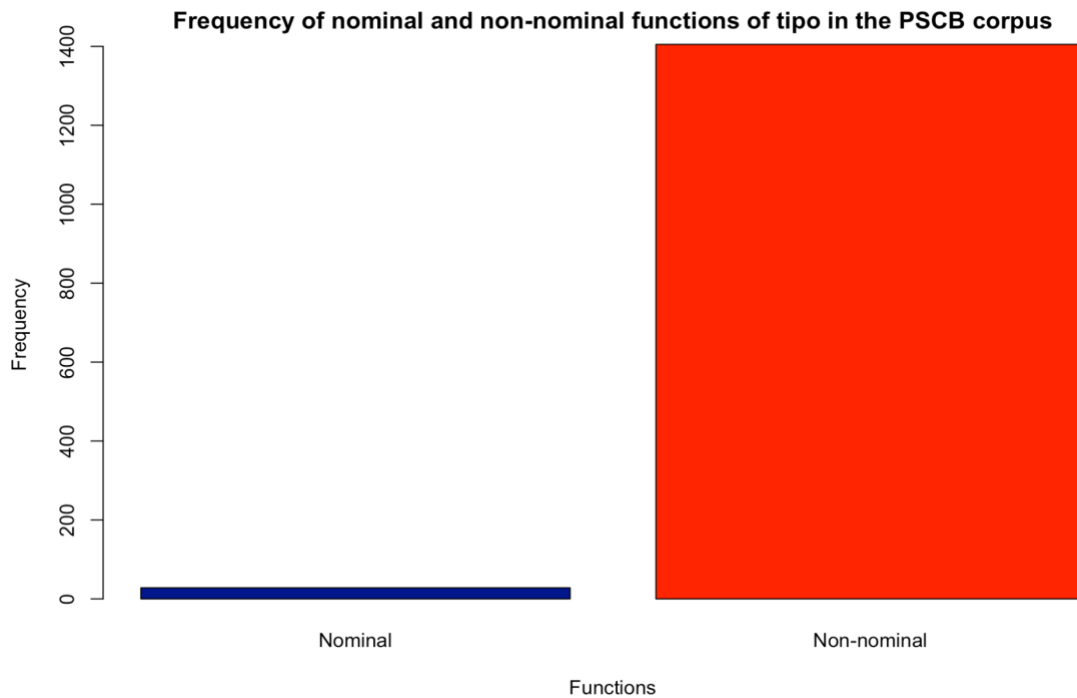


Figure 3.1. Frequency of nominal and non-nominal functions of *tipo* found in the data from the PSCB corpus.

Results from this initial analysis show a high frequency in the occurrence of non-nominal forms of *tipo* in the data. Although the results provide an overview of the apparent widespread use of non-nominal forms of *tipo* in the *Carioca* dialect, questions remained as to whether the expansion of these forms is a recent phenomenon. In an attempt to address that question, a comparative analysis of the PSCB corpus and two other corpora of oral discourse in Rio de Janeiro – *Projeto Norma Linguística Urbana Culta*

(NURC, Rio de Janeiro) and *Programa de Estudos sobre o Uso da Língua* (PEUL, Rio de Janeiro)²⁸ was conducted²⁹.

The *Norma Lingüística Urbana Culta* (NURC) project was created as an extension of the *Proyecto de Estudio Coordinado de la Norma Lingüística Culta de las Principales Ciudades de Iberoamérica* ('Project of coordinated study of the cultured linguistic norms in the main cities of Iberoamerica'). The project initially focused on the speech of educated individuals from five main cities in Brazil, with written data later being added to the corpus (Projeto Norma Urbana Culta, n.d.). The corpus, available online, includes audio recordings and transcriptions of interviews with upper-class *Cariocas* conducted in the 1970s and some follow-up interviews from the 1990s. Conversely, the *Programa de Estudos sobre o Uso da Língua* project ('Study Project on the Use of Language') (PEUL) focused on the investigation of variation in the *Carioca* dialect. Participants' educational level ranged from elementary to high school levels (de Paiva & Gomes, 2016). The goal was to use quantitative research methods to analyze stratified data and gain a better understanding of linguistic variation in that dialect. The PEUL corpus consists of interviews conducted in the 1980s and early 2000s with *Carioca* children, teenagers, and adults (de Paiva & Gomes, 2016).

²⁸ This is currently the most recent corpus of *Carioca* speakers made available to researchers in Brazil.

²⁹ Both were the result of research projects conducted through the Federal University of Rio de Janeiro (UFRJ).

Data from both corpora were retrieved for the investigation of frequency of nominal and non-nominal uses of *tipo* (see Section 3.4 for details on the analysis). From the NURC corpus, 47 transcripts (corresponding to interviews with individuals 20 to 80 years old collected in the 1970s and 1990s) were examined. In the case of the PEUL corpus, examples were extracted from 16 interviews conducted in 1999 and 2000, with participants' age ranging between 9 to 48 at the time of data collection. The end goal of this comparison was to examine the frequency of use of nominal and non-nominal forms of *tipo* in the past 50 years and gain access to the possible progression of use of those forms over time³⁰. Results show a significant increase in the number of non-nominal functions compared to its nominal function. Figures 3.2 (frequency of nominal and non-nominal tokens) and 3.3 (within-group comparison of percentage of nominal and non-nominal forms) provide a clear picture of the progression of the changes in the use of *tipo* in Rio de Janeiro over time.

³⁰ The extremely limited number of corpora focusing on the *Carioca* dialect limited the breadth of the comparative analysis.

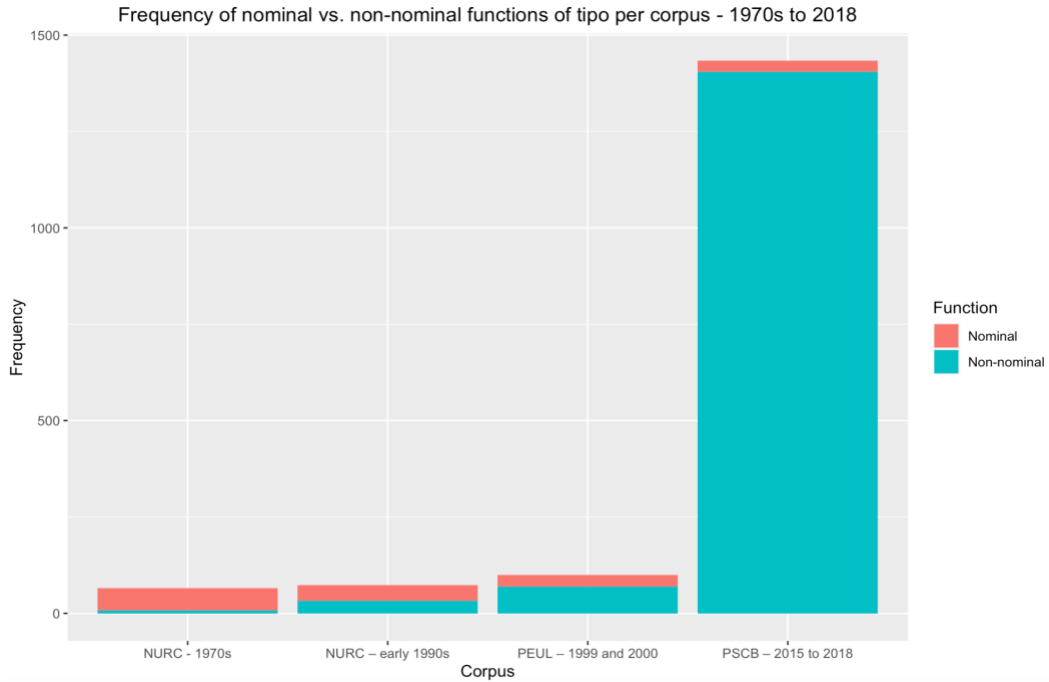


Figure 3.2. Frequency of tokens of nominal and non-nominal functions of tipo by corpus.

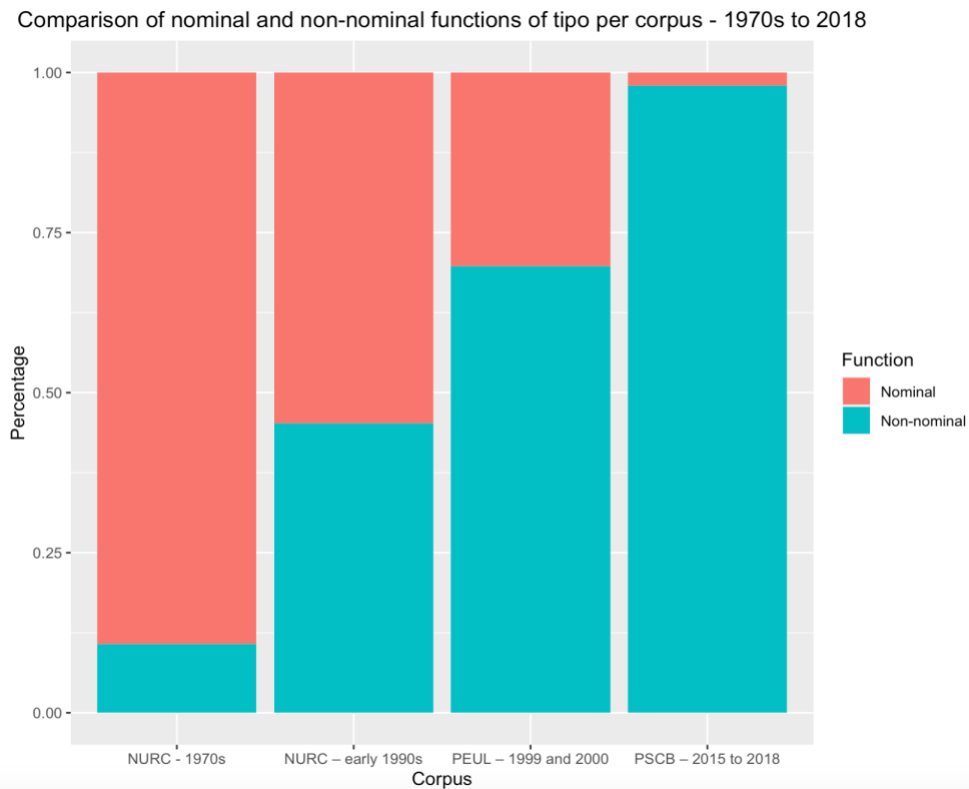


Figure 3.3. Comparison of the percentage of nominal vs. non-nominal functions of tipo per corpus (1970 - 2018).

Results show a progressive increase in the number of non-nominal forms of *tipo* in oral discourse in Rio de Janeiro. These findings, although important, should be seen with some caution as, unlike the PSCB corpus, both the PEUL and the NURC corpora include participants from different age groups; in addition, the NURC corpus consists solely of individuals that had attended post-secondary educational institutions unlike the PEUL and PSCB corpora. To address some of these differences and to get a clearer picture of the progression of non-nominal forms of *tipo* in the speech of *Carioca* teenagers, the frequencies of both nominal and non-nominal forms found in the PSCB corpus were analyzed vis-à-vis the teenage speech data retrieved from the PEUL corpus (the only one to include teenage participants). Data were retrieved from seven interviews with children and teenagers between the ages of 9 and 17. Due to the fact that there is a considerable discrepancy in the number of participants and the number of tokens, the frequency of use of nominal vs. non-nominal forms in each corpus was compared. Figure 3.4 provides a comparison of the total number of tokens per function in each corpus while Figure 3.5 offers a visual representation of the percentage of nominal vs. non-nominal forms per corpus (within-group comparison).

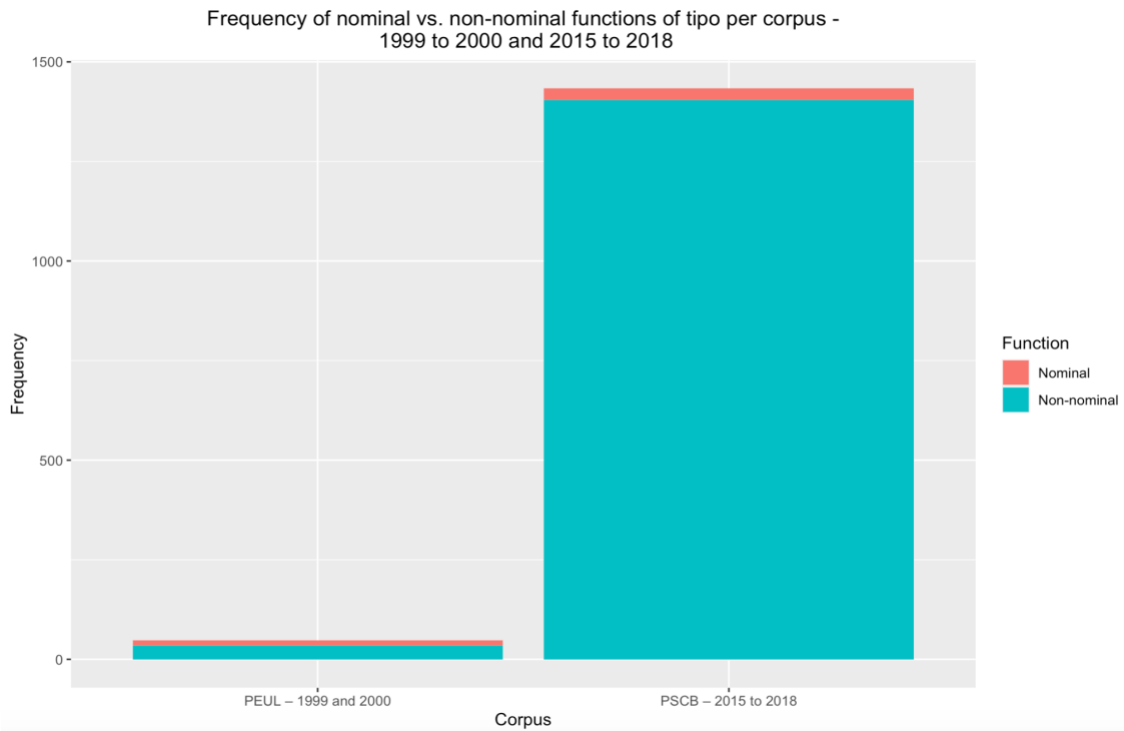


Figure 3.4. Number of nominal vs. non-nominal functions of tipo produced by teenagers in the PEUL and PSCB corpora.

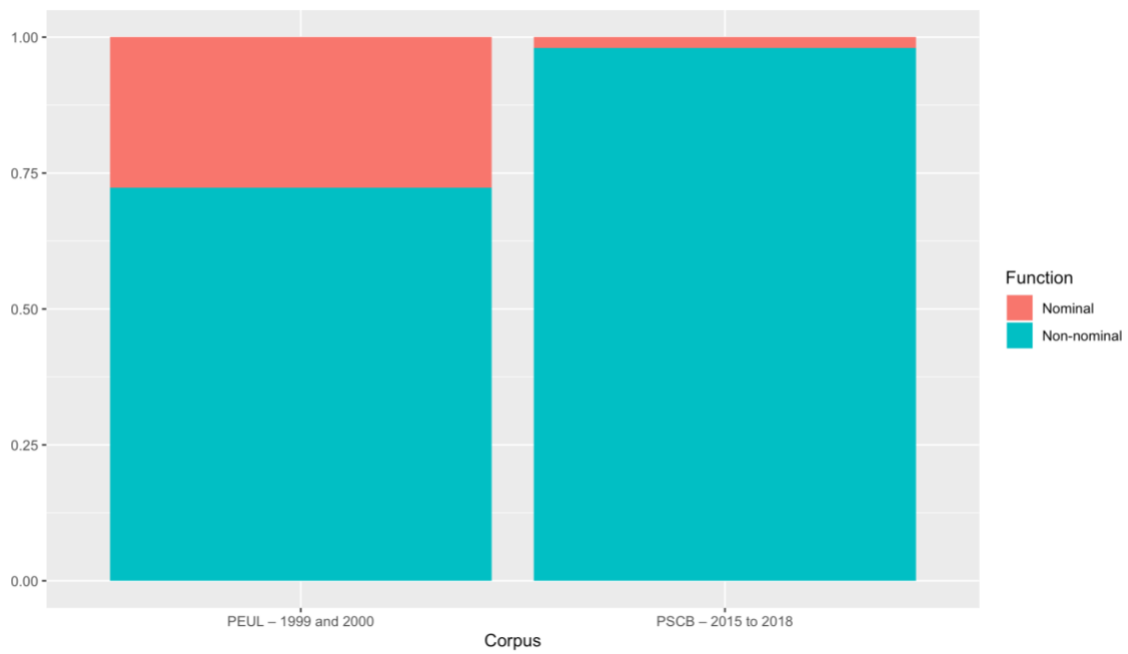


Figure 3.5. Percentage of nominal vs. non-nominal functions of tipo by teenagers identified in the PEUL and PSCB corpora.

As Figure 3.5 shows, results indicate not only that non-nominal forms of *tipo* were already more frequently used by teenagers in 1999/2000 when compared to the nominal form but also that the frequency of non-nominal forms is increasing at a fast pace, suggesting that *tipo* is undergoing grammaticalization. In addition to the increase in frequency, the identification of both nominal and non-nominal forms in the data (*divergence*) and its probable *deategorialization* (changing from an open class category to a closed-class category) (see Section 3.2) also supports the argument that the word is grammaticalizing as postulated by Bittencourt (1999).

3.5 Patterns of Use of *Typo*

One of the topics that has generated discussion when it comes to grammaticalized linguistic elements is that of the apparent “freedom” of use of grammaticalized elements. Some scholars have argued that grammaticalized elements can appear anywhere in a sentence or that their use is random or even ad hoc (Siegel, 2002). This possible lack of constraint has been described by some regarding the word *like* in English as well as the word *tipo* in Portuguese (see, for example, de Sá, 2017, Romaine & Lange, 1991, and Siegel, 2002).

In addition to being labeled as a word that represents vernacular poverty (see Section 1.5) and even a linguistic epidemic (Sut, 2006), *tipo* has also been defined as a word that substitutes almost any noun and can seemingly be inserted anywhere (de Sá,

2017). Despite being perceived as unconstrained, research on the grammaticalization of linguistic items such as the word *like* in English has shown that their distribution and functions are rule governed (see D'Arcy, 2005 and Tagliamonte, 2005, for a discussion on the topic).

To examine whether non-nominal uses of *tipo* are rule governed, i.e., to uncover possible rules governing its use in discourse, an analysis of linguistic predictions consisting of grammatical elements found preceding and following *tipo* was conducted³¹ (Tagliamonte, 2005) using authentic language data retrieved from the PSCB corpus. The goal of investigating linguistic constraints is twofold:

- To identify structures in which *tipo* is found
- To identify possible systematicity in the use of its non-nominal forms

A Praat script was used to extract sentences and relevant information from the textGrids that contained the transcriptions of the interviews. Data were imported into a Microsoft Excel table for the analysis. The table contained the following information: speaker (code), sex, age, school where the interview was held, pre-*tipo* word,

³¹ Henceforth these will be referred to as *pre-tipo* and *post-tipo*.

tipo/possible co-variants, post-*tipo* word, utterance, grammatical function of pre-*tipo* word, and grammatical function of post-*tipo* word.

Next, an analysis framework adapted from Tagliamonte's (2005) investigation of patterns of use of *like*, *just*, and intensifiers was used. The analysis was done in stages, starting with pre-*tipo* elements, followed by post-*tipo* elements. When found in sentence-initial position, the classification of a pre-*tipo* element was not possible due to the fact that no element preceded it; in the case of post-*tipo* sentence, the example was coded as ___ S (in which ___ stands for *tipo* and S for sentence). Several rounds of analysis were conducted:

- initially, an analysis of macro elements was conducted. This included, for instance, looking at clauses and sentences (e.g., preceding a coordinate clause, such as in *Eu não gosto de funk mas, tipo, gosto de brincar assim* 'I don't like funk music, but, like, I like to play like this.'^{32 33});

- once the analysis of macro elements was concluded, the investigation focused on micro elements. This included, for example, identifying if *tipo* precedes a noun phrase (such as in *Eu nunca quebrei nada, tipo, nenhum osso, nada* 'I've never

³² Although an analysis of patterns of use of *tipo* in relation to coordinate and subordinate conjunctions was conducted, the results will not be included in this dissertation as further analysis of the data is needed for conclusive findings to be presented.

³³ From participant RDJ_104_M_15 (PSCB).

broken anything **like no bone**, nothing'³⁴), or if it precedes a verb (such as in *Só que ele tem doze anos. Aí, tipo, discute, briga, depois a gente pára* 'But he is twelve. So, **like**, [we] **argue, fight**, then we stop'³⁵).

To avoid ambiguity and the possibility of inconsistent classification, the analysis focused on the linguistic element that *immediately* preceded and followed *tipo* with the exception of *preceding a sentence* and *preceding a clause*, i.e., sentence-initial and clause-initial positions.

The data for the analysis consist of 1,496 examples which contained the word *tipo* spontaneously produced by participants. For more information on the data and on the participants, please, see Section 3.2 and 3.3.1, respectively.

3.5.1 Results

Results from the analysis of the pre-*tipo* grammatical categories failed to provide a consistent identifiable pattern as there was a considerable dispersion amidst different categories, with the highest frequency and pattern being that of \emptyset ___ (preceded by no linguistic element, i.e., in sentence-initial position). Results from the analysis of the grammatical categories that follow *tipo*, however, show that the non-nominal form is most often found in pre-sentential position (___S, 64.5% of the tokens), pre-clausal

³⁴ From participant RDJ_162_F_14 (PSCB).

³⁵ From participant RDJ_107_F_12 (PSCB)

position (___Clause, 13.20%), and preceding noun phrases (___NP³⁶, 11.74%) (see Figure 3.6)³⁷.

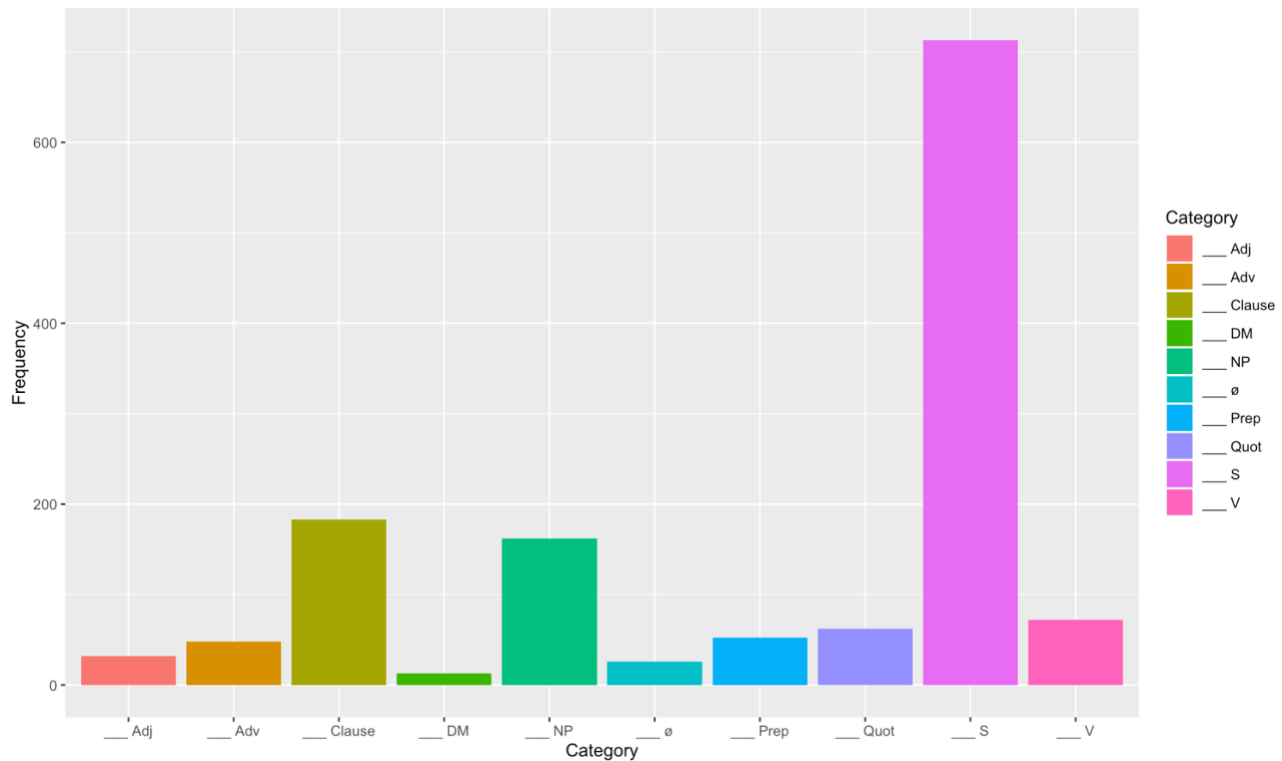


Figure 3.6. Frequency of grammatical categories following *tipo*³⁸.

³⁶ For the purpose of this analysis, NPs were classified according to the parameters established by Tagliamonte (2005) plus the inclusion of personal pronouns, which, unlike the aforementioned study, were found in the data analyzed in this study.

³⁷ The most frequent grammatical categories that follow the non-nominal forms of *tipo* were classified as follows: ___ S, preceding a sentence; ___ Clause, pre-clausal; ___ NP, preceding a noun phrase; ___ DM, preceding a discourse marker; ___ V, pre-verbal; ___ Quot, preceding a direct quote; ___ Prep, preceding a preposition; ___ Adv, pre-adverbial; ___ Adj, preceding an adjective; ___ ø, sentence-final.

³⁸ In regard to PrepP, the non-nominal form was distinguished from the canonical form based on the following characteristics: (a) inability to be inflected; (b) optionality (its absence would not deem the sentence ungrammatical); (c) being followed by either *de* ('of') as well as other prepositions (such as *para* 'to', *a* 'to', and *com* 'with') but not forming a prosodic unit with the following PrepP.

Results suggest systematicity in the use of *tipo*, corroborating findings from previous studies on linguistic constraints of grammaticalized items (D'Arcy, 2005; Tagliamonte, 2005). The results from this initial exploratory analysis not only offer insight into the structural constraints in which *tipo* is used in the *Carioca* dialect but also provide the foundation for the subsequent functional analysis, detailed in Section 3.5.

3.6 Functional Analysis

To better understand the grammatical and discourse-pragmatic functions that *tipo* is performing in the speech of teenagers in Rio de Janeiro, an investigation based on the functional analyses previously conducted by Bittencourt on the Brazilian noun *tipo* (1999), Voghera on the Italian noun *tipo* (2013), D'Arcy on the English word *like* (2017), and Rosenkvist and Skärlund (2013) on the Swedish *typ* was conducted. The end goal was to gain an insight into current functionally distinct uses of *tipo* in the vernacular of *Carioca* teenage speakers. Results from this analysis may lead to a better understanding of both its pragmatically motivated functions and its non-discursive functions (such as a noun, a conjunction, or a preposition).

To explore the possible functional polysemy of *tipo*, a qualitative analysis (Bittencourt, 1999; D'Arcy, 2017; Lima-Hernandes, 2005; Voghera, 2013) was conducted. Of the initial 1,496 examples that contained non-nominal forms of *tipo*, 63 were excluded from the analysis due to uncertainty over the function it was performing. For

instance, in example 33, the speaker loses her train of thought and does not conclude her sentence. In cases such as this one, the classification was deemed uncertain and the extract removed from the analysis. In total, 1,433 examples were analyzed. Relative frequencies of the different functions being performed by *tipo* were also analyzed.

- (33) *Eu gritei tipo uma.. eu esqueci até o que que*
 I screamed like one I forgot even the what that

eu tinha gritado
 I had screamed
 'I screamed **like**, one, I even forgot what I had screamed about' (RDJ_117_F_13)

Both quantitative and qualitative analyses of the 1,433 utterances that contained *tipo* were conducted. Qualitatively, the functional analysis was conducted following previous frameworks adopted for the examination of grammaticalizing words (D'Arcy, 2017; Lima-Hernandes, 2005; Rosenkvist and Skärlund, 2013; Voghera, 2013) and, in agreement with the description of the processes it entails, considered the functional changes that grammaticalizing elements undergo as following a unidirectional path (see Section 3.2). The functional analysis involved the examination of the following dimensions: *meaning* (e.g., meaning 'type,' or 'approximately,' or 'as'); *syntactic role*; and

discourse role. Functions were classified in terms of the part of the speech³⁹ (including discourse and pragmatic functions). For instance, when classifying *tipo* as performing the function of a preposition, the analysis included the examination of the syntactic construction (in this case, the presence of a nominal element functioning as its complement).

In regard to *pragmatic* and *discourse functions*, it is important to clarify that the position taken in this study is one that sees discourse and pragmatics playing a role on how grammatical categories are used by speakers, thus making their inclusion necessary for a proper classification to be conducted (for more on this topic, see Voghera, 2013). This framework has also been shown to be useful in the identification of the most frequent usages in the data (Voghera, 2013).

Quantitatively, the analysis focused on the frequency of the identified functions and their distribution by sex, as, impressionistically, female participants seemed to use *tipo* much more frequently than their male counterparts.

3.6.1 Functional Analysis - Results

Results from the analysis show that *tipo* is performing several functions in oral discourse, including its canonical function as a noun. The co-existence of canonical and grammatical and pragmatic functions is yet another element that supports the

³⁹ This model has also been shown to reduce overclassification of functions (Voghera, 2013).

argument that the lexical item may be undergoing a process of grammaticalization as posited by Bittencourt (1999) (see Section 3.2).

3.6.1.1 Noun.

As discussed in previous sections, *tipo* is a masculine noun⁴⁰ in Portuguese, meaning ‘kind,’ ‘type,’ ‘class,’ ‘model,’ and, as a noun, it can be inflected for number and is often found in NP1 of NP2 (NP1 de NP2) constructions (*tipo + de +NP2*). Its nominal function, the most frequent in the 1970s and 1990s, was one of the least frequent in the PSCB data, corresponding to just over 2% of the occurrences (see Figure 3.4). Examples 34, 35, and 36 illustrate its nominal function.

(34) *Ah vários tipos*
 Oh many **types**
 ‘Oh, several **types**.’ (RDJ_115_F_12)

(35) *A gente gosta praticamente do mesmo tipo de música*
 we like practically of the same **type** of music
 ‘We basically like the same **kind of** music.’ (RDJ_126_F_13)

(36) *Eu escuto todo tipo de música mas tem aquelas*
 I hear every **kind of** music but have those
músicas que eu não gosto

⁴⁰ Although examples of *tipo* performing the function of referential noun (with the meaning of ‘person’) have been identified in the literature on the grammaticalization of *tipo* in BP (see Lima-Hernandes, 2005), this function was not found in the current data.

songs that I not like
 'I listen to all **kinds of** music, but there are those songs that I do not like.'
 (RDJ_279_M_12)

As the examples above show, *tipo* as a noun can be inflected for number (34) and modified by a quantifier (35, 36).

3.6.1.2 *Preposition.*

The preposition function of *tipo* is one of the three most recurrent functions in the dataset. As a preposition, *tipo* indicates similarity and comparison and can introduce an exemplification what has been previously said. As a preposition, *tipo* is followed by a nominal element that functions as its complement. When *tipo* performs this function, it can be paraphrased as 'such as' or 'for example' as shown in examples 37, 38, and 39.

