A Sociolinguistic Study of Discourse Marker *Assim* in Brazilian Portuguese

by

Gabriel Goto Escudero

A thesis submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfilment of the requirements of the degree of

MASTER OF ARTS

Department of Linguistics
University of Manitoba
Winnipeg

Examining Committee

Supervisor:

Dr. Verónica Loureiro-Rodríguez
Department of Linguistics
University of Manitoba

Members:

Dr. Nicole Rosen
Department of Linguistics
University of Manitoba

Dr. Robert Hagiwara
Department of Linguistics
University of Manitoba

Abstract

This thesis aims to understand the linguistic and social factors that impact the use of Brazilian Portuguese discourse marker *assim* 'like that' in Natal, an economically neglected city with a low prestige dialect, and Rio de Janeiro, the second-most populous Brazilian city with a prosperous economy and a dialect associated with high status and power. A total of 930 tokens of discursive uses of *assim*, a stigmatized feature in Brazilian Portuguese, were collected from 105 individual interviews conducted in 1993 for the Discurso & Gramática corpus (Votre & Oliveira, n.d.; Cunha, 1998) and were analyzed based on linguistic (clause placement, syntactic categories adjoined, conversation topic) and social factors (age, sex, location). Overall, the most common placement of *assim* is clause-medial, and χ^2 statistical analysis shows that males prefer placing *assim* after nouns, while females prefer using it after verbs. The placement of *assim* between a determiner and a noun and between the subject and the verb are unusual, confirming previous conclusions (Silva & Macedo, 1992). Moreover, *assim* never appears before an auxiliary verb, even though this syntactic constraint was not mentioned in previous literature. Mixed-effects statistical analysis shows that speakers from Rio de Janeiro use *assim* more frequently at the end of the clause, with this use being led by 18-20-year-old males, and that females from both locations prefer to use *assim* with topics that prompt them to express opinions. This thesis contributes to the existing literature on discourse markers, offering a glimpse into the connection between prestige and language variation in Brazilian Portuguese.

Acknowledgments

I am deeply indebted to my advisor, Dr. Veronica Loureiro-Rodriguez, for the patience and guidance during a long time preparing my work. Also, words cannot express my gratitude to the members of the committee, Dr. Nicole Rosen and Dr. Rob Hagiwara, who generously provided knowledge and expertise to improve my research. Adding to that, I'm extremely grateful to the Department Chair, Dr. Jila Ghomeshi, who believed in my qualities and provided moral support. And a special thanks to the Government of Manitoba, which provided me with funds to pursue my program.

I am thankful to my classmates Sabrina Bellonia and Vivian Xu for their company in my studies and their suggestions for the elaboration of the project. Also, I'd like to acknowledge the beneficial assistance from Dr. Sky Onosson, who provided the necessary statistical knowledge to proceed with the research. I sincerely thank my colleague Precious Affia, who always had faith in my success and motivated me to finish this thesis.

Lastly, I'd like to mention that I could not have undertaken this journey without my family: my mother, Maria Claudia, who accepted having her child being far away for a long time to follow his dream, and my wife, Carolina, who trusted my talent, had the patience to observe this long process of writing a thesis, giving the more than needed emotional support, and my children, Gabriel and Melissa, who, not even precisely knowing what I was working on, gave the love and care, with their admiration inspiring me to get to this special milestone.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Discourse markers play a crucial role in the sequence of speech (Schiffrin, 1987) even though they do not add new information to the utterance, as they do not carry propositional meaning (Vincent, 1983, cited in Silva & Macedo, 1992). Brazilian Portuguese discourse markers manage discourse structure and are very common, but are considered nonstandard features and thus are not discussed in traditional prescriptive grammar books (Silva & Macedo, 1996). Their use has been associated with insecurity and disfluency, and even public speakers and career coaches advise people not to use discourse markers to avoid being perceived as lacking confidence (Polito, 2016). Such negative attitudes towards discourse markers are so institutionalized that explicit commentaries deeming the use of discourse markers unacceptable can be found in government documents, such as the hiring process guidelines from the Brazilian Institute of Geography and Statistics and the curriculum of primary education from the Brazilian Federal District (Freitag, 2007).

In Brazilian Portuguese (BP), the discourse marker *assim* 'like that' is commonly used in casual speech (Silva & Macedo, 1992). Researchers have analyzed the discourse functions, as well as the grammaticalization of *assim* (Lopes-Damasio, 2008; Martellota et al., 1996). However, the social factors impacting its use are understudied, and research has produced inconclusive and often contradictory results (Silva & Macedo, 1992; 1996; Strapazzon, 2018), especially concerning the role played by speakers' age and sex. Furthermore, to date, no studies have analyzed the use of *assim* quantitatively by applying statistical models.

The present thesis aims to fill these gaps in the literature by conducting a variationist analysis of the use of *assim* in two geographically separate Brazilian cities (Natal and Rio de Janeiro) with distinct dialects and differing socio-economic realities. Rio de Janeiro is the country's former capital and has historically been inhabited by influential political figures and occupied by the industrial sector. Consequently, the dialect spoken in this city, known as the *carioca* dialect, is a prestigious variety associated with high status and power (Benchimol, 1992). In contrast, the *nordestino* dialect spoken in Natal is perceived as having very low prestige, a consequence of being historically neglected by the regional development policies of the federal government, which has resulted in slow economic growth and a great degree of poverty (Cano, 2017; Araújo, 2017). My research, then, analyzes the use of *assim,* a stigmatized feature in

Brazilian Portuguese, in the speech of male and female speakers from cities representing opposing socio-economic and linguistic realities. Specifically, this thesis aims to answer the following research questions:

1) What are the linguistic factors that determine the use of *assim*?

2) Do speakers' location, sex, and age help explain the variation in the placement of *assim* in the clause? and

3) Does the use of *assim* vary according to the topic of the conversation?

This thesis is structured into six chapters. Chapter 2 presents an overview of the syntactic placement and sociolinguistic patterns of discourse markers, with a particular focus on *assim*. Chapter 3 describes the methodology and the statistical models used to analyze the data. Chapter 4 presents the distributional analysis of the data and the results from the statistical analysis. Chapter 5 offers the interpretation of the results, and finally, chapter 6 presents the conclusions of this research and offers some ideas for further study.

Chapter 2: Literature Review

This chapter describes what previous studies have found regarding discourse markers. Section 2.1 first defines the concept and provides an overview of its functions, then discusses the syntactic placements and sociolinguistic patterns associated with the use of discourse markers, with a focus on English *like*. Section 2.2 discusses previous research on discourse marker *assim* in Brazilian Portuguese, similarly indicating its syntactic placements and the social factors associated with its use.

2.1. What are discourse markers?

Although diverse definitions of discourse markers (henceforth DMs) exist in the literature, there is a general agreement that they are features that signal a relationship between two adjoined segments (Fraser, 1999; Lopes, 2016). DMs are primarily found in spontaneous oral speech (Silva & Macedo, 1992) to manage and organize the structure of discourse (Lopes, 2016; Pichler, 2013). Due to their inherent pragmatic nature, DMs are syntactically optional (Fraser, 1999) and thus have often been described as fillers (Martellota et al., 1996) or grammatically superfluous (Redeker, 1991). DMs have been identified in diverse languages such as Spanish (Stenström & Jørgensen, 2009), French (Sankoff et al., 1997), English (e.g. D'Arcy, 2005, 2017; Andersen, 2001), Sierra Miwok (Schourup, 1985), Hebrew (Maschler, 2001), Bislama (Meyerhoff & Niedzieslki, 1998), Japanese (Lauwereyns, 2002), Swedish (Kotsinas, 1994), Norwegian (Hasund, 2003), Dutch (Haddican & Zweig, 2012), German (Umbach & Ebert, 2009), and European (Lopes, 2016) and Brazilian Portuguese (Silva & Macedo, 1992).

Many studies have attempted to identify the diverse pragmatic functions of DMs. One of the most common functions specified in the literature (Fox Tree, 2010) is creating coherence between two adjacent phrases, which has led some authors to state that DMs play a grammatical role in discourse (Schiffrin, 1987). DMs also assist in the turn-taking organization, with some authors seeing them as turn-initiators (Fung & Carter, 2007) and others as turn-relinquishers (Duncan, 1972). Finally, DMs also contribute to building solidarity in conversations, with some authors claiming that there is a higher frequency of use of DMs when speaking to a friend than to a stranger (Redeker, 1990; Jucker & Smith, 1998), and others suggesting the opposite: that is, that DMs are used less often in conversations between partners or close relatives (Ostman, 1981). As for the circumstances where the DMs occur, some authors have found a greater frequency of

use of DMS in relaxed conversations (Broen & Siegel, 1972), while others suggest that the use of DMs is influenced by the speaker's anxiety derived from provocative topics (Lalljee & Cook, 1975).

2.1.1. Syntactic placement of DMs

The syntactic characteristics of DMs have received less scholarly attention. Overall, however, researchers have identified patterns in the placement of DMs inside the sentence structure. In an early study, Traugott (1995) describes DMs as syntactic adjuncts: features adjoined to another phrase instead of being the head of its functional category. Following this approach, Brinton (2006) and D'Arcy (2017) describe DMs as features that target CPs, canonically the left periphery of the sentence, in the initial position. Other studies, however, have found DMs located at the end of the sentence, such as in Tojolabal Mayan, where speakers use these features as a clitic meaning 'anyway' (Brody, 1989). Finally, an instance found in clause-medial position appears in Estonian (Campbell, 1991), where the interrogative *es* is used between the first and second constituents in a sentence.

DM *like* is probably one of the most studied in English *(*D'Arcy, 2017; Andersen, 2001). *Assim* has long been translated as *like* in the literature (Silva & Macedo, 1992) and has also been shown to be present in a myriad of occurences in the sentence (Lopes-Damásio, 2008), which suggests a syntactic similarity to *like*. Many studies on *like* have dealt with its syntactic placement and rule-governed qualities. Overall, *like* appears more often in the initial and medial positions in a sentence (Underhill, 1988; Tagliamonte, 2005; D'Arcy, 2005; 2017), while sentence-final *like* is restricted to particular English dialects, such as those spoken in Northern Ireland (Schweinberger, 2013; Diskin, 2017) and Scotland (Miller & Weinert, 1995). Except for the clause-final placement in these particular dialects, *like* tends to be adjacent to a phrasal construction with an established left-bracket condition applied within the clause, preceding a constituent dominating a focused element (Ross & Cooper, 1979). While some studies have analyzed the use and frequency of *like* in syntactic environments based on its placement next to a particular category of words, and have demonstrated its high frequency before nouns and verbs (Underhill,1988; Andersen, 2001; Tagliamonte, 2005), others have identified functional categories such as complementizer and tense as the classification for the syntactic slot where *like* would be inserted (D'Arcy, 2005; 2017).

2.1.2. Sociolinguistic patterns of DMs

Age and sex are the most commonly identified social factors impacting the use of DMs. Whenever a new form of DM gets attention from media, education, and linguists, its use is always associated with young speakers (Tagliamonte, 2005). English *like* is most frequently used by adolescents (Andersen, 2001; Dailey-O'Cain, 2000), even though there are cases found in children as young as seven years old (Levey, 2006) and other instances where it is found to be shared by all age groups in a speech community (D'Arcy, 2017). In apparent-time studies, the higher frequency of a linguistic feature by the younger age groups with a concurrent drop in the older group can be interpreted as age-grading (Chambers, 1995; Tagliamonte, 2005), meaning that its use is associated with a particular period of a person's life, or as language change occurring in both the individual and the community (Labov, 2001).

While some studies have found no differences in the use of DMs between males and females (e.g. *like* in British and American English; Andersen, 2001; Dailey-O'Cain, 2000), other studies attest that females use DMs more frequently than males, as seen in Canadian English *like* (Tagliamonte, 2005), Spanish *en plan* (Jørgensen, 2009), and French *comme* (Sankoff et al., 1997). Furthermore, research on *like* has shown that, while males prefer using *like* in the clause-medial position, especially in the NP context (Levey, 2006), females use *like* clause-initially more often (D'Arcy, 2005; 2017).

Finally, some studies have demonstrated that social class and ethnicity also play a role in using DMs. For example, regarding social class, Spanish *en plan* and British English *like* are found more frequently in speakers from the upper class (Jørgensen, 2009; Andersen, 2001). As for ethnicity, *like* is scarcely used in places with a great density of ethnic minorities, while it is widespread within "white adolescent speech" (Andersen, 2001).

2.2. DM *assim* in Brazilian Portuguese

Assim ‘like that', derived from a Latin phrase made by the preposition *ad* and the manner adverb *sic* (Ferreira, 1983), fits the general description of DMs: it relates to the segments by which it is surrounded, it is used for interactive purposes (Castelano & Luquetti, 2012), and although it does not add new information to the sentence, it has a pragmatic value (Silva & Macedo, 1992). Like other DMs, *assim* serves as a marker of hesitation and complementation that helps process, explain, or clarify information (Silva & Macedo, 1992). *Assim* allows the speaker additional time to think (Urbano, 1999; Castelano & Luquetti, 2011) while signalling the

hearer that the following word might not be exact (Rosa, 1992; Longhin-Thomazi, 2006). *Assim* also allows the speaker to keep their turn in conversations, maintaining the hearer's attention (Souza, 2010) and preventing interruptions (Martelotta et al., 1996).

However, *assim* has unique characteristics that set it apart from other DMs in Brazilian Portuguese. While in Brazilian Portuguese other forms of hesitation (particularly non-lexicalized fillers, such as *aa, éee,* and repetitions) occur freely in discourse, *assim* has some syntactic constraints. Also, while other DMs have their placements restricted to sentence boundaries (e.g., *né?* 'isn't it?', *sabe?* 'you know?', *entendeu?* 'got it?'), *assim* may occur within the sentence (Silva & Macedo, 1992). Furthermore, *assim* has multiple pragmatic functions (Lopes-Damasio, 2008; Souza, 2010), a quality not shared by other DMs in BP.

