



Is the Brazilian Portuguese phrase *estar indo*, i.e. *be going [to]*, being grammaticalized into an auxiliary, and if that is the case, is it a borrowing from English?

A comparative study of the grammaticalization of Early Modern English *be going to* and the Present-Day Brazilian Portuguese *estar indo*.

KRISTINE LAUVLAND CORREA

SUPERVISOR

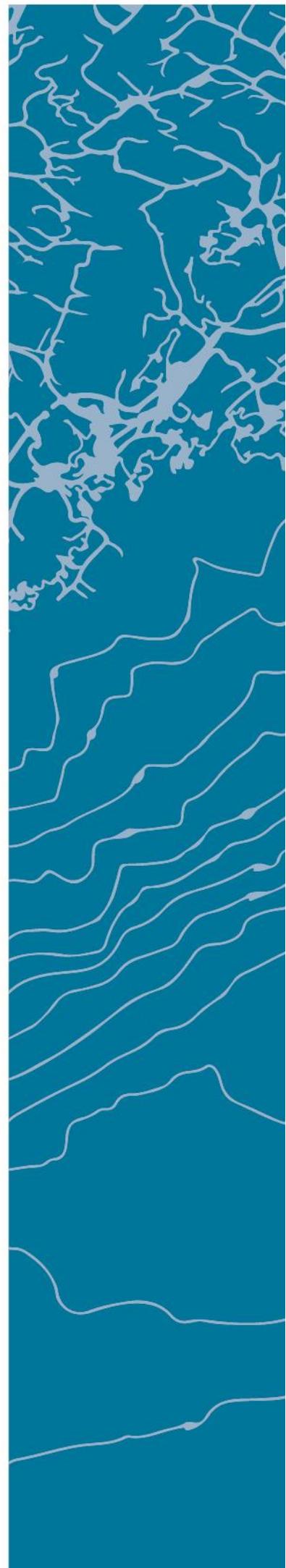
Professor Allison Louise Wetterlin

Master's thesis

University of Agder, 2018

Faculty of Humanities and Education

Department of Foreign Languages and Translation



Abstract

The aim of this thesis is to compare the grammaticalization of English *be going to* and Brazilian Portuguese *estar indo* (i.e. the equivalent to English *be going to*) in order to investigate whether the phrase is grammaticalizing in Brazilian Portuguese, and furthermore, to discuss whether the grammaticalization is influenced by English. To discuss these issues, the thesis consults research and corpora that illustrate the use and grammaticalization of the phrase in the two languages.

Data from the BYU corpora and *Twitter*¹ show that Brazilian Portuguese *estar indo* is used in situations where it appears to be a (purposive) future auxiliary, but it is quite rare and it is not accepted in Standard Brazilian Portuguese. Similar to the use of *be going to* in Early Modern English, the auxiliary use of *estar indo* is most frequent with action and/or activity verb collocates (e.g. to work, to buy), but the present thesis has found evidence for *estar indo* occurring with mental verbs (e.g. to think).

The study also explores whether the grammaticalization of *estar indo* is a consequence of the global spread of English and its subsequent influence on Brazilian Portuguese, concluding that contact between English and Brazilian Portuguese could have reinforced the use of *estar indo*, but it is unlikely that *estar indo* is a direct borrowing from English or that the grammaticalization is triggered by extreme language contact with English.

¹ An online, interactive social media network.

Acknowledgment

I want to thank Allison Louise Wetterlin for the conversations, the excellent feedback and the guidance, but primarily I want to thank you for helping me through this.

To my daughter Akamai, thank you for being the light in my life and for inspiring me with your ceaseless enthusiasm and endeavour to communicate in all languages you are exposed to. You are my everything and I love you!

To my husband, thank you for inspiring this thesis with your vernacular, and thank you for cleaning and for pretending to understand why I have been moody these past months.

To my mum, my family and friends, I could not have done any of this without you. I am forever grateful.

To the University of Agder, thank you for employing amazing professors and staff. Thank you to all the wonderful lecturers that let me bring my infant daughter to classes. Thank you for all the competence, the opportunities and experiences you have given me in Norway and Australia.

List of abbreviations:

AE	American English
AUX	Auxiliary
BE	British English
BP	Brazilian Portuguese
BYU	Corpora from Brigham Young University
EP	European Portuguese
FEM	Feminine
FL	Foreign Language
FPP	Future Present Progressive
HC	Helsinki Corpus
ICT	Information and Communication Technology
MASC	Masculine
NOW	News on the Web
OBC	The Old Bailey Corpus
PBE	Proceedings of the Old Bailey
PBP	Present-Day Brazilian Portuguese
PL	Plural
SG	Singular

Stages of English (adapted from Hopper and Traugott, 2003, p. xx):

OE	Old English (c. 600-1125)
ME	Middle English (c. 1125-1500)
EMdE	Early Modern English (c. 1500-1750)
ME	Modern English (c. 1750-1950)
PDE	Present-Day English (c. 1950 -)

Table of contents

.....	1
Chapter 1: Introduction and Preliminaries	1
Introduction.....	1
Preliminaries.....	2
1.1. Terms and the differences between the variety of progressives, future auxiliaries and futurates.....	2
1.2. Data and Methodology.....	5
1.3. Aims and Organisation	6
Chapter 2: Grammaticalization: Relevant literature and research, key terminology and grammaticalization theory	7
2.1.1. What is grammaticalization, and what characterises a grammatical item?	7
2.1.2. Some essential mechanisms in grammaticalization.....	9
2.2. Clines	13
2.3. Reanalysis	14
2.4. Frequency.....	15
Chapter 3: Some preliminaries on <i>be going to</i> as a progressive, the progressive aspect and BP progressives.....	17
3.1. A couple of notes on language contact (between BP and English) and borrowing grammar ..	17
3.2. Differences between BP and European Portuguese progressives and different progressive aspects.....	18
3.3. Different terms referring to <i>be going to V</i> or <i>estar indo V</i>	20
Chapter 4: Data and Methodology.....	21
4.1. A comparison of EMdE data in Traugott (2012) and BP data.....	21
4.2. The BYU corpora.....	23
4.3. The Twitter corpus	24
4.4. Searching criteria and selection of data.....	27
Chapter 5: Data Analysis	30
5.1. Comparison of BP <i>estar indo</i> and PDE <i>be going to</i> in the BYU corpora	30
5.2. Brief overview of the frequency of <i>estar indo</i> compared to the periphrastic future in the BYU corpora	32
5.3. BP collocates in the Twitter corpus.....	35
5.4. Temporal adverbials in the Twitter corpus	38
5.5. Key findings	40
Chapter 6: Discussion	42
6.1. Grammaticalization of English <i>be going to</i> : Why “be going to” and not “go”?	42
6.1.1. Grammaticalization of V-ende/ V-ing periphrasis.....	44

6.1.2. Language contact between OE, Celtic and Latin	45
6.1.3. Different perspectives on the grammaticalization of <i>be going to</i>	47
6.2. Grammaticalization of BP <i>estar indo</i>	50
6.2.1. Processes of change in BP and the phonetic erosion of <i>ESTAR</i>	52
6.2.2 Future verb phrases in Brazilian Portuguese and grammaticalization of IR.	55
6.3. English in Brazil	59
6.4. Innovation versus language contact in the grammaticalization of English <i>be going to</i> and BP <i>estar indo</i>	62
6.5. A brief summary of some differences between BP <i>estar indo</i> and English <i>be going to</i>	63
Chapter 7: Conclusions	65
7.1. Implications for future research	67
Bibliography	68
Corpora	72
Dictionaries	72
Appendix	73

List of Figures

Figure 1: Twitter search for data	26
Figure 2: Example of Twitter search result	26
Figure 3: Varieties of <i>estar</i> in Rocha's corpus (Rocha, 2012, p. 17).	28
Figure 4: Hopper and Traugott's revised schema of the development of AUX "be going to" (2003, p.93).	48

List of tables

Table 1: Illustration of decategorialization from major category: <i>Noun</i> , to minor category: <i>Pronoun</i> .	11
Table 2: Illustration of decategorialization from major category: <i>Verb</i> , to minor category: <i>Auxiliary</i>	12
Table 3: Explanation of Rebracketing in Grammaticalization	15
Table 4: Overview of collocates in Traugott's research (2012)	22
Table 5: Overview of the data found in BYUP 2015.	23
Table 6: Traditional and phonetically eroded <i>estar</i>	29
Table 7: Examples of PDE "be going to" and PBP "estar indo" as lexical items and grammatical items in the BYU corpora.	31
Table 8: Data from BYU on the frequency of <i>Estou indo</i> + infinitive.	33
Table 9: Data from BYU on the frequency of <i>IR</i> + infinitive, i.e. the most frequent future verb phrase	34
Table 10: BP collocates in the Twitter corpus	35
Table 11: Temporal adverbials	36
Table 12: bracketing A of (29)	37
Table 13: bracketing B of (29)	37
Table 14: Phonetic erosion of <i>ESTAR</i> : Data from BYU showing the frequency of <i>tô</i> , <i>tá</i> and <i>tamo</i>	53

Chapter 1: Introduction and Preliminaries

Introduction

Grammaticalization and comparative linguistics are the two corner stones of the present thesis, thus it has one foot placed in theoretical and cognitive linguistics and the other foot placed in descriptive linguistics. In order to discuss whether Brazilian Portuguese (BP) *estar indo* is being grammaticalized, the study investigates the mechanisms that took place in the early stages of the grammaticalization of Early Modern English (henceforth: EMdE) *be going to* and compare the findings to BP *estar indo*. It is important to keep in mind that *grammaticalization* is a narrow field and it did not become a topic in its own right until the 1980s (cf. Hopper and Traugott, 2003, p.30). In addition, the terminology is complex and it is continuously questioned and developed. Therefore, chapter 2 accounts for some of the basic terminology and mechanisms within grammaticalization research.

Grammaticalization of future verb phrases and *-ing* progressives in English can be discussed and compared with several languages. Orozco (2014), for example, shows that newer grammatical constructions can be more frequent in both oral and written language, but still they are denounced by prescriptivists and deprioritised in language learning textbooks. By referring to corpora studies both Rocha (2012) and Orozco (2014) show that the *periphrastic future* in BP and Spanish (see explanation in section 1.1.), is currently used more often than the traditional *morphological future*, but still the *morphological future* is prioritised in language learning textbooks (Orozco, 2014).

This is just one of many studies on the grammaticalization of future verb phrases and *-ndo/-ing* progressives (see e.g. Caucullos, 2000, Orozco, 2014, amongst others). Hopefully, this study will shed light on some of the complexity of grammaticalization, which may have implications for e.g. foreign language teaching. The first step, however, is to explain how this study employs grammaticalization theory as a research framework.

To set the stage for addressing the grammaticalization of *be going to* and *estar indo*, section 1.1. explains the many terms referring to progressives, future auxiliaries and *futurates*, and how such terms are employed in the study. The remaining sections of chapter 1 will introduce the data and the methodology used in this thesis and give a detailed description of its aims and organization.

Preliminaries

1.1. Terms and the differences between the variety of progressives, future auxiliaries and futurates.

To avoid confusing the many different future and progressive verb phrases, this section briefly summarizes how some of them differ from *be going to* and *estar indo*. The *future present progressive* (henceforth: FPP), for instance, must not be confused with *estar indo V*. How the FPP differ from *be going to/estar indo* can be observed in the following examples:

(1)

Amanha meu gordinho tá comprando os livros para mim.

‘Tomorrow my hubby’s buying the books for me.’

(2)

Amanha meu gordinho tá indo comprar os livros para mim.

‘Tomorrow my hubby’s going to buy the books for me.’

In the examples above the temporal adverb “tomorrow” places the action “to buy books” in the future. Without temporal adverbs, the verb phrase in (1) would be interpreted as a focalized progressive in both BP and English, i.e. that the husband is buying books at the moment of the utterance. In (2), however, the BP verb phrase is ambiguous without the temporal adverb, in the sense that the husband could be walking to the book store at the time of the utterance, but he could also be going to buy them in the (near) future. Since *be going to* has undergone full grammaticalization in Present-Day English (PDE), the verb phrase “my hubby’s going to buy the books” is most likely to be interpreted as a future verb phrase. Still, some of its original ambiguity is traceable, because the sentence without temporal adverb can be interpreted as “he is going to the book store, in order to buy the books, as we speak”.

Rocha (2012) and Nesselhauf (2007) explain that the FPP is a futurate in the sense that it is the future use of the present progressive which is not traditionally a future marker. As seen in the examples above, the FPP is productive in both BP and PDE. Nonetheless, In BP, the FPP and *estar indo* are more or less equally ambiguous, whereas *be going to* in English has grammaticalized fully as a future auxiliary and is thus less ambiguous than the English FPP. How the FPP and other futurates and future verb phrases in BP may influence the use and grammaticalization of BP *estar indo* is discussed in chapter 6.

Example (2) also illustrates that it is necessary to distinguish between the present progressive *be going* and the future auxiliary *be going to*, thus henceforth the future auxiliary *be going to* + main verb will be written *be going to V* (PDE), and *estar indo V* (BP). The term *progressives* will be used for all verb phrases with the progressive form of the verb, i.e. the English *-ing* participle, or the BP *-ndo* participle, with or without one or more auxiliaries, as exemplified below:

(3) a. ‘Seeing the bird’

b. *Vendo o passaro*

(4) a. ‘I’m going’

b. *Estou indo*

(5) a. ‘He’s been working’

b. *Estava trabalhando*

In Portuguese the *-ndo* participle is usually referred to as *gerúndio*, directly translated into *gerund* in English, but to avoid confusion, *gerund* will only be used when referring to the nominal use of the *-ing* progressive, and for reasons of space it will not be discussed here.

Another issue that influences the use of BP *estar indo*, is that there are at least ten different future verb tenses and futurates in BP and *estar indo* would only be one of them. This of course was not the situation in EMdE, in which future tense was expressed through the use of modal auxiliaries and temporal adverbs. The varieties of future verb phrases and futurates in BP range from **morphological futures** to **periphrastic futures**. Morphological futures express future tense, and pronominal subjects in Spanish and standard Portuguese, through the morphology of the verb and auxiliaries are not employed, e.g. standard BP *pagará* is translated to PDE ‘he/she will pay’. Periphrastic future verb tenses form phrases with the auxiliaries IR (i.e. GO, and this is the most frequently used future verb phrase in both BP and Spanish), *haver que*.i.e. ‘be/have to’, *ter que* i.e. ‘have to’, *querer* i.e. ‘will/want to’ (cf. Rocha, 2012, among others).

Finally, it is necessary to highlight the differences between *estar indo V* and the BP future verb progressive referred to as *gerundismo*, and these differences are exemplified in (6).

:

(6)

Amanha meu gordinho² vai estar comprando os livros.

[Tomorrow my hubby goes (i.e. will) be buying the books]

‘Tomorrow my hubby will be buying the books.’

Gerundismo is akin to the future continuous tense in English, i.e. will + INF be + *-ing* progressive, except that the English modal auxiliary “will,” is replaced by the future auxiliary IR, i.e. GO in BP. Notice that both the future continuous and gerundismo encode the progressive aspect *and* places the action in the future simultaneously. In BP, it is made up by both the high-frequency periphrastic future marker, i.e. present tense IR/GO, and the periphrastic *-ndo* progressive and the auxiliary *estar/be* is not conjugated and remains in the infinitive.