(37) *Eu nunca quebrei nada tipo nenhum osso nada*
 I never broke anything **like** none bone nothing
 'I have never broken anything **like** no bone, nothing.' (RDJ_162_F_14)

(38) *É tipo um interrogatório*
 is **like** an interrogatory
 'It's **like** an interrogatory.' (RDJ_207_F_12)

(39) *Eu acho que é tipo uma explicação*
 I think that is **like** an explanation
 'I think that it is **like** an explanation.' (RDJ_209_M_15)

3.6.1.3 Conjunction.

When performing the function of a conjunction, *tipo* connects two clauses: a main clause and a comparative clause that is explanatory or illustrative. As a conjunction, *tipo* can be replaced by the word *como* 'as'. In example 40, the participant is talking about a movie and uses *tipo* to introduce a clause that explains what made it popular among teenagers.

- (40) *É tipo uma realidade que existe no mundo*
Is like a reality that exists in the world

'It is like a reality that exists in the real world.' (RDJ_203_F_11)

In the example below, the speaker provides an explanation of what she means by 'arguing' with her friends as a way to minimize possible misunderstandings. The explanatory clause is introduced by *tipo*.

- (41) *Mas tipo é uma coisa tipo briga e*
but like is one thing like argue and
volta a se falar
return to each other speak

'But, like, it is something **like** arguing and going back to talking to each other.'
(RDJ_154_F_14)

3.6.1.4 Adverb.

As an adverb, *tipo* has an approximative meaning and precedes a quantifying phrase (D'Arcy 2005; Voghera, 2014). For instance, in example (42) the speaker discusses the number of times the students practice sports, and she uses *tipo* preceding twice and once a week, thus indicating an approximation. This meaning is also found in example (43), when participant 115 states that she would spend a long time with her brother - approximately one hour.

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

'In high school we will train **like** twice a week, once a week.' (RDJ_264_F_14)

- (43) *É, aí eu ficava tipo uma hora sentada com ele*
is then I stayed **like** one hour sitting with he

'Yes, then I would spend, **like**, an hour sitting with him.' (RDJ_115_F_12)

3.6.1.5 Sentence-final adverb.

One unexpected finding, which has not yet been mentioned in the literature, was that of *tipo* performing the function of sentence-final adverb, which could be translated as 'so

to speak.’ Although not previously found in BP, this function has been shown to be present in the English use of *like* (D’Arcy, 2005, 2017). Of the functions identified in the data, sentence-final adverb is one of the least frequent, with less than five examples being found.

Examples found in the data (such as 44 and 45 below) seem to suggest that this ‘so-to-speak’ function may be used to downplay what was previously said by the speaker.

(44) *Não tipo... eu moro não escondida, dá pra saber, tipo*
 no like I live not hidden can to know **like**
 ‘No, like... I live not in hiding, it is easy to know [where I live], **like**.’

(RDJ_239_F_14)

(45) *É, ela é muito competitiva, tipo*
 Yes she is very competititve **like**
 ‘She is very competitive, **like**.’ (RDJ_276_F_13)

3.6.1.6 Quotative complementizer.

The use of *tipo* as a marker of reported speech has, to date, been rarely mentioned in the literature (Laurentino, 2016) while the quotative function (Fleischman & Yaguello 2004; Romaine & Lange 1991) has been identified in studies of language change and grammaticalization such as in the case of the English *like* and Italian *tipo* (Blyth, Recktenwald, & Wang, 1990 and Voghera, 2013).

Results from the current analysis show frequent examples of *tipo* as a quotative complementizer. In the occurrences identified, *tipo* often introduces a direct quote that reflects an approximation of what was uttered rather than a reproduction of the utterance verbatim. Unlike the findings of the analysis of *tipo* introducing direct speech in Italian, which shows that *tipo* often introduces utterances containing common knowledge, maxims, and commonplaces (Voghera, 2013), examples from the PSCB corpus indicate that *tipo* functions as a signal of a close approximation of what was said.

For instance, in example (46), the participant describes how her friend reacted to her behaviour on her birthday, and she describes that her friend, being puzzled, felt like asking what she was doing and the reason for her odd behaviour. This, however, seemed to be her own view or thoughts of what her friend said rather than her friend's exact words.

- (46) *No meu aniversário foi engraçado porque eu fiquei muito*
on my birthday was funny because I stayed very
inquieta, e ela ficou tipo, "O que você tá fazendo?"
uneasy and she **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?"' (RDJ_215_F_13)

A similar example is presented in (47) when another participant discusses how she reacted to her mom yelling at her. The repetition of the word *mãe* several times suggests that this is an approximation of what she said rather than an exact account of the words uttered during the exchange with her mother.

- (47) *Ela entrou e começou a gritar, eu fiquei tipo,*
 she entered and started to yell I **stayed like**
"Mãe, pára, mãe, mãe, pára!" e começou a falar
 mom stop mom mom stop and started to speak
tudo que eu faço em casa, tipo, "Ela não
 all that I do in house **like** she not
lava a louça!" Eu ficava vermelha. Eu lavo, tá?
 wash the 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?’ (RDJ_340_F_13)

In some examples, the segments of discourse which *tipo* introduce seem to report the speaker’s thoughts or inner monologue rather than being a direct reproduction of what was said, such as in example (48).

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

'But I was **like**, "My sister and I are not the same person." Why would you want to write the same text?' (RDJ_226_F_13)

The identification of several constructions in which *tipo* introduces direct speech led to a separate analysis which aimed to examine the structure of these constructions as well as their overall frequency in the data. Results are presented in Figure 3.7.

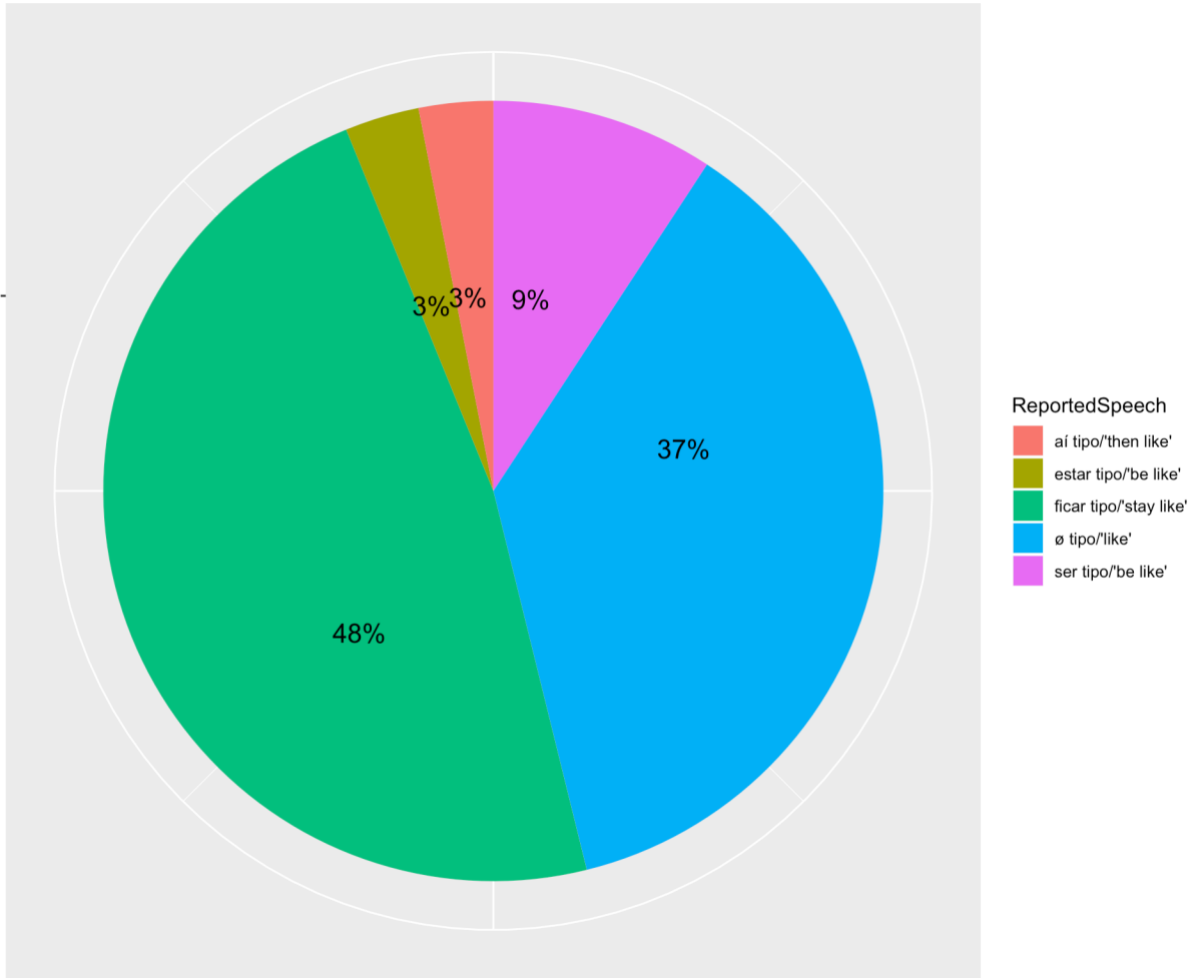


Figure 3.7. Distribution of *tipo* in reported speech constructions.

Results show that *tipo* is most frequently either preceded by the verb *ficar* ('stay') in the simple past tense or used in isolation ($\emptyset + \textit{tipo}$) when introducing direct speech.

The examples below illustrate these constructions:

- *aí + tipo*, 'then like'

(49) *Aí tipo, "Ai meu Deus do céu!"*
 Then like Oh my God of heaven

'Then, like, "Oh, my God!" (RDJ_274_F_11)

- *estar + tipo*, 'stay/be like'

(50) *Pra ela, ela tá falando normal, mas ela tá tipo,*
 To her, she is speaking normally but she **stays like**

"EI! AQUELA MENINA, EU NÃO GOSTO DELA!"
 hey that girl, I not like her

'To her, she is speaking normally, but she **is like** [yelling], "HEY! THAT GIRL, I DO NOT LIKE HER!" (RDJ_264_F_14)

- *ficar + tipo*, 'stay like'

(51) *No meu aniversário foi engraçado porque eu fiquei muito inquieta*
 On my birthday was funny because I stayed very restless

e ela ficou tipo, "O que você tá fazendo?"
 and she **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?" (RDJ_215_F_13)

- \emptyset + *tipo*, 'like'

(52) *Ela chega a me botar de castigo tipo, "Ah, não*
 She gets to me put in punishment **like** Ah no

eu não posso mais sair de casa"
 I not can more leave the house

'She even gets to ground me like, "Oh, no, I cannot leave the house anymore.'

(RDJ_116_M_14)

- *ser + tipo*, 'be like'

(53) É **tipo**, "Dorme lá em casa?"
 is like sleep there at home

'It is like, "Sleep over at my place?'" (RDJ_262_F_14)

Regarding distribution by sex, the use of *tipo* introducing a direct quote was more commonly found in the speech of female participants when compared to that of male participants, corresponding to 96.83% of occurrences, indicating that the use of *tipo* as a quotative complementizer may still be relatively recent.

3.6.1.7 Discourse Marker⁴¹.

Of the functions *tipo* is performing, that of a discourse marker⁴² is unquestionably the most frequent, making a more detailed discussion of this function necessary.

Research on linguistic innovation and grammaticalization (see 2.3) has brought significant attention to discourse markers (DMs). Traditionally referred to as colloquialisms (Salmon, 1975) "linguistic Cinderellas" ("familiar, drab, hard-worked, and lacking in morphological, phonological, and etymological glamour") (qtd. n Brinton, 1996, p.1), and "disfluent" phenomena (Crible, 2018), they have also been

⁴¹ Here, discourse markers refer to both markers and particles (D'Arcy, 2005; Brinton, 1996) as elements that guide the interpretation of the utterances without affecting their propositional meaning (see Voghera, 2013).

⁴² A distinction between the use of *TIPO* as a discourse marker and as a discourse particle will be discussed later in this section.

considered one of the elements used by young speakers that could result in irreversible damage to (as reported in Bittencourt, 1999).

The multifunctionality and lack of propositional meaning of DMs gained considerable attention of scholars after Schiffrin's (1987) seminal work on eleven DMs in English (*and, but, because, so, now, or, then, oh, well, y'know, and I mean*). DMs are ubiquitous in spoken language (Tagliamonte, 2005) and have been studied extensively. Also called *pragmatic markers* (Brinton, 1996), *discourse particles* (Kroon, 1995; Schoroup, 1985), *pragmatic particles* (Ostman, 1983), *pragmatic connectives* (van Dijk, 1979) and *discourse connectives* (Blakemore, 1989) among others, DMs have also become of interest to researchers due to their essential role in interpersonal communication.

Although terminologies have abounded, a consensus about which term to be used has yet to be reached due to their polysemic and multifunctional nature. The definition has been considered both complex and classification controversial, resulting in different criteria being proposed and different linguistic approaches being used by scholars to investigate these elements. For instance, Schiffrin (1987) refers to these elements as *discourse markers* and stresses their contribution to the maintenance of coherence within the multilayered interaction of which conversation is comprised, i.e., they provide valuable information on how units of talk that precede and follow them should be interpreted together. Fraser (1996), on the other hand, has argued that the term *pragmatic marker* comprises both discourse and pragmatic markers as the first is

seen as a subtype of the latter (with *pragmatic markers* having specific procedural meanings that could be inferred through context). Van Dijk (1979) posits that there are differences between semantic and pragmatic connectives while Brinton (1996) uses the term *pragmatic markers* to stress how these features guide interpretation rather than having a propositional meaning (Beeching, 2016).

The terminological diversity reflects both the wide range of linguistic approaches that have been employed for their studies and the multiplicity of functions which these elements are said to fulfill. These include discourse connectors, turn-takers, confirmation-seekers, intimacy signals, topic-switchers, hesitation markers, boundary markers, fillers, prompters, repair markers, attitude markers, and hedging devices (Beeching, 2016; Brinton, 1996; D'Arcy, 2006; Fraser, 1996; Schiffrin, 1987). DMs have been described as expressions with no or little propositional meaning (Brinton, 1996; Schiffrin, 1987) that play both textual and interpersonal roles (Beeching, 2016; Brinton, 1996). Some DMs have been shown to be sociolinguistically marked (Tagliamonte, 2005, 2016), and researchers have not only investigated their use and diffusion but also their development, from their canonical meaning to their somewhat ambiguous and polysemic forms as DMs (Beeching, 2016).

Despite the negative perception and stigmatization of these items by some (de Sá, 2017; Sut, 2006), research has shown that DMs perform discourse-pragmatic functions. DMs have been described as “sequentially dependent elements which

bracket units of talk" (Schiffrin, 1987, p.31) that express relation or have structural functions. DMs are a diverse class of elements that have been shown to function as "clue words" to discourse structure (Redeker, 1990), provide significant contribution to how discourse is structured, and also to constrain the interpretation of a proposition and its relevance within the discourse structure (Blakemore, 1987).

Discourse markers provide listeners with information on how an utterance relates to the previous utterance(s) and/or the discourse (Fraser, 1988) and help speakers hold the floor (Brown, 1977), among other functions (Pichler, 2013). Their presence is 'optional' in the sense that removing them from an utterance does not deem the utterance ungrammatical nor does it deeply affect its propositional meaning; however, they do perform pragmatic functions by signaling how the upcoming utterance should be interpreted within the context of the prior discourse (Fraser, 2009). Consider, for example, examples (54) in which the same utterance is either introduced by a DM or not, and example (55), in which different DMs suggest different implications for the answers (Schourup, 1998).

(54) "I opened the package".
"What was in it?" / "**Well**, what was in it?"

(55) "Are you happy?"
"**Oh**, yes." / "**Well**, yes." (Schourup, 1998)

Despite the diversity of elements that have been included in the category of DMs, Brinton⁴³ (1996) provides a description of what could be considered some of the characteristics/features of prototypical DMs. DMs are:

- more frequently found in spoken rather than written discourse (see also Schiffrin, 1987; Tagliamonte, 2005, 2016; Watts, 1989);
- multifunctional (see also Schiffrin, 1987; Tagliamonte, 2016);
- pervasive in spoken discourse (see also D'Arcy, 2005);
- stigmatized;
- frequently phonologically reduced (see also Schiffrin, 1987);
- not easily placed in traditional word classes;
- part of a separate tone group;
- optional (their absence does not make a sentence ungrammatical) (see also Fraser, 1988 and Schiffrin, 1987);
- mainly found sentence-initially (see also D'Arcy, 2005; Fraser, 1990);
- elements with little or no propositional meaning (see also Schiffrin, 1987);
- derived from an array of grammatical categories and, thus, marginal in nature;

⁴³ Brinton (1996) and Fraser (1996) make a distinction between DMs and pragmatic markers. The distinction will be discussed in the upcoming paragraphs.

- found “outside the syntactic structure of loosely attached to it and hence have no clear grammatical function” (Brinton, 1996, p.34).

Researchers have also suggested that DMs are more frequently produced by women, being perceived by some as gender-specific (Brinton, 1996, p. 35; Tagliamonte, 2016, p. 26). Although questions have been raised regarding this argument, findings have indicated that some DMs are more commonly found in the speech of female speakers (Brinton, 1996; Tagliamonte, 2016).

Despite the difficulty in classifying DMs due to their multifunctionality, scholars have proposed that DMs can be categorized into types/ classes. Fraser (2009) states that there are four types of pragmatic markers namely:

- a. basic pragmatic markers, which indicate the illocutionary force the speaker is conveying;
- b. commentary pragmatic markers, which indicate comments. They are subdivided into:
 - i. assessment
 - ii. hearsay
 - iii. evidential
 - iv. (non)deference
 - v. manner-of-speaking

- c. discourse markers, which signal the relationship between an utterance and the previous utterance(s). They are subdivided into:
 - i. inferential (e.g., so, hence, consequently)
 - ii. contrastive (e.g., alternatively, but, even so)
 - iii. elaborative (e.g., and, alternatively, on top of it all)
- d. discourse structure markers, which indicate elements of the organizational structure of the discourse. They are subdivided into:
 - i. attention
 - ii. topic
 - iii. discourse management

Regarding the pragmatic functions performed by DMs, Brinton (1996) argues that the two main functions are:

1. Textual functions – structuring meaning as a text
 - Initiating and closing discourse
 - Marking topic shifts
 - Indicating new and old information
 - Constraining the relevance of adjoining utterances
 - Turn-taking (organizing exchange as a cohesive discourse)
 - Repair-making (performing an important role in achieving coherence in oral discourse)

2. Interpersonal function – “the expression of the speaker’s attitudes, evaluations, judgments, expectations, and demands, as well as of the nature of the social exchange, the role of the speaker and the role assigned to the hearer” (Brinton, 1996, p. 38) (see also Schiffrin; Schourup 1983)

From the perspective of discourse as a structure (Schiffrin, 1987), i.e., of how it is composed of linguistic constituents, the term discourse marker is frequently used to refer to a *possibly* syntactically unconstrained class of elements that are distinguished by both their encoded meaning and their function in discourse. DMs are considered functional elements that belong to an array of syntactic categories by some researchers (see e.g., Schourup, 1999) and optional for others (see e.g., Fraser, 1988). For instance, Fraser (1988) argues that DMs are syntactically optional as their removal does not affect the grammaticality of the utterance in which they are found; they would also be optional as they do not affect the semantic relationship of the other elements that comprise the utterance (Brinton, 1996).

The difficulty to syntactically define DMs has led some researchers to claim that these elements are syntactically free as they can *seemingly* occur anywhere within an utterance. Moreover, syntactic definition of these elements has been considered difficult (Schiffrin, 1987). On the one hand, this argument raises questions about the ability of studies to identify rules that may govern the distribution of DMs; on the other hand, the

argument seems to fail to take into account the inherent systematicity of languages (see D'Arcy, 2005). If DMs could surface anywhere in a sentence, that would mean that these elements would fail to follow the structure of languages, which are rule-governed (D'Arcy, 2005). Due to the unlikeliness of DMs being completely unconstrained, scholars have investigated whether they could be rule bound. Results have revealed that DMs are not syntactically free (D'Arcy, 2005; Tagliamonte, 2005). For instance, Underhill (1988) argues that the English DM *like* often introduces a constituent while Wolgemuth's (2003) research shows that *like* in Canadian English is mostly found before noun phrases.

Following the classification proposed by Fraser (1988) and Brinton (1996) (see also D'Arcy, 2005 and Tagliamonte, 2016), this paper will refer to *discourse marker* as the discourse-pragmatic element at the left periphery of the sentence (i.e., the one at the leftmost position which marks the relationship of the proposition with the previous discourse) while *discourse particles* will be used to refer to the element that performs discourse-pragmatic functions in other positions within the utterance. This classification follows the description of prototypical features of DMs as discussed by Brinton (2016) and Fraser (1990) as described above.

Results from the functional analysis show that the most frequent function *tipo* performs in the speech of teenagers is that of clause-external discourse marker followed by that of the clause-internal discourse particle.

Discourse Marker – Tipo as a discourse marker is the most frequent function found in the PSCB corpus. Comprising over 60% (N=874) of the total examples in the PSCB corpus, *tipo* is often found sentence-initially. Performing this function, it guides the interpretation of the upcoming utterance by the hearer within the context of the foregoing discourse. In example (56), *tipo* precedes some extra information that is being shared by the speaker so as to clarify what was originally said, thus providing relationship between consecutive utterances.

(56) *Tipo, eles brigam aí uma semana depois esqueceram*
like they argue then one week after forgot

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

(RDJ_164_F_15)

As a DM, *tipo* is found in utterance-initial position and is shown to constrain the relevance of the upcoming utterance. It is also found to precede statements that are intended to illustrate what was mentioned in the previous utterance. Despite the fact that this function is quite recurrent, example (57) shows a speaker who uses *tipo* with that finality while also using *por exemplo* ('for example') as an apparent reinforcement of how the listener is to interpret the upcoming utterance.

(57) *Aí eles tipo, e as brigas foram tipo muito rápida*
 then they like and the fights were like very fast
porque todo mundo chega e separa ... E a maioria
 because all world arrives and separates ... and the majority
tipo, por exemplo a Paula e a garota. Elas voltaram*
like for example the Paula and the girl they returned
a se falar
 to each other talk

'Then, they, like, and the fights were like really short-lived because everyone gets together to separate them... and the majority... **like**, for example, Paula and the girl. They got back to speaking to each other.'

(RDJ_264_F_14)

Recurrently *tipo* is used in sentence-initial position preceding an example or illustration of that refers to a topic under discussion. *Tipo* was also shown to be used sentence-initially when the speaker is attempting to take the floor and to indicate the beginning of a speaker's turn. In example 58, the first speaker (Ana*) starts explaining that she and her sister were given an award by the school just like Julia* was. However, as Ana tries to explain why she and her sister did not get to make use of the award, Julia, who is her best friend, takes the floor and proceeds to tell the full story.

(58) Ana: **Tipo**, *eu passei também, mas eu não fui.*
Like I passed too but I not went

'Like, I was also selected, but I did not go.'

Júlia: *Tipo assim elas passaram, só que ... explicar do começo*
Like they passed only that ... explain from the start
 "Like, they were selected, only that... Explaining from the start."

3.6.1.8 Discourse Particle.

As a discourse particle, *tipo* is shown to perform discourse-pragmatic functions in positions other than sentence-initial. As a DP, *tipo* performs interpersonal functions (such as establishing solidarity or common ground) (D'Arcy, 2017) in addition to introducing subjective information regarding the topic under discussion. For instance, in example 59, *tipo* seems to highlight the upcoming information (Underhill, 1988), indicating that the focus of the utterance is to come.

(59) *E meu pai tipo já chega do trabalho tarde,*
 and my dad **like** already arrives from work late
ai quando eu acordo meu pai já saiu, então
 then when I wake up my dad already left so
eu quase não vejo ele
 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.' (RDJ_262_F_14)

In example 60, however, the DP seems to be signalling equivalence between *falando assim* ('speaking this way') and the description of what 'this way' entails.

(60) *Falando assim tipo como se a gente não*
speaking this way like as if we not

tivesse feito nada do trabalho e eles dois tivessem
have done nothing of work and they two had

feito tivessem colocado nosso nome
done had put our name

'Speaking this way **like**, as if we had not done any of the work and they had both added our name.' (RDJ_307_F_13)

The use of *tipo* as a discourse particle has been also found in the media and even written text such as in comic strips. Figures 3.8 and 3.9 illustrate how some perceive *tipo* as seemingly being used anywhere in the sentence⁴⁴.



Figure 3.8. 'The terrible acute tiponitis'.

'She, like, walked into the bar and I, like, looked at her. Like, she was, like, cool and, like, we started, like a, like, conversation, like, cool, like like... like like like like like like,' (Retrieved from <https://adaoiturrgarai.wordpress.com>)⁴⁵

⁴⁴ Including utterance initial (functioning as a DM)

⁴⁵ I would like to sincerely thank Adão Iturrgarai and Maurício de Sousa Produções for their support and for giving permission for the use of their work in this dissertation.



Figure 3.9. 'Acute tiponitis'

'A wine of, like, powerful tannins with, like, notes of cinnamon and, like, pepper!' 'Heidegger, like, did not have, like, any influence on, like, existentialism!' 'There isn't, like, in comic strips, nothing, like, before Popeye and, like, nothing after, like, Popeye.' (Retrieved from <https://adaoiturruugarai.wordpress.com/>)

While at times the pervasive word is mocked as in the comic strips above, at other times it is shown as an example of one of the salient linguistic elements in the speech of preteens and teenagers (Figure 3.10).



Figure 3.10. 'So?' 'What's up?' 'Like!' 'Cool!'⁴⁶ (De Souza, 2004)

⁴⁶ The *Turma da Mônica* has been the most popular comic book in Brazil since the 1960s. It remains one of the most read comic books in the country as it depicts relatable stories about a tight group of children and preteens. This image was part of a series of stories about the older boys in the group (who are between 10 and 11) believe they have transitioned into adolescence and, therefore, should act like teenagers.

The functions identified in the analysis not only support the findings from prior studies on the multifunctionality of *tipo* but also provide important information about its functions as a *sentence-final adverb* and *quotative complementizer* which, to date, had not been identified in studies which examined *tipo* in the *Carioca* dialect.

3.7 Functions by Gender

One of the remaining questions regarding the different functions that *tipo* is performing in the speech of teenage *Cariocas* referred to the possible interaction between the functions found in the data and gender. This question results from the review of the literature that shows that young women lead in the use of innovative forms in oral discourse (Tagliamonte, 2016; Tagliamonte & D'Arcy, 2009). To examine the possible relationship between the different functions of *tipo* and *gender*, a Chi-square test⁴⁷ was performed.

Results show a significant relationship between the functions that *tipo* is performing and *gender* ($X^2(6, N = 1433) = 44.60, p < .001$), with female speakers being more likely to use the non-nominal forms of *tipo* (e.g., as a quotative complementizer or a DM) when compared to male speakers. These findings are similar to that of previous

⁴⁷ The Chi-square test, used to test the relationships of categorical data, tests the likelihood of the observed distribution being due to chance.

studies which have shown that young female speakers are often in the forefront of language innovation and change (Tagliamonte, 2004, 2016; Tagliamonte & D'Arcy, 2009).

3.8 Additional Findings

The functional analysis also yielded interesting findings beyond those that were initially identified of being of potential interest. The following subsections present a synopsis of these findings, including clear examples that support the argument of systematicity in the use of *tipo* in utterances.

3.8.1 Co-variants

Findings show that three *tipo* constructions are regularly used by the participants. Of the three types, two are periphrastic, i.e., consist of multi-word expressions which replace a single word (Haspelmath, 2000). Results from an analysis of construction and pragmatic and grammatical functions suggest that the co-variants are interchangeable and are found in similar syntactic constraints in the tokens analyzed. With the exception of sentence-final adverb, which is restricted to *tipo* + \emptyset , all forms were shown to perform the same grammatical and pragmatic functions.

- a. **Tipo** – in the most frequent of the three constructions, *tipo* is found performing all the functions identified in the functional analysis (i.e., a noun, preposition,

conjunction, adverb, sentence-final adverb, quotative complementizer, discourse marker, or discourse particle). In this case, *tipo* is not part of a periphrastic construction (as illustrated in example 61).

- (61) *Aí tipo a pessoa tá assim do seu lado,*
 then like the person is this way on your side
e tá você escrevendo lá.
 and are you writing there
 ‘Then, **like**, the person is right by your side, and you are there, writing.’
 (RDJ_101_F_14)

- b. *Tipo Assim* (*tipo* + adv of manner *assim* ‘this way, like this’) (also see Bittencourt, 1999) - the periphrastic construction *tipo assim* was the second most frequent form found in the data. *Tipo assim* was shown to perform all functions performed by *tipo* (see example 62) with the exception of sentence-final adverb and quotative complementizer.

- (62) *Não, entre a gente não, mas tipo assim, uma pessoa*
 no between we no but like one person
vim de fora e vai ouvir um funk e
 comes from out and will listen to a funk and
aí vai começar com aquelas letra
 then will start with those lyrics
 ‘No, not among us, but, **like**, an outsider comes and listens to funk music, and then those lyrics start...’ (RDJ_305_F_15)

- c. *Tipo que* (*tipo* + particle *que* ‘that’, ‘what’) - of the three forms found in the data, *tipo que* was the only one not to have been previously identified by researchers, making this the first account of its use in vernacular Brazilian Portuguese.
- Although not as pervasive as *tipo* or *tipo assim*, *tipo que* was shown to be highly favoured by male speakers. *Tipo que* also performed most of the pragmatic and grammatical functions performed by *tipo* + \emptyset with the exception of sentence-final adverb and quotative complementizer. For instance, in (63), *tipo que* is functioning as a DM and, in (64), as a DP.

(63) *Tipo que eu tenho um quintal aqui, aí tipo que tem*
like I have a backyard here then **like** have
um vizinho aqui morando no meio do meu quintal!
 a neighbour here living in middle of my backyard
 ‘**Like**, I have a backyard here, then **like** there is a neighbour living in the middle of my backyard!’ (RDJ_281_M_13)

(64) *Meu irmãozinho gosta muito de tipo que tênis de rodinha*
 my little brother likes a lot of **like** sneakers of wheel
 ‘My little brother really likes **like** roller shoes.’

These findings are especially relevant to this study as *layering* (the existence of forms that express the same meaning) is one of the five principles of grammaticalization (Hopper, 1996). The presence of co-variants in the speech of *Carioca* teenagers is another factor that corroborates the argument that *tipo* is grammaticalizing.

3.8.2 *Tipo and Subject Doubling*

In addition to the findings previously discussed, the analysis also identified a recurrent structural pattern in the speech of participants. The phenomenon, subject doubling, has been identified in several languages such as Canadian French, Parisian French, Andean Spanish, Dutch, and, more recently, English (Auger, 1998; Zahler, 2014; Lipski, 2010; de Vogelaer & Neuchermans; 2002, Tagliamonte, 2019, respectively) among others. Subject doubling has been defined as a form of left dislocation or a pronominal appositive, leading to the subject being present twice in a clause (see Auger, 1998 and Tagliamonte, 2019 for more on these definitions). In most constructions, subject doubling consists of an initial NP that is followed by a pronoun that stands for the NP (65).

- (65) My parents they are cool. (overheard in a chat among a group of
NP Pron teenagers in British Columbia, Canada)

Although subject doubling has not yet been investigated in BP, numerous examples were found in the PSCB data, indicating that the construction is very productive in the *Carioca* dialect. Subject doubling is so recurrent in the data that

occurrences were tagged in Praat from the beginning of the transcription stage due to its pervasiveness. Examples 66, 67, and 68 illustrate this phenomenon.