Examples (1-2) show DM *assim* placed within the sentence, connecting two segments. If removed, there would be no disruption at the semantic level, which supports the claim that *assim* only has an interactive function.

(1)	*Eu*	*tambem*	*vejo*	*assim*	*a*	*politica*	*como*	*um*	*problema*
	I	also	see	[like]	the	politics	as	a	problem

'I also see [like] politics as a problem' [Jorge Luis/M/21-31/Rio]

(2)	*Ele cantava*	*um*	*canto*	*assim*	*triste*
	He was.singing	a	song	[like]	sad

'He was singing a [like] sad song' [Italo/M/21-31/Natal]

2.2.1. Syntactic placement of assim

Research has shown that the syntactic placement of *assim* is rule-governed. Even though Lopes-Damasio (2008) explicitly states that *assim* has free positioning in discourse, the author does not successfully demonstrate all possible placements in a sentence, thus confirming that certain syntactic slots are not available for *assim.* Silva & Macedo (1992, 1996) could not find any instance of *assim* used between determiner and noun or between subject and verb, and Lopes-Damasio (2008) did not include examples of *assim* between determiner and noun either. Silva & Macedo (1992) further claim that what distinguishes *assim* from other forms of hesitation is that it cannot be used in those two contexts.

Strapazzon (2018) concluded that *assim* can be placed at the beginning of a sentence, at the end of a sentence, or in the middle of a sentence, usually within a clause. Prototypically,

considering its original use as a conjunction, *assim* is attested at the beginning of a sentence (Strapazzon, 2018); speakers can use it before a declarative sentence, as a sole utterance, alongside adverbs or other DMs (Lopes-Damasio, 2008), or between a raised topic noun and the sentence itself (Lopes-Damasio, 2008; Souza, 2009). At the end of a sentence, being the last word spoken in a turn, some authors classify it as an attenuator (Souza, 2009; Strapazzon, 2018; Lopes-Damasio, 2008). When *assim* is within a clause it is usually described as a complement introducer (Silva & Macedo, 1992), with speakers primarily placing it between a noun or a verb and their respective complements (Silva & Macedo, 1992; 1996; Longhin-Thomazi, 2006).

2.2.2. Sociolinguistic patterns of assim

Few studies have analyzed *assim* from a sociolinguistic perspective (Silva & Macedo, 1996; Strapazzon, 2018). As the use of DMs carries a social stigma in BP (Freitag, 2007), researchers assumed that *assim* would be more frequent in the speech of social groups that show linguistic insecurity in interactions, such as adolescents, females, and people with a lower educational level (Silva & Macedo, 1996). An early study (Silva & Macedo, 1996) in Rio de Janeiro (a different corpus from the present thesis) confirmed that females produced more instances of *assim* than males and that young speakers used *assim* more than older speakers. This analysis did not find that level of education played a role in variation. However, more recently, a study of the speech from Chapecó in the southern region of Brazil (Strapazzon, 2018) found that older males with high educational levels used *assim* more than younger females with lower academic levels. Although Strapazzon (2018) does not elaborate on the differences between his findings and Silva & Macedo's (1996), it is possible to observe that geographical location plays a role in such variation. This hypothesis is confirmed by an early study by Silva (1999). She found that the use of *assim* presented a sharp distinction between northern and southern regions, with speakers from the south using *assim* more frequently.

My thesis expands on the existing sociolinguistic literature on *assim* by looking at speakers from two cities that differ in economic conditions and geographical location (Rio de Janeiro and Natal) and considering speakers' age, sex, and topic of conversation. Furthermore, my study considers all syntactic positions identified in the literature for *assim*.

Chapter 3: Methodology

The goal of this thesis is to identify the linguistic and extralinguistic constraints that explain the use of DM *assim* in Natal and Rio de Janeiro. Specifically, it aims to answer the following research questions: (1) What are the linguistic factors that determine the use of *assim*?, (2) Do speakers' location, sex, and age help explain the variation in the placement of *assim* in the clause?, and (3) Does the use of *assim* vary according to the topic of the conversation? This chapter describes the methodology used to answer these questions. Section 3.1. describes the linguistic corpus where the data comes from, and section 3.2 details how the tokens were compiled. Section 3.3 concerns the exclusion criteria for selecting tokens of *assim*, while sections 3.4 and 3.5 list the linguistic and extralinguistic factors considered for the analysis, respectively. Finally, section 3.6 explains the statistical model applied for this study.

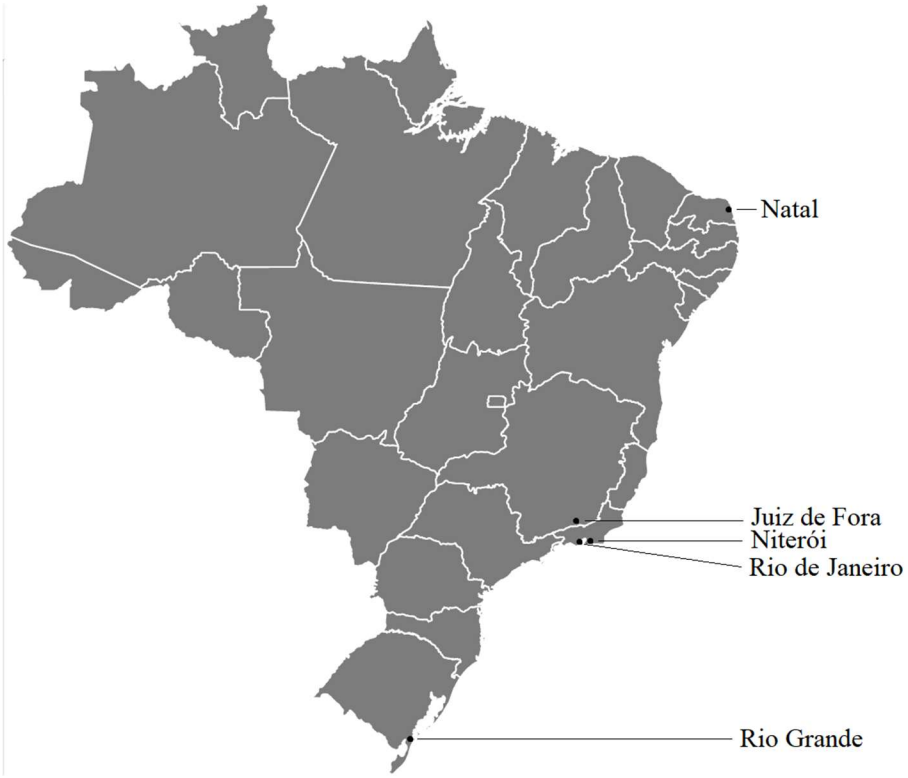
3.1. Corpus description

This study uses data from the *Discurso & Gramática* corpus (henceforth D&G), which was selected because it includes data from different places in Brazil - unlike other Brazilian corpora, which are restricted to one location only. This corpus also follows a sociolinguistic methodological principle that recommends an informal environment in interviews (Labov, 1972), posing questions that make the speaker talk about their life, beliefs, and experiences (Meyerhoff, 2006, p. 30).

Using corpus data to study DMs is common practice. For example, in BP, Martelotta et al. (1996) used the D&G corpus in their study of grammaticalization of *assim*, while Cunha (2000) and Cunha & Tavares (2007) drew from this corpus to analyze other discursive elements. In English, Andersen (2001) created the Bergen Corpus of London Teenage Language to study the use of *like* as a pragmatic marker, which, just as in the D&G corpus, was composed of recorded interviews from 1993 and stratified speakers by gender, age, and location. Likewise, for her investigation of discourse markers in teenage talks, Tagliamonte (2005) built a corpus of interviews with 26 speakers from Toronto (10-22 years old). This corpus later expanded and integrated with other collections (Tagliamonte, 2012) to form the Toronto English Archive, which involves speakers of a broader age range, keeping its restricted locality. A sample of the latter corpus was also used by D'Arcy (2005) in her comprehensive examination of discourse feature *like*.

The D&G corpus was built in 1993 by members of the Functional Linguistics research group from the Federal University of Rio de Janeiro in cooperation with the Fluminense Federal University and the Federal University of Rio Grande do Norte. The corpus consists of 171 speakers from five cities, namely Rio de Janeiro, Natal, Rio Grande, Juiz de Fora, and Niterói, with their geographic locations illustrated by Figure 1. All speakers are students, with ages ranging from 5 to 31. During the interviews, they were asked to 1) narrate a story which they experienced firsthand, 2) retell an event that happened to someone else, 3) describe a place where they would spend time, 4) give instructions on how to do something, and 5) give their opinion on a topic.

Figure 1. Brazil divided by states, with the location of the cities in the D&G corpus



Note. Adapted from *Brazil Blank Map* by F. Menegaz, 2007, Wikimedia Commons (https://upload.wikimedia.org/wikipedia/commons/1/10/Brazil_Blank_Map.svg). CC BY-SA 3.0

One limitation of the D&G corpus is that prosodic information and audio files are unavailable, although pauses and vowel lengthening are reflected in the transcriptions. Having this information might help when potentially ambiguous interpretations of *assim* arise, as

different stress patterns indicate other functions (e.g. oxytone [a.'sĩ] usually has an adverbial meaning, while paroxytone-like ['a.sĩ] is more likely to function as a discursive maker). However, the intonational curve on its own is not always sufficient to resolve ambiguous readings (Silva, 1999, p. 340), and other strategies must be used. For instance, the corpus uses parenthetic quotes indicating that *assim* refers to a gesture made by the speaker, which points to *assim* being used as an adverb. When such information is not available, looking at the larger context of the answer in which *assim* appears helps identify potentially ambiguous tokens. Thus, transcriptions should suffice to identify uses of *assim* as a discourse marker.

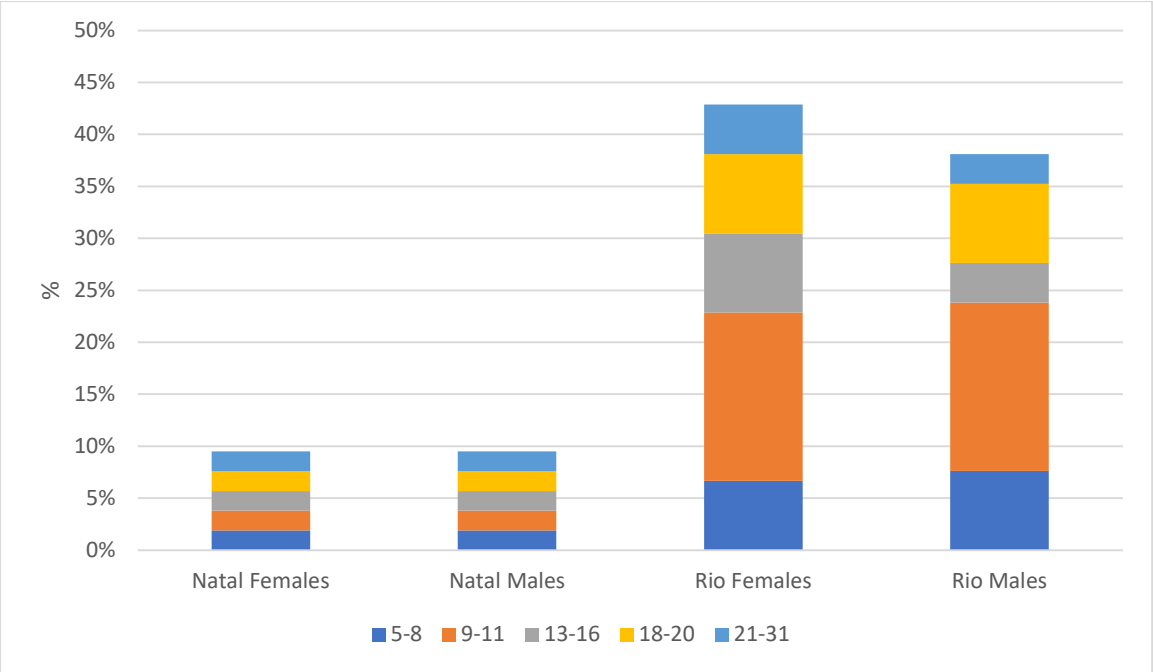
Another shortcoming regarding the D&G corpus is the speakers' restricted age range, preventing us from analyzing the use of *assim* in individuals older than 31. Finally, the fact that this corpus was built 28 years ago does not allow us to look at the current use of *assim*. However, this corpus being constructed in 1993 enables a close comparison with seminal works on either BP *assim* and English *like* from around the same time (Andersen, 2001; Tagliamonte, 2005; Barbieri, 2007; Silva & Macedo, 1996). For further analysis regarding the development of *assim* and possible linguistic change, collecting new data would be necessary.

3.2. Data collection

This study uses all speakers from Natal (n = 20) and Rio de Janeiro (n = 85) included in the D&G corpus. The distribution of speakers by age, gender, and location is displayed in Figure 2. These two locations were chosen for their geographical distance, their different socio-economic conditions (Ferrari, 2005), and their prototypical representation of the Northeast and Southeast dialects, respectively (Nascentes, 1955). According to the D&G research group, the fact that the number of speakers from Rio de Janeiro is higher than those from other locations simply reflects its larger population[1]. That being said, speakers from Rio de Janeiro produced fewer words (average of approximately 1,200 words per speaker) than those from Natal (about 8,000 words), resulting in an overall more similar number of words for each city.

[1] Information taken from the corpus section of their website: https://discursoegramaticablog.wordpress.com/corpus/ (retrieved on June 04, 2021). According to Brazil's 1991 Census, Rio de Janeiro had a population of approximately 10,000,000 and Natal 700,000.

Figure 2. Distribution of speakers by age, sex, and location



Even though Brazil presents a vast range of dialects (Figure 3), extralinguistic differences enable the perception of a prestigious one; it is acknowledged that it is the *carioca* dialect spoken in Rio de Janeiro. They argue this dialect is cosmopolitan and influential as it takes foreign terms and spreads local coinages (Nascentes, 1953). This status has been long-established since Brazil was still a colony of the Portuguese empire (Alencastro, 1997).