This thesis will argue that the verb phrase *be going to/estar indo* is part of an ongoing trend in both English and BP, where the use of different periphrastic progressives expands the register of future verb phrases. These views are also reflected in literature by Nesselhauf, who writes that the “[t]he potential role of the ‘future’ use of the [the present progressive] construction (as in I’m flying home next week) has, (...), been neglected so far,” (2007, p.191). Rocha (2012) also comments that the increasing use of progressive constructions in BP has expanded the register of future verb phrases.

The discussion (in chapter 6) will emphasise the connection between the relatively high frequency of BP progressives and *estar indo* V, and that the high frequency of English progressives too have influenced the grammaticalization of the EMdE *be going to* V. The connection between *-ing* progressives and EMdE *be going to* is elaborated upon in section 6.1 and how *estar indo* fits in with the expanding number of modal auxiliaries and future verb phrases in BP is discussed in section 6.2., but before discussing these issues any further it is necessary to compare the development of EMdE *be going to* and BP *estar indo* more thoroughly. The following section elaborates upon the methodology and data used to compare the constructions.

² *Gordinho* is the masculine diminutive of ‘fat’, i.e. ‘little fatty/fat boy,’ but is frequently used as an informal and colloquial moniker for one’s partner.

1.2. Data and Methodology

To investigate whether BP *estar indo* is being grammaticalized, the thesis compares data on EMdE *be going to* and data on BP *estar indo*. The data is selected and analysed through both synchronic and diachronic corpora studies. The diachronic studies consist of data analysis and research on the grammaticalization of English *be going to* by primarily Traugott (2012), but findings by Jakobi (2004), and others will also be mentioned. Hence, I will refer to their research and selection of data and corpora when accounting for methodology and data on *be going to*. Traugott (2012) has selected data from the Helsinki corpus (1500-1720), and the first fifty years, i.e. 1674-1723 of Proceedings of the Old Bailey (POB), which is part of the larger Old Bailey Corpus (OBC) (Traugott, 2012, pp. 231-232), and then these findings are compared to data on the BP *estar indo*.

The data on BP *estar indo* are selected from different corpora. Some is found in the BYU corpora³, but the Portuguese BYU corpora do not gather data from sources after 2015, nor social media platforms, so in addition data from the social media network *Twitter* has been gathered. Since *estar indo* V still is in the very early stages of grammaticalization, the data from *Twitter* has made it possible to account for the productivity of *estar indo* V in contemporary colloquial BP (i.e. 2017/2018).

The mini corpus of *estar indo* + main verb collocates that was collected for this thesis is named “the Twitter corpus of the BP *estar indo* auxiliary” (henceforth: The Twitter corpus). The methodology used for gathering data from Twitter was developed by Rocha in 2012, when she researched the previously mentioned (see example 1 and 2) Brazilian Portuguese Future Present Progressive (BP FPP) for her master’s thesis.

As will be explained in section 3 and 4, there are many similarities between the BP data and the EMdE data analysed by Traugott. The collocates are mostly action verbs and some indication of physical movement is often conveyed through the context. *Be going to/estar indo* is sometimes ambiguous and could be read as both a future auxiliary and the use of the present progressive of GO/IR usually indicates that some form of physical movement is involved.

³ See: BYU, o corpus do português, web/dialectm 1 billion words, created 2015-2016. Retrieved from <https://www.corpusdoportugues.org/web-dial/> (31.10.2017).

1.3. Aims and Organisation

The aim of the thesis is to examine whether BP *estar indo* is in the process of grammaticalizing into a future auxiliary by comparing research on the grammaticalization of future *be going to* in EMdE and Present-Day BP *estar indo*. In addition, the thesis aims to investigate whether language contact between PDE and BP is the root cause of the grammaticalization of BP *estar indo* and how future auxiliary *be going to/estar indo* is interrelated with the grammaticalization of *-ndo/ -ing* progressives.

The study is divided into seven chapters. The aim of chapter 2 is to review some of the literature on grammaticalization and to explain key terminology that is especially relevant for the thesis. Examples from grammaticalization processes in English and BP will be used to illustrate some of the terminology, and the selected literature consists of Hopper and Traugott (2003), Traugott (2012), Killie (2007, 2014a, 2014b), Diewald (2010, 2011), Jakobi (2004), Rocha (2012), Nesselhauf (2007, 2010), Zilles (2007), Orozco (2007, 2014), and others. Chapter 3 briefly summarizes some views on language contact, and in addition the chapter accounts for terms referring to *be going to*, progressives and differences between these verb phrases in European Portuguese (EP) and BP.

Data from different corpora, which document the usage and frequency of *be going to/estar indo* in PDE, EMdE and BP are analysed in chapters 4 and 5. The synchronic and diachronic development of English *be going to* and *-ing* progressives, are compared to the development of BP *estar indo* and its interrelation with *-ndo* progressives, and the findings are discussed in chapter 6. The discussion also aims to consider sociolinguistic aspects and features of vernacular BP, which influence the use of *estar indo*, and makes it differ from the development and use of PDE *be going to*.

Chapter 2: Grammaticalization: Relevant literature and research, key terminology and grammaticalization theory

This chapter gives a preliminary outline of grammaticalization theory and an overview of the terminology that is necessary for understanding the grammaticalization of *be going to*, and for discussing whether BP *estar indo* is grammaticalizing. The following section begins by explaining how grammaticalization is defined. Section 2.1. accounts for some key terms and mechanisms, and problem areas within grammaticalization. The remaining three sections explain the terms *grammaticalization clines*, *reanalysis* and *frequency*.

2.1.1. What is grammaticalization, and what characterises a grammatical item?

The fundamental question that needs to be addressed is how *grammaticalization* is defined. A simple explanation of the term, is that grammaticalization is a research framework for studying the processes in which lexical items become grammatical items and furthermore how grammatical items develop new grammatical functions (Hopper and Traugott 2003). The processes in which lexical items develop grammatical functions is often referred to as *primary grammaticalization*, while the processes in which grammatical items develop new grammatical functions is referred to as *secondary grammaticalization* (see: Killie 2014b). Some of the mechanisms that take place in these processes are: *reanalysis*, *semantic bleaching* (also called *desemantization*), *generalization*, *extension*, *morphological reduction*, *phonetic erosion*, *increased frequency*, and more.

Furthermore, there are several problem areas within grammaticalization research that have implications for this study. Hopper and Traugott's (2003) textbook on grammaticalization is thorough and comprehensive, but still, it has been very difficult to find concise definitions of basic fields and terms, such as e.g. *grammar* and *semantics*. Killie (2014a) has pointed out that the lack of a concise definition of the term *grammatical item* makes it difficult to account for the basic mechanisms within grammaticalization. Furthermore, perhaps the most extensively debated question is the *hypothesis of unidirectionality*, i.e. the notion that grammaticalized items move along a cline and go through *stages* in a unidirectional manner (see Trudgill 2016a, among others), and this elaborated upon to some extent, but not discussed in great detail.

As mentioned in the previous paragraph, the definition of ‘grammatical’ and ‘grammatical form/item’ is somewhat diffuse and unclear (cf. Diewald 2010, Hopper and Traugott 2003, and Killie 2014a). The Concise Oxford Dictionary of Linguistics defines grammatical in terms of the juxtaposition between grammatical and lexical (Matthews 2007), and to a certain extent so do Hopper and Traugott (2003), although they use the term *content* item for *lexical* item, and *function* item for *grammatical* item (Hopper and Traugott 2003, p. 4). Thus, ‘grammatical’ is defined as the binary opposition to ‘lexical,’ but there is no description of how grammatical items are characterised.

Since the aim is to investigate whether BP *estar indo* has been attributed the features of a grammatical item instead of those of a lexical item, it is essential that we know what the characteristics of a grammatical item are. Diewald (2011) argues that traditionally there are three important features that identify grammatical items;

- 1) obligatoriness (see explanation and example below)
- 2) paradigmization, i.e. the development of a paradigm in which the grammaticalized item is a component
- 3) relational meaning (see final paragraph).

This thesis defines *obligatoriness* and *obligatorification* as the obligatory use of one form over another. The present progressive in English, for example, is used obligatorily when expressing that something is occurring at the moment of the utterance, as exemplified in the sentence below.

(7) What are you doing right now?

- I’m studying.

Notice that in PDE, the simple present would be perceived as ungrammatical in the same sentences (i.e. “What do you do right now?⁴ – I study.”), thus the present progressive is obligatory. In Spanish, however, the present progressive is not obligatory since both the present progressive and the simple present would be possible (7) (see Cacoullos, 2000). There is, however, nothing obligatory about the auxiliary use of BP *estar indo*, which could suggest that it is pointless to investigate a possible grammaticalization of *estar indo* any further, but the

⁴ “What are you doing right now” could also be a pragmatically understood as “are you busy/free right now”.

notion of obligatoriness as a key feature of grammatical items has been criticised (see Hopper and Traugott, 2003, p.32).

Lehman (2015) explains that different grammatical items display varying degrees of obligatoriness and paradigmaticization. He stresses that obligatoriness is not necessarily a criterion for grammatical items, nor an absolute criterion for grammaticalization (cf. Diewald, 2011, Hopper and Traugott, 2003). Thus, *relational meaning* stands out as the key feature accepted and agreed upon by several linguists. Relational meaning is for e.g. when pronouns make deictic or anaphoric references to nouns, and by referring to Diewald (2011, 2010), Killie (2014b) also emphasises *relational meaning* as a key feature of grammatical items. Contrary to obligatoriness, there is reason to investigate whether *estar indo* encodes relational meaning, and this is investigated by referring to data in chapter 4.

2.1.2. Some essential mechanisms in grammaticalization.

The present section focuses on mechanisms that have taken place in the grammaticalization of English *be going to*, such as *semantic bleaching*, *generalization*, *decategorialization* and *loss in morphosyntactic properties*. Throughout the chapter some mechanisms are given more attention than others because they are more relevant to the thesis.

The first interrelated mechanisms or changes that need to be accounted for are the ones that constitute grammaticalization, and according to Heine there are four such mechanisms (2002, p. 2). Namely, *desemanticization*, *extension or context generalization*, *loss in morphosyntactic properties* and *phonetic erosion*. Luckily, English *be going to* can be used to illustrate most of these mechanisms.

The first mechanism/changes accounted for by Heine, is desemanticization, also called ‘semantic bleaching’ or ‘semantic reduction’ i.e. loss in meaning content. The lexeme GO, for instance, initially contained a specific meaning (i.e. to walk), but through diachronic and synchronic processes the meaning content became ‘bleached,’ and currently ‘go’ could be interpreted as *to leave*, *to move*, *to walk/stroll/run*, *to go by buss/car/etc.*, *future auxiliary*, etc. (see Hopper and Traugott 2003, Jakobi 2004).

Generalization of meaning differs from semantic bleaching. Generalization refers to the notion that lexical items, which are prone to grammaticalize, already are very general or “basic” (Hopper and Traugott, 2003, pp. 101-103). It is more probable that items that grammaticalize are general and undergo semantic bleaching, than narrowing of meaning (ibid, p. 103). *Go* is a

typical example of a lexical item that is general or superordinate (ibid, p. 101), as seen when *go* is compared to specialized (or subordinate) terms such as *strolling*, *walking*, *peregrinating*, *trekking*, and many others. In other words, *stroll* is much less likely to grammaticalize than *go*.

Heine's second interrelated change or mechanism is *extension*, or *context generalization* (cf. Heine, 2003, Zilles, 2007, Trudgill, 2016). For instance, *be going to* is no longer restricted to contexts that indicate physical movement or intention. 'He is going to fetch water' indicate physical movement with the intention of fetching water, but 'He is going to like it' indicate neither physical movement nor intention. BP *estar indo*, on the other hand, occurs currently in situations where it is ambiguous and it is mostly restricted to contexts indicating physical movement and/or intention. Heine's third stage is *loss in morphosyntactic properties* includes the loss of independent word status, i.e. cliticization or affixation, and the fourth and last stage is *phonetic erosion*, or *phonetic reduction* (see. Heine, 2003, Zilles, 2007, Trudgill, 2016). Both phonetic erosion and loss in morphosyntactic properties will be exemplified later on, but first the thesis will consider whether these changes occur in a unidirectional manner.

The hypothesis of unidirectionality, is the view that from a diachronic perspective, processes and mechanisms involved in grammaticalization take place in a unidirectional manner (see Hopper and Traugott, 2003, p. 99), i.e. phonetic erosion e.g. would not occur before semantic bleaching according to this hypothesis. The corpus studies elaborated upon in chapters 4 and 6 will show that phonetic erosion of *estar* occurred *before* the grammaticalization of *estar indo*, and this could imply that *estar indo* is *not* grammaticalizing in a unidirectional manner. There are, however, counterexamples of unidirectionality and *estar indo* could be another instance that challenges the hypothesis. Another study that challenges the hypothesis of unidirectionality is Trudgill's research on grammaticalization and phonetic erosion in vernacular English and this is elaborated upon in the discussion.

The processes that are emphasized in the hypothesis of unidirectionality are *generalization and decategorialization* (Hopper and Traugott 2003). While generalization already has been accounted for in the previous paragraphs, decategorialization is explained by referring to the following figures. Decategorialization is the notion that grammaticalized items move along clines (this is elaborated upon in the next section), and typically from a major category to a minor category, for instance, from nouns to pronouns (see example a), or from verbs to auxiliaries and then from auxiliary to clitic/affix (see example b).

The phrases in the tables underneath are taken from the lyrics of the song *É* by the Brazilian songwriter Gonzaguinha. These examples also illustrate the primary grammaticalization of the noun “a gente” (literally: ‘the people’) to the 1st PL pronoun equivalent to English ‘we’ (cf. Zilles, 2007). This pronoun is mentioned to illustrate how BP verb phrases are becoming increasingly more periphrastic. This study also argues that the use of *estar indo* is triggered by the switch towards increased periphrasis in BP and the rise of periphrastic progressives and more variety in future expressions.

(8). Examples of decategorialization from major to minor category

a. Major category: Noun

Table 1: Illustration of decategorialization from major category, noun, to minor category, pronoun.

Major category	Intermediate category	Minor category
Noun	Ambiguous and inclusive. Either noun or pronoun depending on the listeners interpretation.	Pronoun
<i>A gente</i> não tem jeito de babaca [The people don’t have the manner of (a) jerk] ‘The people aren’t stupid’	<i>A gente</i> não tem jeito de babaca [We, the people, don’t have the manner of (a) jerk] ‘We/the people aren’t stupid’	<i>A gente</i> não tem jeito de babaca [We don’t have the manner of (a) jerk] ‘We aren’t stupid’
<i>A gente</i> meaning ‘the people’	<i>A gente</i> , meaning ‘the people, including the speaker,’ as in: “We/us, the people”)	<i>A gente</i> meaning ‘We’

b. Major category: Verb

Table 2: Illustration of decategorialization from major category: Verb to minor category: Auxiliary

Major category <i>Content item</i>	Intermediate category	Minor category <i>Function item</i>
Verb	Ambiguous	Auxiliary
A gente <i>vai</i> para Noruega [The people go to Norway] 'We are going to Norway'	A gente <i>vai</i> viajar para Noruega. [The people go to travel to Norway] 'We are going to travel to Norway'	A gente <i>vai</i> gostar da viagem. [The people go to like the trip] 'We are going to like the trip'
<i>Vai</i> meaning present tense 3SING 'go'.	<i>Vai</i> meaning either purposive present tense 3SING 'go'.	<i>A gente</i> meaning 'We'

The phrases in the examples above illustrate the shift in BP, from expressing tense, aspect, mood and pronouns through affixes these features are now expressed through independent morphemes that have been grammaticalized from lexical items, and even as grammatical items they continue to grammaticalize, and the increased frequency of these periphrastic expression now characterises oral vernacular BP (cf. Zilles, 2007, amongst others).