(66) *O meu pai ele não me entende*
 the my dad he no me understands
 'My dad doesn't understand me.' (RDJ_105_M_15)

(67) *Eu e minha mãe a gente ficou desesperada.*
 I and my mom we stayed desperate
 'My mom and I we were desperate.' (RDJ_305_F_15)

(68) *A minha mãe ela me prendeu depois que eu cheguei aqui*
 the my mom she me keep after that I arrived here
 'My mom she has kept me in (not allowed me to go out) after I got here.'
 (RDJ_121_F_14)

More relevant to this study, findings showed the use of *tipo* in some subject-doubling structures. Although a very small number of NP + *tipo* + pronoun construction was found in in the data, most of the examples seem to suggest that *tipo* may be licensing pronoun doubling (rather than a pre-clausal NP + a pronoun that stands for the NP) in the subject position of sentences. Example 69 illustrates.

(69) *Ele tipo ele sempre vai me buscar, me levar*
 he like he always goes me pick up me take
 'He like he always picks me up and takes me.' (RDJ_115_F_12)

In fact, nearly all the examples of subject doubling constructions in which *tipo* was found consisted of those in which the subject is 1 PS (i.e., 'I'). Examples 70, 71, and 72 illustrate these structures.

(70) *Eu tipo eu nem posso namorar*
 I like I nor can date
 'I, like, I can't even date.' (RDJ_164_F_15)

(71) *Eu tipo eu tenho muito peso por causa do meu tamanho*
 I like I have much weight for reason of my size
 'I, like, I weigh a lot because of my size.' (RDJ_154_F_14)

(72) *Eu tipo eu nunca fiquei eu nunca fui muito de tirar nota baixa*
 I like I never stayed I never was very of take grade low
 'I like I was never sent to summer school, I was never one to get low grades.'
 (RDJ_115_F_12)

Although questions could be raised about the possibility of these structures being simple reformulations or restarts, the absence of hesitation markers, pauses, or partial production of words suggests otherwise. Additionally, the systematicity and frequency of the presence of *tipo* primarily with 1PS subjects suggest that speakers may be using *tipo* to allow for subject doubling in structures in which subject doubling might

not be originally allowed⁴⁸. Findings indicate that the use of *tipo* in such constructions is not ad hoc but rather reflects a productive structure in the *Carioca* dialect. However, since no previous research has investigated subject doubling or the aforementioned construction pattern using *tipo* in BP, further research is necessary to better understand the phenomenon.

3.9 Summary

This chapter aimed to provide an initial investigation of the current state of *tipo* in Rio de Janeiro. The several layers of analysis addressed the proposed research questions by examining frequency, systematicity, and function.

The initial findings offer an account of the frequency of use of *tipo* by teenage speakers, indicating that its non-nominal forms are now the most frequent in the conversational discourse of teenage speakers. More importantly, results also show a steady increase of non-nominal forms in the past 50 years. A subsequent analysis of linguistic predictions aimed to uncover patterns of use of *tipo* in utterances and examine possible systematicity in its use. Results show that the use of *tipo* is not ad hoc but rather follows a very regular pattern as to where it can be used in discourse. With the knowledge that its use is rule governed, an analysis of the grammatical and discourse-pragmatic functions it is performing was conducted. Results reveal eight functions,

⁴⁸ This is especially true when we consider that frequently these structures commonly consist of NP + pronoun.

which were classified in terms of the part of speech. Of those, quotative complementizer and sentence-final adverb were of special interest as the first has not been fully investigated in Brazilian Portuguese, and the latter had not yet been identified.

Finally, results also revealed *layering*, with three different forms of *tipo* being found in the dataset (*tipo*, *tipo assim*, and *tipo que*), and all three performing mostly the same functions, with the exception of being a noun. Of the three, *tipo que* is the only one which has not been previously identified in Brazilian Portuguese. In addition, the possible use of *tipo* to license the use of two pronouns in subject-doubling structures requires further analysis as this may indicate yet another functional expansion of the word.

Chapter 4 Production

Results from the analyses described in Chapter 3 corroborate previous findings suggesting that *tipo* is undergoing grammaticalization and performing grammatical and pragmatic functions beyond those of a noun. Results show an increase in generality (Bybee, 2003b) and use in contexts that have gone from more specific (a noun) to more general (such as a discourse marker).

During the transcription of the data, notes on the production of *tipo* raised questions regarding possibly interesting elements in how speakers were producing the word as it performs various functions in conversational speech. These observations were considered of relevance to this study as researchers have argued that grammaticalized linguistic items often undergo phonological reduction/erosion (Bybee, 2003b, 2003c; Heine, 2003). To address them, an analysis of the acoustic properties of *tipo* as it performs different functions was conducted. The goal was to examine the possible correlation between acoustic realization and the grammatical and discursive functions it performs. This innovative approach was designed to test acoustic correlates to support grammatical and pragmatic observations. Ultimately, results from this investigation may provide valuable information about the process of phonological erosion in grammaticalizing linguistic elements as postulated in the theory of grammaticalization.

This chapter is organized as follows: first, I present a discussion on phonological erosion within the theory of grammaticalization. Then, information on the data collection and the participants is discussed, followed by the description of the acoustic analysis conducted, and the results.

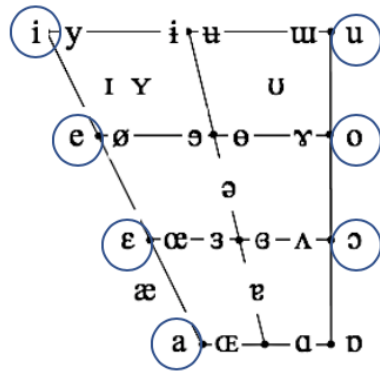
4.1 Phonological reduction/erosion

Research on speech production has revealed that the production of utterances varies considerably. Sound changes such as epenthesis (addition of a segment such as in *psicologia* /psikolo'ʒiv/ → /pɨsikolo'ʒiv/, 'psychology') and reduction (e.g., vowels in unstressed syllables moving to a more central position such as the reduction of the vowel [u] to a schwa [ə] in the American English word *campus* /'kæmpəs/) are often found in natural speech (Clark, Yallop, & Fletcher, 2008). Segmental features of speech (consonants and vowels) have been shown to be susceptible to reduction and even deletion (removal of segment, such as the deletion of schwa in English) (Hooper, 1978; Kohler, 1991).

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, a study by Silva (1997) shows that 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 /'nad/ (de Matos & Sandalo, 2006). In their study, de Matos and Sandalo (2006) argue that 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' /autori'dadzi/ being produced as /autori'dadz/ (de Matos & Sandalo, 2006). Results of the analysis of 336 lexical items indicate that elision of the final vowel was found in 220 of them (de Matos & Sandalo, 2006). In addition to word position, reduction (and eventual devoicing and elision) of vowels in unstressed syllables have been shown to be commonly found in most languages, including Brazilian Portuguese (Nobre & Ingemann, 1983). Nobre and Ingemann (1983) and Kenstowicz and Sandalo (2016) discuss the vowel reduction in unstressed syllables in BP, which goes from seven vowels (in stressed syllables) to three vowels (in word-final position) (see also Barbosa & Albano, 2004; Câmara Jr., 1972). Figure 4.1 provides a summary of the vowel system in BP in both environments.

Vowel System - Stressed Syllable



Vowel System - Word-final

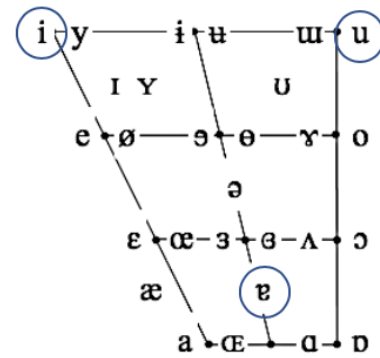


Figure 4.1. The vowel system of Brazilian Portuguese in stressed syllables and word-final position (vowels are circled).

Regarding word duration, previous research has suggested that duration might be influenced by parts of speech. In his study, Dilts (2013) shows the reliability of length of lexical items when compared to other parts of speech. The author postulates that it is possible to argue for a continuum of parts-of-speech from less to more susceptible to phonetic reduction (Dilts, 2013). 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, 2003a, 2003c). For instance, results of a study by Fowler and Housum (1987) show significant reduction of words that have been used before in the discourse. The researchers argue 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, 2003b, 2003c). 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. As Meillet (1912) postulates, this phonological erosion would reflect the fact that function words do not usually carry focus. Pagliuca (1994) concisely states that:

As a lexical construction enters and continues along a grammaticalization pathway, ... it undergoes successive changes ... broadly interpretable as ... a unidirectional movement away from its original specific and concrete reference and to increasingly abstract reference. Moreover, ... material progressing along a pathway tends to undergo increasing phonological reduction and to become increasingly morphologically dependent on host material (qtd. in Janda & Joseph, 2003, p. 79).

Loss of phonetic substance has been associated with the process of grammaticalization (such as in the case of *going to* --> *gonna*) (Heine, 2003, 2014; Kokorniak & Fabiszak, 2014; Meillet, 1912). As the item undergoes semantic bleaching and loses its canonical properties (*decategorialization*), speakers are able to use it in contexts that it was not (or could not be) previously used, thus making it possible for the word to acquire new syntactic and morphological functions (Hopper & Traugott, 2003). The linguistic item would then become more general and frequent, eventually becoming predictable. In their study, Podlubny, Geeraert, and Tucker (2015) show that

the production of *like* in Western Canadian English differs based on the lexical and grammatical functions it performs in speech. Their results from an analysis of 884 tokens show acoustic differences, with vowel duration being a predictor of the function the word *like* was performing.

Although researchers have postulated the universality of a co-evolutionary process of change in meaning (from less to more abstract) and form (from a full to a phonologically reduced form) which would take place during grammaticalization (Lehmann, 1995), some have disputed those claims. For instance, Bisang's (2004) study on languages spoken in East and Southeast Asia shows that the co-occurrence of these processes during grammaticalization is not necessarily present in those languages. Similarly, results from Schiering's (2010) cross-linguistic study led the author to propose that diachronic phonological reduction is more present in stress-based languages than in syllable-based languages. Although some researchers (see e.g., Schiering, 2010) subscribe to the idea that phonological reduction should not be included as one of the main processes of grammaticalization, empirical data has shown phonological reduction/erosion to be frequently found in grammaticalizing/grammaticalized linguistic items.

In line with previous research on grammaticalization (e.g., Bybee, 2003b; Heine, 2003), this study includes phonological reduction/erosion as one of the main processes that linguistic items undergo during grammaticalization cross-linguistically (Heine,

2003). By including phonological erosion as a common process of grammaticalization, it is possible to postulate that the process of phonological reduction of grammaticalized items (in tandem with the functional and morphosyntactic processes that have been shown to take place during grammaticalization) may also be present in Brazilian Portuguese. Having found no previous study that combines both functional and acoustic analysis of *tipo* in BP, this analysis will provide the first multilayered investigation of the acoustic realization of the functions performed by that word within the theory of grammaticalization. In addition, the examination of possible phonological erosion of the grammaticalized forms when compared to its nominal (canonical) one may provide insight into the process(es) this reduction entails.

Based on the results from the frequency and functional analyses conducted, which show a steady increase in frequency of the use *tipo* in conversational discourse, an acoustic analysis of tokens of *tipo* was conducted, which aimed to answer the following research question:

4. Are different functions of *tipo* acoustically distinct? If so:
 - a. Are there durational differences between tokens of nominal and non-nominal functions of *tipo*?
 - b. When comparing nominal to non-nominal functions of *tipo*, what are the differences (durational, quality) at the segmental level?

The next sections provide information on the data, participants, and analysis.

4.2 Participants

Participants were 76 middle school students (female, N = 50, male N = 26) between the ages of 11 and 14 (see Chapter 2) who took part in adapted Labovian sociolinguistic interviews (Labov, 1984) conducted between 2017 and 2018⁴⁹. All participants were monolingual native speakers of the *Carioca* dialect of Brazilian Portuguese born and raised in Rio de Janeiro.

4.3 Data Collection

Due to the background noise found in the audio files of the interviews collected in 2015 and 2016 (see Section 1.6), the data for this analysis were comprised of the audio recordings of 38 interviews conducted during three field trips between 2017 and 2018. A ZOOM H4n Pro Handy Recorder and two compatible SHURE WH20 cardioid headset microphones⁵⁰ were used (see Section 1.6.1) in addition to a ZOOM Q4 Handy Video Recorder, which recorded both audio and video data. The interviews were recorded in 16-bit Waveform Audio File Format (.wav) with a sample rate of 44.1 kHz. The audio files are comprised of two separate audio channels, one for each interviewee in the paired interviews. It is important to clarify that, although the data from 2017 and

⁴⁹ The few participants who were 11 at the time of data collection were close to turning 12 as all students in their grade must be at least 12 years before the middle of the school year.

⁵⁰ The choice of headsets was based on the fact that they are adjustable and non-intrusive, allowing participants to feel more comfortable during the interviews.

2018 were not collected under desirable laboratory conditions, all the recordings were of high quality. Audio files were loaded into Praat (Boersma & Weenink, 2017) (see Section 1.6.2), and the recordings of each participant was analyzed individually.

The sociolinguistic interviews conducted during 2017 and 2018 varied in length between 24 and 48 minutes. In total, 23 hours and 27 minutes of new interviews were recorded. As with the previous data collection sessions, interviews were scheduled taking into account students' availability and following the schedule suggested by the school administration and teaching staff. The procedures for the data collection were the same as detailed in Section 1.6.1.

4.4 Data

The data for this analysis correspond to all the 38 new interviews conducted between 2017 and 2018 and which are part of the PSCB corpus. As previously discussed (see Section 1.6.2), audio recordings of the interviews were transcribed and annotated in Praat (Boersma & Weenink, 2017) by the author, a speaker of the *Carioca* dialect.

Separate tiers were used for the transcription of each participant's speech following orthographic conventions. As previously stated, colloquialisms were kept to a minimum in the transcripts but were included in detail in the annotation tier.

Interjections, filled pauses (e.g.; uh, ah), and restarts were transcribed.

In addition to the two initial tiers created during the transcription of the audio files (see Section 1.6.2), three additional tiers were added prior to the commencement of the analysis. The final five tiers were comprised of the following information:

1. orthographic transcription of utterances;
2. mark-up of all the instances of *tipo* and inclusion of the function it is performing in the utterance;
3. segmentation⁵¹ of *tipo* (marking-up of the segmental boundaries was as follows: /t/, /i/, closure, /p/, and /o/, which is usually raised and realized as [u])⁵²;
4. mid-point measurement of vowels (measurement was done manually for each vowel marked in a point tier);
5. general observational notes from the transcriber (see Figure 4.2 below).



Figure 4.2. Example of some general notes indicating possible elements of interest.

⁵¹ Segmentation was adapted from the one discussed Drager's (2015) and Podlubny, Geeraert, and Tucker's (2015) analysis of *LIKE*.

⁵² Prior to the analysis, segmentations and measurements (4) were verified by a second phonetician.

To minimize possible identification errors, in this study deletion is marked in the data as not only a perceptual (auditory) inability to identify a segment but also the inability to identify it in a spectrogram and wave form.

To examine the phonetic realization of the word *tipo* as it performs lexical, grammatical, and discursive functions, 1,064 tokens were extracted from the 2017 and 2018 interviews of the PSCB corpus. A Praat script was created for the automatic extraction of measurements (segment duration, token duration, and F1, F2, and F3 formant measurements of vowels) as well as for the extraction of the information contained in each tier (see above). The use of the script aimed to reduce the chances of possible human error in the extraction of measurements.

Of the initial 1064 tokens, 44 tokens of *tipo* as an adverb and as a sentence-final adverb were excluded from the analysis due to the reduced number of tokens per function (e.g., sentence-final adverb, 6 tokens). In total, 1,020 tokens of *tipo* were analyzed. The tokens examined consist of *tipo* performing the six main functions identified in the functional analysis (see Chapter 3), namely noun (NOUN), preposition (PREP), discourse marker (DM), discourse particle (DP), conjunction (CONJ), and quotative complementizer (QUOT).

4.5 Exploratory Analysis

An exploratory acoustic analysis that aimed to investigate the phonetic realization of the functional categories instantiated by *tipo* was conducted in R (R Core Team, 2013).

To look for potential patterns of interest in the dataset, plots were generated using the *ggplot2* package (Wickham, 2016) in R prior to conducting statistical analysis. It is important to stress that visible differences do not necessarily translate into statistically significant differences; however, visually perceptible differences can help researchers identify potential areas of interest when conducting an exploratory analysis. Due to time constraints and the scope of this analysis, a decision was made to limit the investigation to the following analyses:

- token duration by function
- segment duration by function
- F1 and F2 formants of the high vowels [i] and [u]⁵³
- frequency of elision of the [i] and [u] vowels
- F1 formant of each vowel by function
- F2 formant of each vowel by function

Formant measurements and durational and functional data were plotted. Mixed effects models using *lme4* (Bates, Mächler, Bolker, & Walker, 2015) for R were used to examine the possible correlation between factors, including correlation between duration (of each segment and every token) and function as well as the correlation

⁵³ See Section 1.3.

between F1 and F2 measurements of the vowels and the function being performed by *tipo*.

4.6 Results

The following subsections present the results of several analyses conducted which aimed to examine the acoustic characteristics of *tipo*. For each individual analysis, boxplots and/or bagplots were created, and the *lmer* function from the *lme4* package was used to build mixed effects models. The *confint* function of the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) was used to estimate 95% confidence intervals. All models included *participant* as a random effect. Non-significant predictors were excluded from the models individually based on the closest t-value to zero.

Therefore, the results presented are from models that do not include non-significant predictors. The following predictors were considered when structuring the models: *age* (11 to 16, for a total 6 factors), *gender* (female/male), and *function* (the six most frequent functions identified in the data, namely conjunction, discourse marker, discourse particle, noun, preposition, quotative complementizer).

4.6.1 Full and Elided Vowels

To gain insight into the duration of the nominal and grammaticalized forms of *tipo*, an initial analysis that aimed to examine the duration of vowels was conducted. Of the total number of vowel segments (N = 2,040), 39.75% were elided (N = 811). Nearly half

of the /i/ segments (47.17%, N=481) and over 30% of the of the /o/ segments (32.35%, N=330) were elided⁵⁴ (see Figure 4.3).

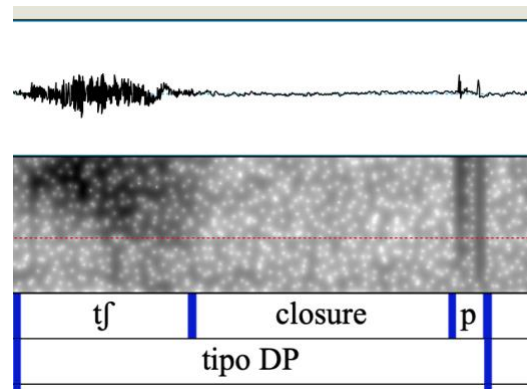


Figure 4.3. Elided /i/ and /o/ vowels in *tipo*.

Table 4.1 provides the percentage of elided and non-elided vowels in the six main functions of *tipo* identified in the data⁵⁵.

⁵⁴ This number includes the coalescence of the vowels /o/ and /a/ in the construction *tipo assim*, which no longer shows phrasal boundary (N = 173), suggesting the possible fusion into a single word.

⁵⁵ The analysis also focused on the possible devoicing of vowels; however, devoicing was only found in two tokens of *tipo*.

Table 4.1. *Elided and non-elided /i/ and /o/ vowels by function.*

<i>Function</i>	<i>Non- elided /i/</i>	<i>Percent non- elided /i/</i>	<i>Elided /i/</i>	<i>Percent elided /i/</i>	<i>Non- elided /o/</i>	<i>Percent non- elided /o/</i>	<i>Elided /o/</i>	<i>Percent elided /o/</i>	<i>Total</i>
NOUN	26	92.9%	2	7.1%	24	85.7%	4	14.3%	56
CONJ	20	29.9%	47	70.1%	44	65.7%	23	34.3%	134
DM	294	58.4%	209	41.6%	307	61.0%	196	39.0%	1006
DP	117	45.9%	138	54.1%	183	71.8%	72	28.2%	510
PREP	60	53.1%	53	46.9%	82	72.6%	31	27.4%	226
QUOT	22	40.7%	32	59.3%	50	92.6%	4	7.4%	108
Total	539	-	481	-	690	-	330	-	2040

Results indicate that elision of the /i/ vowel is quite frequent in most grammaticalized forms when compared to the nominal form (7.14%) while the vowel /o/ is not elided as frequently in any of the functions investigated. Based on the data extracted from Praat), /o/ is more commonly elided in the DM and CONJ functions of *tipo*. A *generalized linear mixed effects model* (Bates et al., 2015) was created to examine the likelihood of full /i/ and /o/ vowels (non-elided) being found in each function. Results were plotted for the visualization of the predictions (see Figure 4.4). The main goal was to calculate the average differences in log-odds between predictors, which indicate a

higher (positive log-odd results) or lower (negative log-odd results) probability of a full vowel being found in different functions.

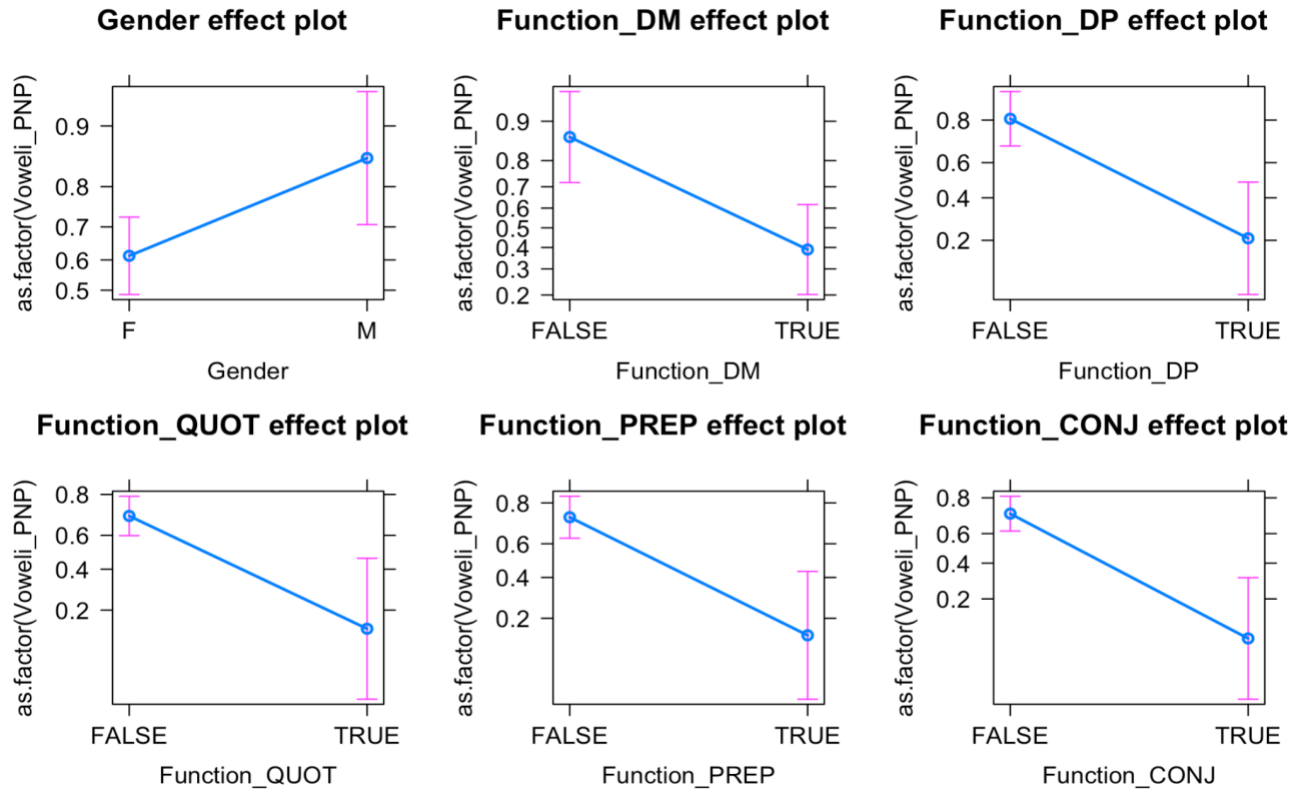


Figure 4.4. Visualization of predictions from the /i/ generalized mixed effects model.

For the /i/ vowel, the intercept (base value of 2.98) suggests that the probability of a full (non-elided) /i/ vowel being found in the NOM function is 95.17%⁵⁶. The probability increased, on average, by 1.31 log-odds (98.65%) if the speaker was *male*. The chances of a full /i/ vowel being found in other functions decreased substantially

⁵⁶ Log odds were converted to percentage using the following equations (1) Odds = exp[LO] : (2) p = Odds/(1+Odds)

with the CONJ being reduced by 3.43 log-odds, a reduction in probability to 39.17% (68.57% for males (intercept: 2.89 lo + male: 1.31 lo + conjunction: -3.42 lo)). Results can be found in Table 4.2. The closest function to the predictions for nominal *tipo* was the DM, which was still substantially different with full /i/ only predicted to be produced in 65.93% of cases for girls (intercept: 2.98 lg + -2.32 lo) and 87.76% for boys (intercept: 2.98 lo + -2.32 lo + male: 1.31 lo).

Table 4.2. *Generalized linear mixed effects model results for the vowel /i/ presence in tipo.*

	<i>Estimate</i>	<i>Std. Error</i>	<i>z-value</i>	<i>Pr(> t)</i>	<i>CI₉₅</i>
(Intercept)	2.98	0.85	3.50	0.0004	1.31 : 2.03
Gender M	1.31	0.51	2.55	0.01	1.51 : 4.98
Function DM TRUE ⁵⁷	-2.32	0.83	-2.80	0.005	-4.29 : -0.90
Function DP TRUE	-2.75	0.84	-3.30	0.0009	-4.74 : -1.32
Function QUOT TRUE	-2.70	0.88	-3.06	0.002	-4.75 : -1.15
Function PREP TRUE	-2.83	0.85	-3.32	0.0008	-4.84 : -1.36
Function CONJ TRUE	-3.42	0.88	-3.87	0.0001	-5.47 : -1.87

Next, another *generalized linear mixed effects model* was created to examine the likelihood of a full (non-elided) /o/ vowel being present as *tipo* performs different

⁵⁷ TRUE stands for full vowels while FALSE stands for elided vowels.

functions. Just as with the /i/ model, results were plotted for the visualization of the predictions (see Figure 4.5). Average differences in log-odds between predictors indicate either a higher or a lower probability of full /o/ vowels being found in the different functions *tipo* performs.

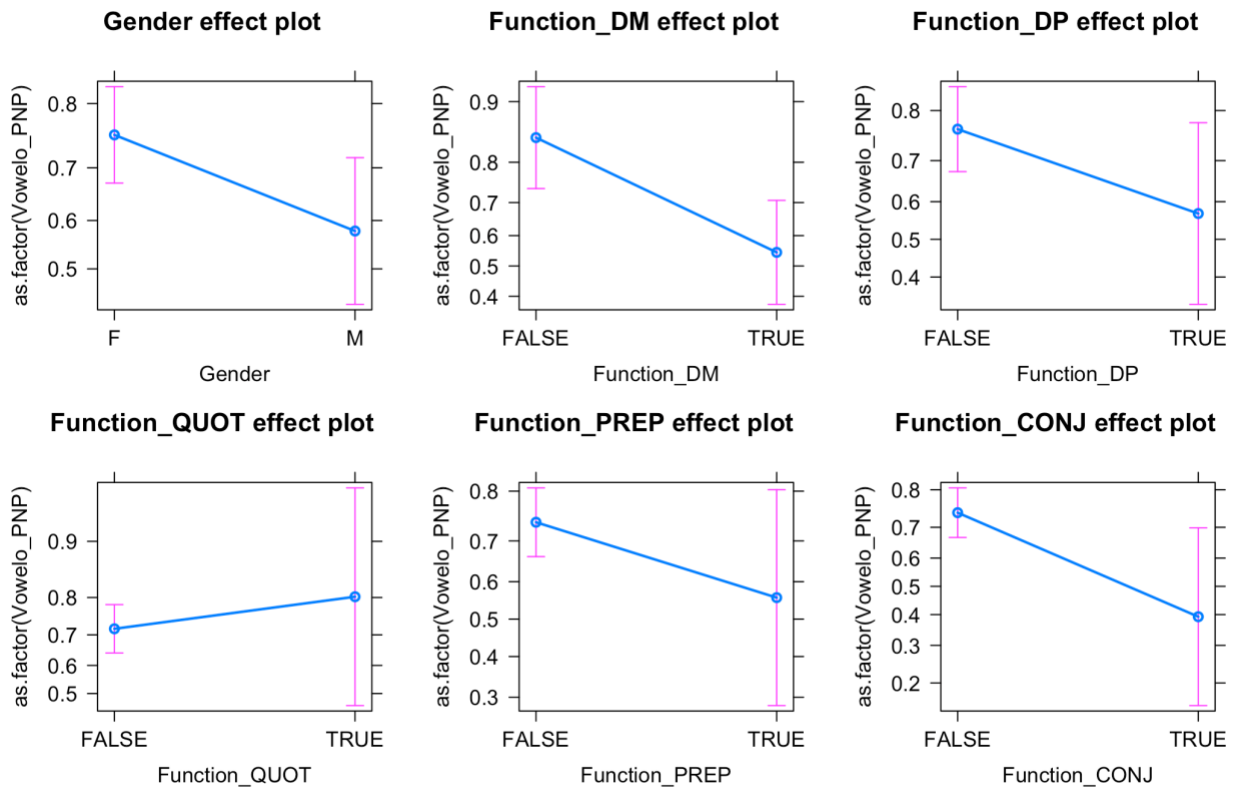


Figure 4.5. Visualization of predictions from the /o/ generalized mixed effects model.

For the /o/ vowel, the intercept (base value of 2.27) suggests that the probability of a full (non-elided) /o/ vowel being found in the NOM function is 90.64%. The probability decreased, on average, by 0.80 log-odds (81.31%) if the speaker was *male*.

The chances of a full /o/ vowel being found decreased substantially with the DM being reduced by 1.53 log-odds, a reduction in probability to 67.70% (48.50% for males (intercept: 2.27 lo + male: -0.80 lo + DM: - 1.53 lo)). The chances of a full /o/ being found also decreased substantially with the CONJ being reduced by 1.49 lo, a reduction in probability to 68.57% (49.50% for males (intercept: 2.27 lo + male: -0.80 lo + CONJ: - 1.49)). Results can be found in Table 4.3.

Table 4.3. General linear mixed effects model results for the vowel /o/ presence in *tipo*.

	<i>Estimate</i>	<i>Std. Error</i>	<i>z-value</i>	<i>Pr(> t)</i>	<i>CI₉₅</i>
(Intercept)	2.27	0.62	3.66	0.0002	1.14 : 3.65
Gender M	-0.80	0.37	-2.15	0.03	-1.58 : -0.06
Function DM TRUE ⁵⁸	-1.53	0.61	-2.53	0.01	-2.89 : -0.43
Function CONJ TRUE	-1.49	0.67	-2.23	0.02	-2.95 : -0.24

Next, average durational measurements were used to generate a plot for a visual inspection of the durational differences among the six main functions of *tipo* in the data (see Figure 4.6). The predictors analyzed were *gender* and *function*. Analysis of the plots suggested that the total duration of *tipo* as it performs its nominal function is considerably longer than that of other functions. Additionally, the plot indicates that the

⁵⁸ TRUE stands for full vowels while FALSE stands for elided vowels.

tokens generated by male speakers are longer than those from female speakers (see Figure 4.6).