Figure 3. Brazil's dialectal division, according to Nascentes (1953)

Note: The bold line represents the core division of north and south speech, with the sub-varieties illustrated by the colour scheme. Names in italics are demonyms for the Brazilian states from where the variety came. Adapted from *Brazil Blank Map* by F. Menegaz, 2007, Wikimedia Commons (https://upload.wikimedia.org/wikipedia/commons/1/10/Brazil_Blank_Map.svg). CC BY-SA 3.0

Since 1763, still in the imperial regime, Rio de Janeiro was assigned as the capital of the State of Brazil. Given the social exchanges with imperial people and the European environment that came with them, the dialect spoken in Rio de Janeiro was quickly elevated to a prestigious status, as many influential people lived there and showed their authoritative speech to the rest of the country (Alencastro, 1997). Rio de Janeiro was the capital of Brazil only until 1960, but its influence by power, presented at political, economic, and cultural levels, remains.

The *carioca* dialect's prestige in the country qualified its establishment as the standard linguistic variety for cultural events. According to two theatre conferences in 1937 and 1956, this dialect was assigned in plays, poem readings, and erudite song performances (Rodrigues, 1986). Therefore, theatrical productions favoured actors and singers from Rio de Janeiro to be part of most of the stage plays and musicals throughout Brazil, and artists from other regions were trained to emulate the dialect.

The diffusion of this dialect was also facilitated using broadcast media. The first radio station was founded in Rio de Janeiro, transmitting news, educational programs, and cultural events to the whole country. The largest national television network (Rede Globo) is currently based in Rio de Janeiro. Carvalho (2004) found that Brazilian television is a source of prestigious speech, even influencing Portuguese speakers from neighbouring countries, like Uruguay. Professionals appearing on radio and television are mainly from Rio de Janeiro, making its dialect dominant in broadcast media programs (Carvalho, 2004).

The city of Natal had a different historical path. Just after the Portuguese conquest, the societal configuration of the region consisted of a few white males that explored agricultural activities with their enslaved Black and Indigenous people (Maia, 1998, as cited in Pereira, 2007). The city's first years were marked by poverty and slow economic growth (Cascudo, 1999). After 1633, the Dutch occupied the area for nearly twenty years, causing numerous conflicts which resulted in death and destruction (Mello, 2010). These factors resulted in a lack of prestige in those lands, regarded as unimportant to the Portuguese government (Mariz & Suassuna, 2002). The development and growth of Natal happened after World War II due to its favourable geographic position at the corner of the continent, bringing significant military personnel to the city at that time (Pereira, 2007).

The speech from Natal is associated with the *nordestino* 'northeastern' dialect, as the city is located in that respective region. One characteristic phonological process from the northeastern dialect is opening vowels, which does not happen in dialects from the south of the country (Aragão, 2015). Because of a substantial linguistic prejudice in Brazil, most people from the south hearing speakers from the northeast area regard them as "naïve, uneducated, or inelegant" (Maia, 1986, p. 210).

Transcriptions were downloaded from the corpus website, and participants' responses to the interviewer's questions were entered into an Excel spreadsheet. I then extracted all 1,370 occurrences of *assim* and coded them for speaker name, age, sex, location, interview topic and the syntactic criteria of their position in the clause.

3.3. Exclusion criteria

This thesis concerns the discursive function of *assim* only; thus, non-discursive uses (n = 264) were excluded from the analysis (section 3.3.1.). However, some discursive uses were excluded (n= 176), as specified in section 3.3.2.

3.3.1. Non-discursive uses

Non-discursive uses of *assim* are those where it functions as an adverb: its standard syntactic function (Romaine & Lange, 1991; Andersen, 2001) as described in dictionaries and prescriptive grammars (Bechara, 2009). The following items illustrate the uses that emerged from the adverbial function of *assim*.

1) Anaphoric uses. This use is regarded as the standard anaphoric adverbial function of *assim*, the word being roughly translated as 'like this' (6):

(6) *Achava que aqui deveria ser* ***assim*** *também no Brasil.*
I.thought that here it.might be **like.this** as.well in Brazil.
'I thought that it might be like this here in Brazil as well' (referring to a situation in the USA discussed earlier) [Gerson/M/18-20/Natal]

2) Gesture references. The deictic origin of *assim* makes it helpful to refer to the speaker's action while speaking. This function is another example of traditional use (7).

(7) *Coloca* ***assim*** *(gestos).*
You.put **like.this** (gestures).
'Put it like this' (participant makes a gesture) [Wesley/M/5-8/Natal]

3) Idiomatic expressions. Idiomatic expressions with *assim* are used as adverbs with adversative (*mesmo assim* 'even so') or temporal (*assim que* 'as soon as') meanings (8-9).

(8) *Mas* ***mesmo assim*** *continuava a amizade com a gente.*
But **even so** he.kept the friendship with the people.
'But even so, he kept friendship with us.' [Gerson/M/18-20/Natal]

(9) *Eu fui conhecendo esse pessoal* ***assim que*** *eu cheguei lá.*
I was meeting these people **as.soon.as** I got there
'I met these people as soon as I got there.' [Ítalo/M/21-31/Natal]

4) Approximative adverbs. Instances of *assim* with the meaning 'about' preceding numerals (10):

(10) *É muito difícil ter* ***assim*** *dois meses, três meses de intervalo*
It's very difficult to.have **about** two months, three months of interval.
'It's hard to have about two, three months' break.' [Rosemeire/F/18-20/Natal]

3.3.2. Phrases with discursive uses

The following criteria are based on previous literature on general extenders (Tagliamonte, 2005; Dubois, 1992), quotatives (Romaine & Lange, 1991), and move introducers (Souza, 2009).

1) Quotatives. *Assim* is used as part of a phrase, with a verb used for quotations. It is applied to introduce a reconstructed speech, as shown in (11).

(11) *Aí ela disse **assim**: "eu posso levar você lá".*
So she said **like**: "I can take you there".
'So, she was like: "I can take you there."' [Rosemeire/F/18-20/Natal]

2) General extenders. *Assim* is used as part of a phrase at the end of a sentence, "serving to extrapolate from what has previously been said" (Dubois, 1992, p. 179-80). It is crucial to emphasize that when *assim* is used as a general extender, although serving supra-sentential purposes, it modifies a noun. Essentially, *assim* functions here as an anaphoric adverb[2] and, therefore, would be excluded either way from the study. An example is presented in (12).

(12) *Acho que é onde o pessoal se reúne para debate, qualquer coisa **assim**.*
I.think that is where the people refl meet to debate, anything **like.that.**
'I think it is where people meet to debate or something like that.'
[Gerson/M/18-20/Natal]

3) Move introducers. All instances in which *assim* is used in combination with the verb *ser* 'be' (the phrase *é/foi assim* 'It is/was like'), operating to introduce a move, which is a unit in conversational analysis, at the beginning of sentences (Souza, 2009). Although the whole phrase can be categorized as a discourse marker because of its bleached lexical meaning and linking nature, this use of *assim* has been excluded because the word individually functions as an adverb in this fixed phrase, modifying the verb. *Assim* as a general extender also serves a similar purpose, but in this case, it modifies a noun rather than a verb. However, in both cases, *assim* is used as a non-discursive feature within a discursive phrase. An example is shown in (13).

[2] According to Bechara (2009), *assim* is still an adverb even when modifying a noun (p. 241), although as a general rule a word with this function would be categorized as an adjective.

(13) *É* ***assim****, imediatamente* *não.*
It's **like**, immediately not
'It's like, not immediately.' [Rosemeire/F/18-20/Natal]

3.4. Linguistic factors

The present study centers on the variation of the position of *assim* in the clause. For this reason, I follow the categorization proposed by Diskin (2017), who distinguished the variants of DM *like* based on this condition, labelling them as clause-initial, -medial, or -final. Her method involves prosodic and contextual signs, differentiating initial from final positions when *like* is used between clauses. For Diskin (2017), a "short period of speaker silence" (p. 149) determines the position of *like* in the clause. The clause-initial variant is preceded by silence.

Conversely, the clause-final variant is followed by silence. The clause-medial variant does not use silence as a condition but relies on semantic embedding in the discourse. Diskin (2017) was inspired by Denis & Tagliamonte (2016), who used the same convention to classify the DM *you know* according to its position in the utterance.

Following Diskin's (2017) analysis of like, this thesis uses the classification of *assim* based on its position in the clause, as follows.

1) **Clause-initial *assim*.** This variant occurs at the beginning of the turn or is preceded by silence[3] when it appears between clauses. Example (14) illustrates this variant as follows.

(14) ***Assim*** *no* *finalzinho* *da* *história não* *se* *falava mais nela.*
[Like] at.the very.end of.the story not INDET talked more in.her.
'[Like] at the end of the story, people did not talk about her anymore.'
[Gerson/M/18-20/Natal]

Example 14 shows the variant preceding the whole sentence. However, I classified another circumstance as clause-initial, which is not addressed in Diskin's (2017) study. *Assim* can be used inside the sentence but preceding a complement clause. In this case, this variant could theoretically be interpreted as clause-medial for its syntactic structure and placement regardless of the speaker's silence (see item 3). However, the clause-initial type is applied to align with the explanation discussed in D'Arcy (2017), who argued that the clause-initial *like* targets a CP

[3] All silences are conventionally transcribed as an ellipsis (…) in the corpus.

projection (p. 90) in matrix or subordinate contexts. Therefore, to have a comprehensive analysis of the occurrences of the word and enable comparison of similar phenomena, clause-initial *assim* has two subtypes: a) sentence-initial, as exemplified in (14), and b) sentence-medial, when preceding a complement clause, shown below in (15):

(15) *As músicas* ***assim*** *que eu não conheço.*
The songs [like] that I don't know
'The songs [like] which I don't know.' [Vladmir/M/13-16/Natal]

2) **Clause-final *assim***. This variant occurs at the end of the turn or followed by a silence when it appears between clauses (16).

(16) *Aí teve a história da conexão* ***assim****.*
Then there.was the story of.the connection [**like]**
'Then there was the story of the connection [like].' [Ítalo/M/21-31/Natal]

In my data, I coded for the syntactic categories preceding *assim* in clause-final position. The syntactic categories may be noun, verb, adjective, adverb, or preposition, determiner, auxiliary, degree word, coordinator, complementizer, or another discourse marker. Example (16) above shows *assim* preceded by a noun. Examples (17-18) show the use of *assim* preceded by a verb and an adjective, respectively.

(17) *Como eu posso explicar* ***assim****?*
How I can explain [like]
'How can I explain [like]?' [Jorge Luiz/M/21-31/Rio]

(18) *A cantina é bem grande* ***assim***
The canteen is very big [like]
'The canteen is very big [like]' [Flavia M./F/9-11/Rio]

In cases where another DM follows clause-final *assim*, I still considered it a clause-final variant. DMs are syntactically optional, so they do not formally incorporate the clause, keeping the preceding *assim* positioned at the end of the structure. Furthermore, DMs do not contribute to the propositional meaning of the head of the phrase (Pichler, 2013), and do not change the circumstance where *assim* is placed in the clause. An example of clause-final *assim* followed by a DM is presented in (19).

(19) *Minha mãe me contou quando eu era criança bem*
My mom me told when I was child very
pequena ***assim*** *né?*
little [like] isn't.it
'My mom told me when I was a tiny child [like]' [Angela/F/9-11/Rio]

3) **Clause-medial *assim***. The categorization of this variant occurs when *assim* is placed within a clause, embedded in both directions, as (20) shows.

(20) *Aparece uma luz* ***assim*** *bem bonita.*
Appears a light [like] very beautiful
'It shows [like] a very beautiful light.' [Rosemeire/F/18-20/Natal]

This example differs from the sentence-medial clause-initial *assim* because the following complement phrase is not a clause. When *assim* occurs in clause-medial positions, it introduces a complement (Silva & Macedo, 1992) and has the attribute of scoping in both directions (Lopes-Damasio, 2008). BP being a head-initial language, it is implied that *assim* follows the head of the phrase. Therefore, the preceding element was labelled as noun, verb, adjective, adverb, preposition, determiner, auxiliary, degree word, coordinator, complementizer, or another discourse marker in the same approach used for the clause-final variant. Examples (21-23) illustrate circumstances where clause-medial *assim* appears after a verb, an adjective, and an adverb, respectively.

(21) *Podia até repartir* ***assim*** *cada um compartimento*
Could even divide [like] each one section
'You could even divide [like] each of the sections' [Rosemeire/F/18-20/Natal]

(22) *Ela morava num lugar tão esquisito* ***assim*** *pra sair.*
She lived in.a place so weird [like] to go.out.
'She lived in a place so weird [like] to go out.' [Rosemeire/F/18-20/Natal]

(23) *As pessoas estão totalmente* ***assim*** *perdidas*
The people are totally [like] lost
'People are totally [like] lost' [Rafaela/F/21-31/Rio]

3.5. Extralinguistic factors

Following previous sociolinguistic research on discourse markers, I included the following information for each token, as provided in the D&G: speaker sex (Silva & Macedo, 1996; Tagliamonte, 2005), speaker age group (Silva & Macedo, 1996; Tagliamonte, 2005), speaker location (Andersen, 2001), and topic of the interview (Holmes, 2013).

1) Speaker age group: The D&G corpus does not provide speaker's date of birth, but rather the school grade each individual was attending at the time data was collected. Thus, for each Natal and Rio de Janeiro speaker I can only provide an age range corresponding to the grade they were attending, as coded in the D&G: 5-8, 9-11, 13-16, 18-20, and 21-31 year-old.

2) Speaker sex: male and female.

3) Speaker location: Natal and Rio de Janeiro.

4) Topic: personal narrative, retold narrative, place description, instruction, and opinion.