In short, the mechanisms mentioned in this section are just some of the changes an item is exposed when it is grammaticalized. Although they are not absolute criteria for grammaticalization, Hopper and Traugott (2003) illustrate how grammaticalized items from many different language groups go through these stages in a unidirectional manner. There are also some counter-examples of unidirectionality, e.g. that grammaticalized forms skip stages or mechanisms. The grammaticalization of English *be going to*, however, has been viewed as unidirectional, and the following paragraph illustrates how *be going to* has undergone changes in different stages of a grammaticalization cline.

2.2. Clines

Hopper and Traugott (2003) describe how clines or continuums represent stages of grammaticalization. A grammatical *cline*, a *continuum* or a *chain*, “is a metaphor for the empirical observations that cross-linguistically forms tend to undergo the same kinds of changes or have similar sets of relationships, in similar orders” (Hopper and Traugott, 2003, p.6). The following example retrieved from Hopper and Traugott, illustrate primary grammaticalization in which a lexical item is typically at the one end of the cline and a grammatical form at the other end, but a grammatical item is the first stage of the cline in secondary grammaticalization.

(9) Content item > grammatical item > clitic > inflectional affix

(Hopper and Traugott, 2003, p. 7)

Another example is a verb-to-affix cline:

(10) Full verb > auxiliary > verbal clitic > affix

(Hopper and Traugott, 2003, p. 111)

Be going to and *to have* are used in the following examples to illustrate the cline in (10):

(11) Cline with HAVE

full verb/content item > auxiliary > clitic

I have a dog > *I have had* a dog > *I’ve* had a dog

(12) Cline with BE GOING TO

full verb/content item > purposive auxiliary/ambiguous > future auxiliary >

She is going to the store > She’s going to buy it > She’s going to like it >

Phonetically eroded auxiliary

she’s gonna like it.

One can also argue that in (African) American vernacular English (AAVE), the construction has undergone morphosyntactic reduction, since the phrase “I am going to” often is abbreviated “Imma” or “I’mma,” which is pronounced /'aɪmə/, or ['a:mə]⁵, as e.g. in the song title “Imma Be” by Black Eyed Peas (2010). In other words, “going to” is in certain contexts, phonetically and morphosyntactically reduced to the affix “a”.

Although it is unlikely that this abbreviation will be accepted in Standard English, it is still worth mentioning that the construction has reached the ending point of the cline, i.e. affixation, in vernacular American English. BP *estar indo*, however, is still mostly used as a content item, and it has only recently started to be productive as a purposive auxiliary. Therefore, it is necessary to consider how reanalysis of EMdE *be going to* occurred, and when the future auxiliary use became more frequent than the lexical use. Hence, the terms *reanalysis* and *frequency* as processes of grammaticalization are the topics in focus.

2.3. Reanalysis

Be going to is often used to exemplify *reanalysis* (cf. Hopper and Traugott, 2003). Reanalysis involves *rebracketing* of elements in a certain construction. For instance, the lexical *be going to* is bracketed as *be*+ progressive GO as the main verb phrase, followed by preposition ‘to’ and either an adverb of space or a compliment verb phrase. Reanalysis also includes *reassignment* of morphemes to different semantic-syntactic categories (Hopper and Traugott, 2003, p. 51). Both rebracketing and reassignment of morphemes are illustrated in the examples within the table below:

Table 3: Explanation of Rebracketing in Grammaticalization

<p>Original bracketing of <i>be going to</i>:</p> <p>[I am going [to fetch water]]</p> <p>Auxiliary <i>be</i> + progressive aspect of main verb <i>going</i> + purposive preposition <i>to</i>.</p> <p>Reading: “I am physically going at the moment of the utterance and the purpose of this ‘going/walking activity’ is to fetch water.”</p>	<p>Rebracketing:</p> <p>[I [am going to] fetch water]</p> <p><i>Be going to</i> is reassigned as a future tense marker.</p> <p>Reading: “I’ll fetch water.”</p>
--	---

⁵ Cf. Imma, Wiktionary, Retrieved from <https://en.wiktionary.org/wiki/Imma> 24.04.18

Reanalysis of *be going to* took place in between EMdE and Modern English (MdE), i.e. in the 1700s (cf. Traugott, 2012, Jakobi, 2004, Wu, He & Feng, 2016), but it became fully grammaticalized in the late 17th century and it is from this period that it is frequent in corpora (Wu, et.al., 2016), and the study of *frequency* and its implications for studies on grammaticalization is elaborated upon in the following section.

2.4. Frequency

In this study, *frequency* is examined and assessed through data analysis of findings from different corpora (see Ch. 4). Increased frequency, i.e. how often a word is used, is usually part of the grammaticalization process (cf. Killie 2014a, Hopper and Traugott 2003). Another aspect of research on frequency is that “statistical evidence for the frequency of forms is a valuable tool in providing empirical evidence for unidirectionality” (Hopper and Traugott, 2003, p. 126).

Furthermore, much of the literature on grammaticalization is dedicated to examining frequency effects on grammaticalization (cf. Hopper and Traugott, 2003, p. 127). By referring to Bybee and Thompson (1997), Hopper and Traugott suggest that there are two effects of token frequency, which is understood as the number of times an item occurs in e.g. a text (2003, p.127). The two effects are labelled the *Reduction Effect* and *the Conservation Effect* by Bybee and Thompson in 1997 (Hopper and Traugott, 2003, pp. 127-128).

The Conservation effect shows that frequent irregular forms keep their irregularity, while less frequent irregular forms tend to become more regular (Hopper and Traugott, 2003, p. 128). *The Reduction Effect* is especially interesting because it shows that frequent items undergo phonetic erosion more easily and faster than less frequent items. The auxiliary *be* is the perfect example of both *the Reduction Effect* and *the Conservation Effect*. The present tense *be* conjugation paradigm is the most irregular of all PDE verbs, and the *be* auxiliary has undergone phonetic reduction and cliticization, as seen in example (5):

(13)

a. She’**s** playing b. You’**re** not eating I’**m** studying

There is a certain amount of overlap between the emergence of *be going to V* as a construction, and the high frequency of *-ing* progressives. As mentioned briefly in the previous sections, this thesis argues that the increased *frequency* of *-ing* progressives enabled the grammaticalization of EMdE *be going to*. The increased frequency, grammaticalization and reanalysis of the

English *-ing* periphrasis as a progressive marker was evident in standard Modern English, but it has roots in OE and EMdE (Killie, 2014a, Kranich, 2010).

In short, both *be going to* and *-ing* progressives became increasingly frequent and grammaticalized in the 17th century (cf. Jakobi, 2009, Traugott, 2012, Wu, et.al., 2016). In her article in the grammaticalization of the *-ende/-ing* periphrasis, Killie (2014a) concludes that the grammaticalization of the *-ing* periphrasis could have been reinforced by English contact with written Latin. In the same way, section 6.3. discusses if contact with English influenced the grammaticalization of *-ndo* progressives and *estar indo* in BP, but first, the following chapter considers some preliminary issues regarding language contact in general and specifically *be going to* as a progressive.

Chapter 3: Some preliminaries on *be going to* as a progressive, the progressive aspect and BP progressives.

To discuss the grammatical properties of English *be going to* and compare them to BP *estar indo*, it is necessary to clarify what is meant by terms such as *progressive*, *periphrasis*, *future verb*, etc. In the introduction there is a brief overview of some terms referring to different progressives, future verb phrases and futurates, but this chapter will elaborate upon terms used specifically about *be going to/estar indo*. In addition, the first sections summarise relevant theory on language contact, different varieties of Portuguese, contact between BP and English, and other issues with implications for how the development of *be going to (V)* and *estar indo (V)* is studied here.

The first section focuses on linguistic borrowings and language contact. The second section explains important differences between the morpho-syntactic expression of the *progressive aspect* in BP and EP (European Portuguese), and the difference between *focalized progressives* and *durative progressives*. The third sections accounts for different terms that are used about *be going to* and/or BP *estar indo*.

3.1. A couple of notes on language contact (between BP and English) and borrowing grammar

Borrowing grammar is a rare phenomenon. According to Garcia (2010) and Rocha (2012), the FPP is a grammatical loan recently imported from English, but neither discuss why BP would have borrowed this grammatical construction from English. Contrary to grammar, lexemes are frequently spread and borrowed. Most modern languages have some borrowed lexeme(s) from English, or anglicisms. Norwegians, for instance, say “bacon” instead of “svinesvor,” and Mexicans say “rentar” (“to rent”) instead of “alquilar” to mention a couple examples.

Furthermore, research on language contact tells us that lexica are usually taken from the superstrate language and keeps the grammar from the substrate language (cf. Zhiming, 1998). Although it is unusual to borrow grammar it has been known to happen, and it is not unlikely that BP could borrow grammar since there are many English loanwords in BP (cf. Montes, 2016). The language contact between BP and English, however, is not extreme and it does not resemble the instances of creolization that have taken place during the era of colonialization. In

addition, Brazil is one of the few countries that has implemented official policies to restrict the use of English (cf. Tsuda, 2010).

If BP is borrowing grammar from English, is it then due to extreme contact between the languages because of migration and Brazilian diaspora? Cacoullos (2000) emphasises that bilingual speakers of Spanish and English in Columbia and the USA often use Spanish *-ndo* progressives in the same manner as *-ing* progressives are used in English. Unlike BP, Latin-American Spanish has been in extreme contact with English for the past two centuries. Firstly because of American annexation of Mexican territory (1846-1848), and secondly because of the high number of Spanish speakers in the USA, which is currently more than 50 million (cf. the Guardian, 30 June 2015). Changes in BP, however, cannot be attributed to the same type of extreme contact with English, and the number of BP/English bilinguals is very small compared to that of Spanish/English bilinguals.

Nevertheless, the use of BP progressives *is* very similar to the use of English progressives, and the examples in the following section illustrate how the syntax of BP progressives is even more similar to English progressives, than to European Portuguese progressives.

3.2. Differences between BP and European Portuguese progressives and different progressive aspects.

Examples (14) and (15) below elucidate the differences between European Portuguese (henceforth: EP) progressives and BP progressives. In addition, they exemplify different present progressive aspects. In order to investigate why the BP progressives, including *estar indo*, are more prone to grammaticalization than EP progressives, this section also discuss different progressive aspects, in order to explore whether the progressive aspect influences the grammaticalization of the periphrasis.

Killie (2014a) uses Bertinetto's⁶ terms *focalized progressive* and *durative progressive*, and these terms are also employed in the present study. The verb phrases in examples (14a) and (15a) are *durative progressives*, i.e. continuous events over a longer period of time (Killie, 2014a, p. 362), and (14b) and (15b) illustrate the *focalized progressive*, which occurs at a single point in time (Killie, 2014a, p. 362). The examples (14) and (15) also shows how contrary to both PDE and BP, the present progressive aspect in European Portuguese is not expressed

⁶ Killie's reference (2014a): Bertinetto, Pier Marco. 2000. The progressive in Romance, as Compared with English. In Dahl (ed.), 559–604

through the use of *-ndo/ -ing* participle, but through the auxiliary *estar* (to be) + *a* (preposition 'to') + infinitive:

(14) European Portuguese present progressive aspect:

- a. *Sabe que estás a trair.*
[Knows that are to betray.]
'He knows that you are you cheating on him.'
- b. *Está a beber água na cozinha.*
[Is to drink water in the kitchen]
'He is drinking water in the kitchen'

(15) BP present progressive aspect:

- a. *Ele sabe que voce está traindo.*
'He knows that you are cheating [on him].'
- b. *Ele está tomando água na cozinha.*
He is drinking water in the kitchen.

Examples (14a) and (15a) show that the BP verb phrases are more periphrastic than the EP verb phrases. In (14), the European Portuguese subject pronouns are omitted and conveyed through the inflectional morphology, while the BP subject pronouns are not. One could argue that the pronoun 'ele' (i.e. 'he') in example (15) is emphatic, or even redundant, but Zilles research on the development of new BP pronouns show that BP pronouns are losing their emphatic properties as the verb phrase is reorganised and subject pronouns are conveyed through independent morphemes, and even phrases, instead of affixation (cf. Zilles, 2007, among others). This implies that subject pronouns are no longer dropped and BP verb phrases become more periphrastic.

In sum, although the periphrasis of BP and EP progressives differ, the progressive aspect, i.e. the semantic content of the progressives are the same. Rocha (2012) demonstrates that BP progressives are more prone to grammaticalization than EP progressives, and this will also become more evident after the data analysis in chapters 4 and 5. This could be due to the fact the BP progressive is morpho-syntactically more similar to the English, and thus more susceptible to English influence, but the grammaticalization can also be triggered by the fact that BP verb phrases in general are becoming more periphrastic.

3.3. Different terms referring to *be going to V* or *estar indo V*

It is important to be precise about how the terms referring to *be going to (V)* or *estar indo (V)* are understood and used in this thesis, therefore this section will explain what is meant by "phrase", "construction", "progressive", and "-ing periphrasis

- **The phrase** refers to *be going to (V)* or *estar indo (V)*, which is often ambiguous, but can be interpreted as an unambiguous lexical item (as in (16a)), or it may have undergone full (primary) grammaticalization and thus function as an unambiguous future auxiliary (as in (16b)).

(16) a. He *is going to* the bus.⁷

b. He *is going to* like it.⁸

- **The construction** is the term used for *be going to V* or *estar indo V* when it has undergone full primary grammaticalization, as seen in (16) b.
- The V-ende/V-ing **progressive** refers to *aspect*, i.e. semantic content as well as the morpho-syntactic structure. As with the term *the periphrasis*, *the progressive* is employed as a reference to all *v-ing periphrases*, including *be going to V*.
- The **-ing periphrasis**, which also encompass *be going to V*, refers to the morphosyntactic expression, but not the semantic content. The *-ing* periphrasis must not be confused with the term BP *periphrastic future*, which is explained in chapter 6.

⁷In a), "*is going to*" is a lexical item, i.e. it expresses the progressive aspect, and it does not function as a future auxiliary.

⁸In b), "*is going to*" is a grammatical item, i.e. it functions as an unambiguous future auxiliary. In other words, one cannot physically "go to like something".

Chapter 4: Data and Methodology

To compare English *be going to* and BP *estar indo* both qualitative and quantitative methods are used. Online corpora enable quantitative research on PDE *be going to*, but data on *be going to* from the 16th and 17th century is difficult to come by. BP *estar indo* is primarily observable in oral language, and the frequency is very low in written language. Because EMdE AUX *be going to* and BP AUX *estar indo* are rarely seen in written language, quantitative methods are not suitable for studying the productivity and frequency of the constructions.