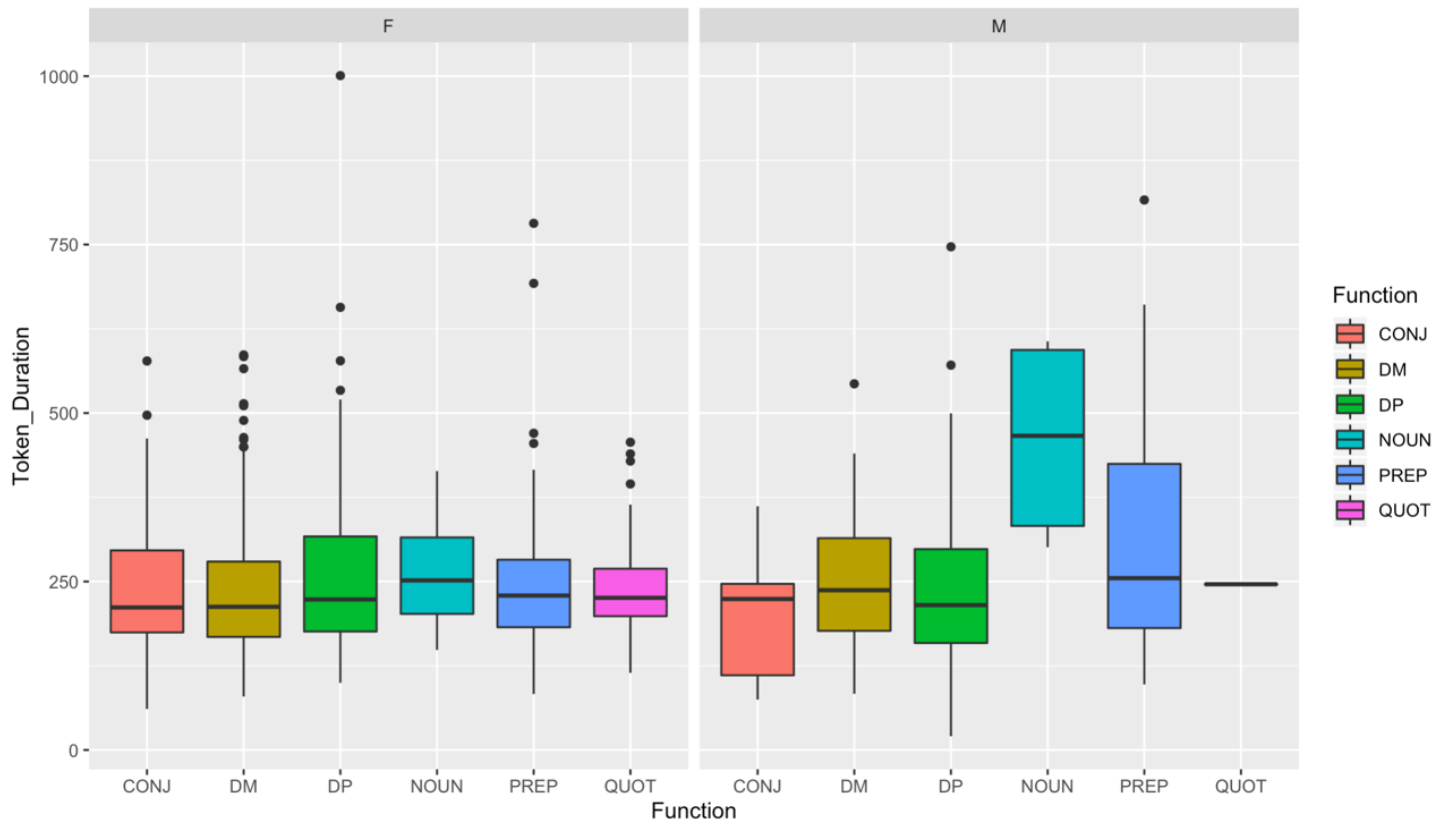


Figure 4.6. Boxplot of token duration by function and gender⁵⁹.

Results from a mixed effects model corroborates findings from the visual examination of the plot. On average, *tipo* tokens produced by male speakers are significantly longer (31.8ms) than those produced by female speakers [$\beta = -31.83$, $SE = 15.62$, $df = 67.84$, $t = -2.04$, $p = .01$, $CI_{95} = -1.69 : -62.52$]. Additionally, nominal tokens of *tipo*

⁵⁹ F refers to female participants and M to male participants.

were shown to be significantly longer than those of *tipo* performing non-nominal functions. Durational differences between the intercept NOM function and CONJ and DM functions were found to be statistically significant. Results are presented in Table 4.4.

Table 4.4. Total duration of *tipo* by function and gender.

	<i>Estimate</i>	<i>Std. Error</i>	<i>df</i>	<i>t-value</i>	<i>Pr(> t)</i>	<i>CI₉₅</i>
(Intercept)	260.65	9.08	75.90	28.69	<2e-16	242.88 : 278.41
Gender M	31.83	15.62	67.84	2.04	0.01	1.32 : 62.50
Function_DM_TRUE	-16.09	6.48	1004.92	-2.48	0.01	-28.89 : -3.43
Function_CONJ_TRUE	-25.79	12.79	986.56	-2.12	0.04	-50.86 : -0.73

It should be noted that although the results above show significant durational differences, some might be attributed to differences in segments produced by male and female speakers (see Figure 4.6 above). To address that question, a second *linear mixed effects model* was created which aimed to examine the possible interaction between *gender* and *NOM function*. Results show significant interaction between the predictors, with the nominal function of *tipo* produced by male speakers being, on average, 114.97ms longer than the intercept [$\beta= 114.97$, $SE= 56.69$, $df= 919.41$, $t= 2.03$, $p= 0.04$, $CI_{95}= 4.32 : 227.76$].

To better understand the durational differences among the tokens, the duration of each segment of *tipo* in each of the six functions was examined. The purpose of the analysis was to gain a better understanding of what the process of phonetic reduction/erosion in grammaticalization entails. This investigation was considered especially relevant as erosion has been identified as one of the main processes that grammaticalizing linguistic elements undergo. This approach, which has not yet been used in grammaticalization research, may provide a clearer picture of the elements that comprise what scholars and researchers have come to identify as reduction/erosion within the theory of grammaticalization.

Plots were created for a visual examination of the data before statistical analyses were conducted. *Linear mixed effects* models included *participant* as a random effect and *gender*, *age*, and the *functions* of *tipo* as predictors. The models also considered the possible interaction between the predictors *gender* and the different functions under analysis.

4.6.2 *Affricate*

The analysis started with the examination of the first segment in *tipo* (the affricate [tʃ]). Figure 4.7 indicates substantial durational differences between the affricate found in the nominal form when compared to the other five functions. Most notably, the affricates produced by male speakers are considerably longer than those produced by female speakers.

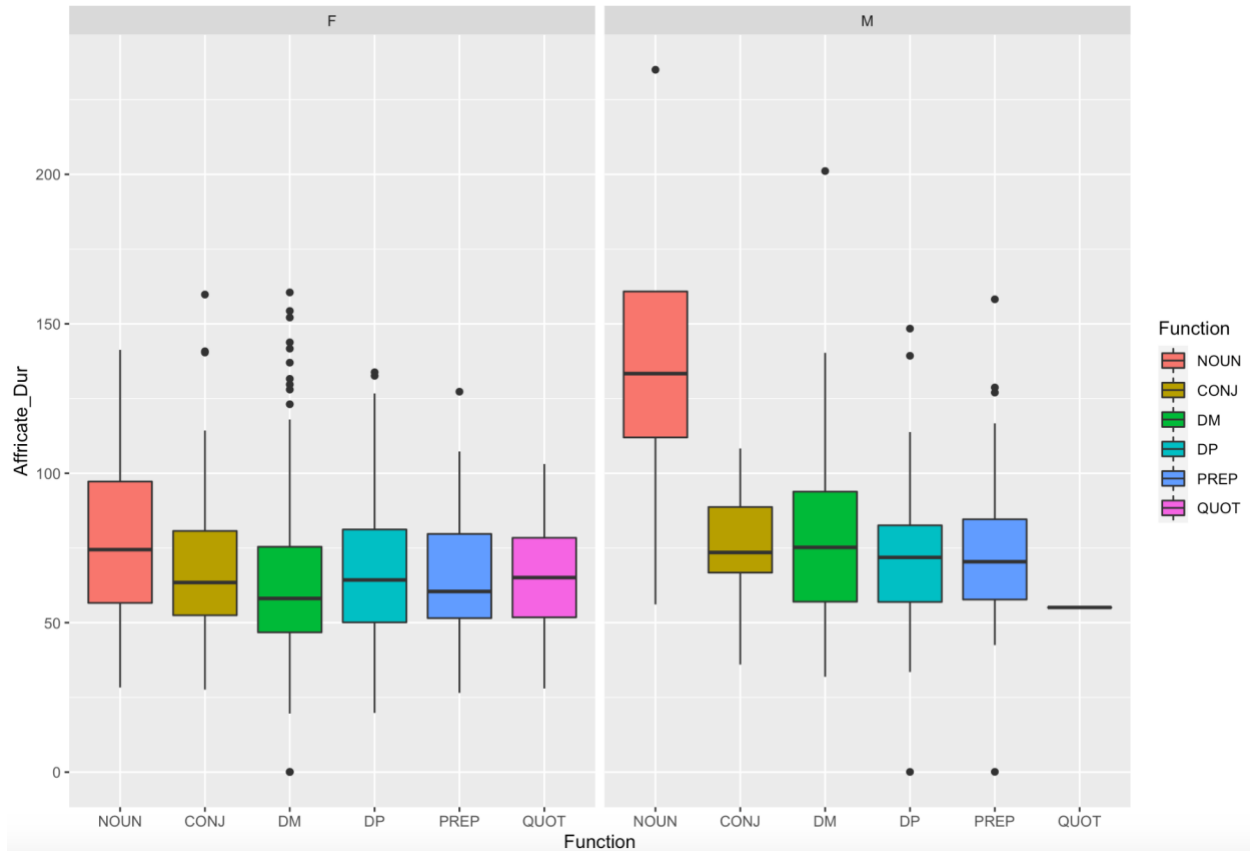


Figure 4.7. Overall duration of the voiceless affricate [tʃ] by function by gender.

This possible durational difference was confirmed by the results from a linear mixed effects model that indicated that the affricate is significantly longer in the NOUN form when compared to the other functions by, on average, 13.93ms [$\beta= 13.93$, $SE= 4.90$, $df= 997.33$, $t= 2.84$, $p= 0.004$, $CI_{95}= 4.28 : 23.50$]. Results also show an interaction between the duration of the affricate /tʃ/ and *gender*, with the segment produced by male speakers being significantly longer in the NOUN form when compared to the grammaticalized forms under investigation [$\beta= 50.55$, $SE= 13.06$, $df= 928.54$, $t= -4.68$, $p=0.0001$, $CI_{95}=$

25.09:76.27]. No other predictor significantly affected the durational differences of the affricate.

4.6.3 *Stop*

Next, an analysis of the stop consonant /p/, which aimed to examine the duration of the segment in *tipo* as it performs its six most frequent functions, was conducted.

Measurements used consisted of the combination of closure and VOT measurements.

Closure in tokens in which the /i/ vowel was elided was measured from the end of visually identifiable frication of the /tʃ/ segment in a spectrogram where there is minimal deviation from zero (indicating a period of silence) until noise is visually identifiable (such as the release burst of the stop) (see Figure 5.1), and, in tokens in which /i/ was not elided, from the last identifiable acoustic characteristic (periodic wave) of the preceding vowel until noise is visually identifiable. In the case of VOT, measurements included the release burst until the beginning of the periodic wave of the following vowel. The combination of measurements aimed to provide a general idea of the length of the consonant⁶⁰. Figure 4.8 shows possible durational differences in the realization of the stop as *tipo* performs different functions. Visual inspection indicates, once again, noticeable durational differences between the stop consonant of the nominal form of *tipo* produced by male and female speakers.

⁶⁰ Although the analyses of VOT and closure separately could provide more detailed information about the processes that the stop consonant might be undergoing, this investigation was beyond the scope of this study.

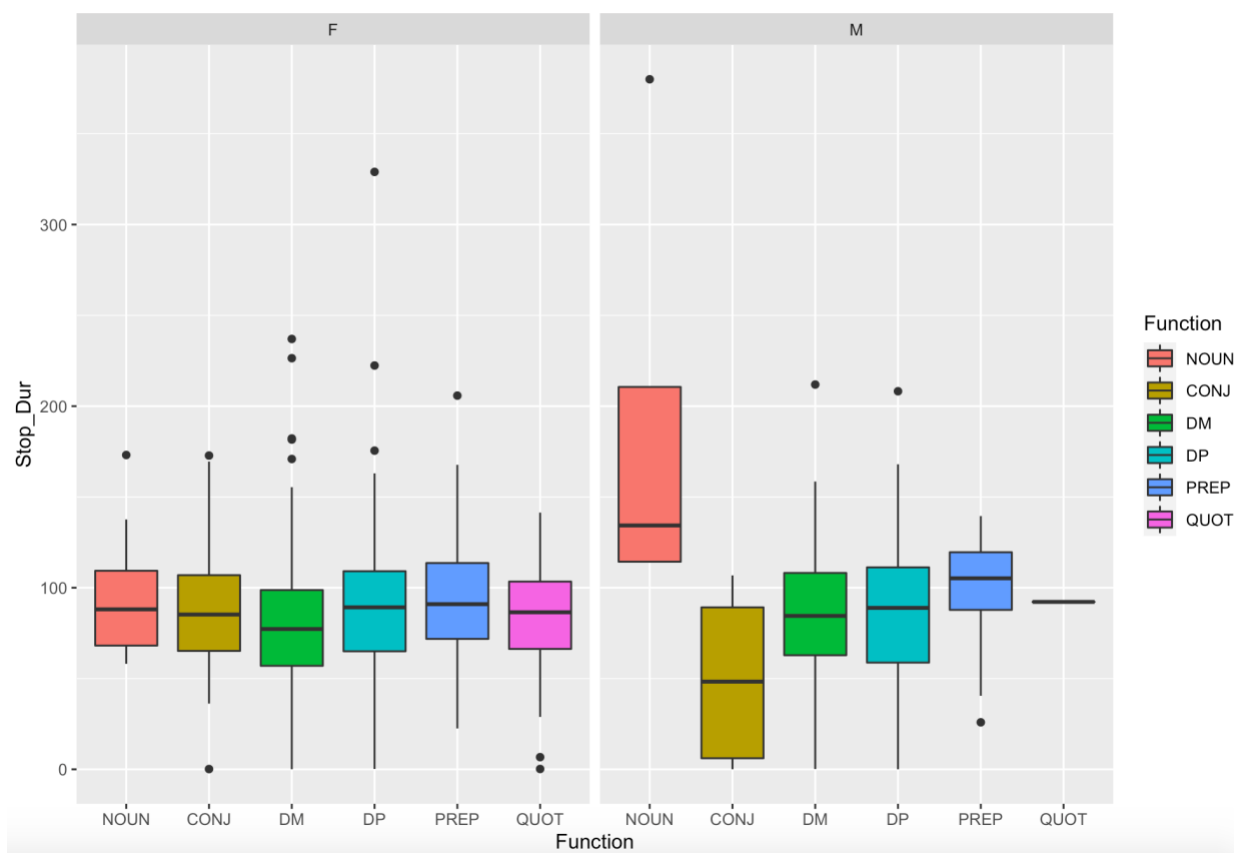


Figure 4.8. Overall duration of the voiceless stop /p/ by function and gender.

A linear mixed effects model was used to investigate if differences were statistically significant. Results show that duration of the /p/ segment is significantly shorter in the DM function, being, on average, 9.39ms shorter than the intercept [$\beta = 9.39$, SE= 2.53, df= 1000.09, $t = -3.70$, $p = 0.0002$, CI₉₅= -14.36 : -4.44]. Results also indicate an interaction between *gender* and NOUN and *gender* and CONJ functions. The stop consonant produced by male speakers in the nominal form was significantly longer than that produced by female speakers, being, on average, 67.49ms longer [$\beta = 67.49$, SE= 19.46, df= 922.01, $t = 3.47$, $p = 0.0005$, CI₉₅= 29.32 : 106.73] while, in the case of the CONJ

function, male speakers' stop consonant was shorter by, on average, 31.48ms [$\beta = -31.48, SE = 11.67, df = 1003.22, t = -2.70, p = 0.007, CI_{95} = -54.34 : -8.70$]. No other predictor significantly affected durational differences.

4.6.4 Vowels

Having examined phonological reduction from the perspective of elided vowel segments, a subset of the dataset which excluded elided vowels was created to allow for the analysis of duration and formant measurements of the vowels that were present in the tokens.

Plotting of vowel durations showed a possible durational difference in the /i/ and /o/ vowels when comparing the nominal function to the six grammaticalized functions (see Figure 4.9).

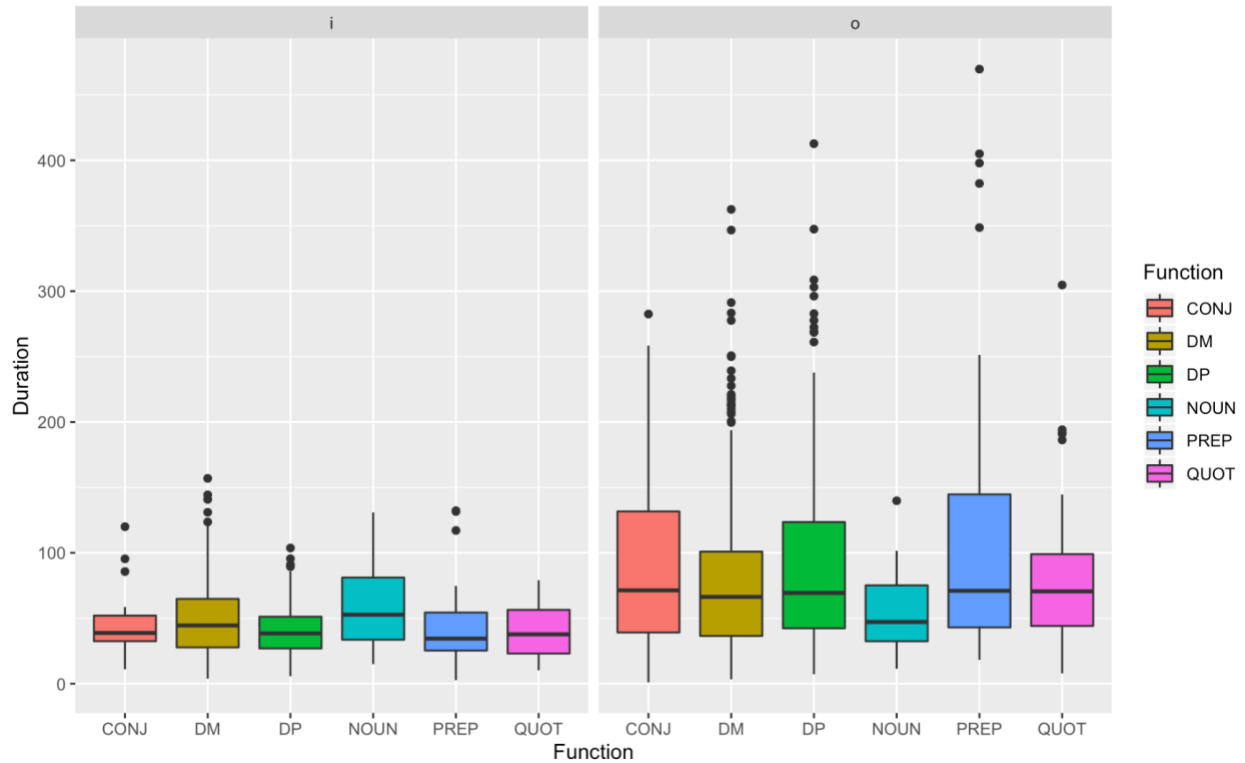


Figure 4.9. Overall duration of /i/ and /o/ (which is realized as [u] in BP) by function.

As the results from the analysis of the duration of consonants indicated significant durational differences between the affricate and stop segments produced by male and female speakers, a second set of plots was created to examine if vowels displayed a similar pattern. Figure 4.10 presents the plotting of the /i/ vowel by gender, while Figure 4.11 shows the plotting of the /o/ vowel.

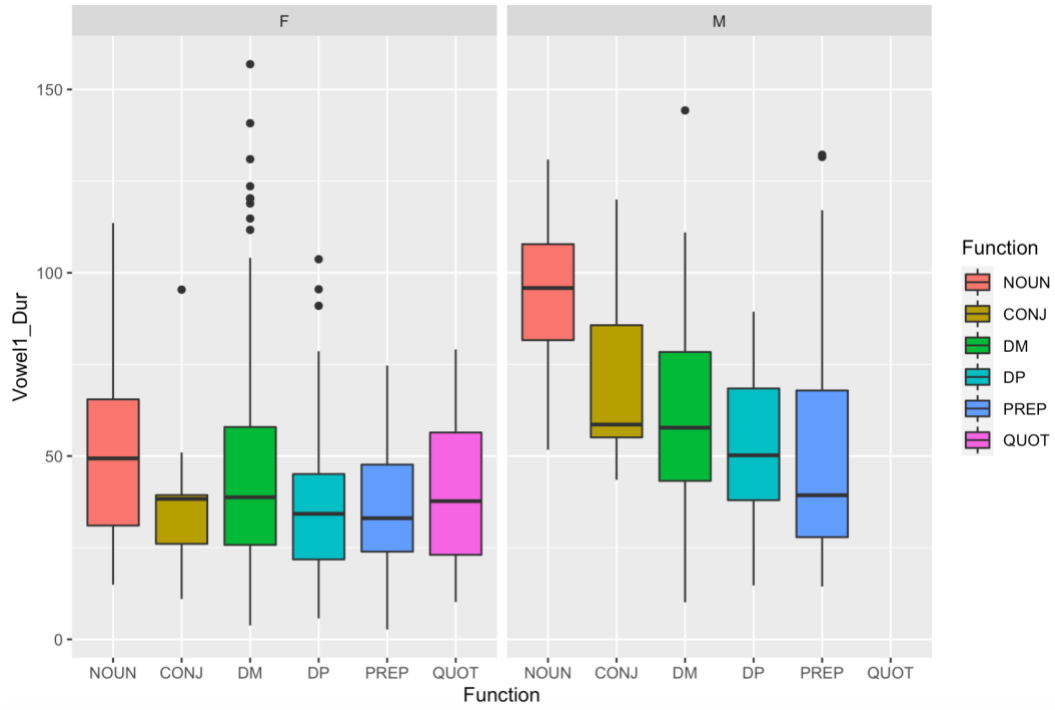


Figure 4.10. Duration of the /i/ vowel by gender and function.

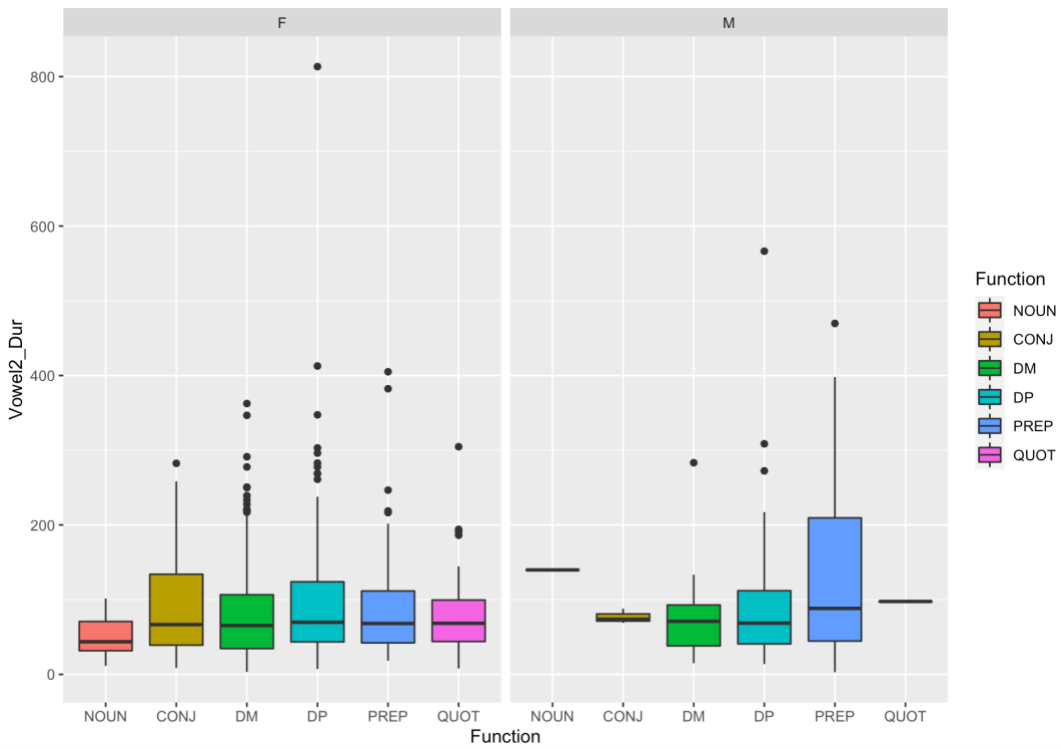


Figure 4.11. Duration of the /o/ vowel by gender and function.

Once again, visual inspection of the plots suggests durational differences between male and female speakers⁶¹. Linear mixed effects models were run to examine the durational differences of the segment by function and also to look for a possible interaction between the duration of the segment, *gender*, and *function*. Results of the model show that the interaction between *function* and *gender* is not significant. However, results show significant differences in duration of the /i/ segment produced by male speakers in general, with /i/ being, on average, 14.86ms longer than that produced by female speakers [$\beta= 14.86$, SE= 5.42, df= 66.36 t= 2.74, p= 0.008, CI₉₅= 4.19 : 25.44]. Additionally, results show that /i/ is significantly shorter in all functions when compared to the intercept NOM. Results are presented in Table 4.5.

⁶¹ It is important to note the fewer number of tokens of *tipo* produced by male speakers.

Table 4.5. Duration of the /i/ vowel by function and gender.

	<i>Estimate</i>	<i>Std. Error</i>	<i>df</i>	<i>t-value</i>	<i>Pr(> t)</i>	<i>CI₉₅</i>
(Intercept)	46.43	5.41	411.87	8.57	<2e-16	35.87: 57.00
GenderM	14.86	5.42	66.36	2.74	0.008	4.19 : 25.44
CONJ_TRUE	-24.55	5.53	1012.83	-4.44	1.01e-05	-35.46 : -13.77
DM_TRUE	-15.79	4.84	1011.12	-3.26	0.001	-25.35 : -6.36
DP_TRUE	-22.04	4.97	1011.76	-4.44	1.02e-05	-31.86 : -12.36
PREP_TRUE	-24.95	5.18	1012.87	-4.82	1.66e-06	-35.16 : -14.86
QUOT_TRUE	-18.45	5.71	1010.99	-3.23	0.001	-29.73 : -7.31

In the case of /o/, results from the model show that *gender* is a significant predictor of durational differences, with male speakers producing segments that are, on average, 31.36ms longer than female speakers [$\beta= 31.60$, SE= 15.54, df= 337.17, $t= 2.02$, $p= 0.04$, CI₉₅= 1.14 : 61.62]. Interaction between *function* and *gender* was also found to account for the durational differences of the /o/ vowel in *tipo* performing the DM function produced by male speakers, which is, on average, 43.58ms shorter than that of female speakers [$\beta= -43.58$, SE= 15.75, df= 1009.36, $t= -2.77$, $p= 0.005$, CI₉₅= -74.65 : -12.96]. Table 4.6 presents the results from the model.

Table 4.6. Duration of the /o/ vowel by function and gender.

	<i>Estimate</i>	<i>Std. Error</i>	<i>df</i>	<i>t-value</i>	<i>Pr(> t)</i>	<i>CI₉₅</i>
(Intercept)	76.10	7.43	194.93	10.24	< 2e-16	61.66 : 90.55
GenderM	31.36	15.54	337.17	2.02	0.04	1.14 : 61.62
Gender M:Function_DM_T ⁶²	-43.58	15.75	1009.36	-2.77	0.005	-74.65 : -12.96
GenderM:Function DP_T	-39.10	16.89	1007.64	-2.32	0.02	-72.66 : -6.27

Next, a cross-comparison of the F1 and F2 measurements of the vowels /i/ and /o/ in the tokens was conducted. Extreme outliers were removed as datapoints were likely measurement errors; x and y measurements were set manually (F1 - between 800Hz and 250Hz, and F2 - between 2700Hz and 700Hz)⁶³. Plotting can be found in Figure 4.12.

⁶² T stands for TRUE.

⁶³ In total, 40 measurements were removed.

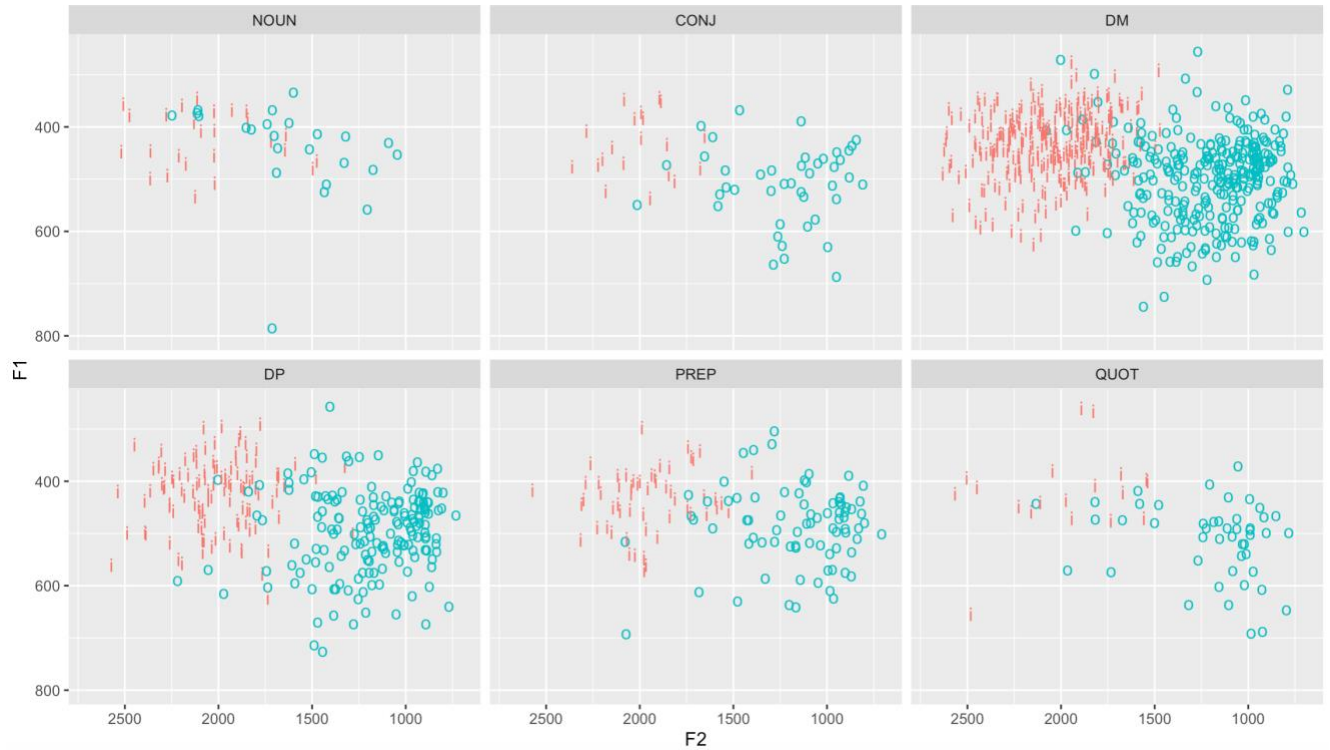


Figure 4.12. Plotting of F1 and F2 measurements (in Hz) of /i/ and /o/ vowels by function.