3.6. Statistical analysis

Following Tagliamonte et al. (2016) and Diskin (2017), I applied a generalized linear mixed-effects model to my data. In their study of quotative *be like* in cities in Canada and New Zealand, Tagliamonte et al. (2016) used this model and found patterns based on the interaction between predictors. Diskin (2017) used this model for clause-final *like* to test the relative effect of independent variables against the presence or absence of *like*, with the addition of random effect for the speaker to better account for interspeaker variation (p. 149). Thus, this model fits my data because the study aims to find the social predictors' strength in the variation in positions of *assim* in the clause. The mixed-effects model suits a condition where there are many subjects in a naturally unbalanced set, as all sociolinguistic data is drawn from "less than ideally distributed datasets" (Tagliamonte, 2012, p. 139). Moreover, using the individual as a random effect, the model considers that the use of the variable might differ in the frequency of specific speakers, which is not the pattern predicted from the group as a whole (Johnson, 2009). Therefore, this model disregards outliers.

The mixed-effects model was performed using the *glmer* package (Bates et al., 2015) in RStudio. All dependent variants (clause-initial, clause-final, and clause-medial *assim*) and independent factors (age, sex, location, and interview topic), with all likely interactions, were

included in the formula. I also applied this model by grouping speakers for sex and location, creating the following four subsets: a) only male speakers from both Natal and Rio de Janeiro, b) only female speakers from both Natal and Rio de Janeiro, c) both male and female speakers only from Natal, and d) both male and female speakers only from Rio de Janeiro.

Additionally, as clause-medial *assim* is prevalent in the corpus and appears inside many different phrases, the statistical significance of the distribution of this variant was also calculated. Because I am dealing with nominal data and comparing categories (i.e. using the cross-tabulation of sex and each of the other factors at a time: age groups, interview topics and preceding elements), I used an χ^2 test, an established model applied by previous research on discourse markers (e.g., Andersen, 2001; Barbieri, 2007). Selecting clause-medial *assim* and separating the data for each location, showing their sex pattern in frequency, makes it possible to state each distribution, checking the significance of their correlation by the χ^2 test. The frequency is normalized at one thousand words, following Andersen (2001) and Barbieri (2007).

Chapter 4: Results

This chapter describes the overall distribution of *assim* in the data and provides the statistical analysis performed to derive these results. Section 4.1 reports the distributional patterns of *assim* in the corpus, based on speaker location, sex, age, the topic of conversation, and placement in the clause. This section also presents the distribution of clause-final and clause-medial *assim* according to the preceding word and a cross-tabulation of linguistic and extralinguistic factors. Section 4.2 presents the results from the mixed-effects model and the χ^2 for the clause-medial variant.

4.1. Distributional analysis

4.1.1. Overall distribution of assim

Assim tokens (n = 930) are relatively balanced across the two cities, with speakers from Natal producing 54% (n= 502) of the tokens and those from Rio de Janeiro, 46% (n = 428). The distribution of tokens by speaker sex (Table 1) is relatively balanced as well, with 42% (n = 393) of total tokens being produced by males and 58% (n = 537) by females.

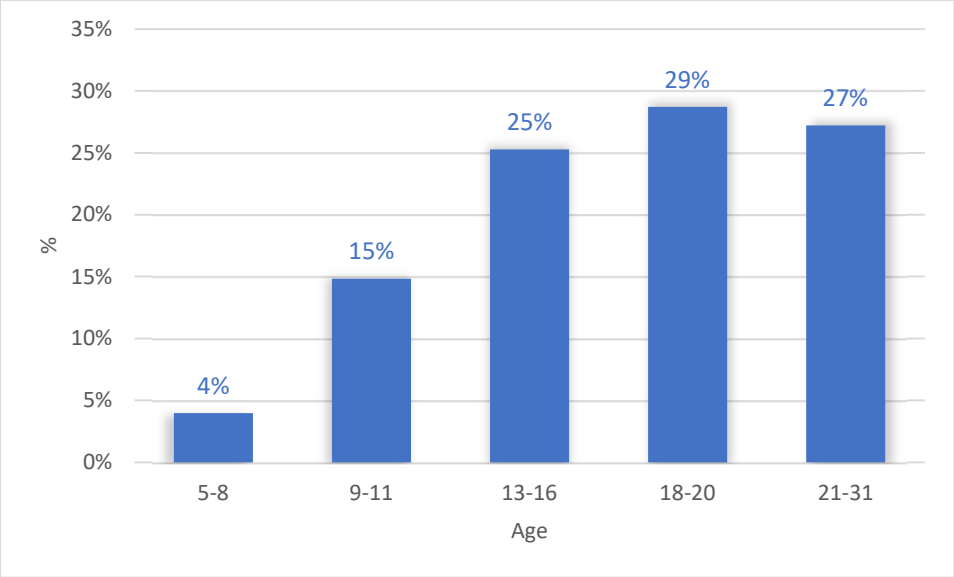
Table 1. Distribution of total tokens (n = 930) by location and sex

	NATAL	RIO DE JANEIRO
MALES	198 (21.29%)	195 (20.97%)
FEMALES	304 (32.69%)	233 (25.05%)
TOTAL	502 (53.98%)	428 (46.02%)

As for the distribution of *assim* by speaker age (Figure 4), the two youngest groups, 5-8 and 9-11 year-old, have the lowest production, with 4% (n = 37) and 15% (n = 138) of the tokens, respectively. This pattern is consistent with the frequency of *assim* per thousand words in each group (2.0 tokens per thousand words in the 5-8 year-old group, and 3.2 per thousand words in the 9-11 year-old group). However, two differences emerge in comparing the overall distribution of *assim* and the frequency of tokens per thousand words. First, the 13-16 year-old group has the highest rate of frequency per thousand words, with 4.9 tokens, contrasting with the percentages by age, where the 18-20 year-old group has the highest proportion (29%, n = 267). Moreover, there is a sharp decline in the 21-31 age group rate at 3.3 tokens per thousand words, placing them very close to the frequency found in the 9-11 year-old group. This frequency differs

from the proportion of *assim* in the 21-31 year-old group, where it maintains a high percentage of 27% (n = 253).

Figure 4. Overall distribution of assim (n = 930) by speaker age



Regarding the topics discussed in the interviews (Figure 5), personal narratives and stating opinions account for more than half of the total tokens, with 28% (n = 258) and 25% (n = 229), respectively, followed by retelling accounts and describing places, with 17% each (n = 163 and n = 157, respectively). Lastly, giving instructions has the least frequent use of *assim*, with 13% (n = 123) of the tokens.

Figure 5. Overall distribution of assim (n = 930) by topic

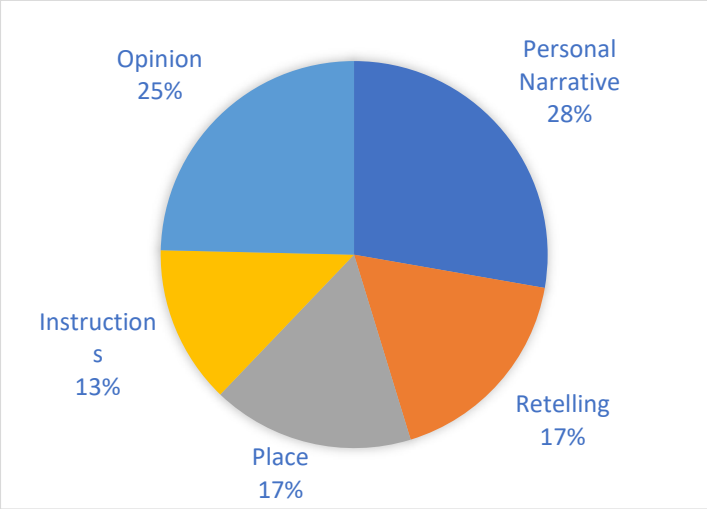


Figure 6 shows the distribution of *assim* according to its placement in the clause. *Assim* appears more often in the medial position (within the clause structure), with 59% (n = 545) of the total tokens. The rates of initial and final placements were mutually similar, with 20% (n = 187) and 21% (n = 198) respectively.

Figure 6. Overall distribution of assim (n = 930) by placement in the clause

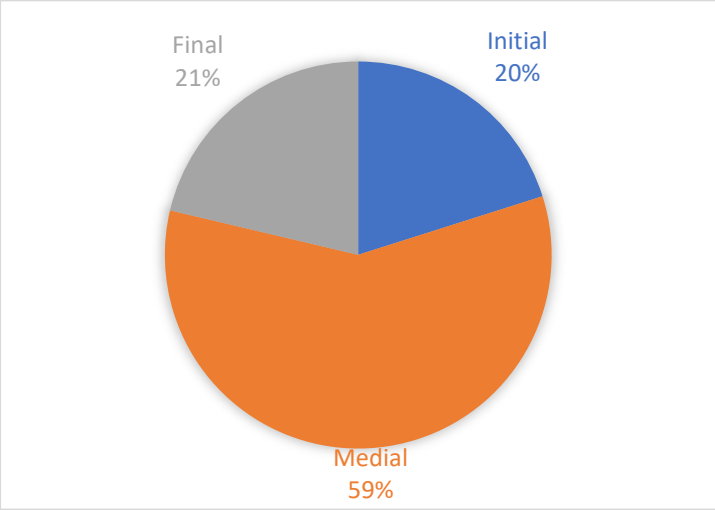


Figure 7 shows the distribution of clause-medial *assim* according to its preceding element. Nouns and verbs are prevalent in this context, representing 83% of the cases (40%, n = 220, and 43%, n = 232, respectively). Nouns and verbs are also the most common syntactic categories preceding *assim* in clause-final position, with 49% (n = 97) and 26% (n = 51) of the tokens, respectively (Figure 8).

Figure 7. Distribution of clause-medial assim (n = 545) by preceding element

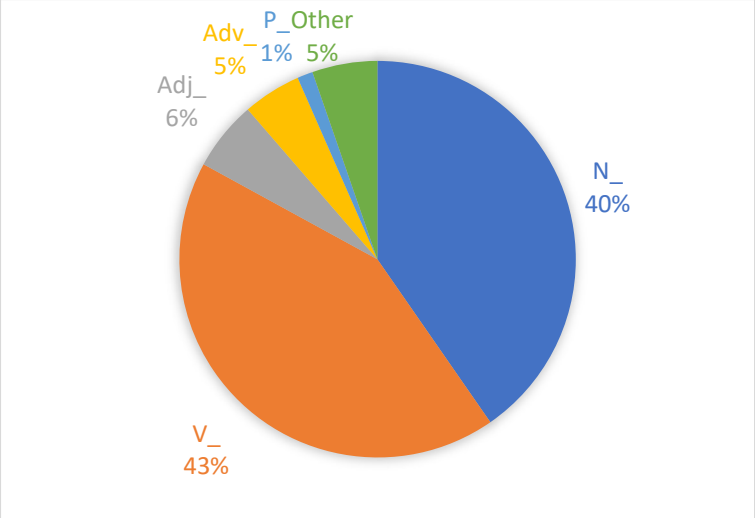
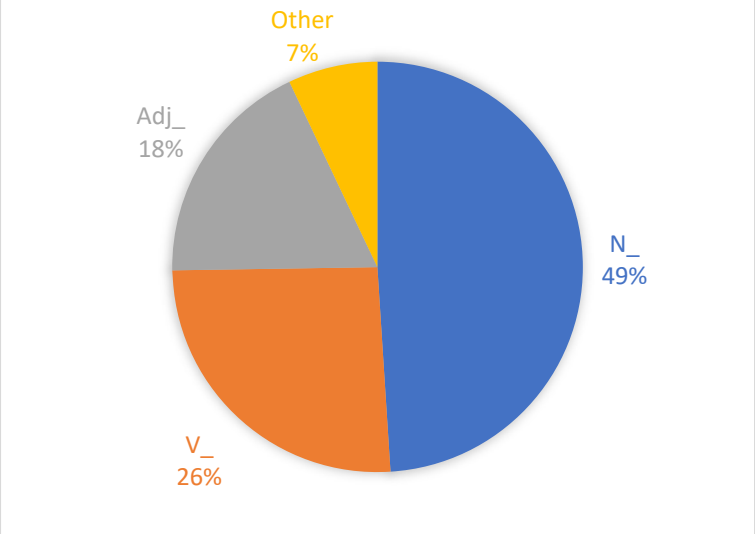
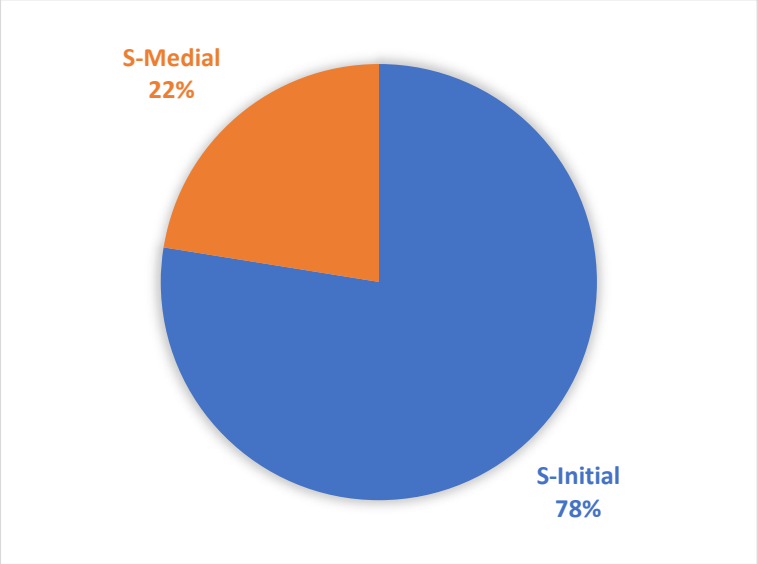


Figure 8. Distribution of clause-final assim (n = 198) by preceding element



Regarding clause-initial placements of *assim* (Figure 9), the distribution shows that most tokens are used at the beginning of a sentence (78%, n = 145), while tokens of clause-initial *assim* that are placed within a sentence structure (i. e., a dependent or relative clause) are less common (22%, n = 42).