Thus, this thesis consults a range of different corpora in order to do a qualitative study of the grammaticalization. The Twitter corpus makes it possible to investigate how much BP *estar indo* has grammaticalized, and corpora from BYU (Brigham Young University) are used firstly, to find examples of BP AUX *estar indo*, and secondly to examine the frequency of PDE *be going to* and BP *estar indo*, and finally to compare instances of future AUX *estar indo* with examples of the periphrastic future verb phrase (i.e. *vou comprar*/ 'I will buy') in BP. The following sections give short descriptions of the corpora used.

4.1. A comparison of EMdE data in Traugott (2012) and BP data

The EMdE data, which is gathered from written transcripts of oral language c.1600-1750, will be very different from the BP data. The BP data is neither transcribed oral language, nor writing. Instead it is spontaneous digitally orthographic utterances, much more similar to casual, informal, vernacular and unmonitored conversation, or as McWhorter (2013) labels it: “fingered speech”. The EMdE data shows that AUX *be going to* occurred in standard oral English, while the BP data shows that AUX *estar indo* occurs in vernacular, non-standard BP. Furthermore,

The English *be going to* has become the archetype example of primary grammaticalization, i.e. a lexical phrase becoming a grammatical phrase (cf. Hopper and Traugott, 2003). It has also become very frequent since its emergence in Early Modern English (EMdE) (cf. Traugott, 2012, Jacobi, 2004), and in fact, all, *-ing* progressives have become increasingly frequent in Present-Day English (cf. Nesselhauf, 2007, 2010).

Nevertheless, contrary to data on PDE *be going to*, it is virtually impossible to find samples of unmonitored and casual EMdE. One could try to make a corpus from the dialogues in certain

novels, but it would still be written language and the entire speech would be fiction, which is not quite as authentic as spontaneous oral language. The Helsinki Corpora (HC) and the Proceedings of the Old Bailey (POB) are treasure chests of EMdE and MdE (Modern English) spoken language. These corpora include data from criminal proceedings, so although the speech is more constrained than informal gossiping among peers, it is still authentic oral speech. According to the POB webpage, there is data from 197,745 criminal trials conducted at the central criminal courts of London, dating from 1674-1913.

Traugott defines *linguistic context* as the linguistic environment, including pragmatic inferences, larger discourses in both written and spoken language, and it is not restricted to the clause in which the item of interest appears (2012, p. 232). In other words, it is not defined as the physical setting, but the **linguistic** environment. Three conclusions are highlighted here since they are especially relevant for this thesis (adapted from Traugott, 2012, pp. 241-242):

- I. *Critical contexts* enable grammaticalization even though the contexts often are pragmatically ambiguous
- II. Reanalysis has taken place in *Switch contexts*, as e.g. when AUX *be going to* occurs with a mental verb, or in other instances where pragmatic meaning has become semantic meaning.
- III. Conventionalization contexts leads to generalization, increased productivity and the grammaticalized construction becomes more frequent.

Over the centuries, English *be going to* has appeared in all of the linguistic contexts described above. Consequently, *be going to* has been exposed to a wide range of micro-changes, which has shaped the use of the construction. Furthermore, Traugott (2012) has focused on the following collocates appearing with AUX *be going to* in the POB corpus, and in section 4.5. they are compared to the BP collocates in the Twitter corpus:

Carry, buy, see (in the sense visit), look at, see after, do, work, marry, arraign, pay, 'bilk' (cheat).

Table 4: Overview of collocates in Traugott's research (2012):

There are more main verb collocates, and “[t]here are possibly seventy-eight relevant examples of BE going to V in POB 1674-1723 (‘possibly’ because in some cases whether a collocating lexeme is a verb or a noun is undecidable)” (Traugott, 2012, p. 238). Furthermore, she writes that twenty-six are still ambiguous, viz. they could be read as both physically *going* and as

future auxiliaries, and “[f]orty of the seventy-eight examples appear to be of the auxiliary, i.e. to have the future meaning” (Traugott, 2012, pp.238-240).

The EMdE samples are characterised as more formal speech than the BP samples in the Twitter corpus. Although (most of) the EMdE speakers are non-elite, the court room setting is formal, and thus they probably adjust their utterances to resemble formal, standard English. In other words, EMdE *be going to* had already grammaticalized to the extent of occurring in formal speech, but BP AUX *estar indo* appears only in casual, vernacular language.

4.2. The BYU corpora

As mentioned, this thesis will also be using the BYU corpora, which contains both English and Portuguese corpora from written texts. According to their webpage “[The BYU corpora] are probably the most widely-used corpora currently available” (BYU, overview: <https://corpus.byu.edu/overview.asp>¹⁰). The largest corpus is the Google Books corpus of American English, which contains 155 **billion** words. Some texts are formal such as academic essays and articles, news channels and newspapers (both broadsheet and tabloids), books (fiction, non-fiction, literally any type of book imaginable), etc. and some texts are of more informal character, such as blogs, personal webpages, etc.

The Portuguese BYU corpus that was created in 2015-2016 contains data that was gathered in 2013-2014. It contains 1 billion words found in digital sources from Brazil, Portugal, Angola and Mozambique.

⁹ BYU, overview. Retrieved from <https://corpus.byu.edu/overview.asp> (25.04.18)

¹⁰ BYU, overview. Retrieved from <https://corpus.byu.edu/overview.asp> (25.04.18)

Table 5: Overview of the data found in BYUP 2015. Adapted from BYU¹¹.

Country	CODE	General (may also include blogs)			(Only) Blogs			Total		
		Words	Web pages	Web sites	Words	Web pages	Web sites	Words	Web pages	Web sites
Brazil	BR	319,435,592	286,712	25,351	336,244,918	321,305	35,248	655,680,510	608,017	60,599
Portugal	PT	136,144,529	184,512	12,082	190,503,822	221,338	9,005	326,648,351	405,850	21,087
Angola	AO	17,877,399	19,178	1,240	17,255,595	21,233	418	35,132,994	40,411	1,658
Mozambique	MZ	16,936,743	19,236	1,065	15,070,829	17,910	404	32,007,572	37,146	1,469
TOTAL		490,394,263	509,638	39,738	559,075,164	581,786	45,075	1,049,469,427	1,091,424	84,813

The corpora site provides search functions that make it possible to exclude countries and compare words from the different varieties of Portuguese. Since only the BP sources are relevant for this study, the actual word count for the corpora used contains 655 million words. This Portuguese corpus is referred to as the *New* BYU corpus. The corpus from 2004-2006 contains 45 million words from the 1200s to the 1900s and this corpus is referred to as the *Old* Portuguese corpus. The Old corpus contains words from Portugal and Brazil, but there are slightly more words from Portugal, probably due to the fact that the BP (and Brazil for that matter) did not exist until the late 15th century.

4.3. The Twitter corpus

Since there are only a few instances of *estar indo* in the BYU corpora, it has been necessary to gather data from other sources in order to investigate the use of the phrase in BP. Therefore, the present study has made a corpus solely for the purpose of investigating the grammaticalization of BP *estar indo*. The findings in this new corpus have been surprising and may indicate that BP *estar indo* is further along in the grammaticalization process than the initial assumption.

¹¹ BYU, o corpus do português, web/dialectm 1 billion words, created 2015-2016. Retrieved from <https://www.corpusdoportugues.org/web-dial/> (31.10.2017).

To explore the use of *estar indo* it is necessary to consider the social and linguistic contexts in which it appears. When *estar indo* is used on Twitter it is in a very different context than the formal court room context of EMdE *be going to*, and the type of language that is used on Twitter, and within social media in general is very different from formal writing.

The American linguist John McWhorter claims that texting, which is the term used for communication through ICT, is fingered speech and *not* writing (see McWhorter, 2013). In a talk (or mini-lecture) he gave at a TED conference in 2013 he also says that “[linguists] have actually shown that when we are speaking causally and in an unmonitored way, we tend to speak in word packets of maybe seven to ten words” (McWhorter, 2013, 02:40). He then shows examples of “texting dialogues” and it is evident that the speakers or “texters” use short and abbreviated utterances, pragmatic particles, markers of empathy, and they are constantly developing new discourse markers.

Until September 2017, the *tweets*, i.e. the digital utterances on Twitter, were limited to 140 characters, or approximately 25 words, and although that is bigger than the word packets in oral conversation, it is considerably smaller than e.g. hand-written letters that were used in long-distance communication before ICT. Thus, I categorise tweets as casual fingered speech, and not standard writing.

The methodology that consists of using Twitter to gather data was first developed by Rocha (2012). She collected data from Twitter and created a mini corpus documenting the use of BP *present progressive verb* phrases as futurates. Inspired by her work and the efficacy of her methodology, this study has made a similar corpus containing *estar indo* phrases instead of FPP phrases. There are 193 *estar indo* phrases that can be interpreted as future auxiliaries. The Twitter corpus can be found at: <https://typecraft.org/tc2/ntceditor.html#3885>, and it is also attached in the appendix.

The *Twitter data* is collected simply by searching for phrases within the social network. The *tweets* (i.e. statements of 140 characters) are publicly posted and they are accessible for all users of the network, and underneath is an example of a search:

Figure 1: Twitter search for data

Advanced search

Words

All of these words

This exact phrase

Any of these words

None of these words

These hashtags

Written in

The search usually consisted of a temporal adverb and a phrase with AUX *estar indo* + collocate, and a typical result would look like this:



Figure 2: Example of Twitter search result

The previous tweet could be translated into:

“My internet is garbage luckily already¹² in a little while I am going to sleep,” or: “The internet connection is horrible, luckily I’m going to sleep in a little while”.

¹² “Ja” is a temporal adverb and it usually means “already”. It also occurs as a conjunction in the expression “Ja que,” which is usually translated to “since” or “as”. The temporal adverb “daqui a pouco” (“in a little while”) indicates that “to indo” (“I’m going to”) is the future AUX, but adding “ja” makes it ambiguous, thus in a sense it is as if the two temporal adverbs eliminate one another, and instead of expressing either future or present tense it adds to the ambiguity.

Since the language style on Twitter is colloquial and vernacular, which is the linguistics environment in which BP AUX *estar indo* is the most frequent, the searching criteria is withdrawn from vernacular BP as well, and that is the topic of the following section.

4.4. Searching criteria and selection of data.

In PDE there are three different conjugations of present tense *be*, i.e. *am*, *are*, *is*. Including the clitics ‘*m*, ‘*re*, ‘*s*, there are six possible forms. In standard BP, however, the six conjugations encode the six different personal pronouns through the morphology of the verb. This means that in formal or standard BP the pronouns are dropped, i.e. *I am* would be just *estou*:

(17) a. *estou, estás, está, estamos, estáis, estão.*

b. ‘I am, you are, he/she/it is, we are, you are, they are.’

There are also abbreviated forms, e.g. *estou (I am)* is abbreviated *tô (i.e. I’m)*. Furthermore, the paradigm of *estar* in vernacular BP is more complex than the paradigm in standard BP. To account for the frequency of e.g. “I’m going” one would have to make searches for at least four distinct varieties of “I am going,” namely *estou +indo*, *tô +indo*, *tou +indo*, *to +indo*. Zilles (2007) accounts for how the pronouns and inflectional morphology of second person singular and first-person plural (see example 26) have been grammaticalized and phonetically eroded. How the new pronouns influence *estar indo* will be discussed further in chapters 5 and 6.

Rocha (2012) used the standard BP paradigm of *estar*, labelled *canonical orthography*, and some of the contracted forms, labelled *internet orthography* (see the figure 3, adapted from Rocha, 2012, p. 17).

Figure 3: Varieties of *estar* (Rocha, 2012, p. 17).

Table 4: Search terms for the FPP data acquisition

	Canonical orthography	Internet orthography (<i>-es /loses its first syllable</i>)
1st SG	estou *ndo ¹²	tou, tô, tow *ndo
2nd SG	está, estás *ndo	tá *ndo
3rd SG	está *ndo	tá *ndo
1st PL	estamos *ndo	tamos *ndo
2nd PL	estão, estais *ndo	tão *ndo
3rd PL	estão *ndo	tão *ndo

The present thesis, however, includes the grammaticalized pronoun *a gente* (*the people*), which functions as the 1st PL and takes the auxiliary *tá* (cf. Zilles, 2007). Thus, instead of reusing Rocha's paradigm of *estar*, this study will use the varieties of *estar* in table 10. The pronoun *a gente* is illustrated in the example below, and *tá* is the most frequent form of present tense *estar* and may occur with 2nd SG, 3rd SG and 1st PL.

(18) a gente **ta indo comer** sushi **amanha**

[the people's¹³ going to eat sushi tomorrow]

'we're going to eat sushi tomorrow'

¹³ The BP pronoun *a gente* takes the verb conjugation as 3rd SG, therefore the direct translation would be "the people is".

Table 6: Traditional and phonetically eroded *estar*

	Traditional/Prescriptive <i>estar</i>	Phonetically eroded <i>estar</i> in oral, digital and/or vernacular BP
1st SG	estou indo + collocate	to, tou, tô, tow indo + collocate
2nd SG	está, estás indo + collocate	tá, ta indo + collocate
3rd SG	está indo + collocate	tá, ta indo + collocate
1st PL	estamos indo + collocate	tá, ta, tamos, tamo indo + collocate
2nd PL	estão, estais indo+ collocate	tão ¹⁴ indo + collocate
3rd PL	estão indo + collocate	tão indo + collocate

For practical reasons, only two search sets were employed when data was gathered for the Twitter corpus, i.e. *to, tô indo + collocate* (*I'm going to*) and *tá, ta indo + collocate* (*you/he/she/it/we are going to*) and a couple of results are listed in examples (19) and (20) below:

(19) **amanhã tô indo** fazer minha matricula no aaz

‘tomorrow I’m going to do my matriculation at aaz’

(20) E quando a crush¹⁵ avisa **que tá indo estudar** na sua faculdade ano **que vem**

‘And when the crush (i.e. the person you are very attracted to) warns you/tells you that she is going to study at your faculty next year’

In sum, the present chapter has described the corpora and searching criteria that are used in this study to investigate the grammaticalization and frequency of EmdE and PDE *be going to* and BP *estar indo*. The study emphasises that the BP data and searching criteria are derived from vernacular language, whereas the EmdE and PDE data are from written texts or transcribed formal speech, and in the following chapter the data is analysed.

¹⁴ «Tão» is polysemous. It can be translated to the adverb “so” or “very,” e.g. a) *Tão injusto!* i.e. ‘So unfair!’, or, it can be translated to the abbreviated 3rd PL of ESTAR/BE, e.g. b) *Tão sendo injustos*, i.e. ‘They’re being unjust’.

¹⁵ “o/a crush” is an example of the many frequently used loanwords from English. The singular, feminine indefinite article *a* indicates that the “crush” in question is female.

Chapter 5: Data Analysis

The present chapter compares data that shows the frequency of PDE *be going to* and BP *estar indo*, and the BP *periphrastic future* (akin to the PDE future construction with modal AUX *will* as in e.g. “I will study”). One of the aims of the analysis is to discuss what type of collocates that appear with BP *estar indo* and what the collocates and temporal adverbs indicate about the grammaticalization of *estar indo*.