Figure 4.13 provides a clearer visualization of the measurements used in the cross-comparison.

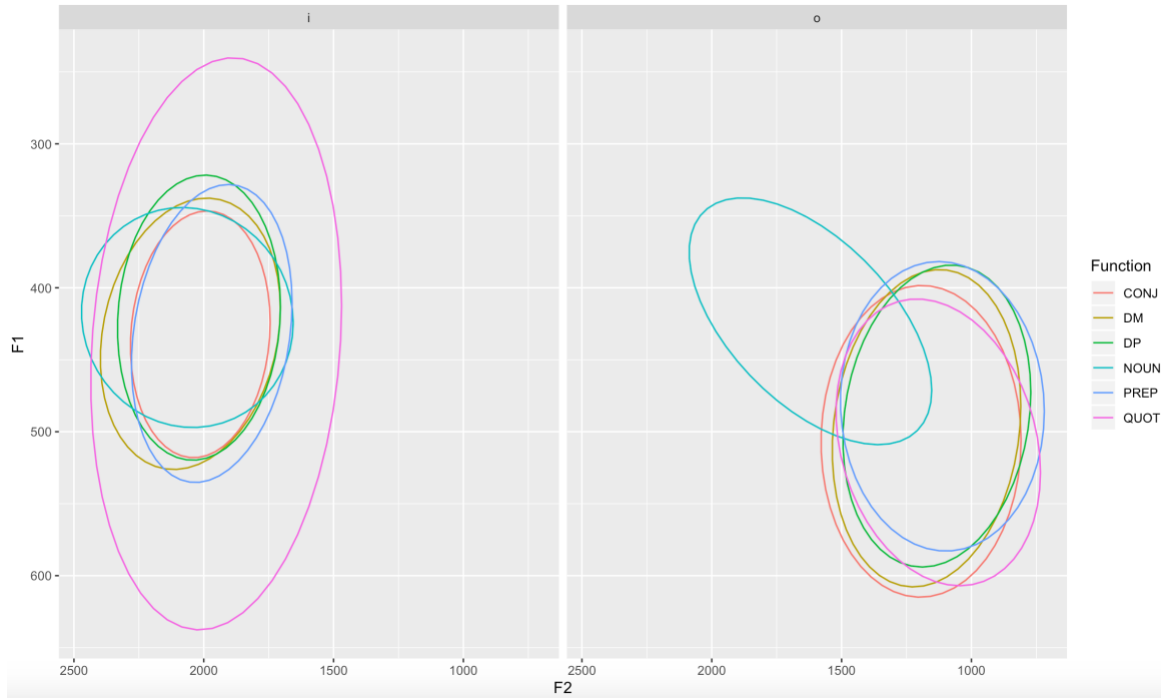


Figure 4.13. Plotting of F1 and F2 measurements of the /i/ vowel (left) and /o/ vowel (produced as [u]) (right) by function.

Visual inspection suggests different patterns with the two vowels with the /o/ vowel in grammaticalized functions of *tipo* being more retracted and lower than that of the nominal form. Results from a linear mixed effects model show non-significant differences in the F1 measurement of the high front vowel /i/ in the six functions under analysis [intercept $\beta= 427.68$, SE= 40.93, df= 471.64, $t= 10.45$, $p= <2e-16$, CI₉₅= 28.27 : 74.89]. The following predictors were considered when structuring the model: F1, *gender* (female/male), and *function* with *participant* considered as a random effect.

In the case of the /o/ vowel, results reveal that function plays a significant role in vowel height, with the F1 measurements of all grammaticalized functions being higher

than the intercept NOUN, equating to a lower tongue body position⁶⁴. For instance, in the DM function, the F1 measurements of the /o/ vowel is, on average, 70.99Hz higher than that of the NOUN function [CI₉₅= -28.81 : 134.08]. In addition, male speakers were also shown to produce significantly lower F1 values (higher tongue body position) (β = -76.53, CI₉₅= -91.19 : 11.76). Table 4.7 presents the results of the model.

Table 4.7. *F1 measurements of /o/ by function and gender.*

	<i>Estimate</i>	<i>Std. Error</i>	<i>df</i>	<i>t-value</i>	<i>Pr(> t)</i>	<i>CI₉₅</i>
(Intercept)	447.53	21.47	556.65	20.84	2e-16	405.48 : 489.27
Function CONJ	76.16	25.92	677.48	2.94	0.003	25.68 : 126.94
Function DM	70.97	21.82	664.87	3.25	0.001	28.51 : 113.90
Function DP	68.48	22.38	667.78	3.06	0.002	24.88 : 112.24
Function PREP	59.35	23.84	673.30	2.49	0.013	12.98 : 106.36
Function QUOT	63.09	25.51	677.30	2.47	0.013	13.47 : 113.32
Gender M	-75.99	15.05	67.96	-5.05	3.73e-06	-105.30 : -46.56

⁶⁴ It is important to keep in mind that /o/ in word-final position is raised significantly when compared to other languages (see Chapter 1), resulting in an allophone closer to [u].

Next, an analysis of the F2 frequency of each vowel was conducted. In regard to the vowel /i/, results from a linear mixed effects model revealed non-significant differences in retractedness in the six functions under analysis. Similarly to the results of the F1 model, gender was shown to play a significant role in retractedness, with male speakers having a considerably lower F2 frequency when compared to female speakers [$\beta = -133.97$, SE= 46.75, df= 65.29, $t = -2.87$, $p = 0.005$, CI₉₅= -225.32 : -42.74], equating to a more retracted vowel. Next, a mixed effects model designed to examine the possible correlation between F2 measurements, gender, and function of the vowel /o/ was conducted. Results reveal function to be a significant factor in vowel retractedness [intercept CI₉₅= 1484.53 : 1740.55] (see Table 4.8).

Table 4.8. F2 measurements of /o/ by function.

	<i>Estimate</i>	<i>Std. Error</i>	<i>df</i>	<i>t-value</i>	<i>Pr(> t)</i>	<i>CI₉₅</i>
(Intercept)	1612.49	65.54	562.62	24.60	< 2e-16	1484.53 : 1740.55
Function CONJ	-375.64	78.35	681.35	-4.79	2.01e-06	-529.17 : -222.66
Function DM	-410.07	66.07	671.81	-6.21	9.46e-10	-539.14 : -281.00
Function DP	-417.42	67.68	673.30	-6.17	1.20e-09	-549.99 : -285.27
Function PREP	-426.49	71.89	676.81	-5.93	4.76e-09	-567.32 : -286.12
Function QUOT	-458.20	77.10	682.02	-5.94	4.48e-09	-608.84:-307.51

Results indicate substantial differences between the /o/ vowel in the nominal form and the grammaticalized forms; more specifically, they suggest that the /o/ vowel is retracting/retracted in the grammaticalized functions when compared to the canonical form, which is fronted. This is a somewhat unexpected result as phonological reduction is often associated with centralization; furthermore, /o/ also seems to be lowering in the grammaticalized forms. It is possible that the retraction and lowering of the final vowel which clearly differentiates the NOUN function from the grammaticalized ones results from the process of grammaticalization; however, it is also important to consider that these results might have been partially affected by the low number of nominal tokens found in the data.

4.7 Summary

Results from these analyses reveal significant differences in the acoustic realization of *tipo* as a noun when compared to the five other functions investigated. Results show not only differences in the production of *tipo* by male and female speakers (with nominal tokens of *tipo* being longer when produced by male speakers when compared to female speakers) but also total durational differences between the nominal and the grammatical and pragmatic functions under analysis.

An analysis of the segments of *tipo* provided an even more detailed view of the differences between its nominal and expanded functions:

1. Elision of vowels - results show that the elision of the /i/ vowel is frequent in all functions of *tipo* when compared to its nominal form while the /o/ vowel is not elided as the high vowel;
2. The affricate /tʃ/ is significantly longer in the nominal form of *tipo* when compared to the other functions while the voiceless stop /p/ is significantly shorter in the discourse marker function of *tipo* when compared to its nominal form;
3. The /i/ and /o/ vowels produced by male speakers were significantly longer than that produced by female speakers. Moreover, results also reveal that the high front vowel /i/ is significantly shorter in all functions when compared to the nominal form;
4. Results from the analysis of F1 formants of vowels show that the F1 of the /o/ vowel in all grammaticalized functions was significantly higher than the nominal function, indicating that the vowel is lower in those functions. Regarding the investigation equating to a lower tongue body position

The systematicity identified in the results raised questions about whether speakers may be able to discriminate different realizations of *tipo*. These questions led to the design of

a discrimination task experiment that aimed to examine participants' ability to perceive durational differences between tokens of *tipo*.

Chapter 5 Perception

This chapter presents the details of an exploratory perception study of *tipo* conducted with teenage students at one of the public schools in the city of Rio de Janeiro where the production study took place. As discussed in Chapter 4, initial results from the acoustic analysis of the tokens of *tipo* indicated that the elision of its vowels (specially /i/) was common in grammaticalized functions when compared to its nominal form. In light of these findings, questions were raised regarding how the production differences between nominal and non-nominal forms might affect listener perception. To answer those questions, participants were invited to participate in an auditory perception experiment.

The chapter is organized as follows: first, a brief discussion on sound perception and discrimination is presented; next, a description of the experiment is discussed; and finally, the results from the analysis are reported.

5.1 Sound perception and discrimination

As discussed in Section 4.1, phonological attrition/erosion is considered one of the main processes that linguistic elements undergo during grammaticalization (Heine, 2003; Heine & Kuteva, 2007). The loss of phonetic substance, although not limited to grammaticalization, has been shown to be correlated to the increase in frequency use resulting from the grammaticalization of linguistic items (Bybee, 2003c; Heine &

Narrog, 2010). This process (within the context of grammaticalization) is of special relevance to this study due to the findings from the acoustic analysis of tokens of *tipo* presented in the previous chapter. Results revealed that the elision of vowels is more commonly found in the five grammatical and pragmatic functions of *tipo* identified in the dataset (*conjunction, discourse marker, discourse particle, preposition, and quotative complementizer*) when compared to its nominal form. Having found significant differences in the elision/non-elision of vowels among the functions investigated, questions about speakers' possible ability to discriminate the presence from the elision of vowels (and, consequently, *tipo*'s nominal function from its non-nominal functions) remained⁶⁵. Answering that question may provide valuable information about whether speakers perceive the different functions *tipo* performs as "different words". Results may also offer a valuable contribution to the field of grammaticalization by providing initial insight as to how grammaticalized functions of linguistic elements are stored and perceived by speakers. To the best of my knowledge, no perception study of acoustic differences among grammaticalized functions of a linguistic item has been conducted to this date.

⁶⁵ The discrimination task was designed prior to the conclusion of the acoustic analysis. At the time of data collection, only elision (and non-elision) of vowels had been examined, which limited the author's ability to include vowel quality and consonant duration in the task.

Production differences relies on listeners' ability to parse the speech stream. Researchers have conducted studies which aimed to shed a light on how humans process sounds as well as their ability to decode language and perceive differences and similarities between/among different sound patterns (see Liberman et al., 1967, Massaro, 2014, and Pisoni, 1973 for seminal work on speech perception).

Perception and discrimination of consonant and vowel sounds have been extensively researched (see, for example, Gerrits and Schouten, 2004; Pisoni, 1973, 1975; and Liberman et al., 1967). Researchers have argued that there are clear differences in the degree of categorical perception between vowels and consonants (Pisoni, 1973). For instance, Pisoni and Tartter (1981) posit that cue duration may affect categorical perception as vowels would have more uniform formants when compared to stop consonants, in which formant transitions would change more quickly while Gerrits and Schouten (2004) argue that vowels are generally perceived much less categorically than stop consonants (Gerrits & Schouten, 2004).

Although previous studies indicate that speakers can not only produce but also identify vowels in isolation (Potter & Steinberg, 1950), in natural speech vowels are often found in the context of a syllable in which coarticulation with consonants commonly occurs. Studies have shown that vowels are better identified in a consonantal syllabic context (Strange, Edman, & Jenkins, 1979). For instance, Fujimura and Ochiai (1963) examined the effect of coarticulation by comparing participants' ability to

identify vowels in consonantal contexts in comparison to vowels extracted from consonantal contexts and presented in isolation. The authors found that vowels that were presented in a syllabic context were more accurately identified than vowels that had been excised from CVC syllables and presented in isolation. Other researchers have also examined the importance of acoustic properties of syllables in the identification of vowels. In their study, Strange and her colleagues (1976) considered the possible value of the acoustic properties of syllabic structures in leading to better perception and identification of vowels. The authors argue that listeners may rely on information that is distributed throughout the syllable to be able to properly identify a vowel (Strange et al. 1976).

In addition to the important findings regarding the perception of consonants and vowels in syllabic contexts, review of the literature indicates that other phenomena should also be taken into account when investigating elided vowels. For instance, a well-known phenomenon that has been extensively studied in BP is that of epenthesis (addition) of a vowel (usually the /i/ vowel) in consonant clusters that are unacceptable in Brazilian Portuguese due to the language's phonotactic constraints. Vowel epenthesis have been shown to often repair phonotactically illegal environments, those that do not meet the structural requirements of a language (Hall, 2011). Research shows that BP speakers often perceive an 'illusory' /i/ (the shortest vowel in BP) in illegal consonant clusters (Escudero et al., 2009); this illusory /i/ would repair those illegal segmental

sequences such as in the case of the word *pneu* 'tire' /pneu/, which is pronounced [pi'neu] (Dupoux et al., 2011), making repairs in the perception of /i/ in illegal consonant clusters more likely (Dupoux et al., 2011).

If consonantal contexts provide an advantage for vowel identification, coarticulation traces, i.e., traces of the movement and position of the articulators of consecutive segments (Diehl, Lotto, & Holt, 2004), in the consonants preceding the elided vowels may influence speakers' perception of acoustically different tokens in the case of *tipo*. The question, therefore, relied on whether listeners were able to discriminate tokens of *tipo* based on the presence or absence of its vowels and the impact of coarticulation in the discrimination of tokens. More specifically, this experiment aimed to answer the following questions:

1. Does the presence or elision of vowels provide enough information for speakers to be able to discriminate paired stimuli as being either the *same* or *different*?
2. If so, what is the accuracy rate of the discrimination of *same* and *different* stimuli?
3. If not, could coarticulation effects be impacting listeners' ability to accurately discriminate pairs containing different elided vowels or pairs containing elided vs. non-elided vowels?

To answer those questions, an AX (same-different) auditory discrimination task was developed and conducted. The AX design, commonly used in research on speech

perception, consists of two stimuli presented to participants in each trial, with trials being separated from each other by interstimulus intervals (ISI), i.e., by a certain amount of time established by the researcher (McGuire, 2010). Although other experiment designs have also been posited as being suitable for discrimination tasks, the decision to create an AX task resulted from the review of previous research that suggests that this design may reduce auditory memory load (Gerrits & Schouten, 2004). Comparatively, the ABX task design (consisting of three intervals, A and B being different from each other and X being the same as either A or B) has been shown to possibly lead to difficulty in the retrieval of auditory information by the time the last stimulus is presented (Gerrits, 2001). In addition, scholars have argued that the use of three intervals could lead to bias toward selecting the last two intervals (B and X) as the correct ("same") response. In the case of AXB tasks, studies have shown both positive and negative results (Gerrits, 2001; Van Hessen & Schouten, 1999), leading to uncertainty about the fitness of the design to this study.

The AX task design also allowed for the collection of the participants' reaction time for each trial. Reaction-times studies allow for the measurement of how quickly individuals perform linguistic tasks (Whaley, 1978). The fast and transient nature of the processes that comprise language processing means that introspection is not an accessible feature that can be measured using experiments. RT measurements not only make it possible for researchers to observe participants' immediate reaction to the

stimuli to which they are presented but also have been shown to allow for a better understanding of speakers' mental processes (Baayen & Milin, 2010; Whaley, 1978) as they are user specific. For instance, in the case of word recognition of spoken words, frequency and syntactic categories of words have been shown to affect individuals' RT as the recognition of those words has been shown to be the result of speakers' evaluation of a set of possible lexical "candidates" competing to be recognized (Whaley, 1978). As a result, speakers have been shown to identify high-frequency words at a much faster pace than low-frequency words (Whaley, 1978). For instance, in the case of polysyllabic content words, previous research indicates that speakers often recognize them before the end of the word (Marslen-Wilson, 1987). This recognition would rely not only on the acoustic characteristics of the word but also on how phonetically similar that word is to other possible candidates (Pisoni & Goldinger, 1990; Marslen-Wilson, 1987). As an approach to understanding how language processing and word discrimination take place, RT has been used to understand the processing difficulty of the task, with longer RTs being associated with an increase in cognitive load and cognitive processing (Baayen & Milin, 2010).

5.2 Methods

5.2.1 Data Collection

School administration, teachers⁶⁶, and students at a municipal school in the city of Rio de Janeiro received information letters about the study before data collection sessions were scheduled. Similar to the procedures for data collection in the PSCB study, the students interested in participating in the study received a consent form to be signed by their caregivers and the students themselves. Once the signed forms were returned, the school administration decided on the allocated time slots and physical space in the school for the experiment to be conducted. Data collection sessions were scheduled in the school lab as it was located in a quieter area of the school. Participants were verbally briefed about the nature of the experiment (i.e., they were informed that it was a discrimination task consisting of different trials in which they would listen to two stimuli, which could be either the same or different) and were given time to ask questions prior to the commencement of the experiment. The goal was to make students feel comfortable during the experiment and comfortable using the computer.

⁶⁶ I would like to thank the GEO Nelson Prudêncio school administration and teachers for the accommodations made to make the data collection possible.

5.2.2 *Participants*

For this experiment, participants were 98⁶⁷ middle school students (female N = 52, male N = 46) between the ages of 11 and 16 who, at the time of data collection, attended one of the municipal middle schools in the island of Ilha do Governador, Rio de Janeiro, in which the data collection for the PSCB study was also conducted⁶⁸. All participants were monolingual speakers of Brazilian Portuguese born and raised in the city of Rio de Janeiro⁶⁹, and none reported either hearing or speech impairments.

5.2.3 *Equipment*

Two Lenovo Thinkpad T420 laptops running on Windows 7 (2009) belonging to the Social Sciences Research Lab at the University of Saskatchewan⁷⁰ were used to conduct the experiment. The computers were borrowed as they were licensed to use E-Prime. The experiment, created with the support of the Social Sciences Research Lab, was uploaded to both computers before the fieldtrip for data collection. To access the discrimination task, participants wore Sony earbuds. The choice of using earbuds instead of regular headsets was due to the following factors: a) ease of transportation and b) overall affordability⁷¹.

⁶⁷ Unfortunately, due to time constraints, over 60 students were not able to take part in the experiment.

⁶⁸ This is a new group of participants, who took part in the study in 2018.

⁶⁹ All participants were speakers of the *Carioca* dialect of Rio de Janeiro.

⁷⁰ I would like to thank Brianna Groot and the SSRL at the University of Saskatchewan for the support and time given to help me design the discrimination task and for providing all the equipment needed to make the project possible.

⁷¹ The earbuds were disinfected prior to each session.

5.2.4 Stimuli

A token of *tipo* performing its nominal function (see Figure 5.1) produced by a female teenage speaker was extracted from the PSCB corpus (for information on the equipment used to record the interviews, see Chapter 1). The selection of a nominal token of *tipo* resulted from the unlikeliness of vowel elision occurring in the nominal form.

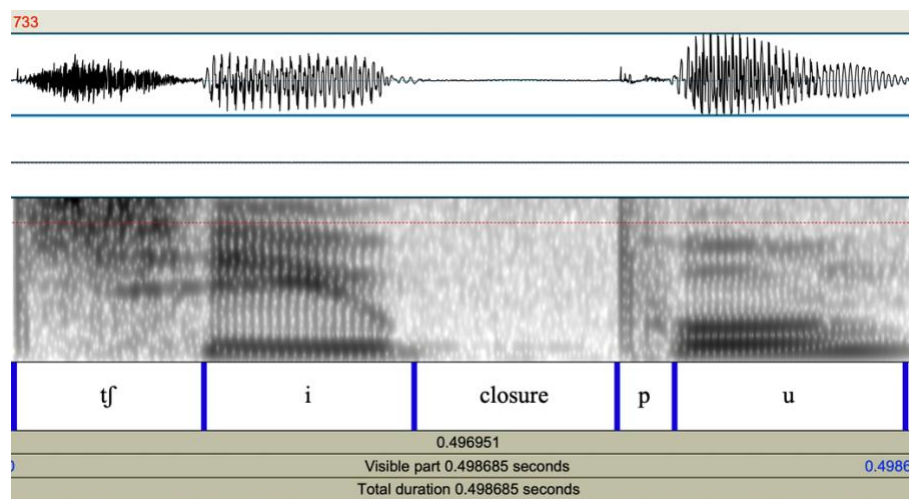


Figure 5.1. Textgrid showing the nominal token of *tipo* used in the experiment.

The selected token was clear and provided easy identification of all segments and their boundaries in the spectrogram. Although some consideration was given to the inclusion of tokens produced by both male and female speakers as stimuli, using only

the token produced by one speaker minimized the number of variables that could affect the discrimination of tokens⁷².

Once selected, the token was extracted from the original audio file and copied three times using Praat (Boersma & Weekink, 2017). The four tokens (the original nominal one and the three copies) were individually saved as *.wav* audio files. Next, three out of the four tokens underwent manipulation in Praat. In the first, the /i/ vowel was excised; in the second one, the /o/ vowel was excised; in the third, both the /i/ and /o/ vowels were excised. Once this process was concluded, the four tokens were saved as separate *.wav* files: *tipo* (full form, realized as /'tʃipu/); *t_po* (elided /i/ vowel, realized as /'tʃpu/); *tip_* (elided /o/ vowel, realized as /'tʃip/); *t_p_* (elided /i/ and /o/ vowels, realized as /'tʃp/)⁷³. The fundamental frequency of each stimulus was the same as that of the original file while the duration varied according to the presence/absence of vowel(s). The stimuli consisting of tokens containing elided and non-elided vowels were representative of actual tokens of *tipo* found in the corpus⁷⁴ (see Figures 5.2, 5.3, and 5.4).

⁷² Variables such as physiological variation (e.g., the difference in vowel tract sizes) between male and female speakers would need to be accounted for as they could affect the participants' responses.

⁷³ It is important to note that there were several tokens of *tipo* identified in the corpus in which neither vowel could be identified auditorily or visually in the spectrogram.

⁷⁴ As discussed in the previous chapter, nearly half of the /i/ segments (47.17%) and over 30% of the /o/ segments (32.35%) were elided, with the nominal version corresponding to only 0.42% of the total of the elided /i/ vowel and 1.21% of the total of elided /o/ vowel.

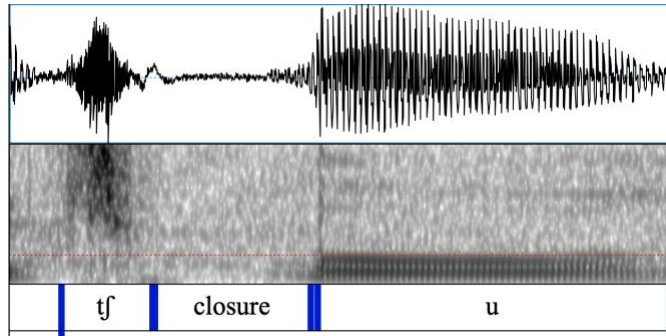


Figure 5.2. Textgrid showing a token of *tj* from the PSCB corpus with an elided /i/ vowel.

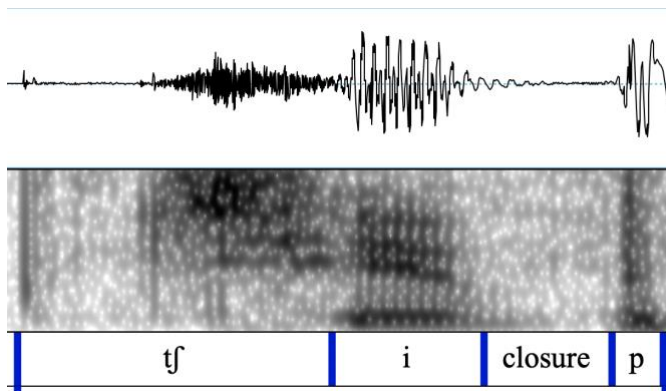


Figure 5.3. Textgrid showing a token of *tj* from the PSCB corpus with an elided /o/ vowel.

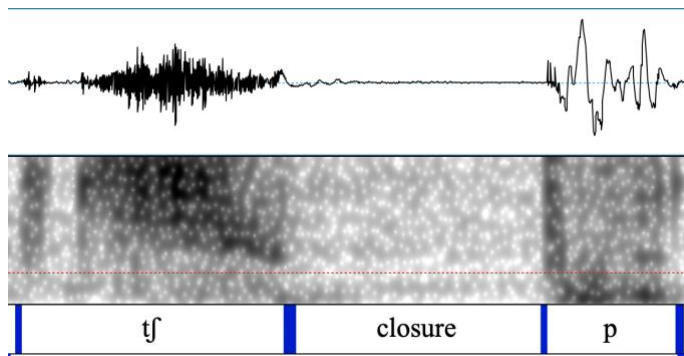


Figure 5.4. Textgrid showing a token of *tj* from the PSCB corpus with elided /i/ and /o/ vowels.

5.2.5 *Experiment Design*

A *same-different* (AX) auditory word pair discrimination experiment was designed using E-Studio 2.0.10.147 for E-Prime 2.0 – 2012 Psychology software tools (Psychology Software Tools, 2012). The goal of the experiment was to compare discrimination accuracy among different pairings of *tipo* tokens as well as participants' reaction time responses as they designated the pairs as either matching or mismatching.

The experiment consisted of two audio intervals in which participants were asked to determine whether stimulus 1 and stimulus 2 were *different* or the *same*. For the experiment, pairs of either *same* or *different* stimuli were created manually. Given the timeline and the increase in complexity during the initial stages of design, pseudo-randomization of the stimuli was used (Bradshaw et al., 2002; Gillam et al., 1988; Wioland et al., 1999), dictating that only the pairings that were manually created were presented to participants. A layer consisting of the randomization of the order in which the paired stimuli would be presented to participants was then included in the experiment design. This allowed for the simplification of the structure of the experiment while maintaining randomization. In addition, the design allowed for the control of the pairings participants evaluated, resulting in the same number of trials for each possible pairing. The pseudo-randomized structure was created as follows:

- as a proxy for what the expected experiment would look like, an initial design was created in an Excel table;

- next, pairs of *same/different* stimuli were entered in EPrime directly.
- to address the fact that there were more *different* than *same* paired stimuli, the task design included the same number of *same* and *different* conditions, i.e., half of the trials consisted of within-category pairs requiring a *same* response while the other half consisted of across-category pairs requiring a *different* response;
- as the goal of the experiment is to investigate speakers' ability to discriminate nominal from non-nominal forms, the paired stimuli consisted of comparisons of the nominal, full form (*tipo*) to both the nominal form itself (*tipo*) and non-nominal forms that contained elided vowels (*tip_*, *t_po*, *t_p_*)
- paired non-nominal forms were added as distractor tokens; this design resulted in an even number of conditions (24 paired stimuli) (see Table 5.1) and the same randomized trials across participants;
- the 24 paired stimuli were then doubled in EPrime (for a total of 48 paired stimuli) that comprised the single experiment;
- randomization of displayed pairs was designed in EPrime; for each trial, the display order of the 48 pairs was randomized.

Table 5.1. *Composition of stimulus pairings used in the discrimination task including distractors*⁷⁵.

<i>tipo stimuli</i>	<i>tip_ stimuli</i>	<i>t_po stimuli</i>	<i>t_p_ stimuli</i>
<i>tipo-tip</i>	<i>tip-tip</i>	<i>tpo-tip</i>	<i>tp-tip</i>
<i>tipo-tipo</i>	<i>tip-tipo</i>	<i>tpo-tipo</i>	<i>tp-tipo</i>
<i>tipo-tp</i>	<i>tip-tp</i>	<i>tpo-tp</i>	<i>tp-tp</i>
<i>tipo-tpo</i>	<i>tip-tpo</i>	<i>tpo-tpo</i>	<i>tp-tpo</i>

Special consideration was given to participants' daily school schedule and ability to take part in the study⁷⁶. To maximize the likelihood of students joining the study and to minimize the chances of participants feeling tired during the experiment, the number of trials was established at 48. Although a higher number of trials could have been presented to participants (see, for example, Connine, Titone, & Wang, 1993, and Sanker, 2019 for studies with larger number of trials), a decision was made to increase the initial sample size (from the original 30 to 50 participants to between 90 and 100), which, consequently, could provide more robust results. Additionally, as participants get tired, they are more likely to make random choices, which could possibly negate the effects.

Auditory memory is time dependent (Gerrits & Schouten, 2004). That makes the decision about the length of the interstimuli interval (ISI), the period of time that

⁷⁵ The bolded pairs indicate the ones that included both nominal and non-nominal forms (the focus of this experiment); non-bolded pairs indicate distractors.

⁷⁶ Data collection sessions took place during school hours.

separates one stimulus from the other, of great importance when designing a discrimination task. As discrimination performance has been shown to increase between 100 and 500 msec, reaching a maximum between 500 and 1,000 msec (Gerrits & Schouten 2004; Pisoni, 1973), for the present task, the interstimuli interval (ISI) within trial was established at 500msec. The intertrial interval was determined by the response time.

A pilot procedure was developed and conducted before data collection sessions took place. Five volunteers who were Brazilian Portuguese native speakers living in Canada took part in the procedure. One change to the study was applied based on the feedback received from volunteers: volunteers stated being confused by the information presented in the instruction screen, leading them to expect either two completely different words (e.g., *tipo - ai*) or the same word (e.g., *ai - ai*) to be presented (rather than the same word possibly being presented in different ways). The initial text was changed to state that participants would hear a word said twice and would need to decide whether they were the same word or different words. Once changes derived from the pilot study were applied, the experiment was considered ready to be presented to participants.

5.2.6 Participants

For the experiment, students from one of the municipal schools on the Island of Ilha do Governador were invited to take part in the experiment. Students who showed

interested in participating received a consent form to be signed by their caregivers and by the students themselves. Once the signed forms were returned, the author and the school staff discussed optimal times for data collection sessions to take place. The experiment was conducted in the school lab as per the decision of the school staff. In total,

5.2.7 Experiment

First, participants were informed of the nature of the experiment (a discrimination task) and that they could ask for breaks if needed. Before the experiment started, text in the initial screen instructed participants to input their demographic information (such as *age* and *sex*) as well as the date in which they took part in the experiment. The initial screen was followed by an instruction page, which was advanced by pressing the space bar. The instruction page consisted of vertically and horizontally centralized text. Duration was set for infinite, requiring key-press to start the experiment and to finish it. The text used in the experiment is presented in Figure 5.5.