Figure 9. Distribution of clause-initial assim (n = 187) by sentence placements



4.1.2. Cross-tabulation of factors (Figure 10) displays the cross-tabulation of syntactic position and speaker sex. Females produce more tokens of *assim* in the medial position than males (35% vs. 24%), which also occurs in the initial position, although to a lesser degree (12% vs. 8%). In contrast, the number of tokens in the final position is similar for both groups (11% each).

Figure 10. Distribution of assim (n = 930) by clause placement and sex

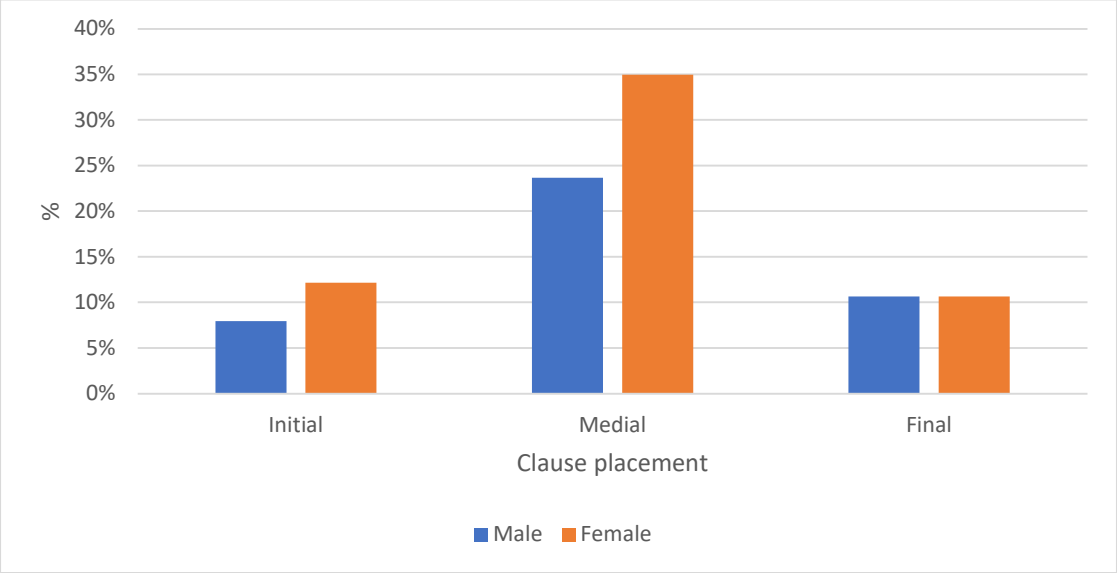


Figure 11 displays the distribution of *assim* in clause-medial position by speaker sex and age. Younger age groups (5-8 and 9-11 year-old groups) use fewer tokens than older groups, echoing the results displayed in Figure 4. Data also shows that clause-medial *assim* is proportionally more frequent among females in the 18-20 year-old group (27%, n = 148), noticeably decreasing its use in the next age group (8%, n = 43). In contrast, males in the 18-20 year-old group rarely use *assim* (8%, n = 41), but increase its use in the 21-31 year-old group (16%, n = 86).

Figure 11. Distribution of clause-medial assim (n = 545) by age and sex

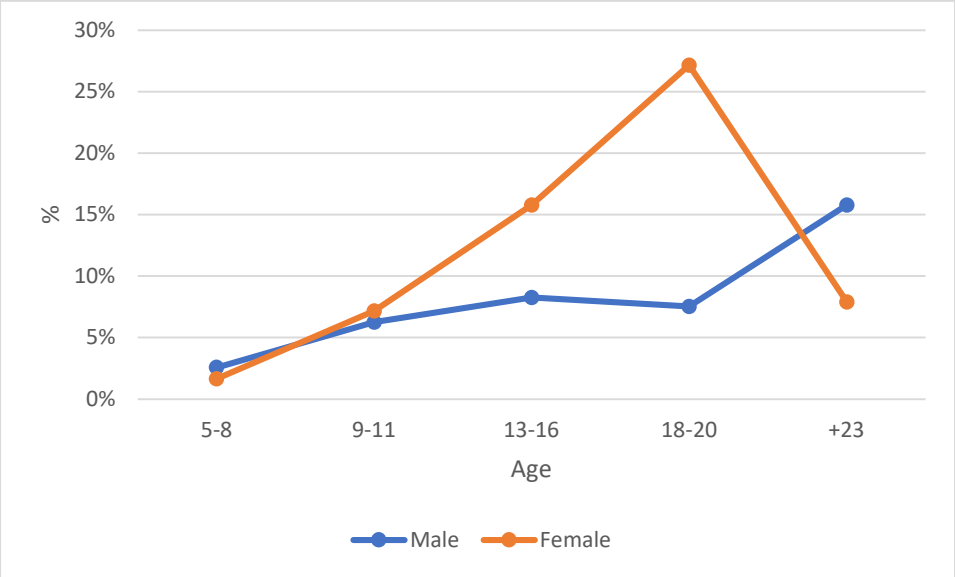


Figure 12 illustrates the distribution of *assim* in clause-medial position according to the preceding element and speaker sex. *Assim* is used after nouns equally by males and females (20%, n = 110 each), while gender differences arise when *assim* is preceded by a verb, with females producing more of these tokens (29%, n = 158) than males (14%, n = 74).

Figure 12. Distribution of elements preceding clause-medial assim (n = 545) by sex

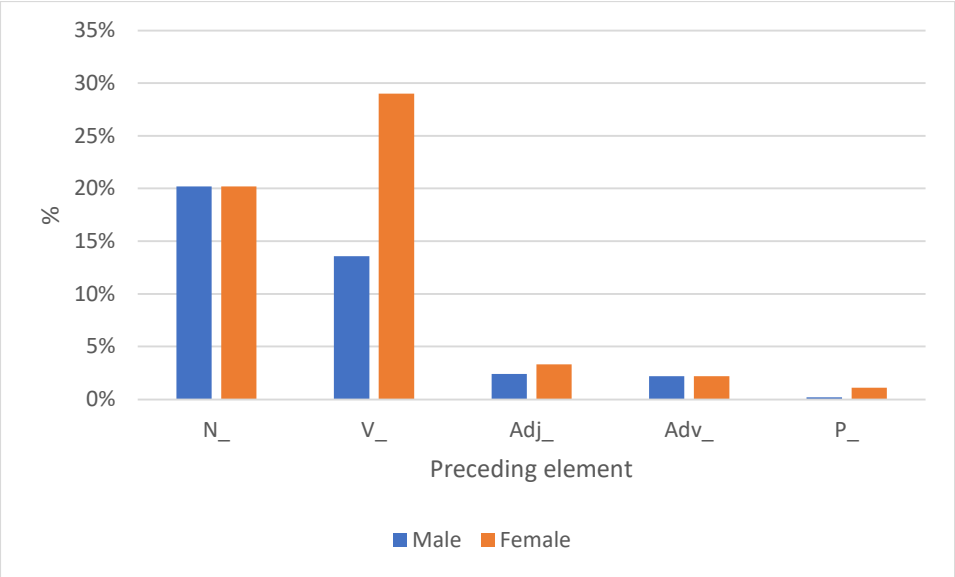


Figure 13 shows that males and females produce a very similar number of tokens of clause-final *assim* after a noun (25% vs. 24%, respectively). This similarity is also present, although at lower rates, for uses of *assim* after verbs and adjectives.

Figure 13. Distribution of elements preceding clause-final assim (n = 198) by sex

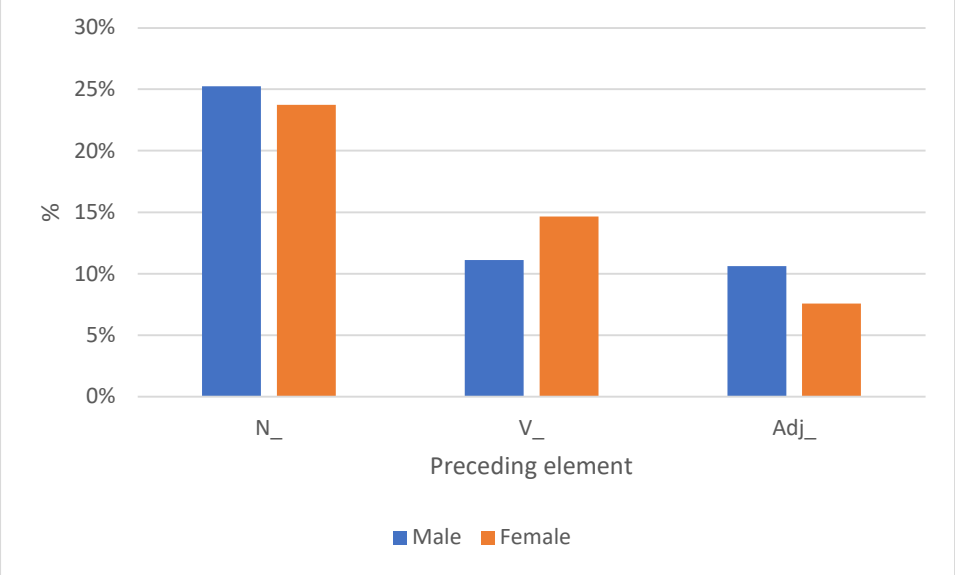
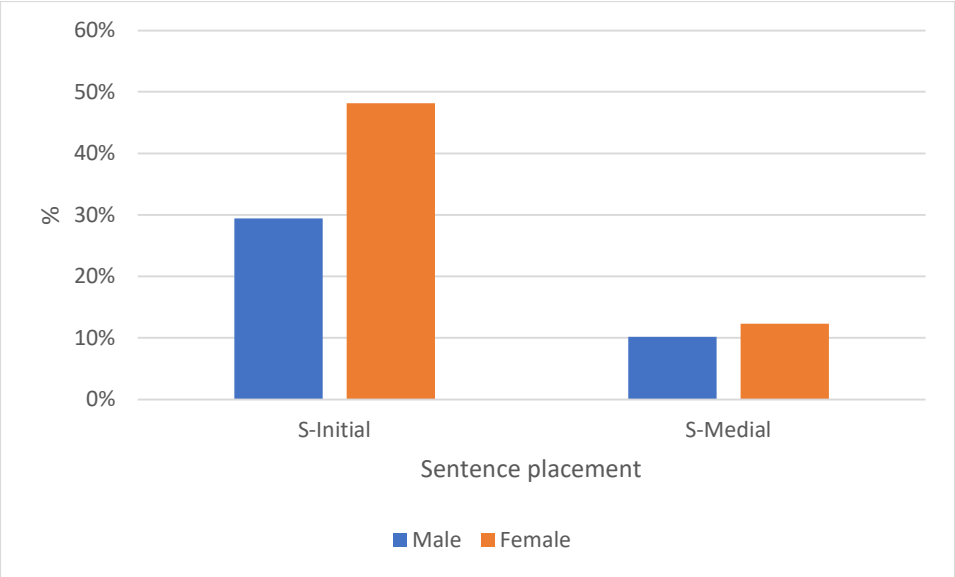


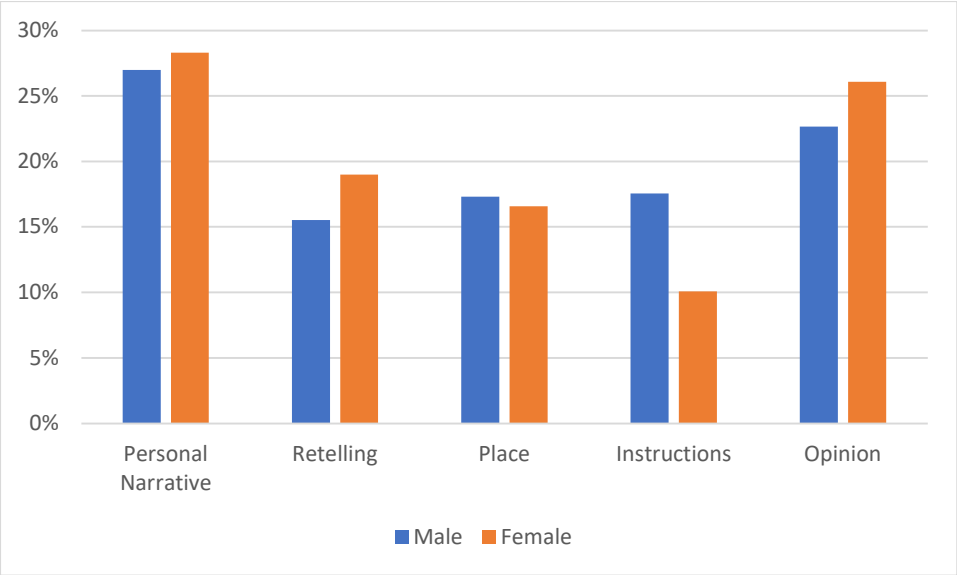
Figure 14 shows that females produce more tokens of clause-initial *assim* at the beginning of sentences (48%, n = 90) than males (29%, n = 55). In contrast, the use of *assim* in sentence-medial position is very similar across genders.

Figure 14. Distribution of clause-initial assim (n = 187) by placement in the sentence and sex



Concerning the interview topics, Figure 15 shows the distribution relative by each sex. Most tokens for each sex appear in personal narratives (28%, n = 152 for females, and 27%, n = 106 for males) and opinions (26%, n = 140 for females, and 23%, n = 89 for males). *Assim* is used at similar lower rates when talking about a place (17%, n = 89 for females, and 17%, n = 68 for males) and retelling (19%, n = 102 for females, and 16%, n = 61 for males). Males' use of *assim* for instructions (18%, n = 69) is similar to their use in retelling and describing a place, while females present their lowest use of assim in this category (10%, n = 54).