5.1. Comparison of BP *estar indo* and PDE *be going to* in the BYU corpora

To assess how far along *estar indo* is in its grammaticalization process the frequency of AUX *estar indo* in the new BYU corpus is investigated, which is then compared to the frequency of the lexical use of *estar indo* (i.e. physically “going somewhere”) and the frequency of PDE *be going to*. As mentioned briefly in the preceding sections, AUX *estar indo* only occurs in oral vernacular BP, thus its frequency in the BYU corpora is very low, so low in fact that it is surprising to find evidence of it at all in the BYU corpora.

Contrary to PDE *be going to*, lexical BP *estar indo* (i.e. the present progressive *estar indo* denoting physical movement) is much more frequent than future AUX *estar indo* V. There are 341557 examples of the PDE phrase “is going to” in the BYU NoW (News on the Web) corpus. Of the 341557, this study did randomly select 200 phrases to investigate the frequency of lexical *be going to* and AUX *be going to*.

Table 7 shows an overview of the data and the frequency of *be going to/estar indo* as lexical items and grammatical items. Notice that the frequency rate is calculated by a random selection of 200 phrases, and a manual count of the appearance of lexical *be going to/estar indo*. Unfortunately, the BYU site lacked the digital tools for assessing whether collocates are main verbs, noun phrases or place adverbials, therefore this was done manually.

Of the 200 phrases in the NoW corpus, there were four instances of lexical *be going to* in PDE, but one is idiomatic, and therefore it is repeated here:

(21) This world is going to the dogs

In example (21), *be going to* is still lexical, but it has a pragmatic meaning rather than a semantic meaning. Thus, out of 200 in total, there are only three instances of lexical *be going to* with a semantic meaning. Two of these sentences are included in the following table:

Table 7: Examples of PDE "be going to" and PBP "estar indo" as lexical items and grammatical items in the BYU corpora.

Lexical and grammatical reading of <i>be going to/estar indo</i> in the BYU CORPORA			
NoW: 14 billion and expanding, i.e. continuously gathering data from News on the Web.		O corpus do Portugues: 1.1 billion. 2016.	
341557 examples of 3SG present tense <i>be going to</i> (i.e. "is going to").		4188 examples of <i>estar indo</i> of which 316 "tá indo" ('s going) 3872 "está indo" (is going)	
Lexical "be going to" in PDE	AUX "be going to" in PDE	Lexical "estar indo" in PBP	AUX/ambiguous "estar indo" in PBP
Frequency rate: Approximately 0,02	Frequency rate: Approximately 0,98	Frequency rate: Approximately 0,975	Frequency rate: Approximately 0,025
" The President is going to those states (...)" ¹⁶	It is going to be humongous and a behemoth ¹⁷	está indo para o São Paulo Is going to São Paulo	quem tá indo morar sozinho Who is going to live alone
the ball is looking like it is going to a certain position ¹⁸	"Tigmanshu is going to be directing me soon. (...)" ¹⁹	Me conta como está indo com o Perfect Tell me how it is going with the (read: Mr.) Perfect	o que ele tá indo fazer lá What he is going to do there

¹⁶ NoW, BYU, 12.04.18. Original source: <https://www.wetinhappen.com.ng/killings-buhari-is-not-interested-in-cosmetic-solutions-says-adesina/> (05.03.18).

¹⁷ NoW, BYU, 12.04.18. Original source: <https://www.concisenews.global/business/oil-marketers-oppose-single-regulator-in-petroleum-bill/> (26.03.18)

¹⁸ NoW, BYU, 12.04.18. Original source. http://www.espncriinfo.com/story/_/id/18058349/george-dobell-ben-stokes-revelling-role-england-go-man (15.11.16).

¹⁹ NoW, BYU, 12.04.18. Original source: <https://timesofindia.indiatimes.com/entertainment/hindi/bollywood/news/ali-fazal-preps-for-milan-talkies/articleshow/63388833.cms> (20.03.18)

In the remaining 196 examples, *be going to* functions as an auxiliary, and one of the most frequent main verbs is BE (as seen in the table above).

In PBP, on the other hand, the frequency is completely opposite: There are only five instances of AUX *estar indo*, in a random selection of 200 phrases, and of these five, two are still ambiguous. There are, however, many examples of *estar indo* being followed by a preposition/conjunction “para”. “Para” is usually translated to “to” or “for” in English, but as a conjunction “para” could also mean “in order to”. The relatively high frequency of “para” could indicate that *estar indo* is establishing itself as a purposive future, which was an important step in grammaticalization of EMdE *be going to* (cf. Traugott 2012)

5.2. Brief overview of the frequency of *estar indo* compared to the periphrastic future in the BYU corpora

The previous section compared the frequency of PDE *be going to* and BPB *estar indo*, but the present section compares data on BP *estar indo* to data on the most frequent future verb phrase in BP, namely the *periphrastic future* (the construction is explained briefly below, and it is discussed further in section 6.2.2). The reason these future verb phrases are compared is to highlight how *estar indo* is used and to further investigate the frequency of *estar indo*.

The BP future verb phrases that are compared are:

(22)

- | | |
|----------------------------------|-----------------|
| a) Estar indo + viajar | b) IR + viajar |
| Be going to + travel (literally) | AUX GO + travel |
| ‘I’m going to travel’ | ‘I’ll travel’ |

In other words, they are both periphrastic verb phrases with either simple present AUX IR (i.e. the simple present tense of GO functioning as a future auxiliary) or progressive *estar indo* (i.e. *be going to* functioning as a purposive, and possibly ambiguous future auxiliary), with the collocate *viajar*, i.e. ‘to travel’. ‘To travel’ is an activity verb indicating physical movement, which implies that *estar indo* (*be going to*) is ambiguous and a purposive future, and not reanalysed as a future auxiliary.

Language change and/or grammaticalization usually occurs in oral vernacular language before it appears in standard written language (cf. Trudgill, 2016, Zilles, 2007). This is also evident in the examples found in the data from the New BYU Portuguese corpus (2015) and the Old BYU Portuguese corpus (2004) in table 7 show the future use of BP *estar indo*:

Table 8: Data from BYU on the frequency of *Estou indo + infinitive*.

Estou indo + collocate (I'm going to)	Frequency in the New BYU corpus (from PBP on the web/vernaculars) ²⁰	Frequency in the Old BYU corpus in genre/historical registers ²¹
Estou indo + viajar I'm going to + travel	41	n/a
Estou indo + pegar I'm going to + get	4	n/a
Estou indo + comprar I'm going to + buy	20	n/a

Some of the examples from the New corpus include the following phrases:

(23) *Tô indo viajar mês que vem*²².

'I'm going to travel next month.'

(24) *Já²³ estou indo comprar umas verduras*²⁴.

'Soon I'm going to buy some vegetables.'

²⁰ BYU, o corpus do português, web/dialectm 1 billion words, created 2015-2016. Retrieved from <https://www.corpusdoportugues.org/web-dial/> (31.10.2017).

²¹ BYU, o corpus do português, genre/historical 45 million words, created 2004-2006. Retrieved from <https://www.corpusdoportugues.org/> (31.10.2017).

²² Kamilla Silva. (25 Feb 2013). *Hey London: Intercambio na Inglaterra*. Blog. Retrieved from <http://www.hey-london.net/2013/02/intercambio-na-ingles-terra-experiencia.html> (31.10.2017).

²³ 'Já'

²⁴ Da minha cosinha (19 Jan 2009). Blog. Retrieved from: <http://www.daminhacozinha.net/2009/01/rag-de-porco-esquecido.html> (31.10.2017).

(25)

(...) *e amanhã estou indo comprar macadamia para fabricar meu proprio queijo*²⁵.

‘(...) and tomorrow I’m going to buy macadamia to fabricate [(i.e. make)] my own cheese.’

The temporal adverbials in the examples above show that *estar indo* functions as a future auxiliary, but it is still a purposive future. In table 8 below there is a brief overview of the frequency of the traditional periphrastic future verb phrase (IR + infinitive). The phrases are more or less equivalent to the phrases used in table 7 and the data is also retrieved from the BYU corpora.

Table 9: Data from BYU on the frequency of IR + infinitive, i.e. the most frequent future verb phrase (see 6.2.2. for further description of future verb phrases in BP).

Vou + collocate (I’m going to/ I will)	Frequency on the web/ vernaculars ²⁶	Frequency in genre/historical registers ²⁷
Vou + viajar I’m going to/ I will + travel	959	41
Vou + pegar I’m going to/ I will + get	1265	13
Vou + comprar I’m going to/ I will + buy	4466	41

As seen in table 8, BP AUX *estar indo V* is rare with activity collocates in written registers, and it is non-existent with mental verb collocates in the BYU corpora. Thus, in order to investigate the grammaticalization of *estar indo* further, the Twitter corpus is consulted.

²⁵ Retrieved from: <https://www.vista-se.com.br/redesocial/10-produtos-que-tem-restos-de-animais-e-voce-nao-sabia/> (31.10.2017).

²⁶ BYU, o corpus do português, web/dialectm 1 billion words, created 2015-2016. Retrieved from <https://www.corpusdoportugues.org/web-dial/> (31.10.2017).

²⁷ BYU, o corpus do português, genre/historical 45 million words, created 2004-2006. Retrieved from <https://www.corpusdoportugues.org/> (31.10.2017).

5.3. BP collocates in the Twitter corpus.

The collocate, i.e. the main verb, occurring with *be going to/estar indo* is one of the best indicators of how much the phrase has grammaticalized (cf. Traugott, 2012, Wu, et.al. 2016). Choosing collocates and gathering data for the Twitter corpus has been a gradual process. Initially the criteria that determined the choice of collocates was that *estar indo* + the collocate had to be relatively frequent on Twitter, i.e. it had to appear more than once or twice, and the BP collocates should be similar to the EMdE collocates in Traugott’s research (2012).

Some of the EMdE collocates are easily translated to BP, but e.g. ‘carry,’ in the sense ‘lift, move, support someone/something’ is not easily translated to BP. ‘Carry’ could be translated to *carregar*, but on Twitter it most frequently means ‘to charge’ or ‘to download.’ ‘Carry’ could also be translated to *levar*, which is more accurately translated to ‘take or bring along’. As a consequence of the problems with finding accurate translations to certain words, the Twitter corpus now presents more than double the amount of collocates that were initially planned. It currently contains 27 collocates, approximately 3000 words and 193 examples of *estar indo V*. The collocates are listed in table 10 below:

Table 10: BP collocates in the Twitter corpus

Activity verbs that often imply movement	Passive activity verbs that imply little or no movement	Semantically bleached verbs and mental verbs
<ol style="list-style-type: none"> 1. <i>andar</i> (to go) 2. <i>viajar</i> (to travel) 3. <i>pagar</i> (to pay) 4. <i>se casar</i> (to marry) 5. <i>trabalhar</i> (to work) 6. <i>comer</i> (to eat) 	<ol style="list-style-type: none"> 14. <i>estudar</i> (to study) 15. <i>dormir</i> (to sleep) 16. <i>ficar</i> (to stay, to be²⁸/become) 17. <i>namorar</i> (to date someone, or to be in a romantic relationship) 	<ol style="list-style-type: none"> 24. <i>tirar</i>²⁹ no sentido ‘passar tempo’ (to take away, in the sense ‘to pass time’) 25. <i>pegar</i> (to pick up, to catch³⁰)

²⁸ Often in the sense “to be with someone”.

²⁹ *Tirar* is another semantically bleached lexeme (see section 2.1.). According to *Priberam* (20.04.18), it can have 33 different interpretations including *to take back, to displace, to clear, to remove, to take away, to take off, to get out, to pull out*. *Tirar* is also the predicate in many idiomatic expressions.

³⁰ This study argues that *pegar* is a semantically bleached lexeme (see section 2.1. for more information on semantic bleaching). According to the Portuguese dictionary *Priberam* (20.04.18), it can have 21 different interpretations including *to grip, to hold, to grasp, to seize, to go ahead with something, to hit (transitive), to get good results, and much more*. In addition, *pegar* is the predicate in many idiomatic expressions.

7. comprar (to buy)	18. trocar (to change)	26. pensar (to think (about))
8. fazer (to do)	19. cortar (to cut)	27. ouvir (to hear, but also used in the sense 'to listen')
9. visitar (to visit)	20. assistir (to watch)	
10. ver (to see, in the sense 'to visit')	21. conhecer (to know, get to know, to see)	
11. furar (to pierce, to make a hole)	22. carregar (to charge, to download)	
12. trair (to cheat on)	23. carregar (to carry)	
13. roubar (to steal, rob, commit fraud, etc.)		

Before gathering data from Twitter, I did not expect to find many examples of *estar indo V*. As seen in table 8, there are only 41 instances of *estar*^{1STSG} *indo* + VIAJAR within the entire BYU corpora. However, I realized that there are infinite combinations of *estar indo V* on Twitter. Most of these verbs are action and/or activity verbs, but there are good examples of mental verbs and verbs that are in the grey area between passive activities and abstract mental states. See the following examples:

(26) *logo ele ta indo pensar em mim*
 'soon he's going to think about mim'

(27) *as pessoa ta indo ouvir musica*³¹
 'the people/the persons are going to hear music'

Although the phrase *ta indo ouvir musica* ('they are going to listen to music') in (27) seems to be an example of an unambiguous future auxiliary, the context of the phrase indicates ambiguity

³¹ The entire digital utterance or tweet was:

miley cyrus ja deu né amiga vamo volta a usa roupa normal nos shows as pessoa ta indo ouvir musica nao compra um vibrador

Directly translated to:

'miley cyrus enough already isn't it girlfriend let's go back to using normal clothes during the performances **the people is going to hear/listen to music** not buy a vibrator'

(see footnote 2). In other words, the people could be physically going to a place to listen to music, or some time in the future ‘the people are going to listen to music’.

The collocate *namorar* is interesting because it is often interpreted as an abstract state, such as “being in love” or “being in a romantic relationship,” which furthermore could indicate that *estar indo* is transitioning from occurring exclusively with activity verbs to occurring with mental verbs. ‘To go on a date’ is a physical activity, but ‘I’m dating (someone)’ is interpreted as an abstract state in both PDE and BP. In other words, it describes a person as “someone in (search of) a romantic relationship,” as seen in the following example:

(24) *falou tambem que daqui a pouco ta indo namorar...*

‘[He or she]said also that in a little while [he or she]is going to be dating ...’

To summarize, the collocates found in the Twitter corpus are more complex and diverse than initially assumed. I hoped to find more or less the same amount of collocates and data in BP that Traugott (2012) had found in EMdE, but I was astonished to find more than double the amount. Some of the collocates are very general in meaning, i.e. they have different meanings (and could have gone through some form of semantic bleaching). However, before discussing whether *estar indo* could be transitioning from occurring exclusively with activity verbs of physical movement, to occurring with mental verbs, it is necessary to consider the level of ambiguity and the linguistic contexts (see section 4.1.) in which *estar indo* appears. Therefore, the following section discusses the temporal adverbials with which *estar indo* occurs.

5.4. Temporal adverbials in the Twitter corpus

Including temporal adverbs to the search criteria ensures that the BP *estar indo* functions as a grammatical item and not a lexical item.