Instructions

*Você irá ouvir uma palavra duas vezes*⁷⁷. ('You will hear a word twice.')

Aperte a tecla de espaço para continuar. ('Press the space bar to continue.')

Aperte 'm' se as palavras soarem iguais. ('Press 'm' if the words sound the same.')

Aperte 'z' se as palavras soarem diferentes. ('Press 'z' if the words sound different.')

Aperte a tecla de espaço para continuar. ('Press 'space' to continue.')



Experiment

Stimulus 1 - blank screen



Inter-stimulus (ISI - 500 ms)



Stimulus 2



⁷⁷ All participants were also verbally told that they would be hearing the same word throughout the experiment.



Final Screen

Obrigada pela sua participação! ('Thank you for your participation!')

Aperte a Tecla de espaço para terminar. ('Press the space bar to finish')

Figure 5.5. *Layout of the different screens that comprised the experiment.*

Participants used a button-press response to indicate whether words were the *same* (*m*) or *different* (*z*) based on the presence or elision of a vowel/both vowels. Results from each session were transferred to a MacBook Pro 2016 computer at the end of the day to prevent data loss. Due to the nature of the experiment design, participant information was completely anonymized in the data, with participants being identified by numbers (01 to 98).

During the experiment, same-different reaction time (RT) measurements were collected to give a better understanding of participants' cognitive processes during the task. In the specific case of this experiment, results may provide information on what contrastive/similar elements may play a role participants' RT to discriminate the paired stimuli.

5.3 Results

Results from the experiment provided information on participants' accuracy in discriminating different paired stimuli but also on the reaction time taken to discriminate each individual pair.

The data from the discrimination task experiment was analyzed in R (R Core Team, 2013). Results from distractors were excluded from the analysis⁷⁸. To look for potential patterns of interest in the dataset, plots were generated using the *ggplot2* package (Wickham, 2016) in R prior to the statistical analysis. This allowed for the examination of possible trends and potential areas of interest in the data. The following analyses were conducted:

- accuracy of discrimination
- reaction time

First, *accuracy of discrimination* and *reaction time* were plotted. Next, linear mixed effects models using *lme4* (Bates, Mächler, Bolker, & Walker, 2015) for R were used to examine the possible correlation between factors, including *reaction time*, *pair type*, *reaction time*, and *accuracy*. The following predictors were considered when structuring the models: *age* (11 to 16, for a total 6 factors), *gender* (female/male), *pair type*; *display*

⁷⁸ Therefore, the results presented here refer only to the pairings that included the nominal form (*tipo-tipo*, *tipo-tip_*, *tipo-t_po*, *tipo-t_p_*, *tip_tipo*, *t_po-tipo*, and *t_p_-tipo*).

order, accuracy, and response time (i.e., reaction time). All models included *participant* as a random effect, and non-significant fixed predictors were individually excluded from the models based on the closest t-value to zero. Consequently, the results presented in this section are solely from models that do not include non-significant predictors.

Frequency count of *accurate* and *inaccurate* responses show that accurate responses (i.e., accurate identification of the paired stimuli as being either the *same* or *different*) corresponded to almost 67% of the responses while the inaccurate responses accounted for slightly over 33% of the responses (see Table 5.2).

Table 5.2. *Total number of accurate and inaccurate responses.*

<i>Accuracy</i>	<i>Responses</i>	<i>Percentage</i>
Accurate	1178	66.78%
Inaccurate	586	33.22%
Total	1764	

Frequency of *accuracy* by *pair type* was calculated. Results show above chance performance in the discrimination of most pairings, with greater accuracy for the *tipo-tipo* pairs. In the case of *different* types, results also indicate that the accurate

discrimination of pairs was above chance (50%). In terms of inaccurately identified pairings, the *tip_-tipo*, *tipo-tip_* pairs stand out for having the lowest frequency of accurate responses regardless of the order in which the stimuli were presented. Results are presented in Table 5.3.

Table 5.3. *Frequency of accuracy by pair type.*

<i>Pair Type</i>	<i>Accurate Response</i>	<i>%</i>	<i>Inaccurate Response</i>	<i>%</i>	<i>Total</i>
<i>tipo-tipo</i>	504	85.71%	84	14.29%	588
<i>tip_-tipo</i>	61	31.12%	135	68.88%	196
<i>tipo-tip_</i>	65	33.16%	131	66.84%	196
<i>tipo-t_p_</i>	162	82.65%	34	17.35%	196
<i>t_p_-tipo</i>	130	66.33%	66	33.67%	196
<i>tipo-t_po</i>	133	67.86%	63	32.14%	196
<i>t_po-tipo</i>	123	62.76%	73	37.24%	196
Total	1,178		586		1,764

A *generalized linear mixed effects model* (Bates et al., 2015) was designed to examine participants' accuracy in discriminating the paired stimuli. The main goal was to calculate the average differences in log-odds between predictors, which indicate a higher (positive log-odd results) or lower (negative log-odd results) probability of participants correctly discriminating the pairs as being the *same* or *different*. Significant results regarding participants' ability to discriminate whether the stimuli in the pairs were the *same* or *different* were found. *Gender, age, and display order* were not significant predictors.

As Table 5.4 shows, the intercept, the average value of a data subset in which the baseline value of the dependent variable when all the predictors (independent variables) are 0, base value of 1.86, suggests that the probability of *tipo-tipo* being accurately identified as being the *same* is 86.58%⁷⁹. The chances of participants accurately identifying the pair *t_p_-tipo* as *different* decreased substantially, being reduced by 1.15 log-odds, a decrease in probability to 0.7108 log-odds (67.06%). In the case of the *tip_-tipo* pair, the chances of the accurate discrimination of the pair as *different* were significantly lower, being reduced by 2.70 log-odds (to -0.83 log-odds), a decrease in probability to 30.30%, while the chances of the accurate discrimination of

⁷⁹ Log odds were converted to percentage using the following equations (1) Odds = exp[LO] : (2) p = Odds/(1+Odds)

the pair *tipo-tip_* was reduced by 2.60 log-odds to -0.73 log-odds, a decrease in probability to 32.41%. The accurate discrimination of the pair *tipo-t_p_* was not found to be statistically significant (-0.24 log-odds). Results can be found in Table 5.4.

Table 5.4. *Glmer model results of accuracy in discriminating the paired stimuli.*

<i>Pair Type</i>	<i>Estimate</i>	<i>Std. Error</i>	<i>z-value</i>	<i>Pr(> t)</i>	<i>CI₉₅</i>
(Intercept)	1.8640	0.1300	14.339	< 2e-16	1.62 : 2.13
<i>t_p_.tipo</i> TRUE ⁸⁰	-1.1532	0.1954	-5.900	3.632e-09	-1.54 : -0.77
<i>tipo.tip_</i> TRUE	-2.5992	0.1991	-13.055	< 2e-16	-2.99 : -2.21
<i>tipo.t_po</i> TRUE	-1.0808	0.1968	-5.493	3.95e-08	-1.47 : -0.96
<i>tip_.tipo</i> TRUE	-2.6969	0.2013	-13.394	< 2e-16	-3.10 : -2.31
<i>t_po.tipo</i> TRUE	-1.3166	0.1931	-6.820	9.11e-12	-1.70 : -0.94

⁸⁰ TRUE stands for accurate responses (whether *same* or *different*).

Next, *response time* (RT) measurements were used to generate plots for visual inspection of RT differences collected during the discrimination task experiment. Responses were plotted to allow for a visual comparison of *response time* across trials. Visual examination of the plot suggests differences in response time across different pairings when comparing reaction times to *accurate* and *inaccurate* classification of each pair as either *same* or *different* (see Figure 5.6).

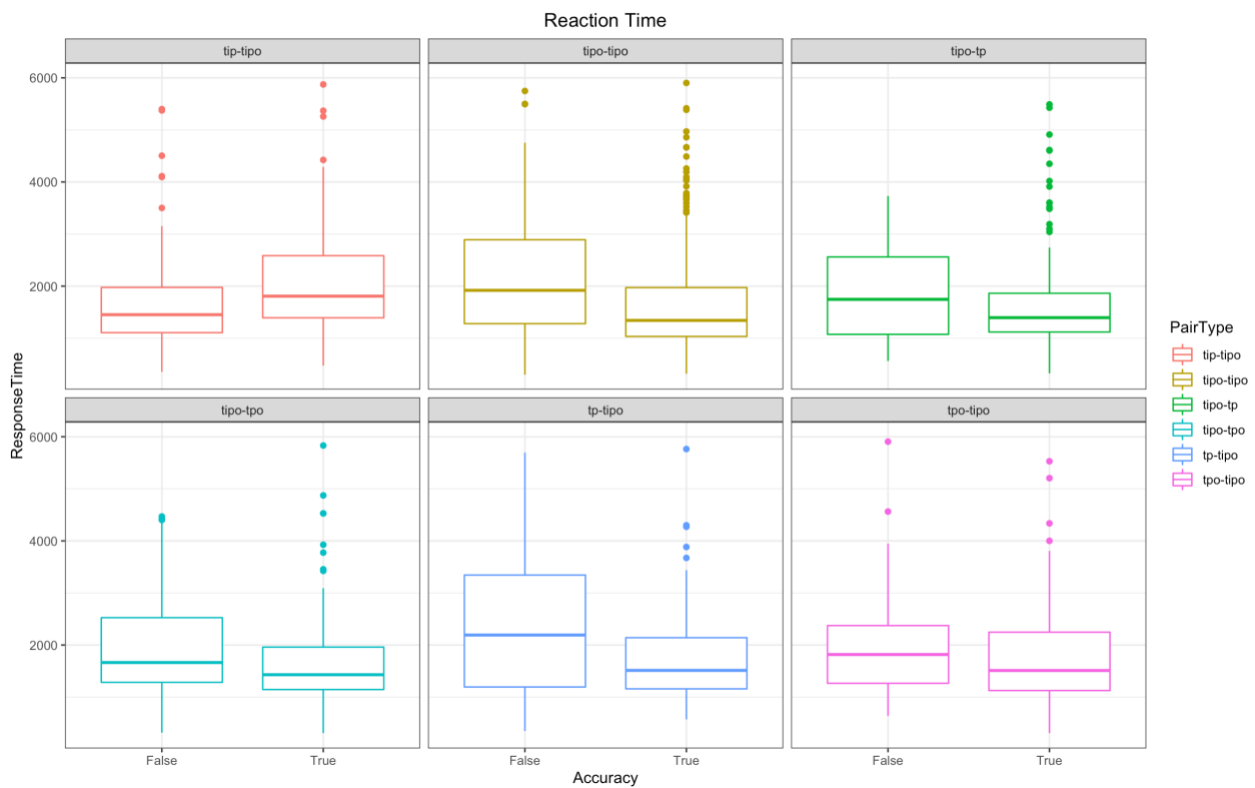


Figure 5.6. Plotting of the average reaction time (ms) for accurate (TRUE) and inaccurate (FALSE) responses.

A linear mixed effects model was created to examine participants' RT for the discrimination of each paired stimuli. This model aimed not only to provide details regarding the time needed for participants to process and compare both stimuli but also to show possible correlation between RT and participants' response (*accurate* or *inaccurate*). The following predictors were considered when structuring this model: *age* (11 to 16, for a total 6 factors), *gender* (female/male), *pair type*; *display order*, *accuracy*, and *response time* (i.e., *reaction time*). with *participant* being a random effect. Non-significant fixed predictors were individually excluded from the models based on the closest t-value to zero.

Results show that accurate responses were significantly shorter than inaccurate responses by, on average, 336ms. Reaction time for the discrimination of *t_p_tipo* pairing was significantly longer than that of the intercept *tipo_tipo* by, on average, 323ms [$\beta= 323.29$, SE= 128.18, df= 1660.55, t= 2.52, p= .01, CI₉₅= 72.56 : 574.11]. Table 5.5 presents the results from the model.

Table 5.5. Reaction time (in ms) across different paired stimuli.

	<i>Estimate</i>	<i>Std. Error</i>	<i>df</i>	<i>t-value</i>	<i>Pr</i>	<i>CI₉₅</i>
(Intercept)	2227.48	135.93	1457.63	16.39	< 2e-16	1961.53 : 2493.35
Accuracy.TRUE ⁸¹	-336.57	87.77	1752.26	-3.83	0.0001	-508.28 : -164.51

⁸¹ TRUE stands for correct response

<i>t_p_-tipo</i>	323.29	128.18	1660.55	2.52	0.01	72.56 : 574.11
------------------	--------	--------	---------	------	-------------	----------------

Reaction time for the discrimination of the intercept *tipo-tipo* took, on average, 2227ms. Comparatively, the RT for the discrimination of the other paired stimuli (with the exception of the *t_p_-tipo* pairing mentioned above) were found to not be statistically significant. However, the results from the analysis show that participants' accurate responses were significantly shorter than those of inaccurate responses by, on average, 336ms ($[\beta = -336.57, SE = 87.77, df = 1752.26, t = -3.83, p = .0001, CI_{95} = -508.28 : -164.51]$). This indicates that, although RT was not shown to be correlated with most specific pairings, participants' RT were faster when they accurately identified the pairings as *same* or *different*.

The results from the discrimination task presented in this chapter suggest that the presence or absence of vowels in the word *tipo* is discernible to speakers. As elision has been shown to be more recurrent in pragmatic and grammatical functions, which *tipo* has expanded to perform due to its ongoing grammaticalization, it is possible to postulate that speakers are able to discern the grammaticalized forms from its nominal form. This, however, is not applicable to all pairings as results from the *glmer* model revealed that speakers' ability to discriminate the paired stimuli *tipo-tip_* and *tip_-tipo* is non-significant. This result deserves special attention as the order of the stimuli was

shown to not affect the perception of the pairing as *different*. Not only do the results show that these tokens are not contrastive to speakers but also suggest that, to them, *tipo* and *t_po* may be allomorph variations and, therefore, not discriminable.

Results also show a correlation between *latency* and *inaccurate response*. *Inaccurate responses* took, on average, 2227ms while *accurate responses* were significantly shorter, being, on average, 1891ms, suggesting that latency might be used as an indication of issues in discriminability of the two stimuli. These results also provide experimental confirmation of previous findings which have identified the fast "*same*" effect in discrimination tasks (in this case, in the discrimination of *tipo-tipo*) when compared to different pairings (i.e., *same* pairings are often identified as such significantly faster than *different* pairings) (Farrel, 1985).

5.4 Summary

The goal of this experiment was to investigate listeners' ability to discriminate acoustically different tokens of *tipo*. By using stimuli that reflected different realizations of *tipo* as it performs both nominal and grammaticalized functions as identified in the dataset (see Chapter 4), this study provides an initial account of speakers' ability to differentiate different forms - and possibly different functions - of that word. Elision and presence of vowels were discernible to speakers, an important finding as elision is recurrently found in grammatical and pragmatic functions while full vowels are often present in *tipo*'s nominal form.

Being the first experiment to address this question from the perspective of grammaticalization, these significant results offer an important insight into speakers' ability to discriminate different functions performed by a grammaticalized linguistic element. However, further testing, including the addition of the different acoustic properties identified during the acoustic analysis, should be conducted to fully investigate the extent to which listeners are able to discriminate the different functions *tipo* performs in the *Carioca* dialect of Rio de Janeiro.

Chapter 6 Discussion

The study described in this dissertation aimed to investigate the current state of grammaticalization of the word *tipo* in the Carioca dialect of Brazilian Portuguese and explore some of the processes *tipo* is undergoing as it expands from its original nominal usage to grammatical and discourse-pragmatic functions in that dialect. To achieve this goal, sequential layers of qualitative and quantitative analyses were conducted: an investigation of the current use of *tipo* (more specifically, a frequency count of nominal and non-nominal functions); a usage-based analysis of *tipo* in conversational speech; a functional analysis of *tipo* in the *Projeto Sociolinguístico Contemporâneo Brasileiro* (PSCB) corpus; an acoustic analysis of *tipo* tokens; and a discrimination task that aimed to explore speakers' ability to discriminate acoustic differences among different tokens of *tipo*. The chapters were laid out to provide access to how the study was designed and conducted, going from a broader view – the examination of the frequency of nominal and non-nominal forms - to a much narrower view – an acoustic analysis and the subsequent discrimination task, which aimed to gain insight into listener perceptions of the different functions that *tipo* performs. The chart below summarizes the sequence of analyses that comprise the framework adopted in this study.



Figure 6.1. *Design of the current study.*

This chapter presents a discussion of the findings from the analyses described in Chapters 3, 4, and 5.

6.1 The Current State of *tipo* in the Carioca Dialect

The innovations in the use of the word *tipo* in Brazilian Portuguese have been previously documented (Bittencourt, 1999; Laurentino, 2016; Lima-Hernandes, 2005); however, to date, *tipo* has largely only been the focus of research in Brazilian Portuguese with respect to its functional expansion. To the best of the author's knowledge, this is the first investigation consisting of a multilayered analysis focusing on semantic, grammatical, pragmatic, phonetic, and psycholinguistic aspects of a single element that traces and compares its functional expansion from its original open class lexical meaning to a grammaticalized form. The results of this study expand our understanding of grammaticalization and its evolutionary processes.

When originally designed, this multi-methodological approach intended to explore the processes often associated with grammaticalization while examining which functions the noun *tipo* is performing in the Carioca dialect of Brazilian Portuguese.

However, the value of this unique research design went beyond the identification of how speakers are using *tipo* and into the structure of the word itself and how it is produced, perceived, and interpreted through objective and quantifiable measurements. The results, summarized in Table 6.1, show evidence of language change in real time.

Table 6.1. *Summary of the results.*

<i>Chapter</i>	<i>Result</i>	<i>Significance</i>
Chapter 3	Expansion in frequency and function	Results, in line with previous research on grammaticalization, show both a progressive increase in frequency of use of <i>tipo</i> in addition to the expansion of its grammatical and discourse functions in conversational speech
Chapter 4	Significant statistical differences in the production of nominal, grammatical, and discourse-pragmatic forms of <i>tipo</i>	Findings show that, in agreement with previous research, <i>tipo</i> is undergoing phonological reduction. However, differences in vowel quality were also shown to be significant when comparing nominal to grammatical

(preposition, conjunction, quotative complementizer) and discourse-pragmatic forms of *tipo* (discourse marker, discourse particle).

Chapter 5	Significant statistical differences in how participants perceived nominal, grammatical, and discourse-pragmatic forms of <i>tipo</i> aurally	Results show that speakers are conscious of the acoustic differences of <i>tipo</i> , indicating that they are not only able to discriminate its nominal form from the grammaticalized ones but also that they are using acoustic features to disambiguate them.
-----------	--	--

The interconnected layers of analyses led to evidence of the grammaticalization of *tipo*. Through the examination of conversational speech, findings show that its nominal form, still in use in the Carioca dialect, is decreasing in frequency, with its grammatical and discourse functions having seen a rapid increase in usage (see Chapter 3), substantiating findings from previous research that indicated that *tipo* was grammaticalizing (Bittencourt, 1999). The overall increase in frequency of forms that no longer carry *tipo*'s nominal properties (in addition to being found in structural constraints beyond those of its nominal form) is consonant with grammaticalization. As

a word undergoes semantic bleaching and expands to perform new functions, these expanded roles may be adopted by members of the speech community, resulting in an overall increase in frequency (Haspelmath, 1999).

In line with previous research that have examined the co-occurrence of grammaticalizing elements with other forms (Bybee & Hopper, 2001; Krug, 1998), this study also took into account string frequency, more specifically the co-occurrence of *tipo* with other words. This led to the identification and/or affirmation of the co-variants *tipo assim* and *tipo que* as well as the periphrastic quotative structure *ficar tipo*. Despite the fact that this step mainly aimed to provide the foundation for the subsequent functional analysis (Chapter 3), findings were still considered a valuable addition to the data analysis as string frequency has been shown to have an effect on word-formation as well as on "constitutionalization of compounds" (Krug, 1998) in which, orthographically, two separate words develop into a single word (Krug, 1998). A cursory examination of Twitter, Facebook, YouTube, and Google show that that may indeed be the case, with the coalescence of the co-variant *tipo assim* often being found in written form as either *tipoassim*, or, most frequently, *tipassim* (see Figures 6.2, 6.3, and 6.4 below).

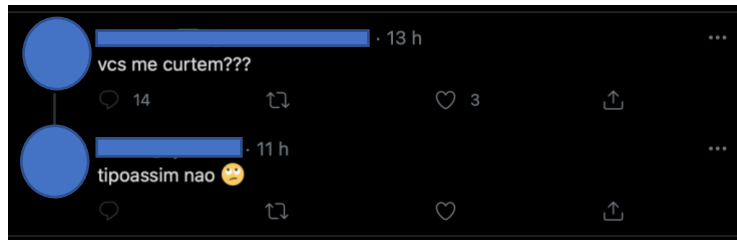


Figure 6.2. Tweet containing *tipoassim* from February 22, 2021.

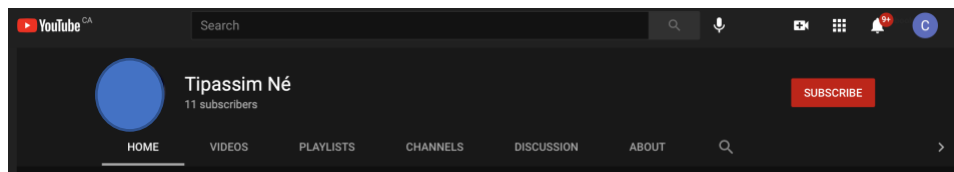


Figure 6.3. YouTube channel *Tipassim Né*.

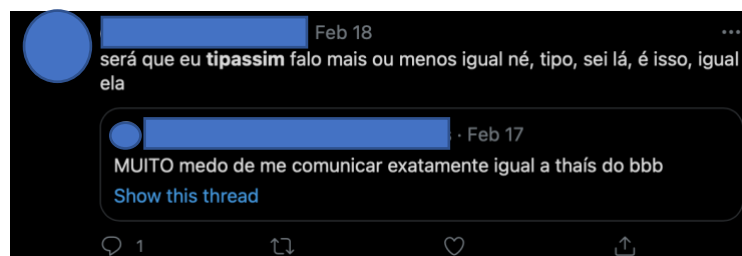


Figure 6.4. Tweet containing *tipassim* from February 18, 2021.

Finally, even though it is not possible to argue that increased word frequency on its own means that a linguistic element is, the results showing the increased usage of *tipo* still offer the first account of the progression of the use of *tipo* by teenage speakers in Rio de Janeiro. Being able to establish its progressive expansion over time was, therefore, an essential step into understanding the changes *tipo* is undergoing.

Having determined that speakers are indeed using non-nominal forms *tipo* with higher frequency, the significant increase (detailed in Chapter 3) raised questions as to how speakers use the word in naturalistic discourse. By examining where *tipo* is found in utterances and the functions it is performing, results shed light on its progression along the grammaticalization cline. The structural and positional regularities observed during the analysis of the PSCB corpus data not only revealed how speakers systematically use *tipo* (such as being frequently found in sentence-initial position), but also refutes the argument that it is used ad hoc, as argued by prescriptivists (see Bittencourt, 1999, for a discussion on this topic). Additionally, the corpus analysis also indicated that some of the functions *tipo* is performing are more recurrent than others, such as its role as a *discourse marker* being more frequent than that of a *conjunction*. This offers insights into the semantic, grammatical, and discourse changes it is undergoing (see D'Arcy, 2005 and Tagliamonte, 2005, for examples on English *like*). Through semantic bleaching and the loss of propositional meaning and nominal properties, speakers have expanded its use in conversational speech and are using *tipo* in innovative ways beyond those of its nominal form.

These findings leave us with an interesting question that is seldom addressed in the grammaticalization literature: is *tipo* currently undergoing grammaticalization or has it grammaticalized? Findings revealed that two of the functions identified in the PSCB corpus (namely *quotative complementizer* and *sentence-final adverb*) were not found

in the corpora of the Carioca dialect from the 1970s, 1990s, and 2000, suggesting an ongoing functional expansion (especially true in the case of *quotative complementizer*⁸², the most frequent of the two). It is difficult to predict how far along the grammaticalization cline *tipo* will go, but it is possible to state that such changes have not yet been finalized. Future research involving diachronic change may provide a historic progression of *quotative complementizer* and *sentence-final adverb* functions (the latter in special due to its low frequency of all functions) and may offer insights about whether these are indeed recent innovations.

This study also offers evidence that the current design not only provides researchers with an opportunity to explore some of the processes inherent to grammaticalization in more depth but may also provide objective measures to identify the grammaticalization of linguistic elements. So far, we have discussed the increase in frequency and functional expansion of *tipo*, both offering evidence of ongoing grammaticalization; however, such findings are limited to examining what *tipo* is doing, leaving questions as to what *tipo* really is. As grammaticalization has been shown to involve different processes such as *semantic* and *phonetic reduction*, it was hypothesized that changes resulting from such processes might result in distinguishable features among the different functions (minimally, between the canonical nominal form and the

⁸² Its use as a *quotative complementizer* also suggests changes to the quotative paradigm in BP.

grammaticalized non-nominal forms). The focus, then, was to explore potential differences in how speakers produced the different functions identified in the dataset, with the possible loss of phonetic substance, commonly associated with grammaticalization, being of special interest.

In agreement with previous studies that indicate that erosion is one of the main processes which elements undergo during grammaticalization (Bybee, 2003a; Bybee & Thompson, 1997; Krug, 2001), differences in total token duration were found to be significant when comparing the nominal and non-nominal functions of *tipo*. As discussed in Chapter 3, *phonological erosion/reduction* is commonly associated with ease of production, leading to durational differences between canonical forms and the expanded functions linguistic elements come to perform (Bybee, 2003a; 2003b). As words grammaticalize, they would become more frequent, resulting in the most frequently used forms eroding faster than less frequent ones (Bybee & Thompson, 1997; Haspelmath, 1999). That loss of phonetic substance would also be due to an increase in predictability of grammaticalized forms, leading to a lesser need for these forms to be phonologically salient (Haspelmath, 1999). In the case of *tipo*, however, differences between nominal and non-nominal forms of *tipo* were not limited to a reduction in total token duration or the elision of vowels. Findings also revealed stark differences in vowel quality between *tipo*'s nominal and expanded grammatical and discourse forms, which, until now, had remained undocumented in BP. Furthermore, the affricate [tʃ]

was found to be significantly longer in nominal *tipo* when compared to the other forms. Similar findings were discussed by Podlubny, Geeraert, and Tucker (2015) in their analysis of *like* in Canadian English, with correlations found between the variability in the degree of diphthongization and the function it performs as well as between segment duration and function.

Such shifts in production between nominal *tipo* and grammatical and discourse-pragmatic *tipo* at the segmental level indicate that, in addition to overall word duration, speakers may make use of other acoustic cues to disambiguate its nominal and non-nominal forms. For instance, the elision of the /o/ vowel was shown to be significant in the data, with speakers eliding that vowel in over 32% of the tokens of *tipo* as it performs functions other than its referential nominal form (see Chapter 4). These findings were not only corroborated by the results from the discrimination task (which shows that speakers are able to discriminate the elision or presence of *tipo*) (see Chapter 5) but also by how the word is being reproduced orthographically, especially in the case of *tipo assim*. A cursory examination of Twitter shows that the elision of the /o/ vowel in *tipo assim* is recurrently found in tweets from Brazilian users (with some even including the hashtag #tipassim). The coalescence of the forms, therefore, is not only represented in how speakers produce the phrase but also on how they write it. Figures 6.5 and 6.6 illustrate.



Figure 6.5. *Tipassim* in tweets from February 21 and February 26, 2021.

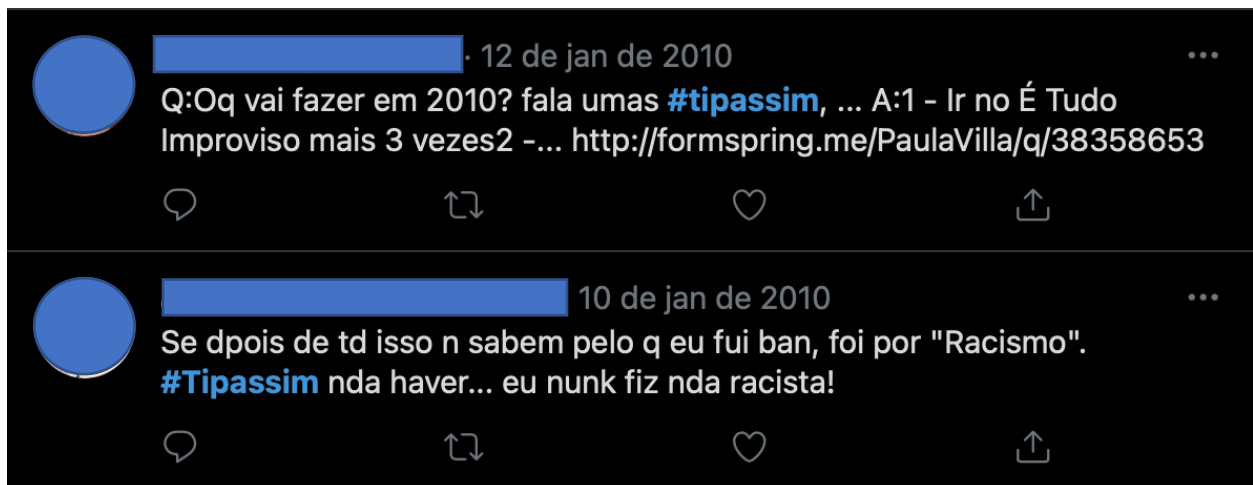


Figure 6.6. Hashtag *#tipassim* showing the elision of the /o/ vowel and coalescence of forms.

The results from the acoustic analysis (Chapter 4) suggest that the term *phonetic erosion* (currently defined as the loss of phonetic segments, of suprasegmental properties, of phonetic autonomy, or as a form of phonetic simplification) (Heine & Kuteva, 2007) may need to be expanded on from its current definition to include

changes at the segment level (such as vowel quality, elision, and segment duration)⁸³. These changes, altogether, seem to result in acoustic differences between nominal and grammaticalized forms of *tipo* that go beyond the encompassing "durational reduction" concept. In light of these findings, the author hypothesizes that a possible *grammaticalization sound shift* could explain the process by which different functions are being acoustically encoded as *tipo* grammaticalizes⁸⁴. *Grammaticalization sound shift* might indicate a possible phonological tendency towards acoustic distinction at the segment level as *tipo* moves along the grammaticalization cline. The changes to the acoustic properties of *tipo* in conjunction with the other processes shown to take place during grammaticalization (such as *divergence* and *generalization*) might both encode and signal that *tipo* is moving along the cline.

If we were to assume that the systematic acoustic differences found between nominal and non-nominal forms of *tipo* are the result of a *grammaticalization sound shift*, it is possible to postulate that speakers may be making use of those cues to guide listeners' predictions of upcoming utterances. This implies that the process of grammaticalization of *tipo* may be intrinsically related to the acquisition of acoustic properties that would become inherent to the grammaticalized forms, thus

⁸³ This argument refers specifically to the erosion of *tipo*.