Figure 15. Distribution of assim (n = 930) by topic and sex



4.2. Statistical analysis

4.2.1. The mixed-effects tests

Table 1 shows the results from running a mixed-effects model and using the speaker as a random effect. As I deal with a linguistic variable with three variants (clause-initial, clause-medial, and clause-final) and this model uses a binomial function, results differ depending on the applied reference level. The reference level refers to R's treatment coding, employed to evaluate factor groups in its estimates (Tagliamonte 2012). The model's scheme compares the variant put as the reference level with the others. The probability of one variant is shown on the 'Intercept' row. A positive value means that the variant selected as the reference has a favourable use in contrast to the others. If the value is negative, the interpretation is that this variant is less likely to occur. But the binary structure that this model is built on impedes the prompt visualization of a preference if the result comes in a negative value because it groups together the other not-referenced variants. Therefore, to have a true visualization of the favourable use of a specific variant, a positive value is required, assessing more than one reference, if necessary.

After testing all three positions (clause-initial, clause-medial, and clause-final) individually as the reference level, adding all predictors (age, sex, location, and topic) with and without interaction, only clause-final *assim* showed significant effects for substantial factors. Apart from this finding, the only combination of factors showing significant results is the

interface of sex and location[4], as Table 2 shows. It is crucial to say that the factors applied in the formula must be suitable in every environment to equally test all variants. For this reason, the factor that accounts for the preceding elements is removed from this step, as it does not apply in clause-initial settings.

Table 2. A mixed-effects model with clause-final assim as the reference level, individual as random, interaction between sex and location

Formula: Dep.var ~ Age + Topic + Sex * Location + (1 | Individual)

Data: *assim*

AIC	**BIC**	**logLik**	**Deviance**
925.2	988.1	-449.6	899.2
RANDOM EFFECTS			
Group	**Name**	**variance**	***SD***
Individual	(Intercept)	0.09758	0.3124
Number of obs: 930, group: Individual, 76			
FIXED EFFECTS			
	ESTIMATE	***SE***	***z*-value**	**Pr(>|*z*|)**
(Intercept)	1.66878	0.36605	4.559	5.14e-06 ***
18-20	0.94768	0.32023	2.959	0.003082 **
13-16	0.17960	0.27103	0.663	0.507557
9-11	0.39051	0.30467	1.282	0.199943
5-8	0.80546	0.51107	1.576	0.115018
Instructions	0.33854	0.31834	1.063	0.287571
Personal Narrative	-0.29759	0.24974	-1.192	0.233415
Opinion	0.33949	0.27711	1.225	0.220528
Retold Narrative	-0.05349	0.30013	-0.178	0.858561
Male	-1.15015	0.35703	-3.221	0.001276 **
Rio	-0.98468	0.31920	-3.085	0.002036 **
Male:Rio	1.51725	0.45039	3.369	0.000755 ***
Signif. Codes : 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1				

[4] The interaction of factors is represented by an asterisk in the formula.

What can be stated from this model's most salient differences is that speakers' social characteristics, i.e. age, sex, and location, can explain the use of *assim*. The age factor shows that the 18-20 year-old group uses more clause-final *assim* than the 21-31 year-old group ($p = 0.003082$). Sex and location indicate that males ($p = 0.001276$) and speakers from Rio de Janeiro ($p = 0.002036$) disfavour the use of clause-final *assim.* The negative estimates in these factors demonstrate the lack of preference in using this variant. Sex has an estimate of -1.15015 for males, and location has -0.98468 for Rio.

On the other hand, the interaction of these two factors – sex and location – shows a considerable likelihood of using the clause-final variant, represented by the positive estimate (1.51725) with a more robust significance level ($p = 0.000755$), explicitly coming from males and speakers from Rio de Janeiro. Furthermore, this result demonstrates that there is a stronger effect in difference between sexes when interacting with location than when considering those factors individually. It means that the location factor strongly affects the pattern of use among male and female speakers. Finally, the interview topics are not relevant for using the variable, demonstrating that only social predictors influence *assim* usage.

The mixed-effects model applied for clause-medial *assim* as the reference level (i.e. using this variant compared to the other ones) is represented in Tables 3 and 4. Table 3 takes into account the data from both locations. In this setting, this variant only shows significance for the 18-20 year-old group ($p = 0.0311$); its negative estimate means that this group is less likely to use clause-medial *assim*. The combined interpretation of Tables 2 and 3 reveals that the 18-20 year-old group prefers using clause-final *assim* and is less likely to use the clause-medial variant. It is important to recall that the formula that resulted in significant values contains an interaction of sex and location, indicating that the age group preference emerges from differences between males and females across cities. Once again, these results confirm that the choice of *assim* variant is not affected by the topic of the conversation.

Table 3. A mixed-effects model with clause-medial assim as the reference level, data from both locations

Formula: Dep.var ~ Age + Topic + Sex * Location + (1 | Individual)

Data: *assim*

AIC	**BIC**	**logLik**	**deviance**
1206.8	1269.7	-590.4	1180.8
RANDOM EFFECTS			
Group	**Name**	**variance**	***SD***
Individual	(Intercept)	0.1796	0.4238
Number of obs: 930, group: Individual, 76			
FIXED EFFECTS			
	ESTIMATE	***SE***	***z*-VALUE**	**Pr(>|*z*|)**
(Intercept)	-0.4342	0.3353	-1.295	0.1953
18-20	-0.6574	0.3049	-2.156	0.0311 *
13-16	-0.2485	0.2815	-0.883	0.3773
9-11	-0.1978	0.2940	-0.673	0.5011
5-8	-0.2107	0.4346	-0.485	0.6279
Instructions	-0.1302	0.2641	-0.493	0.6219
Personal Narrative	-0.0811	0.2201	-0.368	0.7126
Opinion	-0.1923	0.2281	-0.843	0.3993
Retold Narrative	-0.4731	0.2608	-1.814	0.0697 .
Male	0.4333	0.3392	1.278	0.2014
Rio	0.5337	0.2951	1.808	0.0706 .
Male:Rio	-0.5976	0.4245	-1.4008	0.1592
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1				

When testing the model with all four subsets separately, speakers from Natal were the only group that returned a difference with statistical significance. As Table 4 shows, male speakers prefer using clause-medial *assim* (p = 0.03951). However, it is crucial to note that, in general, based on the intercept row, regardless of sex, speakers from Natal are unlikely to use

clause-medial *assim,* according to the negative estimate with a more robust significance code (p = 0.00131).

Table 4. A mixed-effects model with clause-medial assim as the reference level, Natal data only

Formula: Dep.var ~ Age + Topic + Sex + (1 | Individual)

Data: *assim*_Natal

AIC	**BIC**	**logLik**	**deviance**
620.7	667.1	-299.3	598.7
RANDOM EFFECTS			
Group	**Name**	**variance**	***SD***
Individual	(Intercept)	0.0415	0.2037
Number of obs: 502, group: Individual, 19			
FIXED EFFECTS			
	ESTIMATE	***SE***	***z*-VALUE**	**Pr(>|*z*|)**
(Intercept)	-1.27402	0.39648	-3.213	0.00131 **
18-20	-0.21472	0.34053	-0.631	0.52834
13-16	0.56963	0.33415	1.705	0.08825 .
9-11	0.77757	0.69367	1.121	0.26231
5-8	0.17161	0.58721	0.292	0.77010
Instructions	0.08442	0.38121	0.221	0.82475
Personal Narrative	0.41394	0.33246	1.245	0.21310
Opinion	0.28470	0.35095	0.811	0.41725
Retold Narrative	-0.27572	0.38296	-0.720	0.47154
Male	0.59175	0.28742	2.059	0.03951 *
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘’ 1				

To summarize Tables 2 to 4, the social factors of age, sex and location play a vital role in choosing the placement of *assim* in the clause. The interaction of sex and location applied in the formula is the only situation where a significant result occurs. In the 18-20 year-old group, speakers prefer clause-final *assim* (estimate 0.94768, p = 0.003082) and disfavour the clause-medial variant (estimate -0.6574, p = 0.0311), contrasting with speakers in the 21-31 year-old

group (the reference level in this model). Specifically for the data from Natal, the only significant subset, male speakers are likely to use clause-medial, compared with females from Natal, who do not favour its use, impacting the negative intercept result. Topics of conversation turn out not to be significant in any instance.

4.2.2. The χ^2 tests

Results from the mixed-effects statistical analysis suggest a difference between locations when it comes to sex differences. Applying the χ^2 test in a cross-tabulated setting determines the correlation between sex and other factors, such as age, preceding elements, and interview topics in each location. Tables 5 to 7 show the results from these analyses. The frequency of *assim* relative to its total words is incorporated in each group. The reason is to compare this frequency with the overall percentage of the word. Following the methodology applied by Andersen (2001) and Barbieri (2007), the frequency of tokens is normalized at one thousand words. This control allows results to be presented in values approximate to one.

Only Natal shows statistically significant correlations between *assim* usage and the potential predictive factors, which means that the use of *assim* is determined by speaker sex in this city. The variation pattern observed in Natal drastically contrasts with the data from Rio de Janeiro, which shows no significant variation in any factor. The absence of significance in Rio de Janeiro makes it unnecessary to illustrate its respective tables since there is no critical variation to pinpoint. As explained in section 3.6, clause-medial *assim* is the selected type when I intend to demonstrate a pattern from a single variant. It is the most complex feature, with many possible intra-clause positions, and makes up the highest number of tokens in the data (n = 545).

Table 5 displays a significant correlation between age and sex in the use of clause-medial *assim* in Natal. More specifically, the highest proportion is seen from female speakers in the 18-20 year-old group (41.82%, n = 133), dramatically decreasing its usage in the 21-31 year-old group, where the proportion is 6.6% (n = 21), even lower than what is observed from males in the same age group (18.24%, n = 58). This pattern also occurs when looking at the frequency of *assim* per thousand words. The p-value demonstrates that the disparities between males and females across ages in Natal are statistically significant. It means that from one age group to the other, the frequency of usage of *assim* develops differently between males and females (e.g. different rates of increasing or decreasing use depending on the sex group).

Table 5. Distribution of the clause-medial assim in Natal according to age and sex

	N	%	per 1,000 words	Pearson's χ^2 test
Male				$\chi^2 = 98.744$
5-8	9	2.83	2.50	d.f. = 4
9-11	0	0.00	0	Significant at $p < 2.2^e\text{-}16$
13-16	24	7.55	1.57	
18-20	19	5.97	1.10	
21-31	58	18.24	1.45	
Female				
5-8	2	0.63	0.38	
9-11	7	2.20	0.85	
13-16	45	14.15	2.68	
18-20	133	41.82	4.00	
21-31	21	6.60	1.32	
Σ	318	100		

When the interview topic is cross-tabulated with speaker sex, Table 6 shows that Natal also presents a statistical significance in the use of *assim,* represented by males' lower frequency of usage. More specifically, females show a higher proportion of usage when expressing an opinion (28.62%, n = 87), or retelling someone else's story (25.33%, n = 77). In contrast, females have a low proportion in providing instructions, with 7.24% (n = 22). Again, the χ^2 test validates the statistical significance, showing a substantial difference between males and females.

Table 6. Distribution of the discourse marker assim in Natal according to topic and sex

	N	**%**	**per 1,000 words**	**Pearson's χ^2 test**
Male				$\chi^2 = 18.2218$
Personal Narrative	54	27.27	2.81	d.f. = 4
Retelling	35	17.68	2.56	Significant at p = 0.001117
Place	34	17.17	2.45	
Instructions	49	24.75	2.36	
Opinion	26	13.13	2.10	
Σ	198	100		
Female				
Personal Narrative	82	26.97	4.23	
Retelling	77	25.33	4.06	
Place	36	11.84	3.27	
Instructions	22	7.24	2.73	
Opinion	87	28.62	3.93	
Σ	304	100		

Lastly, the other linguistic factor that differs between male and female speakers in Natal is the word that precedes clause-medial *assim*. There is a statistical correlation concerning the use of *assim* in Natal based on the preceding element in a clause-medial environment, as the χ^2 test shows in Table 7. Females show highest uses of *assim* when it is preceded by a verb (33.65%, n = 107), and males' highest use is when *assim* is preceded by a noun (19.5%, n = 62), although females show the same usage level in this environment.

Table 7. Distribution of the clause-medial assim in Natal according to preceding grammatical category and sex

	N	%	per 1,000 words	Pearson's χ^2 test
Male				$\chi^2 = 41.187$
Noun	62	19.50	0.78	d.f. = 8
Verb	22	6.92	0.28	Significant at p = 1.924e-06
Adjective	10	3.14	0.13	
Adverb	7	2.20	0.09	
Preposition	0	0.00	0.00	
Others	11	3.46	0.14	
Female				
Noun	62	19.50	0.78	
Verb	107	33.65	1.34	
Adjective	10	3.14	0.13	
Adverb	11	3.46	0.14	
Preposition	4	1.26	0.05	
Others	12	3.77	0.15	
Σ	318	100		

In sum, there is a significant correlation between sex with other extralinguistic and linguistic factors in the city of Natal. More specifically, regarding age, females favour the use of clause-medial *assim* in the 18-20 year-old group. Furthermore, females from Natal show a preference for their use of *assim* when talking about personal experiences or retelling someone else's. Lastly, in the female group, the element that precedes *assim* within a sentence tends to be a verb. A different situation happens in Rio de Janeiro, as there is no statistical significance among these factors, suggesting a similar use of *assim* by males and females.

Chapter 5: Discussion

This chapter interprets the results from the quantitative and statistical analysis. It draws from previous sociolinguistic literature to answer the three research questions posed earlier: (1) What are the linguistic factors that determine the use of *assim*?, (2) Do speakers' location, sex, and age help explain the variation in the placement of *assim* in the clause?, and (3) Do the different topics of conversation influence the use of *assim*?