The temporal adverbials in the Twitter corpus are listed in the table below.

Table 11: Temporal adverbials

BP	PDE
<i>Daqui a pouco</i> [from here to a little]	'In a little while'
<i>Ja</i>	'soon, "extended now," already'
<i>Ja ja</i>	'soon, in a little while'
<i>((n)o mes que vem</i> [(in) the) month that comes]	'Next month'
<i>((n)a semana que vem</i> [(in) the) week that comes]	'Next week'
<i>Amanha</i>	'Tomorrow'
<i>Logo</i>	'Soon, quickly'

Without temporal adverbials it is much more difficult to determine whether the example of *estar indo V* is a regular future auxiliary, purposive future or ambiguous. The following examples of *estar indo V* in (29) and (30) are not accompanied by temporal adverbials, and they are ambiguous. In example (30), *estar indo V* occurs in a context with another verb that express *conditional mood*, or a hypothetical scenario, i.e. *queria* (*I want* or *I would like to*). The main verb is *pegar* in both (29) and (30), and although *pegar* is difficult to translate, it has been translated to *catch* or *hit* in both examples.

(29) *Bom dia pra quem ta indo pegar 16 hrs de trabalho*

[Good day to whom³² is going to hit³³ 16 hours of work]

'Good morning to whom are going to work for 16 hours'

³² Notice that "para quem" is not considered formal to the same extent as PDE "to whom".

³³ "catch 16 hours of work" or "hit 16 hours of work" are attempts at directly translating "pegar 16 horas de trabalho," but they are not typical PDE expressions. The phrases could, however, be compared to the frequent colloquial (Am)PDE expressions such as "to hit the beach" (go to the beach) or "to hit the sack" (go to bed).

The different bracketing of the phrase in (29), is illustrated in tables 12 and 13 below.

Table 12: bracketing A of (29)

Aux	Main verb	Direct object
Ta indo	Pegar	16 horas de trabalho
<i>Is going</i>	<i>Hit</i>	<i>16 hours of work</i>

Or:

Table 13: bracketing B of (29)

Main verb	Verb complement
Ta indo	Pegar 16 horas de trabalho
<i>Is going (or: is on their way)</i>	<i>To hit 16 hours of work</i>

The phrase in (30) could be an example of ambiguous *estar indo* V, but *tá* is not 3rd SING of *estar*, but a colloquially abbreviated form of the infinitive, i.e. *estar/to be*. In other words, (30) expresses conditional mood and the *estar indo* V construction occurs in a hypothetical context, thus it cannot be interpreted as a future auxiliary. At first glance the tense in (30) seems akin to *gerundismo*, but although *gerundismo* includes infinitive *estar/to be*, the main verb in *gerundismo* is conjugated in the *-ndo* progressive. In addition, the simple present tense of IR functions as a future auxiliary in *gerundismo*.

(30) *Queria tá indo pegar uma praia*

[I would like (past imperfective) to be going to catch a beach]

‘I would have liked (to be going) to hit the beach’

As illustrated by the examples above, temporal adverbials play an essential part in determining the level of ambiguity and the function of *estar indo* in the *estar indo* V phrase.

5.5. Key findings

After its initial grammaticalization in EMdE, *be going to* became fully grammaticalized, and in PDE it is used much more often as an auxiliary than a main verb. This is verified by the data in the BYU corpora, where the frequency rate for PDE AUX *be going to* is 0,98 compared to a frequency rate of 0,025 for PBP AUX *estar indo*. In other words, PDE AUX *be going to* is the most frequent use of *be going to*, while PBP AUX *estar indo* is not frequent at all, and in fact it is astonishing that there is evidence of its existence in the BYU corpora.

However, even though AUX *estar indo* is rare, it is productive! The amount of data in the Twitter corpus corroborates its existence in vernacular and oral/digital language, and the quantity of it exceeded every expectation. There is evidence of its productivity in the BYU corpora, which could suggest that it is gradually more accepted in written language. Before the Twitter corpus was made, it was assumed that BP AUX *estar indo* would only appear together with activity verb collocates, and it was astonishing to find examples of *estar indo* together with mental verb collocates, such as *to think* and *to hear/listen*.

The EMdE data contains 78 examples of *be going to*, of these, there are only 48 examples of AUX *be going to*. Comparatively, in the Twitter corpus, there are at least 150 instances of AUX *estar indo* out of 193 examples of *estar indo* in total. As explained in section 5.1., linguistic contexts have proven to be crucial for understanding the grammaticalization of EMdE *be going to*, and when EMdE *be going to* started to occur with mental verbs it was in a so-called *switch context* where reanalysis already had occurred (cf. Traugott, 2012, p.241). Although the occurrence is rare, the fact that BP *estar indo* occurs with mental verb collocates could indicate that BP *estar indo* is currently in the process of undergoing reanalysis. The example taken from the Twitter corpus below illustrates BP AUX *estar indo* with a mental verb collocate.

(31)

*Deixei Maria das Flores bolada.*³⁴ *Ela tá indo pensar na vida*
[I]left Maria das Flores indignant³⁵. She's going [to] think about life

'I left Maria das Flores in a bad mood. **She is going to think** about life.'

³⁴ The sentence is ambiguous and could either be interpreted as a) the speaker literally left Maria when she was angry, or b) The speaker said or did something (offensive) to Maria that made her angry.

³⁵ "Levar um bolada" is to be hit by a football in the face, and "bolada" is used colloquially in BP as an adjective meaning "angry, annoyed, tired, confused, disappointed, etc.".

There is nothing in the example above that indicates that Maria is physically going or moving with the purpose of “thinking”. In fact, it is (usually) not necessary to move in order “to think,” unlike e.g. “buying milk” or “fetching water” that are actions that usually imply that the subject has to move a certain distance in order to execute the action successfully. Although, the example of *estar indo* in (31) above could be used as an illustration that *estar indo* now occurs in switch contexts, more data is needed, and quantitative research would corroborate how much it has grammaticalized more accurately.

The examples (22) and (23), and other examples from the Twitter corpus show clearly that *estar indo* has developed from *normal use* (cf. Traugott, 2012), i.e. as a lexical item indicating physical movement in the present, to *untypical use*. The untypical use of *estar indo* is now evident in critical contexts, i.e. linguistic contexts where it is primarily ambiguous and in an initial stage of grammaticalization. Furthermore, the fact that AUX *estar indo* occurs with mental verb collocates that do not imply physical movement is another indication that *estar indo* has established itself properly in critical contexts and it may even transition to *switch contexts*.

To summarize, the data verifies that BP *estar indo* is in the process of grammaticalizing, but there is no evidence from the current data analysis that BP *estar indo* will become more frequent and grammaticalize to the same extent as English *be going to*.

Chapter 6: Discussion

The previous chapter emphasised that BP *estar indo* occurs in ambiguous and critical contexts, which according to Traugott (2012) enable grammaticalization. In addition, there are a few examples of BP *estar indo* in switch contexts, i.e. occurring with mental verb collocates. In other words, the data shows that BP *estar indo* is not just prone to grammaticalization, but that it is grammaticalizing. The aims of this discussion are to discuss whether the development and usage of *estar indo* is due to language contact with English, and whether it is likely that BP *estar indo* will undergo full grammaticalization, i.e. that it will grammaticalize to the same extent as PDE *be going to* has.

In the same manner as the data and methodology, this chapter is also organised in such a manner that first the development of English *be going to* and BP *estar indo* are discussed separately, and then the final section and chapter 7 summarizes and compares the differences and similarities. Section 6.1. Discusses the grammaticalization of English *be going to*, and how it is connected with the grammaticalization of the *v-ende/v-ing* periphrasis. Followed by a section about the possible Latin and Celtic influence on the grammaticalization of *v-ende/v-ing* progressive in OE.

Section 6.2. discusses the development of BP *estar indo*, the different future auxiliaries and future verb phrases in BP, and the phonetic erosion of *estar*, and section 6.3. discusses the role of English in Brazil, followed by a summary of the differences and the similarities between English *be going to* and BP *estar indo*.

6.1. Grammaticalization of English *be going to*: Why “be going to” and not “go”?

Traugott emphasises that “Bybee et al. (1994) and Heine and Kuteva (2002), among others, have shown that [the abstract schema ‘motion verb > future’] occurs in many languages worldwide” (2012, p. 240). To illustrate this abstract schema ‘motion verb > future,’ this study emphasises that present tense GO has grammaticalized into a future auxiliary in French, Spanish and Portuguese, but not English. BP and English are the only languages in which the entire progressive phrase *be going to/estar indo* functions as an auxiliary. The aim of this section is

to discuss why simple present tense GO did not suffice in English? Or rather, what triggered the grammaticalization of the entire present progressive phrase?

Jakobi (2004) mentions that language contact between French and Old English/Middle English may have influenced the grammaticalization of *be going to V*. Verb phrases with simple present tense GO as the conjugated auxiliary are labelled *periphrastic futures* in the present thesis (and for more information on periphrastic futures see section 6.2.2.). The grammaticalization of French ALLER, i.e. 'GO' is beyond the scope of this study. Nonetheless, language change is often attributed to language contact, and it is not unreasonable to attribute the semantic bleaching of English GO to French influence. This thesis argues, however, that the grammaticalization of *be going to* cannot be attributed entirely to French influence, because, as emphasised in the previous paragraph, only the simple present tense of French AUX ALLER grammaticalized, whereas the simple present tense of English GO did not grammaticalize into a function item at all.

Traugott (2012), on the other hand, does not emphasize French influence as triggering for *be going to*. Instead she focuses on the linguistic contexts in which grammaticalization of *be going to* takes place. She writes that "if ambiguous contexts occur and persist, as is often the case, they should be counted [as part of the grammaticalization process] since they are an essential part of the ecology of the generalization and further development of the new grammatical construction" (Traugott, 2012, p. 231). In other words, Traugott emphasizes that ambiguous contexts and the ambiguity of the *be going to V* phrase were essential to the grammaticalization of the *be going to*.

Although GO is the lexeme at the heart of the phrase (see e.g. Jakobi, 2004), the progressive form adds to the complexity of the phrase and its grammaticalization process. Since *be going to V* is a periphrastic progressive, I argue that it is tightly connected to the grammaticalization and increased frequency of *-ing* progressive. Hence, for a comprehensive discussion of the grammaticalization of *be going to* it is necessary to account for some of the processes behind the increased frequency and the secondary grammaticalization of *-ing* progressives.

Another way to look at it is that since the grammaticalization of *be going to V* coincided with the increased frequency of *-ing* progressives, it is likely that the grammaticalization and increased frequency of *-ing* periphrases and the grammaticalization of *be going to* reinforced one another, and thus these grammaticalizations are part of the same development. From a holistic viewpoint, the secondary grammaticalization of *-ing* progressives and the primary

grammaticalization of *be going to* could be features of the cyclical tendency towards increased periphrasis in English, which now in PDE could be switching back to increased cliticization that may lead to increased affixation in the future.

Be going to is an example of primary grammaticalization of a lexical item, but it is also strongly connected to the secondary grammaticalization of periphrastic *-ing* progressives. In other words, I argue that *-ing* progressives are especially prone to secondary grammaticalization in both English and BP (cf. Rocha, 2012), and this could be corroborated by the fact that present progressives are more likely to occur as futurates than verbs in the simple present tense. Compare the following examples:

(32)

Next Christmas I'm eating turkey with the rest of you.

(33)

Next Christmas I eat turkey with the rest of you.

The latter sentence comes across as awkward because the simple present occurs as futurate primarily in fixed time tables (e.g. “the bus leaves at 5pm”). As evidenced by Rocha’s thesis (2012) and Nesselhauf (2010), the FPP (Future Present Progressive) has undergone extensive secondary grammaticalization in both PDE and BP, and perhaps in a future study it may be relevant to investigate why *be going to* is categorised as primary grammaticalization and FPP is are categorised as secondary grammaticalization, when they both are *-ing* progressives that have continued to develop grammatical properties. The latter question is beyond the scope here, but the following sections investigate the grammaticalization of *-ing* progressives and whether they were influenced by language contact.

6.1.1. Grammaticalization of V-ende/ V-ing periphrasis

The English V-ende/V-ing periphrasis became visibly productive in EMdE, and the form originated in OE (see Killie 2014, among others), but it is difficult to account for how the grammaticalization started and the exact situations in which it grammaticalized. Did V-ende/V-

ing grammaticalize because of external factors, such as language contact, or, internal factors such as innovation or abductive reasoning³⁶?

Killie (2014a) argues that in OE *V-ende* was not a progressive marker, but an emphatic alternative to the simple tenses (Killie, 2014a, p. 328). She explains how the *-ing* progressive developed from OE *beon/wesan* + *V-ende* periphrasis, and the grammaticalization was evident in first EMdE. She emphasizes that even in PDE it is only the focalized progressive that has undergone obligatorification, since the durative progressive may easily be replaced by simple present or past tense (see the examples below).

(34) She trained for two hours yesterday.

(35) She was training for two hours yesterday.

Killie (2014a) also used data from the HC (Helsinki Corpora), and she concludes that the ME period was a critical context for the progressive, i.e. the type of linguistic environment that enables grammaticalization. In OE the *-ende* periphrasis is frequently compatible with a perfective reading, which could indicate that the grammaticalization into a progressive was due to Latin influence. She refers to previous research in which studies of translations Latin texts in OE. In these translations the *-ende* periphrasis seems to have been used as translations for different Latin periphrastic expressions (Killie, 2014a, p. 373). Her conclusions are commented upon in the following paragraph, which also discusses how contact between Celtic and Old English may have influenced the use of the *-ende/-ing* periphrasis.

6.1.2. Language contact between OE, Celtic and Latin

Several linguists have researched and discussed the role of language contact in the development of the progressive aspect and the *-ende/-ing* progressive in Old English (OE) and EMdE (see: Trudgill, 2011 and 2016b, Killie, 2014, Filppula, M., Klemola, J., & Paulasto, H. 2008). Filppula et.al. emphasizes that the English *progressive form* (PF), but not necessarily the *progressive aspect*, has been attributed to language contact with Celtic (2008, p. 59). In addition, Trudgill mentions in his conclusion that it has been argued that the progressive aspect was borrowed from Brittonic Celtic (2016b, p. 334).

³⁶ According to Hopper and Traugott, abductive reasoning is a type of guessing based on (assumed) universal principles and it is a key process in language learning (2003, p. 43).

Brittonic Celtic and Old British may have been substrata that influenced the grammar of spoken OE, but since these grammatical changes were not adopted in written language it is very difficult to verify this theory with absolute certainty (Trudgill, 2016b, p. 328). Languages such as French and Latin, on the other hand, were superstrate languages that presumably influenced the lexica of English more than the grammar (see e.g. Trudgill 2016b). In his article, Trudgill shows that the history of English gives evidence to the fact that language contact may lead to the transfer of additional grammatical categories and added complexity to the language structure (2016b, p. 334).