⁸⁴ It is important to clarify that this hypothesis is being raised to better explain the results from the analysis of *tipo* and is, therefore, limited to the current study.

disambiguating the nominal form from the other forms. The case of *tipo* seems to be quite complex, going beyond a simple case of either polysemy or homophony. It is possible that the broad shifts *tipo* is undergoing (both lexical-semantic and content-functional) can explain why speakers feel the need to disambiguate nominal from non-nominal functions

If we consider *grammaticalization sound shift* as a process that speakers are using to encode function as *tipo* grammaticalizes and we assume that the process is aimed to disambiguate functions in an attempt to guide speakers' interpretation of upcoming utterances, it would be logical to postulate that listeners would have the ability to discriminate nominal from non-nominal forms of *tipo* based on acoustic cues beyond total token duration. Results from the discrimination task conducted in this study (Chapter 5) reveal that that is indeed the case. Although the task consisted only in testing participants' ability to discriminate tokens with either a vowel or elided vowel(s) (one of the contrasting features between the nominal and grammaticalized forms), significant statistical differences show that listeners were able to consistently perceive the differences aurally, indicating that they are capable of discriminating at least the two broader categories of nominal and non-nominal forms.

Together, the present findings suggest that the nominal form of *tipo* and its grammaticalized forms may be homonyms rather than one single polysemous word, i.e., they would be represented by different lexical units (lemmas) (Levelt, 1989).

Although their meanings are not related to different sources, results from both the acoustic and perception analyses show that speakers and listeners are aware of the acoustic differences between both forms and are able to disambiguate them. This provides an important insight into speakers' cognitive abilities and the nature of language processing as language change takes place. These results raise questions about the interactions among the various parts of the linguistic system and how language change is encoded and interpreted.

This dissertation hopes to contribute to the current body of work on *tipo* in Brazilian Portuguese and to studies on language change and grammaticalization by offering a unique multi-methodological approach that was shown to provide objective and objective and quantifiable evidence of linguistic elements undergoing grammaticalization. Figure 6.7 presents a chart that summarizes the pronunciation and semantic variation of *tipo* in each of its forms.

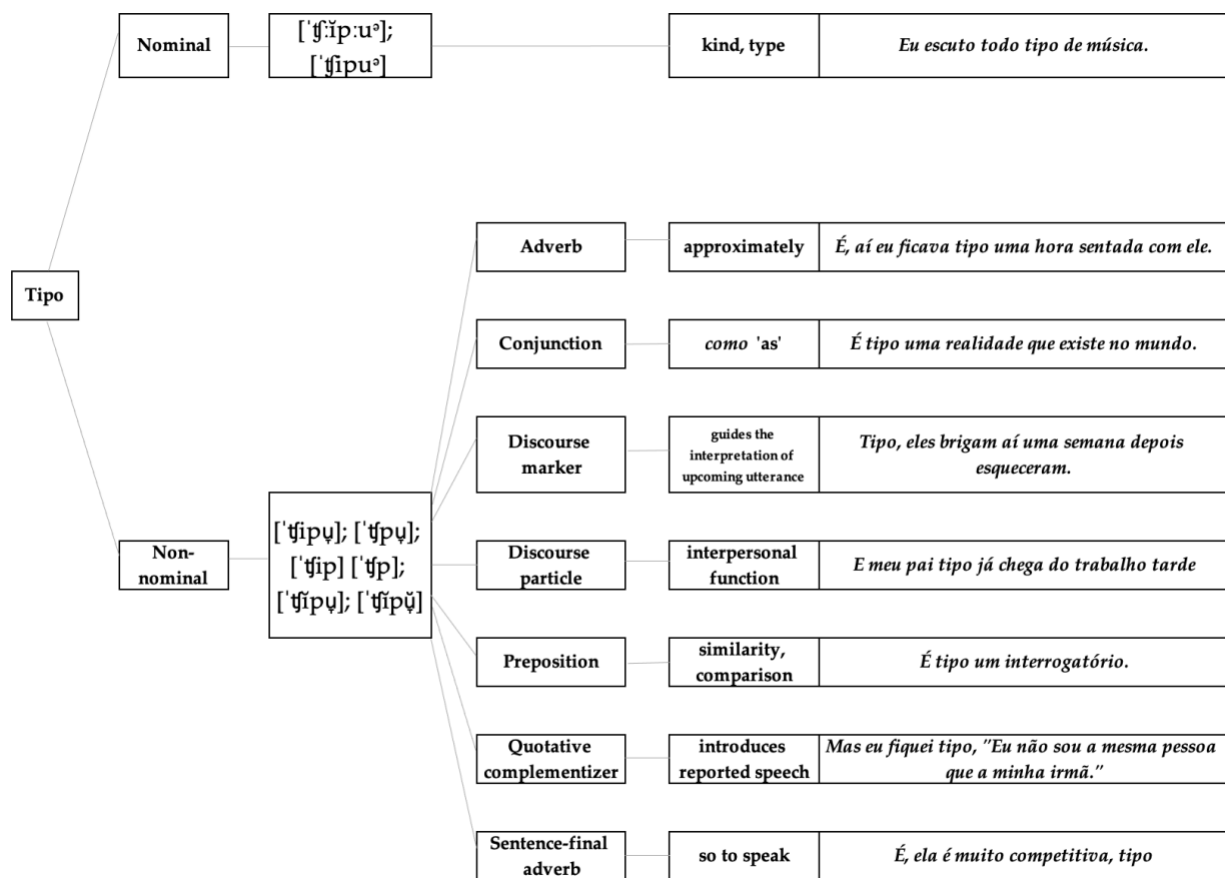


Figure 6.7. The functions of *tipo* in the PSCB corpus including transcriptions, definitions, and examples.

6.2 Limitations

It is important to state that the paucity of studies providing sequential analyses of linguistic elements within the theory of grammaticalization makes the contextualization of some of the findings within the current literature challenging.

Additionally, the results presented in this dissertation are constrained to the age group under investigation and, as a result, do not offer an insight into the current state of *tipo* among the general population of Rio de Janeiro. Future studies should consider

including different age groups to gain access to a more encompassing understanding of the uses of *tipo* in conversational speech.

Regarding the acoustic analysis, it would have been valuable to compare the acoustic realization of each different function of *tipo* to one another as a way to better understand how similar or dissimilar they are⁸⁵. In possession of such results, a discrimination task that includes the most salient acoustic features of the different functions should be conducted to explore speakers' ability to discriminate the different functions of *tipo*.

Finally, as the discrimination task was designed prior to the conclusion of the acoustic analysis, only *elision* of vowels (and resulting durational differences) was included as a contrasting feature. Despite the significant results found, the study would have benefited from including other acoustic differences (such as vowel quality) to better establish participants' ability to discriminate more subtle acoustic features.

6.3 Future Research

As an exploration of the process of grammaticalization and the grammaticalization of the noun *tipo* in the Carioca dialect of Brazilian Portuguese, this study offers both synchronic evidence of the grammaticalization of *tipo* and a possible framework for the objective analysis of linguistic elements undergoing grammaticalization. Its design,

⁸⁵ This analysis was not included in this dissertation due to time constraints but will be addressed in a separate paper.

innovative in nature, reveals change in real time and offers a new approach to establishing and confirming the grammaticalization of linguistic items.

It would be valuable to apply the current framework to examine the equivalents of the word *tipo* in languages in which it has been shown to be undergoing grammaticalization (such as Argentinian Spanish and Italian) (Fernández, 2017; Voghera, 2013, respectively) and in those in which it remains uninvestigated (such as Galician). Results may provide an important insight into whether speakers of other languages are also encoding the expanded functions *tipo* equivalents is performing at the segment level.

Although this study aimed to provide an encompassing investigation of *tipo*, some areas of the analysis have not been explored to their full potential. One such example is the construction *tipo que*, which deserves a more detailed investigation. The presence of *que*, a multifunctional word, alongside *tipo* was unexpected and had not yet been mentioned in previous analysis of *tipo*. Researchers might want to examine its role in this periphrastic construction as a possible new layer of complexity to one of BP's most productive grammatical items.

Additionally, changes in the current quotative paradigm in Brazilian Portuguese should be examined. Although the results of the study account for the use of *tipo* as a quotative complementizer, other structures were identified during the transcription stage of the interviews in the PSCB corpus. The richness of the excerpts found in the

data indicates a very clear change in progress of how speakers are introducing direct quotes, making this a very timely follow-up project. Further research might provide a clearer picture of the current state of the quotative paradigm and offer clues as to what changes are currently taking place in how speakers are introducing direct speech.

Finally, the use of *tipo* in constructions consisting subject doubling (e.g., *Eu e minha mãe a gente ficou desesperada* 'My mom and I we were desperate') in the Carioca dialect was found to be a common phenomenon, especially in utterances in which the subject was the first person singular *eu* 'I' (such as in *Eu tipo eu nem posso namorar* 'I like I can't even date'). Subject doubling was identified in countless examples in the PSCB corpus, indicating that it is a productive structure in that dialect. Having received no attention from researchers to date, an investigation of the role of *tipo* in those constructions (including the investigation of its distribution and function) may offer another layer of information on its role in enabling listeners to better understand utterances.

References

- Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *Journal of the Acoustical Society of America*, 116(5), 3099–3107.
- Aijmer, K. (1997). I think—An English modal particle. In T. Swan & O. J. Westvik (Eds.), *Modality in Germanic Languages: Historical and Comparative Perspectives* (pp. 1–47). Mouton de Gruyter.
- Alami, M. (2015). Pragmatic functions of discourse markers: A review of related literature. *International Journal on Studies in English Language and Literature*, 3(3), 1–10.
- Alkmim, T. (2012). Um texto inaugural: O Visconde da Pedra Branca e o português do Brasil. *Stockholm Review of Latin American Studies*, 8, 21–33.
- Altmann, C. F., Uesaki, M., Ono, K., Matsushashi, M., Mima, T., & Fukuyama, H. (2014). Categorical speech perception during active discrimination of consonants and vowels. *Neuropsychologia*, 64, 13–23.
- Andersen, G. (2000). *Pragmatic markers and sociolinguistic variation: A relevance-theoretic approach to the language of adolescents*. John Benjamins Publishing.
- Assessing the slave trade* (Voyages: The Trans-Atlantic Slave Trade Database). (2013). Emory University. <http://slavevoyages.org/assessment/estimates>
- Auger, J. (1998). Le redoublement des sujets en français informel québécois: Une approche variationniste. *Canadian Journal of Linguistics*, 43(1), 37–63.

- Avelar, J., & Galves, G. (2014). O papel da línguas africanas na emergência da gramática do português brasileiro. *Linguística*, 30(2), 241–288.
- Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31–56.
- Ayrosa, P. (Ed.). (1938). *Vocabulário na língua brasilica: Manuscrito português-tupi do século XVII, coordenado. E prefaciado por Plínio Ayrosa: Vol. XX*. Departamento de Cultura.
- Azevedo, M. M. (1981). *A contrastive phonology of Portuguese and English*. Georgetown University Press.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *Journal of Psychological Research*, 3(2), 12–28.
- Balbi, A. (1826). Introduction à l'atlas ethnographique du globe: contenant un discours sur l'utilité et l'importance de l'étude des langues appliquée à plusieurs branches des connaissances humaines; un aperçu sur les moyens graphiques employés par les différents peuples de la terre; des observations sur la classification des idiomes décrits dans l'atlas; un coup d'oeil sur l'histoire de la langue slave et sur la marche progressive de la civilisation et de la littérature en Russie (Vol. 1). Rey et Gravier.
- Barbosa, P. A., & Albano, E. C. (2004). Brazilian Portuguese. *Journal of the International Phonetic Association*, 34(2), 227–232.

- Barbosa, Pe. A. L. (1951). *Pequeno vocabulário tupi-português com quatro apêndices: Perfil da língua tupi, palavras compostas e derivadas, metaplasmos, síntese bibliográfica*. Livraria São José.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bauer, L. (1994). Productivity. In K. Brown (Ed.), *The Encyclopedia of Language and Linguistics* (3354 - 3357). Pergamon.
- Belchior, E. (2008). Estácio de Sá e a fundação do Rio de Janeiro. *História*, 27(1), 77–99.
- Bentes, A. C., Mariano, R. D., & Ferreira-Silva, B. (2013). Marcadores discursivos e sequências textuais no programa “Manos e Minas”: Uma análise inicial para a tipificação do programa em relação a aspectos textuais-discursivos. *Sociodialeto*, 3(9), 239–265.
- Bergs, A., & Diewald, G. (Eds.). (2008). *Constructions and language change* (Vol. 194). Mouton de Gruyter.
- Bertozzo, A. F. (2014). *De conector a marcador discursivo: “Como,” “que nem” e “tipo” em Chapecó* [M.A. Thesis]. Universidade Federal da Fronteira Sul.
- Biber, D., & Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text-Interdisciplinary Journal for the Study of Discourse*, 9(1), 93–124.

- Bickerton, D. (1971). Inherent variability and variable rules. *Foundations of Language*, 7(4), 457–492.
- Bills, D., & Haller, A. (1984). Socioeconomic development and social stratification: Reassessing the Brazilian case. *The Journal of Developing Areas*, 19(1), 59–70.
- Bindra, D., Williams, J. A., & Wise, J. S. (1965). Judgments of sameness and difference: Experiments on decision time. *Science*, 150(3703), 1625–1627.
- Bisang, W. (1998). Grammaticalization and language contact, constructions and positions. In A. G. Ramat & P. J. Hopper (Eds.), *The limits of grammaticalization* (pp. 13–58). John Benjamins Publishing.
- Bittencourt, V. (n.d.). Tipo (assim) como delimitador de “unidades de informação.” *Revista de Estudos Linguísticos*, 29, 264–269.
- Bittencourt, V. (1999). Gramaticalização e discursivização no português oral do Brasil: O caso “tipo (assim).” *Scripta*, 2(4), 39–53.
- Blakemore, D. (2006). Discourse markers. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 221–240). Blackwell Publishing.
- Blyth, C., Recktenwald, S., & Wang, J. (1990a). I’m like, “Say what?!”: A new quotative in American oral narrative. *American Speech*, 65(3), 215–227.
- Blyth, C., Recktenwald, S., & Wang, J. (1990b). I’m like, “say what?!”: A new quotative in American oral narrative. *American Speech*, 65(3), 215–227.

- Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer* [Computer program] (6.0.36) [Computer software]. <http://www.praat.org/>
- Bolden, G. B. (2009). Implementing incipient actions: The discourse marker 'so' in English conversation. *Journal of pragmatics*, 41(5), 974-998.
- Bonvini, E. (2009). Línguas africanas e português falado no Brasil. In J. L. Fiorin & M. M. T. Petter (Eds.), *África no Brasil: A formação da língua portuguesa* (pp. 15–62). Contexto.
- Bopp, F. (1816). *On the conjugational system of the Sanskrit language-In comparison with that of Greek, Latin, Persian and the Germanic languages*.
- Bradshaw, M. F., Hibbard, P. B., & Gillam, B. (2002). Perceptual latencies to discriminate surface orientation in stereopsis. *Perception & Psychophysics*, 64(1), 32–40.
- Braga, M. L. (2003). Aí e então em expressões cristalizadas. *Cadernos de Estudos Lingüísticos*, 44, 169–177.
- Braga, M. L., & PAIVA, M. D. C. (2003). Do advérbio ao clítico é isso aí. *Português brasileiro: contato lingüístico, heterogeneidade e história*. No prelo.
- Brinton, L. J. (1996). *Pragmatic markers in English: Grammaticalization and discourse functions*. Mouton de Gruyter.
- Brinton, L. J. (2008). *The comment clause in English syntactic origins and pragmatic developments*. Cambridge University Press.
- Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge University Press.

- Browman, C., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, 6, 201–251.
- Bublitz, W. (2011). Cohesion and coherence. In J. Zienkowski, J. Östman, & J. Verschueren (Eds.), *Discursive pragmatics* (Vol. 8, pp. 37–49). John Benjamins Publishing.
- Buchstaller, I. (2001). An alternative view of like: Its grammaticalisation in conversational American English and beyond. *Edinburgh Working Papers in Applied Linguistics*, 11, 21–41.
- Bybee, J. (2003a). Cognitive processes in grammaticalization. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (Vol. 2, pp. 145–167). Lawrence Erlbaum Associates Inc., Publishers.
- Bybee, J. (2003b). Mechanisms of change in grammaticization: The role of frequency. In B. D. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 602–623). Blackwell Publishing.
- Bybee, J. (2003c). *Phonology and language use* (Vol. 94). Cambridge University Press.
- Bybee, J. L. (2011). Usage-based theory and grammaticalization. In B. Heine & H. Narrog (Eds.), *The Oxford handbook of grammaticalization* (pp.1–7). Oxford University Press.
- Bybee, J., & Hopper, P. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. John Benjamins Publishing.
- Bybee, J., & Pagliuca, W. (1985). Cross-linguistic comparison and the development of grammatical meaning. In J. Fisiak (Ed.), *Historical semantics, historical word formation* (pp. 59–83). de Gruyter.

- Bybee, J., Pagliuca, W., & Perkins, R. (n.d.). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.
- Bybee, J., & Thompson, S. (1997). Three frequency effects in syntax. *Annual Meeting of the Berkeley Linguistics Society*, 23, 378–388.
- Byron, D. K., & Heeman, P. A. (1997). Discourse marker use in task-oriented spoken dialog\lambda. In *Fifth European Conference on Speech Communication and Technology*.
- Cacoullos, R. T., & Walker, J. A. (2011). Collocations in grammaticalization and variation. In H. Narrog & B. Heine (Eds.), *The Oxford handbook of grammaticalization* (pp. 225–238). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199586783.013.0018>
- Cagliari, L. C. (2002). *Análise fonológica: Introdução à teoria e à prática: Com especial destaque para o modelo fonémico*. Mercado das Letras.
- Callou, D. (2009). Um perfil da fala carioca. In A. Serpa, C. A. da Costa, C. El-Hani, D. E. L. Ramacciotti, J. T. Cavalcante Filho, & A. B. Novaes (Eds.), *Dons sons às palavras: Nas trilhas da língua portuguesa* (pp. 1–345). Editora da Universidade Federal da Bahia.
- Câmara Jr., J. M. (1972). *The Portuguese language*. University of Chicago Press.
- Câmara Jr., J. M. (1977). *Para o estudo da fonêmica portuguesa* (2nd ed.). Editora Padrão.
- Câmara Jr., J. M. (1986). *Estrutura da língua portuguesa* (16th ed.). Editora Vozes.
- Campbell, L. (1999). *Historical linguistics: An introduction*. The MIT Press.

- Cardoso, D. P. (2009). *Fonologia da língua portuguesa*. Universidade Federal de Sergipe. [http://nead.uesc.br/arquivos/Letras/fonetica_fonologia/Fonologia da L%C3%ADngua Portuguesa.pdf](http://nead.uesc.br/arquivos/Letras/fonetica_fonologia/Fonologia_da_L%C3%ADngua_Portuguesa.pdf)
- Carioca. (n.d.). In *Dicionário Ilustrado Tupi-Guarani*. <https://www.dicionariotupiguarani.com.br/dicionario/carioca/>
- Carvalho, A. M. (2014). Sociolinguistic continuities in language contact situations: The case of Portuguese in contact with Spanish along the Uruguayan-Brazilian border. In P. Amaral & A.M. Carvalho (Eds.), *Portuguese-Spanish Interfaces* (pp. 263-294). John Benjamins.
- Carvalho, A. M., & Lucchesi, D. (2020). Portuguese in contact. In W. L. Wetzels, J. Costa, & S. Menuzzi (Eds.), *The Handbook of Portuguese Linguistics* (pp. 41-55). Wiley Blackwell.
- Castelano, K. L., & Ladeira, W. T. (2010). Funções discursivo-interacionais das expressões “assim”, “tipo”, e “tipo assim” em narrativas orais. *Letra Magna*, 6(12), 1–17.
- Cavalcanti, N. O. (2004). *O Rio de Janeiro setecentista: a vida e a construção da cidade da invasão francesa até a chegada da Corte*. Zahar.
- Cavallieri, F. & Lopes, G. P. (2006). Favelas cariocas: Comparação das áreas ocupadas - 1999/2004. IPP/Prefeitura da Cidade do Rio de Janeiro
- Cavallieri, F. & Vial, A. (2010). Favelas na cidade do Rio de Janeiro: O quadro populacional com base no Censo 2010. IPP/Prefeitura da Cidade do Rio de Janeiro

Censo demográfico 2010: Aglomerados subnormais — Primeiros resultados (pp. 1–259). (2010).

Instituto Brasileiro de Geografia e Estatística - IBGE.

Chambers, J. K. (1989). Canadian raising: Blocking, fronting, etc. *American Speech*, 64(1), 75–88.

Chambers, J. K. (2003). *Sociolinguistic theory: Linguistic variation and its social significance* (2nd ed.). Blackwell Publishing.

Chambers, J. K., & Trudgill, P. (1998). *Dialectology*. Cambridge University Press.

Chomsky, N. (1959). A note on phrase structure grammars. *Information and Control*, 2(4), 393–395.

Chomsky, N. (2002). *Syntactic structures*. Mouton de Gruyter.

Clark, J., Yallop, C., & Fletcher, J. (2008). *An introduction to phonetics and phonology* (Third Edition). Blackwell Publishing.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology*, 19(1), 81–94.

Coseriu, E. (1978). *Sincronía, diacronía e historia: El problema del cambio lingüístico*. Gredos.

Costa, D. B., & de Azevedo, U. C. (2016). Das senzalas às favelas: Por onde vive a população negra brasileira. *Socializando*, 3(1), 145–154.

Cristino, B. (2004). O papel do negro na formação do português brasileiro na visão dos estudiosos dos anos 1920 a 1945. *Centro de Documentação Em Historiografia Da Lingüística*, 2001, 45–60.

- Cristófaros-Silva, T. (1998). *Fonética e fonologia do português*. Editora Contexto.
- Cristófaros-Silva, T., & de Oliveira, M. A. (2001). *Lateral vocalization in Brazilian Portuguese*. 1–8.
- Cunha, C., & Cintra, L. (2000). *Breve gramática do português contemporâneo*. João Sá da Costa.
- da Silva, E. (2011). *Sociolinguistic (re)constructions of diaspora potugueseness: Portuguese-Canadian youth in Toronto* [Dissertation]. University of Toronto.
- da Silva Pacheco, C., dos Santos, T. R. S., & Oliveira, R. S. (n.d.). Análise contrastiva da variação do objeto direto no espanhol cubano e no português brasileiro de um falante bilíngüe. *Revista de Letras*, 10(2), 19–31.
- D’Arcy, A. (2005). *Like: Syntax and Development* [Dissertation]. University of Toronto.
- D’Arcy, A. (2011). Corpora: Capturing language in use. In *Analysing Variation in English* (pp. 49–71). Cambridge University Press.
- D’Arcy, A. F. (2007). Like and language ideology: Disentangling fact from fiction. *American Speech*, 82(4), 386–419.
- D’Arcy, A. F. (2017). *Discourse-pragmatic variation in context: Eight hundred years of LIKE*. John Benjamins Publishing.
- de Alencastro, L. F. (1997). Vida privada e ordem privada no Império. *História da Vida privada do Brasil II*. São Paulo: Companhia das Letras, 12.

- de Almeida, D. M. V., & Marinho, J. H. C. (2012). Dos marcadores discursivos e conectores: Conceituação e teorias subjacentes (About discourse markers and connectors: Concepts and underlying theories). *Gláuks*, 12(1), 169–203.
- de Araújo, D. P. B. (2014). A gramaticalização da partícula lá. *Linguística Textual e Pragmática*, XVIII, 339–352.
- de Brito, E. P. (2016). The Variation of Post-Vowel-S in the Portuguese Dialect of Rio de Janeiro. *Cadernos de Estudos Lingüísticos*, 58(1), 139–151.
- de Carvalho, D. (1990). *História da cidade do Rio de Janeiro* (2nd ed.). Departamento Geral de Documentação e Informação Cultural da Secretaria Municipal de Cultura, Turismo e Esportes. http://www.rio.rj.gov.br/dlstatic/10112/4204210/4101378/historia_cidade_rio_janeiro.pdf
- de Carvalho, M. R. (1987). *Dicionário tupi (antigo)-português*. Biblioteca Digital Curt Nimuendajú.
- de Castilho, C. M. M. (1991). *Os delimitadores no português falado no Brasil* [M.A. Thesis]. Universidade Estadual de Campinas (UNICAMP).
- de Castro, Y. P. (1983). Das línguas africanas ao português brasileiro. *Afro-Ásia*, 14, 81–106.
- de Jong, K. J., Lim, B., & Nagao, K. (2004). The perception of syllable affiliation of singleton stops in repetitive speech. *Language and Speech*, 47(3), 241–266.
- de Lima, S. O. (2005). Elisão de fonemas na cadeia de fala. *Filologia e Linguística Portuguesa*, (7), 185-193.

- de Matos, M. P., & Sandalo, M. F. S. (2006). Síncope vocálica no português brasileiro. *Anais Do 6º Encontro Celsul - Círculo de Estudos Lingüísticos Do Sul*. Celsul - Círculo de Estudos Lingüísticos do Sul, Brasil.
- de Mello, H. R. (1996). *The genesis and Development of Brazilian vernacular Portuguese* [Dissertation]. The City University of New York.
- de Oliveira, M. R., & dos Santos, L. P. (2011). Padrões de uso da expressão “sei lá” no português. *Signótica*, 23(2), 363–384.
- de Paiva, M. C., & Gomes, C. A. (2016). Grupo PEUL: Passado, presente e futuro de uma agenda de pesquisa. *Estudos Lingüísticos*, 58(3), 503–519.
- de Sá, A. R. (2017, January 10). *5 gírias que podem levar a sua imagem pessoal para o precipício* [Professional development website].
administradores.com.br. administradores.com.br/artigos/5-girias-que-podem-levar-a-sua-imagem-pessoal-papa-o-precipicio
- de Vogelaer, G., & Neuchermans, A. (2002). Subject doubling in Dutch: A dialect phenomenon in cross-linguistic perspective. *STUF- Language Typology and Universals*, 55(3), 234–258.
- Dias, A. G. (1858). *Diccionario da lingua Tupy: Chamada lingua geral dos indigenas do Brazil* (Vol. 1). FA Brockhaus.
- Dias, P. (1697). *A arte da lingua de Angola*. Officina de Miguel Deslandes.

- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Annual Review of Psychology*, 55, 149–180.
- Dilts, P. C. (2013). *Modelling phonetic reduction in a corpus of spoken English using random forests and mixed-effects regression*. [Dissertation]. University of Alberta.
- Drager, K. K. (2011). Sociophonetic variation and the lemma. *Journal of Phonetics*, 39(4), 694–707.
- Drager, K. K. (2015). *Linguistic variation, identity construction and cognition*. Language Science Press.
- Du Bois, J. W. (1985). Competing motivations. In J. Hairman (Ed.), *Iconicity in syntax: Proceedings of a symposium on iconicity in syntax* (Vol. 6, pp. 343–365). John Benjamins Publishing.
- Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, 64, 199–210.
- Dweik, B. S., & Qawar, H. A. (2015). Language choice and language attitudes in a multilingual Arab Canadian community: Quebec-Canada: A sociolinguistic study. *British Journal of English Linguistics*, 3(1), 1–12.
- Eckert, P. (1977). Age as a sociolinguistic variable. In F. Coulmas (Ed.), *The handbook of sociolinguistics* (pp. 151–167). Blackwell.
- Eckert, P. (1988). Adolescent social structure and the spread of linguistic change. *Language in Society*, 17(2), 183–207.

- Eckert, P. (2014). Language and gender in adolescence. In S. Ehrlich, M. Meyerhoff, & J. Holmes (Eds.), *The Handbook of Language, Gender, and Sexuality* (2nd ed.). Wiley Blackwell.
- Eckert, P. (2018). *Meaning and linguistic variation: The third wave in sociolinguistics*. Cambridge University Press.
- Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants [Editorial]. *Journal of Phonetics*, 39(SI), 253–260.
- Escudero, P., Boersma, P., Rauber, A. S., & Bion, R. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America*, 123, 1–15.
- Farell, B. (1985). "Same"-"Different" judgments: A review of current controversies in perceptual comparisons. *Psychological Bulletin*, 98(3), 419–456.
- Fernández, J. (2017). The language functions of *tipo* in Argentine vernacular. *Journal of Pragmatics*, 114, 87–103.
- Ferreira, F. (2009). Marcadores de plural no português brasileiro e no crioulo cabo-verdiano. In A. M. Carvalho (Ed.), *Português em contato* (Vol. 2). Iberoamericana Editorial.
- Fleischman, S., & Yagello, M. (2004). Discourse markers across languages: Evidence from English and French. In C. L. Moder & A. Martinovic-Zic (Eds.), *Discourse across languages and cultures* (Vol. 68, pp. 129–147). John Benjamins Publishing.

- Flores-Ferrán, N. (2014). So pues entonces: An examination of bilingual discourse markers in Spanish oral narratives of personal experience of New York City-born Puerto Ricans. *Sociolinguistic Studies*, 8(1), 57–83.
- Flynn, N., & Foulkes, P. (2011). Comparing vowel formant normalization methods. *ICPhS XVII*. ICPhS, Hong Kong.
- 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.
- Franken, M. K., McQueen, J. M., Hagoort, P., & Acheson. (2015). *Assessing the link between speech perception and production through individual differences*. International Conference of Phonetic Sciences, Glasgow, UK.
- Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungarica*, 38(1–4), 19–33.
- Fraser, B. (1990). An approach to discourse markers. *Journal of Pragmatics*, 14(3), 383–398.
- Freitag, R. M. K., Martins, M. A., & Tavares, M. A. (2012). Bancos de dados sociolingüísticos do português brasileiro e os estudos de terceira onda: Potencialidades e limitações. *Alfa*, 56(3), 917–944.
- Friedemann, M. A., & Rizzi, L. (Eds.). (2014). *The acquisition of syntax: Studies in comparative developmental linguistics*. Routledge.

- Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Probus*, 13(2), 247–275.
- Fujimura, O., & Ochiai, K. (1963). Vowel identification and phonetic contexts. *The Journal of the Acoustical Society of America*, 35(11), 1889–1889.
- Fuller, J. M. (n.d.). Use of the discourse marker like in interviews. *Journal of Sociolinguistics*, 7(3), 365–377.
- Gabelentz, G. von der. (1901). *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Weigel.
- Gahl, S. (2008). ‘Time’ and ‘thyme’ are not homophones: The effect of lemma frequency on word durations in a corpus of spontaneous speech. *Language*, 84(3), 474–496.
- Galves, C., Moraes, M. A. T., & Ribeiro, I. (2005). Syntax and morphology in the placement of clitics in European and Brazilian Portuguese. *Journal of Portuguese Linguistics*, 4, 143–177.
- Gaspar, L. (2012). *Línguas indígenas no Brasil* (p. 1) [Government report]. Fundação Joaquim Nabuco. http://basilio.fundaj.gov.br/pesquisaescolar/index.php?option=com_content&view=article&id=832:linguas-indigenas-no-brasil&catid=47:letra-l
- Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, 66(3), 363–376.