5.1. RQ1: What are the linguistic factors that determine the use of *assim*?

Overall, *assim* is more commonly used in the clause-medial position in both Rio de Janeiro and Natal, confirming Strapazzon's (2018) findings in Chapecó. My data also shows a preference by male speakers for using *assim* after nouns (24) and female speakers after verbs (25) in both clause-medial and clause-final positions. Similarly, Cheshire (2005) found that in British English, males tend to use *like* before noun phrases based on the focuser function of the DM, while in Levey's (2006) study, girls favour placing *like* before verb phrases. This sex pattern in the use of DMs, depending on the category of the following word, could be explained by males' orientation towards the referential function of the language (i.e., informational content) and females' orientation towards the affective function (i.e., feelings or emotions) (Holmes, 2013; Cheshire, 2005). In other words, males are more likely to use a DM to highlight a noun because they are more concerned than females with drawing attention to the new information they are introducing in conversations (Erman, 1992), and also because of the referential function inherent to nouns (Levey, 2006).

(24) Nós trabalhamos com varios cartões *assim* pra casamento.
We work with many cards [like] for wedding
'We work with many cards [like] for a wedding.' [Nilson/M/13-16/Rio]

(25) *É muito dificil uma pessoa encontrar assim um casal que seja feliz.*
It's very hard one person find [like] one couple that be.subj happy
'It's tough for a person to find [like] a happy couple.' [Lucia/F/13-16/Natal]

My findings confirm previous claims regarding the unlikelihood of *assim* occurring between a determiner and noun and between a subject and a verb (Silva & Macedo 1992; 1996; Lopes-Damasio, 2008), with less than 1% of total tokens representing these syntactic placements.

Furthermore, no tokens of *assim* before auxiliary verbs (a placement not mentioned in previous studies) were found in my data, suggesting that *assim,* similarly to *like* (Andersen, 2001), is categorically absent in this syntactic slot. This lack of results indicates that *assim* is resistant to being between a filled specifier and the head of a noun phrase or a tense phrase.

5.2. RQ2: Do speakers' location, sex and age help explain the variation in the placement of *assim* in the clause?

Results show that the speaker's location plays a role in the use of *assim,* with speakers from Natal showing less preference for clause-final *assim* and those from Rio de Janeiro favouring this placement. Taking into account that the *carioca* dialect spoken in Rio de Janeiro is considered the standard variety and is the one used on televised media (Massini-Cagliari, 2004), it could be hypothesized that clause-final *assim* represents an innovative form, along the lines of *like* in standard varieties of English (Ferrara & Bell, 1995; Andersen, 2001). Although no statistically significant differences in the frequency of *assim* in clause-medial position were found between locations, the χ^2 tests show a female-led pattern of clause-medial *assim* in Natal, where the low-prestige *nordestino* dialect is spoken. In contrast, no differences between males and females were found in Rio. These findings challenge those of Silva (1999), who found that speakers from northern areas in Brazil use *assim* less than those from southern regions. Moreover, my study goes further by showing regional differences when considering speakers' sex and the placement of *assim* in the clause.

These overall higher uses of clause-medial *assim* and the gender- and location-based distribution of clause-final *assim* lead me to hypothesize that the former is an early variant and the latter a newer one. This hypothesis is inspired by the development and use of DM *like* (D'Arcy 2107). According to D'Arcy (2017), conservative speakers would only use early forms of the feature, *like* in the matrix CP context, while innovative speakers produced a higher number of tokens of this variant and displayed an increasing use of *like* in subordinate CP and TP contexts. *Like* grammaticalized from sentence adverb to discourse marker and then to particle, and their respective positions in the clause (final for the sentence adverb, initial for the marker, and medial for the particle) indicate that *like* went from final to initial and then medial placements. In the case of *assim*, I argue that it went from initial and medial positions to final, from the matrix and subordinate CP and TP contexts to a sentence adverb.

The female-led pattern in Natal and the sex-balanced distribution in Rio de Janeiro found for clause-medial *assim* suggests a spread by diffusion from the prestigious *carioca* dialect spoken in Rio to the undervalued *nordestino* spoken in Natal, similar to the spread of *like* to rural Texas from "educated speakers of standard English" (Ferrara & Bell 1995, p. 285). Labov (2001) presents six stages as a model of the diffusion of a linguistic change from stability. At first, the variant is associated with a reference group, and this use becomes more aligned to a specific gender. After three generations, gender differences start to disappear, thus completing the change. If we consider clause-medial *assim* individually, we can hypothesize that the change in which speakers began to use the DM in this position started in Rio de Janeiro, where in the beginning, there would have been an association of this use with a specific gender. Then the use of *assim* spread to other areas, such as the city of Natal, beginning a pattern of linguistic change in this place too. By the time that *assim* spread to Natal, Rio de Janeiro would be in a later diffusion stage, with no sex differences in using this variant. Meanwhile, based on Labov's (2001) suggested process, Natal should be in an early stage where the sex differences are statistically significant. Another view that may explain the diffusion of *assim* from Rio de Janeiro to Natal is Trudgill's gravity model (1974). According to this model, linguistic innovation affects largert cities first and then spreads to smaller ones. In the present study, the linguistic change would spread from Rio de Janeiro to Natal.

The fact that female speakers in Rio de Janeiro use fewer tokens of clause-final *assim* than their male counterparts may also be explained using Labov's principles of linguistic change (1966, 2001). These principles can lead to introduction of two theories that interpret these findings. The first deals with *assim* being considered in a situation similar to a stable variation. Consequently, the DM has more frequent use of clause-medial assim by females because, as it is an older form of DM, it would lose its stigma, becoming more accepted in the community. The second relates to the idea that there is, in reality, an old change in progress, being clause-final *assim* resisted by females, indicating a change from above, suggesting that this variant is still stigmatized since it is considered a newer form. In cases of stable variation, females have been shown to use fewer stigmatized variants than males (Labov 2001, p. 266). This sex pattern also appears when a stigmatized change is noticeable within the speech community (Eckert, 1989). Due to this, females can also resist it, driven by global prestige norms (Eckert, 1989, p. 250).

DMs in Brazilian Portuguese are highly stigmatized features (Freitag, 2007), which explains why females are less likely to adopt the innovative use of *assim* (i.e. clause-final) in Rio de Janeiro.

Overall, the use of *assim* is more frequent in speakers aged 18-20, with its use slightly decreasing in older participants (21-31 year-old). Similarly, the 18-20 group favours the use of clause-final *assim* to the detriment of the clause-medial variant. However, the younger group (13-16 year-old) does not differ significantly from the oldest group (21-31 year-old). This pattern shows the shape of an inverted v-curve that illustrates age-grading (Downes, 1984, as cited in Tagliamonte, 2012) and is typical of nonstandard forms. This pattern is slightly different from the adolescent peak, where a feature would display the highest frequency of use in the 13-16-year-old group. My data suggests not a change in progress but rather age-grading, that is, use of a feature at a specific stage in life. This outcome supports the hypothesis of clause-final *assim* being an old change in progress, which has similar behaviour to a stable variable, following Eckert's (1989) rationale.

The interpretation that connects the claim of an old change in progress and an age-graded situation is that this feature involves a high degree of social awareness (Labov, 1994, as cited in Tagliamonte, 2012, p. 47). Moreover, being a stigmatized feature, the conformation to societal norms of standardness emerges, explaining the reduced use as speakers get older.

5.3. RQ3: Do the different topics of conversation influence the use of *assim*?

Results show that, in both cities, *assim* occurs more often in personal narratives and opinions. However, a sex pattern is statistically significant in Natal, where females show a higher rate of use of *assim* than males in retelling stories and giving opinions, and a lower frequency in providing instructions. In contrast, no differences between females and males were found in Rio de Janeiro.

The higher use of *assim* in personal narratives and opinions could be related to the subjective nature of these topics. In other words, talking about personal experiences or giving opinions might spark a discourse feature more easily than describing a place or providing instructions, the latter being objective matters by nature. A more subjective topic may require a closer connection and cooperation between interlocutors, and discourse markers, which have interactive functions, may assist in this need. Furthermore, the content shared in personal narratives and opinions are examples of the affective function of language, which females are

more concerned about (Holmes, 2013). These topics encourage epistemic modals and discourse markers (Coates, 1987).

This pattern based on the speaker's sex in Natal does not explain the balanced distribution of *assim* in personal narratives and opinions across sexes found in Rio de Janeiro. This lack of difference could be justified based on the prestigious status of the standard *carioca* dialect. The economic and political power of Rio de Janeiro that resulted in the standardization of their dialect, and the advanced stage that *assim* might have already undergone in the city, can also justify the sex-balanced distribution.

Chapter 6: Conclusion

This chapter summarizes the overall findings of my study and discusses its contributions to the field. It also discusses methodological limitations and suggests recommendations for future research.

This study aimed to analyze the linguistic and social variation of discourser marker *assim* in Brazilian Portuguese, drawing from 105 sociolinguistic interviews collected in Rio de Janeiro and Natal for the Discurso & Gramática corpus, constructed in 1993. Factors observed were the placements of *assim* in the clause, the category of its surrounding words, the speaker's sex, age and location, and the topic of the conversation. As *assim* is more frequently used in the clause-medial position, data showed a preference for male speakers to place the DM after nouns, and for female speakers after verbs. My analysis also found that *assim* resists appearing between a filled specifier and the heads of NPs or TPs. Speakers from Rio de Janeiro favour clause-final *assim* in a male-led pattern, with a peak in the 18-20 year-old group. On the other hand, when looking at the variants individually, clause-medial *assim* presents a female-led pattern in Natal, which also showed a likeliness to use *assim* in topics such as personal narratives and opinions.

Based on the more frequent application in conversations that deal with subjective topics, and considering the different categories of words that appear before *assim,* the assumption of different patterns according to gender reflects the dichotomy of language's referential and affective functions. Finally, these findings can also bring relevant contribution as a supplement to the already stated claim (e.g., Longhin-Thomazi, 2006; Martelotta et al., 1996) of the grammaticalization of *assim*. Even though previous studies demonstrated this development through the functions of *assim*, in this thesis, the placement of the clause brings evidence of a later form as clause-final, which agrees with the innovative use as an attenuator (Lopes-Damasio, 2008). A development of discursive functions through positions in the clause by grammaticalization was already stated for the DMs *like* (D'Arcy, 2017) and *anyway* (Coll, 2009). Hence, another example is present for DM *assim*. It is a crucial reminder that the development path of this DM is hypothesized as changing from clause-medial to clause-final, based on the assigned functions stated in previous studies, the distribution among location and sex, and the syntactic peculiarities of the original meaning of *assim*.

Some of the limitations of this study are a consequence of the corpus's nature: the lack of audio recordings and coding speakers for other social factors, such as class and ethnicity, being

the most salient ones. Even though intonation was a topic already studied by the literature (Silva & Macedo, 1996), research should also analyze speech sound to observe possible phonetic erosion. It is one of the mechanisms for grammaticalization (Heine et al., 1991), and a similar phenomenon was observed for other DMs in BP (Thompson & Onosson, 2021). Additionally, analyzing more social factors will elucidate missing peculiarities and contribute to a detailed result, following Andersen (2001). The shortcomings of the interviews conducted almost 30 years ago and using speakers only up to 31 years old are reasonable justifications for a follow-up study using more recent and expanded data. The synchronic nature of the corpus data only provides a snapshot of a linguistic situation impacted by the social characteristics in Brazil. Therefore, it is necessary to monitor how language deals with the fast-changing social conditions in modern times.

As for recommendations for future study, some unresolved matters and specific gaps were spotted during this thesis's elaboration, and they are a practical approach to expand this investigation. Based on the corpus limitations, it is worth noting that further analysis collecting current data is necessary, as well as the inclusion of age groups older than 31 years old. Ideally, the same speakers should be interviewed again, as well as new speakers in the original age ranges. Additionally, it is crucial to collect data from different locations outside of urban settings to check for similarities with studies like Ferrara & Bell (1995), who compared speech with rural participants. Another matter is to get samples from contexts where unbalanced social powers are manifest, such as courtroom settings (similarly to O'Barr & Atkins, 1980). The use of *assim* as general extender and quotative, excluded by this thesis's criteria for discourse markers, can be the object of separate studies, following what research has observed on sociolinguistic patterns in discourse-pragmatic features (Tagliamonte, 2012). Moreover, an analysis of variation among other BP DMs sharing the same clause-placement of *assim* (e.g., *então* 'then' in clause-initial, and *né* 'innit' in clause-final positions) should add material to this area, and gain substantial results that have been so far disregarded.

References

Alencastro, L. (1997). Vida privada e ordem privada no Império. In L. Alencastro (Ed.), *História da vida privada no Brasil, vol 2: Império - A Corte e a modernidade nacional* (pp. 12–72). Editora Schwarcz.

Andersen, G. (2001). Pragmatic markers and sociolinguistic variation: A relevance-theoretic approach to the language of adolescents. In *(Pragmatics & Beyond, New Series 84)*. John Benjamins Publishing.

Aragão, M. (2015). As vogais pretônicas nos falares nordestinos: dados do ALiB. *IX Congresso Internacional Da Abralin*, 242–256.

Araújo, J. (2017). *Mercado de trabalho e desigualdade: o Nordeste brasileiro nos anos 2000*. [Doctoral dissertation, Campinas State University].

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi: 10.18637/jss.v067.i01

Barbieri, F. (2007). Older men and younger women: A corpus-based study of quotative use in American English. *English World-Wide*, *28*(1), 23–45. https://doi.org/10.1075/eww.28.1.03bar

Bechara, E. (2009). *Moderna gramática portuguesa* (37th ed.). Nova Fronteira.

Benchimol, J. (1992). *Pereira Passos, um Haussmann tropical: a renovação urbana da cidade do Rio de Janeiro no início do século XX*. Secretaria Municipal de Cultura, Turismo e Esportes.

Brinton, L. (2006). Pathways in the development of pragmatic markers in English. In A. van Kemenade & B. Los (Eds.), *Handbook on the history of the English language*. (pp. 307–334). Blackwell.