As mentioned in the previous section, Killie (2014a) discusses whether Latin's position as a dominant written language in OE and ME may have triggered the grammaticalization of *-ende/-ing* progressives. She argues, however, that it is more probable that the OE periphrasis developed its own meaning, which was different from the Latin periphrases (Killie, 2014a, p.374). Her conclusion is that if Latin influenced the construction, it is likely to have done so in a reinforcing manner, but it is unlikely that it triggered the development. The results of this study agree with Killie (2014a) that the *-ing* progressive cannot be attributed to Latin influence alone. Still, there are plenty of other examples of Latin influence on English, as e.g. the relatively high number of lexemes derived from Latin in PDE.

English is often labelled the “modern Latin,” i.e. the global lingua franca of our era (cf. Crystal, 2003), and the high amount of English borrowings in BP, could indicate that English influences BP in the same manner that Latin once influenced OE and ME. On the one hand, global English have such a large influence on BP that it triggers grammaticalization of *-ndo* progressives, but on the other hand, it could be the same scenario as with Latin, that iff³⁷ English influences the grammaticalization of BP verb phrases, it is more likely to do so in a reinforcing manner. Section 6.3. will elaborate upon the role of English in the BP language landscape, and the questions mentioned in the previous paragraph will be taken into consideration there.

However, before addressing to the topic of language contact between English and BP, the study will discuss recent quantitative research that present an alternative perspective on the grammaticalization of English *be going to* (section 6.1.3.). So far, this thesis has considered the *early* grammaticalization of *be going to*, but it has not considered research on the *full grammaticalization* of *be going to*. Recent studies of how *be going to* was fully grammaticalized may shed light on whether it is likely that BP *estar indo* will undergo the same changes and

³⁷ i.e. if, and only if.

develop similar grammatical properties to that of PDE *be going to*. Hence, the next section summarizes and discusses these new perspectives on *be going to*, which also critique some of the diachronic and qualitative corpora studies presented in chapter 4.

6.1.3. Different perspectives on the grammaticalization of *be going to*

Wu, et.al. (2016) critique earlier research on *be going to* for being more qualitative than quantitative. Their article is entitled “Rethinking the Grammaticalization of Future *be going to*,” and they conclude that *be going to* should be categorized as a modal verb instead of a future auxiliary (Wu, et.al., 2016).

When considering earlier research on *be going to* they refer to much of the literature that has been used in the present study, e.g. Bybee and Pagliuca (1987), Traugott (2003), Poplack and Tagliamonte (1999), among others. They comment that the earlier research, which is the basis for much of this thesis, focuses on the initial grammaticalization of *be going to* in Middle English (ME) and EMdE, and that quantitative studies are necessary for a reliable and comprehensive understanding of *be going to*. In other words, Wu et.al. (2016) are concerned with the *full grammaticalization* of *be going to*, and the *full grammaticalization* of *be going to* did not happen until the late 17th century (Wu et.al., 2016).

As mentioned, the purpose of this section is to discuss the arguments by Wu et.al., and to consider the processes in which *be going to* became fully grammaticalized. Furthermore, the research by Wu, et.al. are compared to the findings in the Twitter corpus, in order to discuss whether it is likely that BP *estar indo* will continue to grammaticalize to the same extent as English *be going to* has. According to Wu, et.al. (2016, p. 317) the steps that lead to its full grammaticalization are:

- the application of mental verbs.
- the appearance of non-animate subjects
- the use of passive voice

In the Twitter corpus, there are examples of BP *estar indo* occurring with mental verbs, but the occurrence is so rare that it can hardly be used as evidence to argue that BP *estar indo* will become fully grammaticalized. Furthermore, there are no evidence of *estar indo* occurring with non-animate subjects, nor in the passive voice.

As seen in chapter 4, the present thesis uses corpora from BYU to investigate the use of BP *estar indo* and compare it to PDE *be going to*. For the purpose of this thesis, which has aimed to compare early stages of grammaticalization of *be going to* in English and BP, the HC and POB have been more suitable for describing the diachrony of the earlier stages of *be going to*. Therefore, the historical corpora of the BYU have not been used. Instead, the NoW (News on the Web) corpus and *o corpus do Portugues* have been used to compare *be going to/estar indo* in PDE and PBP. The main reason for choosing NoW and not Google books, which is a much larger corpus with 45 billion words (compared to NoW with “only” 14 billion), is that NoW provides context information that is not quite as easily available in the Google books corpus.

Similar to the methodology in the present thesis, Wu et.al. (2016) also use corpora from BYU, i.e. they use COHA (Corpus of Historical American English) and the Google books corpus. They use the COHA for research and verification of the full grammaticalization of the construction in the latter part of the 1800s, and the Google books corpora for investigating the properties of the construction after it was fully grammaticalized.

Another important issue, which has been elaborated upon in chapters 4 and 5, is that grammaticalization occurs in casual speech and oral language before it occurs in written language. The BYU corpora are excellent for studying and verifying linguistic phenomena in written language, but less useful for studying oral language.

Apparently, *reanalysis* and *analogy* are among the key mechanisms that enabled the full grammaticalization of *be going to* (cf. Wu, et.al., 2016). Unfortunately, I did not account for how reanalysis and analogy operate along different axes in the general introduction to grammaticalization (chapter 2). According to Wu, et.al. a basic understanding for how these mechanisms operate is essential for understanding the grammaticalization of *be going to*, therefore, these mechanisms more thoroughly in the next paragraph.

When discussing the stage in which *be going to* started to occur with mental verbs Wu et.al (2016) refer to Hopper and Traugott’s schema of the development of AUX *be going to*. The revised schema is presented in the figure below (Hopper and Traugott, 2003, p. 93):

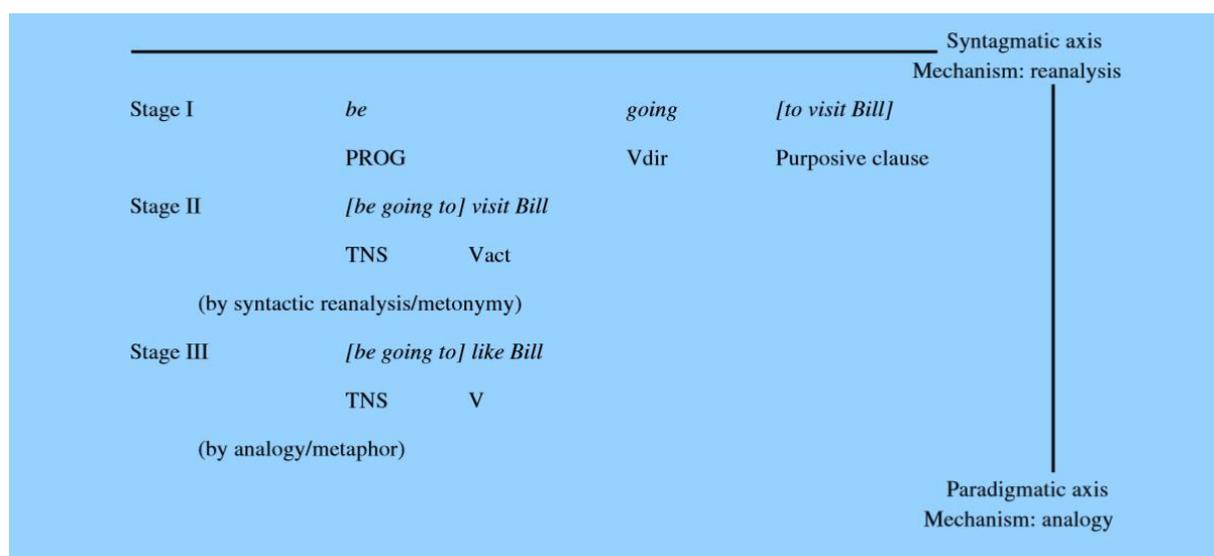


Figure 4: Hopper and Traugott's revised schema of the development AUX "be going to" (2003, p.93).

As seen in the schema, analogy operates along the paradigmatic axis, while reanalysis operates along the syntagmatic axis. Reanalysis, which involves syntactic rebracketing as in step II (also see section 2.3 for a more thorough explanation of rebracketing), is a prerequisite for change through analogy (Hopper and Traugott, 2003, p. 39). The reanalysis in steps I and II are not observable and they do not include morphological or semantic changes, but in step three metaphorical analogizing has enabled morpho-syntactic change and a change in meaning, which enables the occurrence of AUX *be going to* with a mental verb collocate.

English *be going to* has reached the end of the line, both along the syntagmatic and the paradigmatic axis. The same cannot be said about BP *estar indo*, which only recently has started to develop along the syntagmatic axis. Since BP *estar indo* is in the very early stage of grammaticalization, it is necessary to ask what Wu, et.al. (2016) consider the starting point of the grammaticalization of *be going to*.

According to Wu et.al., "*be going to* + place noun" can be considered "the starting point for the research of the grammaticalization" (2016, p. 323). Although it has not been stated explicitly, this is also the viewpoint of the present thesis. In fact, it has taken it as given that *be going to* + place noun is the starting point, but in addition, this study emphasizes that the lexical phrase often continues to co-exist together with the grammatical phrase. In other words, even though a grammatical item is not obligatorified, the item may still have undergone grammaticalization (cf. Hopper and Traugott, 2003, p. 32).

Wu et.al. (2016) do not elaborate upon whether they regard obligatorification as an absolute criterion for full grammaticalized, but that is not necessarily relevant for discussing the grammaticalization of BP *estar indo*, which currently shows no sign of becoming obligatory. On the other hand, there are several questions raised by Wu et.al. that are relevant for this thesis. For instance, if *be going to* is a modal verb, i.e. it expresses modality rather than future tense, could *estar indo* too belong to the category “modal verbs”?

The previous sections argued that *-ing* is closely tied to *be going to*. Since the *-ende/-ing* periphrasis has been labelled a stylistic device (Kranich, 2010), or an emphatic marker in OE and ME (Killie 2014a), could it be that *be going to/estar indo* also are stylistic devices, in this sense that they convey mood rather than aspect?

To summarise, evidence from the BYU corpora show that *be going to* as a lexical item has not disappeared, but as the grammatical construction becomes more frequent, the lexical phrase decreases (cf. table 3 in Ch. 4, and see Wu et.al., 2016). Furthermore, Wu et.al. (2016) highlight that *be going to* expressed motion, direction and intention even when it appeared with mental verbs, such as “to hear” in the 1800s. Could this imply that even though there are examples of BP *estar indo* occurring with mental verbs in the Twitter corpus, it still has not made the transition from critical (i.e. ambiguous) contexts to switch contexts? In other words, it is probably still too early to tell whether *estar indo* has undergone reanalysis.

6.2. Grammaticalization of BP *estar indo*

Since historical processes are always ongoing and furthermore are not deterministic (i.e., they can always be abandoned) we will often not see a complete instance of grammaticalization, only one in progress, (...)

Hopper and Traugott, 2003, p. 32.

The quote above highlights that many instances of grammaticalization are in progress and have been so for centuries, which imply that BP *estar indo* continue to stay in progress instead of grammaticalizing completely. From the data analysed in chapters 4 and 5, we have learned that BP AUX *estar indo* is productive and it even occurs with mental verbs, but this does not mean that it has been reanalysed as a future auxiliary. The research by Wu, et.al. questions the categorization of *be going to* as a future marker and argue instead that *be going to* is a modal verb. Consequently, the present thesis opens up for the possibility that BP *estar indo V*, could be a modal verb as well.

As reflected in the quote above, many items will not grammaticalize completely, and modal verbs are known to be particularly persistent, in the sense that they keep, or at least reflect, their original lexical meaning, sometimes for thousands of years (Traugott and Hopper, 2003, p. 97). Modal verbs are also known for being polysemous, as for instance, the modal auxiliary *will*, which could encode willingness, intention and/or “pure” future (Traugott and Hopper, 2003, p. 97).

In most cases the BP *estar indo* phrase is ambiguous, or even polysemous, in the sense that it can be interpreted as “physically going with the intention of doing something,” or simply expressing “intention,” or that “something is very likely to happen in the future”. The ambiguity of the phrase makes it difficult to determine whether it has undergone reanalysis, but ambiguity has been enabling for the grammaticalization of *be going to* (cf. Traugott, 2012). I argue that when *estar indo* occurs with temporal adverbs such as *amanha* i.e. ‘tomorrow,’ it may have been rebracketed as a future marker.

Whether *estar indo* is grammaticalizing into a modal auxiliary or a future auxiliary, the conclusion to the thesis question is still ‘yes, BP *estar indo* is grammaticalizing’. The other question that remains unanswered is whether BP *estar indo* is a borrowed construction from English. Thus, the next sections will focus on whether the grammaticalization is triggered by innovative forces in BP, or outer influence such as contact with English.

As with most language change, *estar indo* is probably grammaticalizing due to both innovation and language contact. Zilles (2007) comments that the grammaticalization of BP nouns into pronouns is triggered by innovative and sociolinguistic factors within Brazil. She emphasizes, furthermore, that there has been a reorganisation of BP grammar, “[a]s Brazilian Portuguese takes its own course away from European Portuguese, several morphosyntactic changes are in progress, affecting both the paradigm of personal pronouns as well as subject-verb agreement” (Zilles, 2007, p. 16).

A reorganisation of the paradigm of personal pronouns carry massive implications for the paradigms of BP verb phrases, and I argue that the use of *estar indo* is related both to the increased frequency of *-ndo* progressives and the grammatical reorganisation of Brazilian Portuguese. In order to highlight, these innovative and sociolinguistic factors, the following section discusses how the phonetic erosion of *estar* (i.e. ‘to be’) is part of several recent and extensive language changes, and I argue the grammaticalization of *estar indo* is founded in these changes.

6.2.1. Processes of change in BP and the phonetic erosion of *ESTAR*.

[In BP,] there are a number of linguistic processes going on at the same time (...) [and] [t]hese include the long-term syntactic drift away from agreement marking and towards use of overt, preverbal subjects, (...), the semantic shift in the meaning of the form; the phonological reduction of the form; and so on.

Zilles, 2007, p. 51

As seen in the quote above, phonetic reduction of forms, of which the phonetic reduction of *estar* is an example, is part of a bigger process of language change in BP. This thesis argues that this bigger process of change in BP include extensive grammaticalization of subject pronouns, such as *a gente*/'the people' > 1stPL/'we', and secondary grammaticalization progressives, and an overall shift from affixation towards periphrasis.

According to the hypothesis of unidirectionality, phonetic erosion is one of the mechanisms in grammaticalization that take place in the last stages, i.e. after reanalysis and rebracketing (cf. Hopper and Traugott, 2003, Trudgill, 2016a). This section argues, however, that the phonetic erosion of BP *estar* (i.e. *be*), which is the auxiliary in progressive verb phrases, has occurred *before* reanalysis and rebracketing of *estar indo*. This is corroborated by data from the BYU corpora, which is illustrated and elaborated upon in the following paragraphs. Peter Trudgill (2016a) also argues that phonetic erosion can be a trigger to grammaticalization. By referring to examples of grammaticalization in the East Anglian dialect he questions the hypothesis of unidirectionality, which only puts phonetic erosion and cliticization *after* reanalysis and rebracketing (Trudgill, 2016a).

In order to explain the phonetic erosion of *estar*, and how it influences BP progressives, this section gives a short preliminary outline of the semantic content and grammatical properties of *estar*. These aspects are important for understanding the BP *-ndo* periphrasis and *estar indo*. Unfortunately, *estar* is much more complex than English 'to be,' and there is a mountain on literature on the semantic content, morpho-syntactic and grammatical properties of *estar*, thus the present section will only focus on those properties that are relevant to the thesis.