- Gillam, B., Chambers, D., & Russo, T. (1988). Postfusional latency in stereoscopic slant perception and the primitives of stereopsis. *Journal of Experimental Psychology*, 14(2), 163–175.
- Góis, M. L. S., & Martins, A. M. S. (2019). O tupi antigo no português: Algumas questões sobre história, identidade e ensino de linguagem. *Trabalhos Em Lingüística Aplicada*, 58(1), 422–440.
- Goldberg, I. (1922). *Brazilian literature*. Knopf.
- Gonçalves, C. A. (n.d.). Foco e topicalização: Delimitação e confronto de estruturas. *Revista de Estudos Da Linguagem*, 7(1), 31–50.
- Gonzalez, L., & Hasenbalg, C. (1982). *Lugar de negro* (Vol. 3). Editora Marco Zero Limitada.
- Gorski, E., Gibbon, A. D. O., Valle, C. R. M., Dal Mago, D., & Tavares, M. A. (2003). Fenômenos discursivos: Resultados de análises variacionistas como indícios de gramaticalização (Discursive phenomena: Results of variationist analyses as indices of grammaticalization). In C. Roncarati & J. Abraçado (Eds.), *Português brasileiro: Contato linguístico, heterogeneidade e história* (pp. 106–122). 7Letras.
- Gottfried, T. L., & Strange, W. (1980). Identification of coarticulated vowels. *Journal of the Acoustical Society of America*, 68(6), 1626–1635.
- Guy, G. R. (1989). On the nature and origins of Popular Brazilian Portuguese. In *Estudios sobre español de América y lingüística afroamericana: ponencias presentadas en el 45 Congreso*

- Internacional de Americanistas (Bogotá, julio de 1985)* (pp. 227-245). Instituto Caro y Cuervo.
- Guy, G. R. (1993). The quantitative analysis of linguistic variation. In D. Preston (Ed.), *American dialect research* (pp. 223–249). John Benjamins Publishing.
- Guy, G. R. (2003). Variationist approaches to phonological change. In B. D. Joseph & R. D. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 369–400). Blackwell Publishing.
- Guy, G. R. (2013). The cognitive coherence of sociolects: How do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics*, 52, 63–71.
- Guy, G. R., & Ziles, A. M. (2008). Endangered language varieties: Vernacular speech and linguistic standardization in Brazilian Portuguese. In *Sustaining linguistic diversity: Endangered minority languages and language varieties* (pp. 53–66). Georgetown University Press.
- Hall, N. (2011). Vowel epenthesis. In *The Blackwell companion to phonology* (pp. 1576–1596). Blackwell Publishing.
- Halliday, M. A. K., & Hasan, R. (1989). *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective* (2nd ed. edition). Oxford Univ Press.
- Halliday, M. A. K., & Hasan, R. (2014). *Cohesion in English*. Routledge.
- Hansen, M. B. M. (1998). The semantic status of discourse markers. *Lingua*, 104(3–4), 235–260.
- Haspelmath, M. (1999). Why is grammaticalization irreversible? *Linguistics*, 37(6), 1043–1068.

- Haspelmath, M. (2000). Periphrasis. In G. E. Booij, C. Lehmann, & J. Mugdan (Eds.), *Morphology: A handbook on inflection and word formation*. de Gruyter.
- Heine, B. (2003). Grammaticalization. In B. D. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 575–601). Blackwell Publishing.
- 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 Publishing.
- Heine, B., & Kuteva, T. (2004). *The world lexicon of grammaticalization*. Cambridge University Press.
- Heine, B., & Kuteva, T. (2007). *The genesis of grammar: A reconstruction*. Oxford University Press.
- Heine, B., & Narrog, H. (2010). Introduction. In *The Oxford handbook of linguistic analysis* (pp. 1–26). Oxford University Press.
- Herold, R. (1990). *Mechanisms of merger: The implementation and distribution of the low back merger in Eastern Pennsylvania* [Dissertation]. University of Pennsylvania.
- Herold, R. (1997). Solving the actuation problem: Merger and immigration in eastern Pennsylvania. *Language Variation and Change*, 9, 165–169.
- Hinrichs, L. (2011). The sociolinguistic of Diaspora: Language in the Jamaican Canadian community. *Proceedings of the Nineteenth Annual Symposium about Language and Society - Austin Language in the Public Domain*, 54, 1–22.

- Hoffman, M. (2014). Sociolinguistic interviews. In J. Holmes & K. Hazen (Eds.), *Research methods in sociolinguistics: A practical guide* (pp. 25–41). Wiley Blackwell.
- Holm, J. (1987). Creole influence on popular Brazilian Portuguese. In G. Gilbert (Ed.), *Pidgin and creole languages: Essays in memory of John E. Reinecke* (pp. 406–430). University of Hawaii Press.
- Holm, J. (1991). *A theoretical model for semi-creolization*. 9th Conference of the Society for Caribbean Linguistics, Cave Hill, Barbados.
- Holm, J. (1992). Popular Brazilian Portuguese: A semi-creole. *Actas Do Colóquio Sobre Crioulos de Base Lexical Portuguesa*. Colóquio Sobre Crioulos de Base Lexical Portuguesa.
- Hooper, J. B. (1978). Constraints on the schwa-deletion in American English. In J. Fisiak (Ed.), *Recent developments in historical phonology* (pp. 183–207). de Gruyter.
- Hopper, P. J. (1979). Aspect and foregrounding in discourse. In T. Givón (Ed.), *Discourse and syntax* (pp. 213–241). Academic Press.
- Hopper, P. J. (1991). On some principles of grammaticization. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization, Volume 1: Focus on theoretical and methodological issues* (Vol. 1, pp. 17–35). John Benjamins Publishing.
- Hopper, P. J. (1996). Some recent trends in grammaticalization. *Annual Review of Anthropology*, 25(1), 217–236.
- Hopper, P. J., & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, 56(2), 251–299.

- Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization* (2nd ed.). Cambridge University Press.
- Houaiss, A. (1959). Tentativa de descrição do sistema vocálico do português culto na área dita carioca. *Anais do Congresso brasileiro de língua falada no teatro 1*, 217–317.
- Humboldt, W. von. (1825). *Über das Entstehen der grammatischen Formen und ihren Einfluss auf die Ideenentwicklung*. Königliche Akademie der Wissenschaften.
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected Readings* (pp. 269–293). Penguin.
- Janda, R. D., & Joseph, B. D. (2003). On language, change, and language change—Or, of history, linguistics, and historical linguistics. In B. D. Joseph & R. D. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 3–180). Blackwell Publishing.
- Jurafsky, D., & Martin, J. H. (2019). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (Third Edition). Pearson Prentice Hall.
- Kenstowicz, M., & Sandalo, F. (2016). Pretonic vowel reduction in Brazilian Portuguese: Harmony and dispersion. *Journal of Portuguese Linguistics*, 15(6), 1–19.
- Kern, J. (n.d.). *“Como” in commute: The travels of a discourse marker across languages* [M.A. Thesis]. The University of Arizona.
- Kiparsky, P. (2003). The phonological basis of sound change. In B. D. Joseph & R. D. Janda (Eds.), *The Handbook of Historical Linguistics* (pp. 313–342). Blackwell Publishing.

- Kiparsky, P. (2012). Grammaticalization as optimization. In D. Jonas, J. Whitman, & A. Garrett (Eds.), *Grammatical change: Origins, nature, outcomes* (pp. 15–51). Oxford University Press.
- Kohler, K. J. (n.d.). The phonetic/phonology issue in the study of articulatory reduction. *Phonetica*, 48(2–4), 180–192.
- 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 Publishing.
- Kolyaseva, A., & Davidse, K. (2018). A typology of lexical and grammaticalized uses of Russian tip, tipa, po tipo. *Russian Linguistics*, 42(2), 191–220.
- König, E. (1991). *The meaning of focus particles: A comparative perspective*. Routledge.
- Krug, M. (1998). String frequency: A cognitive motivating factor in coalescence. Language processing, and linguistic change. *Journal of English Linguistics*, 26(4), 286–320.
- Krug, M. G. (2001). Frequency, iconicity, categorization: Evidence from emerging modals. In J. L. Bybee & P. J. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing.
- Kuryłowicz, J. (1965). The evolution of grammatical categories. *Diogenes*, 13(51), 55–71.
- Kuznersova, A., Brockhoff, P. B., & Christensen, R. H. B. (n.d.). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26.

- Labov, W. (1966). The effect of social mobility on linguistic behavior. *Sociological Inquiry*, 36(2), 186-203.
- 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*.
- Labov, W. (1991). *Sociolinguistic Patterns* (11th ed.). University of Pennsylvania.
- Labov, W. (1994). *Principles of linguistic change, Voll: Internal factors*. Blackwell.
- Labov, W. (2001). *Principles of linguistic change*. Blackwell.
- Labov, W. (2006). *The social stratification of English in New York City* (2nd ed.). Cambridge University Press.
- Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English: Phonetics, phonology, and sound change*. Mouton de Gruyter.
- Langacker, R.W. (1977). Syntactic reanalysis. In C.N. Li (Ed.), *Mechanisms of Syntactic Change*, (pp. 57–139). University of Texas Press.
- Laurentino, J. J. (2016). *Funções morfossintáticas e discursivas do tipo na fala de jovens natalenses* [M.A. Thesis]. Universidade Federal do Rio Grande do Norte.
- Leiria, L. L. (2000). A ditongação variável em sílabas tônicas finais travadas por /S/. *Organon*, 14(28–29), 133–141.
- Levelt, W. J. M. (1989). *Speaking*. MIT Press.
- Levy, R. F., & Jaeger, T. (2007). Speakers optimize information density through syntactic reduction. *Advances in Neural Information Processing Systems*, 849–856.

- Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431–461.
- Lichtenberk, F. (1991). On the gradualness of grammaticalization. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization I: Focus on theoretical and methodological issues* (pp. 37–80). John Benjamins Publishing.
- Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge University Press.
- Lima-Hernandes, M. C. P. (2005a). *A interface sociolingüística/gramaticalização: Estratificação de usos de tipo, feito, igual e como. Sincronia e diacronia* [PhD Dissertation]. Universidade Estadual de Campinas (UNICAMP).
- Lima-Hernandes, M. C. P. (2005b). Tipo: Deslizamentos funcionais e gramaticalização. *Estudos Lingüísticos*, XXXIV, 1325–1330.
- Lin, J. S., & Stanford, L. M. (1983). The acculturation of Chinese in North America: A sociolinguistic profile of an advantaged immigrant group in Edmonton. *Bilingual Review/La Revista Bilingüe*, 10(1), 47–51.
- Lipski, J. (1973). The survival of a “marked” segment in Portuguese. *General Linguistics*, 13(1), 1–15.
- Lipski, J. M. (2010). Spanish and Portuguese in contact. In R. Hickey (Ed.), *The Handbook of Language Contact* (pp. 550–580). Blackwell Publishing.

- Lobanov, B. M. (1971). Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, 49(2B), 606–608.
- Lobato, L. M. P. (2006). Sobre a questão da influência ameríndia na formação do português do Brasil. *Revista de Estudos Da Linguagem*, 14(2), 11–47.
- Lucchesi, D., Baxter, A. N., & Ribeiro, I. (2009). *O português afro-brasileiro*. Edufba.
- Major, R. C. (1985). Stress and rhythm in Brazilian Portuguese. *Language*, 61(2), 259–282.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71–102.
- Martelotta, M. E., & da Silva, L. R. (1996). Gramaticalização de então. In M. E. Martelotta, S. J. Votre, & M. M. Cezario (Eds.), *Gramaticalização no português do Brasil: Uma abordagem funcional*. Rio de Janeiro, Brazil.
- Martelotta, M. E., & Rêgo, L. M. R. (1996). Gramaticalização de lá. In M. E. Martelotta, S. J. Votre, & M. M. Cezario (Eds.), *Gramaticalização no português do Brasil: Uma abordagem funcional*. Rio de Janeiro, Brazil.
- Massaro, D. W. (Ed.). (2014). *Understanding language: An information-processing analysis of speech perception, reading, and psycholinguistics*. Academic Press.
- Massaro, D. W., & Cohen, M. M. (1983). Categorical or continuous speech perception: A new test. *Speech Communication*, 2, 15–35.
- Mateus, M. H., & d'Andrade, E. (2000). *The phonology of Portuguese*. OUP Oxford.

- Mattos e Silva, R. V. (1993). Português brasileiro: raízes e trajetórias (para a construção de uma história). *Discursos: estudos de língua e cultura portuguesa*, 75-91.
- McGuire, G. (2010). A brief primer on experimental designs for speech perception research. *Laboratory Report*, 77(1), 2–19.
- Meillet, A. (1912). L'évolution des formes grammaticales. *Scientia*, XII.
- Mendes, C. da S., de Medeiros, N. R., & de Oliveira, T. S. (2016). Africanidades na fonética do português brasileiro. *GT1 Africanidades e Brasilidades Em Literaturas e Linguísticas*. I Congresso Internacional e III Congresso Nacional - Africanidades e Brasilidades: Literaturas e Linguística, Espírito Santo, Brazil.
- Mendonça, R. (2012). *A influência africana no português do Brasil*. Fundação Alexandre de Gusmão (FUNAG).
- Meneses, F., & Albano, E. (2015). From reduction to apocope: Final poststressed vowel devoicing in Brazilian Portuguese. *Phonetica*, 72(2–3), 121–137.
- 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>
- Morais e Silva, A. (1789). *Diccionario da lingua portugueza* (Vol. 1–2). Oficina de Simão Thaddeo Ferreira.
- Munn, A., & Schmitt, C. (2002). Bare nouns and the morphosyntax of number. *Amsterdam Studies in the Theory and History of Linguistic Science Series 4*, 225-240.

- N. A. (n.d.). Conheça a história da primeira favela do Rio de Janeiro. *Museu do Amanhã*. <https://museudoamanha.org.br/pt-br/conheca-a-historia-da-primeira-favela-do-rio-de-janeiro>
- n.a. (1953). É o carioca o maior criador de palavras da nossa gíria. *Flan: O jornal da Semana*, 38–39.
- Nascentes, A. (1953). *O linguajar carioca*. Organização Simões.
- Naro, A., & Scherre, M. M. P. (2007). *Origens do português brasileiro*. Parábola Books.
- Netto, W. F. (2011). *Introdução à fonologia da língua portuguesa* (2nd ed.). Editora Paulistana.
- Newmeyer, F. J. (2000). Deconstructing grammaticalization. *Language Sciences*, 23(2), 187–229.
- Newmeyer, F. J. (2006). On Gahl and Garnsey on grammar and usage. *Language*, 82(2), 399–404.
- Nickerson, R. S. (1965). Response times for “same” - “different” judgments. *Perceptual and Motor Skills*, 20, 15–18.
- Nobre, M. A., & Ingemann, F. (n.d.). Oral vowel reduction in Brazilian Portuguese. *Kansas Working Papers in Linguistics*, 8(1), 1–30.
- Norricks, N. R. (2001). Discourse markers in oral narrative. *Journal of Pragmatics*, 33(6), 849–878.
- O que é favela, afinal?* (2009). Observatório de Favelas.

- Oden, G. C., & Massaro, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, 85(3), 172.
- Ogunsanya, E. O. (1971). *Residual Yoruba-Portuguese Bilingualism*. Harvard University Press (unpublished).
- Oh, S. (n.d.). Phonetic realization of vowel reduction in Brazilian Portuguese. *Journal of the Acoustical Society of America*, 143(3), 1968–1968.
- Ohala, J. J. (1974). Phonetic explanation in phonology. *Papers from the Parasession on Natural Phonology*, 251–274.
- Ohde, R. N., Haley, K. L., & McMahon, C. W. (1996). A developmental study of vowel perception from brief synthetic consonant-vowel syllables. *The Journal of the Acoustical Society of America*, 100(6), 3813–3824.
- Oliveira, N. (n.d.). Favelas and ghettos: Race and class in Rio de Janeiro and New York City. *Latin American Perspectives*, 23(4), 71–89.
- OpenStreetMap Contributors. (2020). *Planet dump*.
- Pagliuca, W. (Ed.). (1994). *Perspectives on grammaticalization* (Vol. 109). John Benjamins Publishing.
- Paiva, M. da C. de, & Silva, V. L. P. (2012). Cumprindo uma pauta de trabalho: Contribuições recentes do PEUL. *Alfa : Revista de Linguística (São José do Rio Preto)*, 56(3), 739–770. <https://doi.org/10.1590/S1981-57942012000300002>

- Parkvall, M., & Álvarez López, L. (2003). Português vernáculo brasileiro e a hipótese da semi-crioulização. *Abralin, II*(1), 111–152.
- Pastore, R. E., Friedman, C. J., & Baffuto, K. J. (1976). Comparative evaluation of the AX and two ABX procedures. *The Journal of the Acoustical Society of America, 41*(1), S120–S120.
- Patel, G. (2015). A linguagem da favela parte 2: O dialeto e o estigma (Pereira, Trans.). *Rio on Watch*. <http://rioonwatch.org.br/?p=16516>
- Peixoto, A. (1944). *A obra nova de língua geral de mina*. Agência Geral das Colónias.
- PEUL - Programa de Estudos sobre o Uso da Língua. (n.d.).
[Corpus]. PEUL. <http://www.lettras.ufrj.br/peul/index.html>
- Pezatti, E. G. (2001). O advérbio então já se gramaticalizou como conjunção? *Delta, 17*(1), 81–95.
- Pichler, H., & Levey, S. (2011). In search of grammaticalization in synchronic dialect data: General extenders in northeast England. *English Language and Linguistics, 15*(3), 441–471.
- Pichler, Heike. (2013). *The structure of discourse-pragmatic variation*. John Benjamins Publishing.
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [Dissertation]. Massachusetts Institute of Technology.
- Pilati, E., Naves, R. R., & Salles, H. M. L. (2017). On the syntax of subjects in Brazilian Portuguese: using the ‘SPLIT’ pronominal system as the basis for an alternative analysis. *Diadorim: revista de estudos linguísticos e literários, 19*, 99-139.

- Pinheiro, R. B. (1872). *Apontamentos sobre a picaresca viagem do Imperador de Rasilb pela Europa*.
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253–260.
- Pisoni, D. B. (1975). Auditory short-term memory and vowel perception. *Memory and Cognition*, 3(1), 7–18.
- Pisoni, D. B., & Tash. (n.d.). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2), 286–290.
- Podlubny, R. G., Geeraert, K., & Tucker, B. V. (2015). It's all about, like, acoustics. *Proceedings of the 18th International Congress of Phonetic Sciences*. 18th International Congress of Phonetic Sciences, Glasgow, UK.
- Portolés, J. (1998). *Marcadores del discurso*. Ariel.
- Projeto Norma Linguística Urbana Culta—RJ. (n.d.). [Corpus]. NURC RJ. <https://nurcrj.lettras.ufrj.br/>
- Psychology Software Tools, Inc. (2012). *E-Prime 2.0*. www.pstnet.com
- QGIS Development Team. (2020). *QGIS Geographic information system* [Computer application]. Open Source Geospacial Foundation Project.
- R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <http://www.R-project.org/>
- Ramat, A. G., Mauri, C., & Molinelli, P. (2013). *Synchrony and Diachrony: A dynamic interface*. John Benjamins Publishing.

- Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics*, 29(6), 1139–1172.
- Reed, C. (1975). Reaction times for a same-different discrimination of vowel-consonant syllables. *Perception & Psychophysics*, 18(2), 65–70.
- Ribeiro, C. A. (n.d.). Class, race, and social mobility in Brazil. *Dados - Revista de Ciências Sociais*, 49(4), 833–873.
- Ribeiro, S. S. C., Costa, S. B. B., & Cardoso, S. A. M. (Eds.). (2009). *Dons sons às palavras: Nas trilhas da língua portuguesa*. EDUFBA.
- Rickford, J. R. (1986). The need for new approaches to social class analysis in sociolinguistics. *Language and communication*, 6(3), 215–221.
- Robl, A. (1985). Alguns problemas da influência tupi na fonética e morfologia do português popular do Brasil. *Revista Letras*, 34, 155–179
- Rodrigues, A. de L. S. (2010). Manutenção e supressão de marcadores discursivos no processo de retextualização (Maintenance and suppression of discourse markers in retextualization process). *Estudos Lingüísticos*, 39(1), 930–940.
- Rodrigues, R. N. (2010). *Os africanos no Brasil*. Centro Edelstein de Pesquisas Sociais.
- Romaine, S., & Lange, D. (1991). The use of like as a marker of reported speech and thought: A case of grammaticalization in progress. *American Speech*, 66, 227–279.
- 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. 313–338). John Benjamins Publishing.

- Roth-Gordon, J. (2009). The language that came down the hill: Slang, crime, and citizenship in Rio de Janeiro. *American Anthropologist*, 111(1), 57–68.
- Rousseeuw, P. J., Ruts, I., & Tukey, J. W. (1999). The bagplot: A bivariate boxplot. *The American Statistician*, 53(4), 382–387.
- RStudio Team. (2020). *RStudio: Integrated development for R*.
- Sanches, E. R. (2014). The community of Portuguese language speaking countries: The role of language in a globalizing world. *Workshop*.
- Sanker, C. (2019). Effects of lexical ambiguity, frequency, and acoustic details in auditory perception. *Attention, Perception, & Psychophysics*, 81, 323–343.
- Sankoff, D. (1988). Sociolinguistics and syntactic variation. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey* (pp. 140–161). Cambridge University Press.
- Santos, S. M. (2019). Ongoing grammaticalization process of lexeme “tipo”: A study based on a spontaneous speech corpus of Brazilian Portuguese. *Inventário*, 23, 207–226.
- Sapir, E. (1949). *Language: An introduction to the study of speech*. Harcourt, Brace & World.
- Sayce, A. H. (1900). *Introduction to the science of language: Vol. I* (4th ed.). Kegan Paul, Trench, Trübner & CO.
- Scherre, M. M. P., & Naro, A. J. (1993). Sobre as origens do português popular do Brasil. *Delta*, 9(Especial), 437–454.

- Scherre, M. P., & Naro, A. J. (2010). Sobre as origens estruturais do português brasileiro: crioulização ou mudança natural?. *PAPIA-Revista Brasileira de Estudos do Contato Linguístico*, 11(1), 40-50.
- Schiering, R. (2010). Reconsidering erosion in grammaticalization: Evidence from cliticization. In K. Stathi, E. Gehweiler, & E. König (Eds.), *Grammaticalization: Current views and issues* (Vol. 119, pp. 73–100). John Benjamins Publishing.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge University Press.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime user's guide*. Psychology Software Tools, Inc.
- Schourup, L. (1998). Discourse markers. *Lingua*, 107, 227–265.
- Seara, I. C., Nunes, V. G., & Lazzarotto-Volcão, C. (2011). *Fonética e fonologia do português brasileiro: 2 período*. Universidade Federal de Santa Catarina.
- Siegel, M. E. A. (2002). Like: The discourse particle and semantics. *Journal of Semantics*, 19(1), 35–71.
- Silva, D. J. (1997). The variable deletion of unstressed vowels in Faialense Portuguese. *Language Variation and Change*, 9, 295–308.
- Silva, D. J. (1998). Vowel lenition in São Miguel Portuguese. *Hispania*, 81(1), 166–178.
- Silva, J. C. B. & Cunha, C. S. (2012). Caracterização prosódica dos falares brasileiros: a questão total em Recife, Rio de Janeiro e Florianópolis. *ANTARES: Letras e Humanidades*, 3(6), 282-294.

- Silva, M. (2013, September 27). Tipo assim... *Gazeta do Povo*.
- Silva, M. N. (2010). *A favela como expressão de conflitos no espaço urbano do Rio de Janeiro: O exemplo da Zona Sul Carioca* [Rio de Janeiro, Pontifícia Universidade Católica do Rio de Janeiro]. https://www.maxwell.vrac.puc-rio.br/16168/16168_4.PDF
- Silva, T. C. (2003). *Fonética e fonologia do português: Roteiro de estudos e guia de exercícios*. Editora Contexto.
- Snichelotto, C. A. R., & Görski, E. M. (2011). (Inter)subjetivização de marcadores discursivos de base verbal: Instâncias de gramaticalização (Inter)subjectivation of discourse markers: Grammaticalization instances). *Alfa*, 55(2), 423–455.
- Souza, E. R. F. D. (2009). Gramaticalização dos itens linguísticos assim, já e aí no português brasileiro: Um estudo sob a perspectiva da gramática discursivo-funcional.
- Stathi, K., Gehweiler, E., & König, E. (Eds.). (2010). *Grammaticalization: Current views and issues*. John Benjamins Publishing.
- Stentström, A., & Andersen, G. (1996). More trends in teenage talk: A corpus-based investigation of the discourse items *cos* and *innit*. *Synchronic Corpus Linguistics: Papers from the Sixteenth International Conference on English Language Research on Computerized Corpora*, 189–203.
- Strange, W., Edman, T. R., & Jenkins, J. J. (1979). Acoustic and phonological factors in vowel identification. *Journal of Experimental Psychology*, 5(4), 643–656.

- Strange, W., Verbrugge, R. R., Shankweiler, D. P., & Edman, T. R. (1976). Consonant environment specifies vowel identity. *The Journal of the Acoustical Society of America*, 60(1), 213–224.
- Sut, H. (2006, August 31). *Tipo assim* [Newspaper]. Carta Maior. cartamaior.com.br/?/Editoria/Midia/Tipo-assim/12/11443
- Swan, Toril, & Westvik, O. J. (2011). *Modality in Germanic Languages: Historical and Comparative Perspectives*. Walter de Gruyter.
- Taboada, M. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of Pragmatics*, 38(4), 567–592. <https://doi.org/10.1016/j.pragma.2005.09.010>
- Tagliamonte, S. (2005). So who? Like how? Just what? Discourse markers in the conversations of young Canadians. *Journal of Pragmatics*, 37(11), 1896–1915.
- Tagliamonte, S. (2016). *Teen talk*. Cambridge University Press.
- 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.
- Tagliamonte, S., & D'Arcy, A. (2009). Peaks beyond phonology: Adolescence, incrementation, and language change. *Language*, 85(1), 58–108.
- Tagliamonte, S., & Hudson, R. (1999). *Be like et al. Beyond America: The quotative system in British and Canadian youth*. 3(2), 147–172.
- Tavares, M. A. (1999). A gramaticalização do aí como conector — Indícios sincrônicos. *Working Papers Em Lingüística, UFSC*, 3(1), 129–141.

- Tavares, M. A. (2003). A gramaticalização de e, aí, daí e então: estratificação/variação e mudança no domínio funcional da seqüenciação retroativo-propulsora de informações. Um estudo sociofuncionalista.
- Teyssier, P. (1980). *Histoire de la Langue Portugaise (No. 1864)*. Presses Universitaires de France - PUF.
- Tinsley, T., & Board, K. (2014). *Languages for the future: Which languages the UK needs most and why*. British Council.
- The Portugal News Online. (2007, May 5). Portuguese language gaining popularity [Newspaper]. *The Portugal News Online*. <http://www.theportugalnews.com/news/portuguese-language-gaining-popularity/22235>
- The University of Louisville. (n.d.). The Portuguese Language. *Portuguese Studies*. <https://louisville.edu/anthropology/portuguesestudies/the-portuguese-language>
- Thomas, E. W. (1969). *The syntax of spoken Brazilian Portuguese*. Vanderbilt University Press.
- Thompson, C. P., & Onosson, D. S. (2016). *Projeto Sociolingüístico Contemporâneo Brasileiro*. New Ways of Analyzing Variation 45, Vancouver, B. C.
- Thompson, H. V. G. (2013). *Do léxico à gramática: Os diferentes usos de tipo* [M.A. Thesis]. Universidade Federal do Rio de Janeiro.
- Tinsley, T., & Board, K. (2014). *Languages for the future: Which languages the UK needs most and why*. British Council.

- 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.
- Traugott, E. C. (2010). Grammaticalization. In A. H. Jucker & I. Taavitsainen (Eds.), *Historical pragmatics* (pp. 97–126). de Gruyter.
- Traugott, E. C. (2011). Grammaticalization and mechanisms of change. In H. Narrog & B. Heine (Eds.), *The Oxford handbook of grammaticalization*. Oxford University Press.
- Traugott, E. C., & Heine, B. (Eds.). (1991a). *Approaches to grammaticalization, Volume 1: Focus on theoretical and methodological issues* (Vol. 1). John Benjamins Publishing.
- Traugott, E. C., & Heine, B. (Eds.). (1991b). *Approaches to Grammaticalization, Volume II. Types of grammatical markers* (Vol. 2). John Benjamins Publishing.
- Traugott, E. C., & König, E. (n.d.). The semantics-pragmatics of grammaticalization revisited. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization* (Vol. 1, pp. 189–218u). John Benjamins Publishing.
- Trudgill, P. (2009). Vernacular universals and the sociolinguistic typology of English dialects. In M. Filppula, J. Klemola, & H. Paulasto (Eds.), *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond* (pp. 304–322). Routledge.

- Tuten, D. N., & Tejedo-Herrero, F. (2011). The relationship between historical linguistics and sociolinguistics. In M. Díaz-Campos (Ed.), *The handbook of Hispanic sociolinguistics* (pp. 283–302). Wiley Blackwell.
- Underhill, R. (1988). Like is, like, focus. *American Speech*, 63(3), 234–246.
- Valle, C. R. M. (2000). Marcadores discursivos: Considerações sobre os limites entre a gramaticalização e a discursivização. *Working Papers em Lingüística, UFSC*, 4, 104–113.
- Van Dijk, T. A. (1979). Pragmatic connectives. *Journal of Pragmatics*, 3(5), 447–456.
- Varejão, F. O. A. (2009). O português do Brasil: Revisitando a história. *Caderno de Letras Da UFF - Dossiê: Difusão Da Língua Portuguesa*, 39, 119–137.
- Vasconcellos, J. L. (1901). *Esquisse D'une Dialectologie Portugaise: Thèse Pour Le Doctorat De L' université De Paris (Faculté Des Lettres)*. Aillaud & cie.
- Vasilévski, V. (2012a). *Descodificación automática de la lengua escrita de Brasil basada en reglas fonológicas*. Académica Española.
- Vasilévski, V. (2012b). Phonologic patterns of Brazilian Portuguese: A grapheme to phoneme converter based study. *Proceedings of the EACL, Workshop on Computational Models of Language Acquisition and Loss*. Workshop on Computational Models of Language Acquisition and Loss, France.
- Voghera, M. (2013). A case study on the relationship between grammatical change and synchronic variation: The emergence of *tipo* in Italian. In A. G. Ramt, C. Mauri, & P.

- Molinelli (Eds.), *Synchrony and diachrony: A dynamic interface*. John Benjamins Publishing.
- Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In L. Lehmann & Y. Malkiel (Eds.), *Directions for historical linguistics* (pp. 98–188). University of Texas Press.
- Whaley, C. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17, 143–154.
- Wickham et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. <https://ggplot2.tidyverse.org>
- Wioland, N., Rudolf, G., Metz-Lutz, M. N., Mutschler, V., & Marescaux, C. (1999). Cerebral correlates of hemispheric lateralization during a pitch discrimination task: An ERP study in dichotic situation. *Clinical Neurophysiology*, 110, 516–523.
- Wischer, I. (2008). Grammaticalization of periphrastic constructions. In E. Verhoeven, S. Skopeteas, Y. M. Shin, Y. Nishina, & J. Helmbrecht (Eds.), *Studies on grammaticalization* (pp. 241–250). Mouton de Gruyter.
- Wischer, I., & Diewald, G. (Eds.). (n.d.). *New reflections on grammaticalization*. John Benjamins Publishing.
- 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. Fifth International Conference on Language Resources and Evaluation.

Xavier, H. N., & Magalhães, F. (2003). *The case of Rio de Janeiro* (Understanding Slums: Case Studies for the Global Report on Human Settlements, pp. 1–28). United Nations Human Settlements Programme.

Zahler, S. (2014). Variable subject doubling in spoken Parisian French. *Proceedings of the 37th Annual Penn Linguistics Conference*, 20, 1, 361–371.

Zipf, G. K. (n.d.). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, 15, 1–95.