Brody, J. (1989). Discourse markers in Tojolabal Mayan. *Proceedings of the Chicago Linguistics Society 25*(2), 15–29.

Broen, P., & Siegel, G. (1972). Variations in normal speech disfluencies. *Language and Speech, 5*. 219–31.

Campbell, L. (1991). Some grammaticalization changes in Estonian and theiriImplications. In E. Traugott, & B. Heine (Eds.). *Approaches to grammaticalization. Volume I: Theoretical and methodological issues* (pp. 285–299). John Benjamins.

Cano, W. (2017). Prefácio. In A. Neto, C. Castro, & C. Brandão (Eds.), *Desenvolvimento regional no Brasil: políticas, estratégias e perspectivas* (pp. 11–19). IPEA.

Carvalho, A. M. (2004). I speak like the guys on TV: Palatalization and the urbanization of Uruguayan Portuguese. *Language Variation and Change*, *16*(2), 127–151. https://doi.org/10.1017/S0954394504162030

Cascudo, L. (1999). *História da cidade do Natal*. IHGRN.

Castelano, K., & Luquetti, E. (2011). Gramaticalização e ensino: Uso do operador argumentativo "*assim*" na fala da região noroeste fluminense. *Anais Do 5º Encontro Nacional de Professores de Letras e Artes*. https://doi.org/10.6020/1679-9844/2006

Chambers, J. (1995). *Sociolinguistic theory: Linguistic variation and its social significance*. Blackwell.

Cheshire, J. (2005). Syntactic variation and beyond: Gender and social class variation in the use of discourse-new markers. *Journal of Sociolinguistics*, *9*(4), 479–508. https://doi.org/10.1111/j.1360-6441.2005.00303.x

Coates, J. (1987). Epistemic modality and spoken discourse. *Transactions of the Philological Society*, *85*(1), 110–131.

Coll, M. U. (2009). *'Anyway': A formal approach to the syntax and semantics of discourse markers* [Doctoral dissertation, University of Essex].

Cunha, M. (1998). *Corpus Discurso & Gramática: A língua falada e escrita na cidade de Natal*. EDURFN.

Cunha, M. (2000). *Procedimentos discursivos na fala de Natal: uma abordagem funcionalista*. EDURFN.

Cunha, M., & Tavares, M. (Eds.). (2007). *Funcionalismo e ensino de gramática*. EDUFRN.

Dailey-O'Cain, J. (2000). The sociolinguistic distribution of and attitudes toward focuser like and

quotative like. *Journal of Sociolinguistics 4*(1), 60–80.

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

D'Arcy, A. (2017). *Discourse-pragmatic variation in context: Eight hundred years of like (Studies in Language Companion Series, 187)*. John Benjamins Publishing.

Denis, D., & Tagliamonte, S. (2016). Innovation, right? Change, you know? Utterance-final tags in Canadian English. In H. Pichler (Ed.), *Discourse-pragmatic variation and change in English: New methods and insights* (pp. 86–112). Cambridge University Press.

Diskin, C. (2017). The use of the discourse-pragmatic marker 'like' by native and non-native speakers of English in Ireland. *Journal of Pragmatics*, *120*, 144–157. https://doi.org/10.1016/j.pragma.2017.08.004

Dubois, S. (1992). Extension particles, etc. *Language Variation and Change*, *4*, 179–203.

Duncan, R. (1972). Characteristics of organizational environments and perceived environmental uncertainty. *Administrative Science Quarterly, 17*(3). 313–327. https://doi.org/10.2307/2392145

Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, *1*, 245–267. https://doi.org/10.4324/9780429496288

Erman, B. (1992). Female and male usage of pragmatic expressions in same-sex and mixed-sex interaction. *Language Variation and Change, 4*, 217–234.

Ferrara, K., & Bell, B. (1995). The sociolinguistic variation and discourse function of constructed dialogue introducers : The case of be + like. *American Speech*, *70*(3), 265–290.

Ferrari, M. D. E. M. (2005). *A migração nordestina para São Paulo no segundo governo Vargas (1951-1954): Seca e desigualdades regionais*. [Master's Thesis, Federal University of São Carlos].

Ferreira, A. (1983). *Dicionário de Latim Português*. Editora Porto.

Fox Tree, J. E. (2010). Discourse markers across speakers and settings. *Linguistics and Language Compass*, *4*(5), 269–281. https://doi.org/10.1111/j.1749-818X.2010.00195.x

Fraser, B. (1999). What Are Discourse Markers? *Journal of Pragmatics, 31*, 931–52. https://doi.org/10.1016/S0378-2166(98)00101-5

Freitag, R. M. K. (2007). Marcadores discursivos interacionais: Análise contrastiva entre duas variedades do português falado no Brasil. *Anais Do CELSUL*.

Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics, 28*(3), 410–39.

Haddican, W., & Zweig, E. (2012). The syntax of manner quotative constructions in English and Dutch. *Linguistic Variation, 12*(1), 1–26. https://doi.org/10.1075/lv.12.1.01had

Hasund, I. (2003). *The discourse markers 'like' in English and 'liksom' in Norweigian teenage language: A corpus-based, cross-linguistic study*. [Doctoral dissertation, University of Bergen and Agder University College].

Heine, B., Claudi, U., & Hunnemeyer, F. (1991). *Grammaticalization: A conceptual framework*. The University of Chicago Press.

Holmes, J. (2013). *An introduction to sociolinguistics*. Routledge. https://doi.org/10.2307/329116

Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Linguistics and Language Compass, 3*(1), 359–383. https://doi.org/10.1111/j.1749-818X.2008.00108.x

Jørgensen, S. (2009). En plan used as a hedge in Spanish teenage language. In S. Jørgensen (Ed.). *Youngspeak in a multilingual perspective* (pp. 95–115). John Benjamins.

Jucker, A., & Smith, S. (1998). And people just you know like 'wow': discourse markers as negotiating strategies. In A. H. Jucker and Y. Ziv (Eds.), *Discourse markers: descriptions and theory* (pp. 171–201). John Benjamins.

Kotsinas, U. (1994). *Ungdomsspråk*. Hallgren & Fallgren.

Labov, W. (1966). *The social stratification of English in New York City*. Center for Applied Linguistics.

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society, 1*(1), 97–120.

Labov, W. (2001). *Principles of linguistic change: Volume 2: Social factors.* Blackwell.

Lalljee, M., & Cook, M. (1975). Anxiety and ritualized speech. *British Journal of Psychology*, 66(3), 299–306. https://doi.org/10.1111/j.2044-8295.1975.tb01465.x

Lauwereyns, S. (2002). Hedges in Japanese conversation: The influence of age, sex, and formality. *Language Variation and Change, 14*, 239–259.

Levey, S. (2006). The sociolinguistic distribution of discourse marker "like" in preadolescent speech. *Multilingua, 25*, 413–441.

Longhin-Thomazi, S. R. (2006). Gramaticalização, (inter)subjetivização e modalidade epistêmica: O caso de assim. *Estudos Linguísticos, 35*, 1772–1779.

Lopes-Damasio, L. R. (2008). *A emergência do marcador discursivo assim sob a óptica da gramaticalização: Um caso de multifuncionalidade e (inter)subjetivização.* [Master's thesis, São Paulo State Unversity]. https://repositorio.unesp.br/handle/11449/86593

Lopes, A. C. M. (2016). Discourse Markers. In W. L. Wetzels (Ed.), *The handbook of Portuguese linguistics* (pp. 441–456). John Wiley & Sons.

Maia, V. (1986). Vogais pretônicas médias na fala de Natal. *Estudos Linguísticos e Literários, 5*, 209–225.

Mariz, M., & Suassuna, L. (2002). *História do Rio Grande do Norte*. Sebo Vermelho.

Martelotta, M. E., Nascimento, E., & Costa, S. (1996). Gramaticalização e discursivização de assim. In M. E. Martelotta, S. J. Votre, & M. M. Cezario (Eds.), *Gramaticalização no português do Brasil: Uma abordagem funcional*. Tempo Brasileiro.

Maschler, Y. (2001). Veke'ilu haraglayim sh'xa nitka'ot bifnim kaze ('and like your feet get stuck inside like'): Hebrew kaze ('like'), ke'ilu ('like'), and the decline of Israeli dugri ('direct') speech. *Discourse Studies, 3*(3). 295–326.

Massini-Cagliari, G. (2004). Language policy in Brazil: Monolingualism and linguistic prejudice. *Language Policy, 3*(1), 3–23. https://doi.org/10.1023/B:LPOL.0000017723.72533.fd

Mello, E. (2010). *O Brasil holandês*. Companhia das Letras.

Meyerhoff, M. (2006). *Introducing sociolinguistics.* Routledge.

Meyerhoff, M. & Niedzielski, N. (1998). The syntax and semantics of ‘olsem’ in Bislama. In M. Pearson (Ed.), *Occasional Papers in Linguistics, 20* (pp. 235–243). UCLA

Miller, J., & Weinert, R. (1995). The function of LIKE in dialogue. *Journal of Pragmatics*, *23*(4), 365–393.

Nascentes, A. (1953). *O Linguajar Carioca.* Organização Simões.

Nascentes, A. (1955). Divisão dalectológica do território brasileiro. *Revista Brasileira de Geografia, 17*(2), 213–219.

O'Barr, W., & Atkins, B. (1980). "Woman's language" or "powerless language"? In S. McConnel-Ginet, R. Borker, & N. Furman (Eds.), *Women and Language in Literature and Society* (pp. 93–110). Praeger.

Ostman, J., (1981). *You know: a discourse functional approach*. John Benjamins.

Pereira, M. (2007). *Atlas Geolinguístico do Litoral Potiguar ALiPTG*. Universidade Federal do Rio de Janeiro.

Pichler, H. (2013). *The Structure of Discourse-Pragmatic Variation*. John Benjamins

Redeker, G. (1990). Ideational and pragmatic markers of discourse structure. *Journal of Pragmatics, 14*. 367–81.

Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics, 29*, 1139–1172.

Rodrigues, J. (1986). *História viva.* Global Universitária.

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*(3), 227. https://doi.org/10.2307/455799

Rosa, M. (1992). *Marcadores de Atenuação*. Contexto.

Ross, J., & Cooper, W. (1979). Like, syntax. In W. Cooper & E. Walker (Eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett* (pp. 343–418). Lawrence Erlbaum Associates.

Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M., & Gagnon, L. (1997). Variation in the use of discourse markers in a language contact situation. *Language Variation and Change, 9*, 191–217.

Schourup, L. (1985). Common Discourse Particles in English Conversation. Garland.

Schweinberger, M. (2013). A sociolinguistics analysis of discourse marker LIKE in Northern Ireland. A look behind the scenes of quantitative reasoning. In M. Bieswanger & A. Koll-Stobbe (Eds.), *New Approaches to the Analysis of Linguistic Variability* (pp. 13–39). Peter Lang.

Silva, G. (1999). Anatomia e fisiologia dos marcadores discursivos não-prototípicos. In M. H. Neves (Ed.), *Gramática do Português Falado (Vol VII: Novos Estudos)* (pp. 297–347). Unicamp.

Silva, G., & Macedo, A. (1992). Discourse markers in the spoken Portuguese of Rio de Janeiro. *Language Variation and Change, 4*(02), 235–249. https://doi.org/10.1017/s0954394500000776

Silva, G., & Macedo, A. (1996). Analise sociolinguística de alguns marcadores conversacionais. In A. T. Macedo, C. Roncarati, & M. C. Mollica (Eds.), *Variação e discurso* (pp. 11–49). Tempo Brasileiro.

Souza, E. R. (2009). *Gramaticalização dos itens linguísticos assim, já e aí no português brasileiro: Um estudo sob a perspectiva da Gramática Discursivo-Funcional.* [Doctoral dissertation, University of Campinas]. http://repositorio.unicamp.br/jspui/handle/REPOSIP/269076

Souza, E. R. (2010). Os usos de assim no português falado no noroeste paulista sob a perspectiva da Gramática Discursivo-Funcional. *Estudos Linguísticos, 39*(1). 73–88.

Stenström, A., & Jørgensen, A. (Eds.). (2009). *Youngspeak in a Multilingual Perspective*. John Benjamins.

Strapazzon, L. T. (2018). *Usos de assim na fala e na escrita de chapecoenses* [Master's thesis, Federal University of Fronteira Sul]. https://rd.uffs.edu.br/handle/prefix/2132

Tagliamonte, S. (2005). So who? Like how? Just what? Discourse markers in the conversations of Young Canadians. *Journal of Pragmatics, 37*(11 SPEC. ISS.), 1896–1915. https://doi.org/10.1016/j.pragma.2005.02.017

Tagliamonte, S. (2012). *Variationist sociolinguistics: change, observation, interpretation.* John Wiley & Sons.

Tagliamonte, S., D'Arcy, A., & Rodriguez Louro, C. (2016). Outliers, impact, and rationalization in linguistic change. *Language, 92*(04), 824–849.

Thompson, C., & Onosson, S. (2021). *Phonetic change in the grammaticalization of Brazilian Portuguese tipo*. 2021 CLA Conference. https://cla-acl.ca/programmes/congres-de-2021-meeting.html

Traugott, E. (1995). *The role of the development of discourse markers in a theory of grammaticalization*. ICHL XII. Manchester.

Trudgill, P. (1974). Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography. *Language in Society, 3*(2), 215–246.

Underhill, R. (1988). Like is, like, focus. *American Speech, 63*(3), 234–246.

Umbach, C., & Ebert, C. (2009). German demonstrative so – intensifying and hedging effects. *Sprache und Datenverabeitung, 33*(1). 153–168.

Urbano, H. (1999). Aspectos basicamente interacionais dos marcadores discursivos. In M. Neves (Ed.), *Gramática do Português Falado Vol. 7* (pp. 195–258). Editora da Unicamp.

Votre, S., & Oliveira, M. (n.d.). *A língua falada e escrita na cidade do Rio de Janeiro: Materiais para seu estudo.* Universidade Federal do Rio de Janeiro.