There are two BP verbs that can be translated into PDE *to be*. As in Spanish, BP *ser* and *estar*, both mean *to be*, but they carry different semantic content. In short (and in danger of

oversimplifying), *ser* usually expresses that a state is permanent, but *estar* expresses that a state is transient or it is used in reference to location (see example beneath). Traditionally BP *estar* is also used when the location is permanent³⁸. *Ser*, on the other hand, cannot be used to express location.

(36) *A foca está no banheiro.*

‘The seal is in the bathroom.’

Zagona mentions that both *ser* and *estar* are incompatible with the progressive aspect, i.e. the verbs cannot occur as main verb collocates in BP progressive constructions, and she argues that only AUX *estar* occurs in progressive phrases (2012, p. 302). Bastos, however, argues that the *-ndo* participle of *ser* indeed is possible as the main verb in BP progressives, as seen in the example below.

(37) *Voce tá sendo chata!*

‘You are being annoying³⁹!’

The use of *ser* in the previous example is controversial. *Ser* would traditionally indicate that the subject complement is an intrinsic quality of the subject, but a pragmatic reading of “annoying” in this instance, would lead us to interpret that it is not an intrinsic characteristic of the subject.

Furthermore, in the instance above, “being” is an example of *being/sendo* functioning as a focalised progressive (for more information on focalised progressives see section 5 on the development of the V-*ing* periphrasis). Thus, the compliment in (34) is not an intrinsic description of the subject, but rather an expression of how the behaviour of the subject frustrates the speaker at the moment of the utterance. Indeed, the only main verb *-ndo* participle (i.e. BP *-ing* participle) that is not compatible with the progressive is *estar*. Not only is *estar* the auxiliary in the periphrasis, but *estar* also encodes the progressive aspect in all its tenses (cf. Bastos, 2004, among others).

³⁸ In Spanish, *estar* is still the most frequent verb used for describing permanent locations, but in many situations in BP, the semantically bleached verb *ficar* (i.e. *to stay*) has become more frequent than *estar*. E.g. “Londres fica no sul da Inglaterra” = “London stays/is in the south of England”. Thus, Spanish and BP are developing along different paths in a range of areas.

³⁹ Translation of *chata* from Presbiram, retrieved from <https://www.priberam.pt/dlpo/Traduzir/EN/chata> 30.11.17.

Table 14 illustrates how the abbreviated simple present tense varieties of *estar* have become considerably more frequent since they first occurred in the 1800s (possibly even earlier in oral BP). The corpus of 45 million words has data from the 1300s to the 1900s.

Table 14: Phonetic erosion of *ESTAR*: Data from BYU showing the frequency of *tô*, *tá* and *tamo*

	New BYU corpus (1billion): web/dialects, (in this case from the 1990s-2010s)	Old BYU corpus (45million): in genre/historical registers. In this case from the 1800s and 1900s)
Eroded/abbreviated <i>ESTAR</i> (<i>to be</i>).	Frequency, i.e. number of total tokens.	Frequency, i.e. number of total tokens.
Pronoun paradigm		
Tô I'm	21,743	209
Tá You're (singular) He's/she'a/it's [We] are	102,216	1783
Tamo We're	549	6

In the *Old Corpus*⁴⁰, the instances of abbreviated *estar* from the 1800s and 1900s appear exclusively in BP literature. Furthermore, the instances of abbreviated *estar* from the 1800s are usually from play scripts by Artur Nabantino Gonçalves de Azevedo (1855-1908) Manuel de Oliveira Paiva (1861-1892), in other words, literature that simulate oral language. Azevedo and Oliveira Paiva are also known as authors of Realism/Naturalism from the Northern parts of Brazil, which are characterised by rural, remote and indigent areas.

In addition, most of the tokens appearing from the 1900s were from linguistic studies of oral language in Recife, also Northern Brazil, in the late 1990s. In other words, phonetically reduced *estar* was and still is a feature of spoken BP. Judging from the data, one can assume that the phonetic reduced *estar* traditionally was associated with working class in the northern areas of Brazil, but in contemporary BP the reduced *estar* occurs in oral language all over Brazil and in every layer of Brazilian society.

The majority of instances of reduced *estar* in the New Corpus⁴¹ from the 2010s is also from Brazil. An important reason for the overrepresentation from Brazil, could be that Brazil has the

⁴⁰ BYU, o corpus do português, genre/historical 45 million words, created 2004-2006. Retrieved from <https://www.corpusdoportugues.org/> (31.10.2017).

⁴¹ BYU, o corpus do português, web/dialectm 1 billion words, created 2015-2016. Retrieved from <https://www.corpusdoportugues.org/web-dial/> (31.10.2017).

highest number of speakers, but in addition, progressive verb phrases have grammaticalized in BP, but not in European Portuguese. Thus, *estar* as an auxiliary is much more frequent in Brazil. Nonetheless, reduced *estar* also occurs in Mozambique, Angola and Brazil. Still, one could easily argue that this feature has spread from Brazil to the other countries through Brazilian music media.

Based on Trudgill's research (2016a) and the findings accounted for in the table above, this study argues that the phonetic erosion of *estar* could have triggered the grammaticalization of BP progressives, including *estar indo*. Both Zilles (2007) and Trudgill (2016) argue that sociolinguistics offer both implications and solutions for problem areas in grammaticalization, such as the hypothesis of unidirectionality. The stages in grammaticalization could not be perceived as synchronised from a sociolinguistic perspective (Zilles 2007, p.16). Similarly, Trudgill (2016) suggests that the sequence of the stages proposed by Heine and the cline proposed by Hopper and Traugott is not fixed, but relative to the instance of grammaticalization.

To summarize the sociolinguistic perspective on language change and grammaticalization reflected in the research by Trudgill (2016) and Zilles (2007) present interesting implications for the grammaticalization of *estar indo*, and the study concludes that the sociolinguistic conditions within Brazil could be more important than the external influence from English, and this is investigated further in chapter 6.3.

6.2.2 Future verb phrases in Brazilian Portuguese and grammaticalization of IR.

As mentioned, BP verb phrases are becoming more and more periphrastic (cf. Zilles 2007 and Rocha 2012), and PDE is a language with even more periphrasis than BP, and there are no morphological future verb phrases in English, in any era (at least not to my knowledge).

An issue that could affect the grammaticalization of *estar indo* negatively, is that IR, from which the participle *indo* is derived, already functions as a future auxiliary. Thus, *estar indo* is in a way completely redundant as a future marker in BP. This of course was not the case when *be going to* started to grammaticalize in EMdE (cf. Traugott 2012), which according to Jakobi (2004) initiated in the intersection between increased frequency of *-ing* progressives and contact with French and the grammatical borrowing of AUX ALLER (GO). To highlight the differences between English and BP future tense expressions, this section describes how BP future expressions are used.

The examples below illustrate the differences between future verb phrases in PDE, BP and Spanish. Spanish examples are included to highlight the grammaticalized items and the changes in BP. The examples also serve to illustrate the grammatical reorganisation in BP, which has been commented on by Zilles (2007). As mentioned in the introduction there are many different future verb phrases and periphrastic expression relating to the future, but for reasons of space only a limited selection is included in this study.

The following three examples show the three most frequent future verb phrases in BP (cf. Rocha, 2012). The BP sentence in example (38) is two words shorter than the Spanish sentence, due to the omission of the preposition *a*, which in this case would be translated into English *to*, and the contraction of the preposition *no* and the masculine definite article *o*, in BP. This is an example of how BP is more prone to phonetic erosion and contractions than Spanish.

Periphrastic future:

(38) a. BP: *Vou pensar no caso*

[go⁴² [to] think onthe case]

‘I’ll think about the case’

b. Spanish: *Voy a pensar en el caso*

[go¹⁹ to think on the case]

‘I’ll think about case’

Simple Present Future

(39) a. BP: *Eu faço amanhã*

[I do [Ø-it] tomorrow]

‘I’ll do it tomorrow’

b. Spanish: *Lo hago mañana*

[It [I] ¹⁹do tomorrow]

‘I’ll do it tomorrow’

Morphological Future, i.e. future verb phrase in which both pronoun and tense are encoded through the inflectional morphology (this verb construction does not exist in English).

(40) a. BP: *Fara sol*

⁴² First person singular encoded through inflectional morphology

[Do(future inflectional morphology) sun]

‘It will be sunny’

b. Spanish: Hara sol

[Do(future inflectional morphology) sun]

‘It will be sunny’

In section 6.1., examples (33) and (33) show that the Future Simple Present is very rare in English, but it is very frequent in BP and Spanish⁴³. According to Orozco (2005), Future Simple Present accounts for c. 35% of all future tense expressions in Columbian Spanish, and according to Rocha the periphrastic future accounts for more than 80% of all future tense expressions in BP.

Example (43) illustrates the Future Present Progressive, which according to Rocha (2012) primarily exists in BP, but the present thesis assumes that it also exists in Vernacular Latin American Spanish⁴⁴. Examples (41) and (42) illustrate *o gerundismo*, which uses IR/GO in the same manner as the modal auxiliary ‘will’ is used in English.

(41) Gerundismo

Eu vou estar trabalhando a semana toda.

[I go to be working the whole week.]

‘I will be working the whole week.’

Considering that both *gerundismo* and *estar indo* include the periphrastic progressive, IR and *estar* they seem very similar, but as mentioned in chapter 1, there are some differences. Firstly, *estar* is not conjugated, i.e. it remains in the infinitive, in gerundismo, and the use of gerundismo is controversial and at best only considered accurate in situations where an action compatible with the progressive aspect explicitly occurs at the same time as another action in the future. For instance,

⁴³ As mentioned in 6.1. the English Futurate Simple Present is used when referring to fixed timetables, e.g. ‘the train leaves at five,’ but that is about it.

⁴⁴ In the BYO “Corpus del Español,” the present progressive appears with the collocate “ahorita,” which is a temporal adverb meaning “soon” in Mexican Spanish. See: 2G.MXrevistaesnob.com “(...) pero el lema de esta generación, la que **está llegando ahorita, (...).**” i.e. “(...) but the motto of this generation, the one that is coming soon (...)” (my translation).

(42)

Quando você tá dormindo a noite, eu vou estar pensando em voce.

‘When you are sleeping tonight, I’ll be thinking of you’

Traditional grammarians have paid little attention to the use of *estar indo*, but *gerundismo* is denounced harshly by BP prescriptive grammarians (cf. Rocha, 2012, p. 10). Wikipedia even classifies *gerundismo* as a sub-category of “Vícios de linguagem,” or “Language Vices”⁴⁵. Other sub-categories of “Vícios de linguagem,” include “barbarismos” and “vulgarismos” (i.e. barbarisms and vulgarisms).

The term “barbarism” evokes connotations to colonialism and linguicism. It is well-known that Indigenous languages in Brazil have suffered and many have become extinct because of European colonialism. Still, dialectal features of BP are labelled “barbaric” and “vulgar language vices”. This linguistic prejudice is rooted in eurocentrism and colonialism that has established BP as an “inferior and illegitimate variety of Portuguese and European Portuguese as the ‘correct’ variety” (Massini-Cagliari, 2005, p. 19). Rocha (2012), however, argues that *gerundismo* and other future verb phrase expand the repertoire of verbal expressions, and the Future Present Progressive she researched is illustrated in the following example:

Future Present Progressive

(43) a. BP: Jaja tá chegando

[Soon is⁴⁶ coming]

‘He’s coming soon’

b. Spanish: Ahorita está llegando

[Soon is coming]

‘He’s coming soon’

To summarize, the Periphrastic Future, is by far the most frequent in BP and Spanish, followed by Simple Present Tense Future, and the use of the morphological future is decreasing (see Orozco 2007, 2014, Rocha 2012). The simple present tense of GO grammaticalized into a future marker in French, Spanish and Portuguese, but interestingly the entire periphrastic progressive

⁴⁵ «Vício de linguagem», Wikipedia: https://pt.wikipedia.org/wiki/V%C3%ADcio_de_linguagem#Gerundismo
Retrieved 25. 11.17.

⁴⁶ Third person singular encoded through inflectional morphology

verb phrase *be going to + infinitive* grammaticalized in English, but the simple present tense of GO did not.

There is a general tendency towards more periphrasis in both Spanish and BP (cf. Cacoullos, 2000, Rocha, 2012, Orozco, 2014, and Zilles, 2007). Furthermore, I argue that the cluster of grammatical changes in BP grammar (cf. Zilles 2007) makes BP move even further away from affixation than Spanish. As mentioned earlier, Cacoullos (2000) describes how the increased frequency and grammaticalization of Spanish *-ndo* periphrasis is due to contact between Spanish and English in the Americas, and it is especially influenced by the Spanish-English bilingualism in the USA.

From analysing the Twitter corpus and the examples above I conclude that BP *estar indo* is similar to the early use of *be going to* in EMdE (cf. Traugott 2012) but considering the differences between the future verb phrases in PDE and BP, and the low frequency of *estar indo* compared with other future verb phrases, I argue that it is probable that BP *estar indo* develops in a different direction than the PDE *be going to*. In order to further investigate how English could influence *estar indo*, the following section discusses contact between English and BP.

6.3. English in Brazil

(...): the linguistic consequences of language contact can vary enormously depending on the particular sociolinguistic conditions in which it takes place (Trudgill 2011b).

Trudgill, 2016b, p. 318

As reflected in the quote above, the sociolinguistic conditions within Brazil may have a bigger influence on the grammaticalization of *estarindo* than contact with English. The present section investigates the contact between English and Brazil and why it is assumed that grammaticalizing BP progressive constructions are imported from English (see Rocha, 2012). In addition, it explores some of the sociolinguistic conditions surrounding BP *estar indo*.

BP is heavily influenced by language contact in general. On the one hand, the majority of Brazilians are monolingual speakers of Portuguese, but on the other hand there are still 200 indigenous languages, and there used to be 1500 before European colonialization (see Montes, 2016). Especially, Tupi-Guarani languages have contributed to the BP lexicon (see Rocha, 2012), and is visible in place names such as Taubaté, Curitiba and Parati.

There are also many lexemes from Bantu and Yoruba languages. Brazil has the largest African population outside of Africa, and there are approximately four times more African Americans in Brazil than in the USA (cf. Hillman and D'Agostini, 2013). In addition, the American Indigenous population is still in the process of making contact with the Brazilian society and there is a rising awareness of the need to protect marginalised minority languages. In addition, languages of immigrants such as Japanese, Arabic, German, French, Dutch, Italian, and more have influenced BP (see Massini-Cagliari, 2003, among others).

Brazil is a multi-ethnic and multi-cultural country. Many Brazilians are speakers of other languages than Portuguese, and the search for identity and belonging in the multi-cultural Brazilian society could give English the status of a neutral lingua franca that does not remind people of the conflict, shame and stigma associated with Portuguese as the superior or dominant language. European Portuguese has often been associated with the colonialism and fascism that has marked Brazil's history until the late 1980s, and this could trigger language change that makes BP develop away from European Portuguese.

In the history of Brazil there are many examples of contact between English speakers and Portuguese speakers due to trade, but there are no instances of extreme contact between Portuguese and English due to British or American colonisation or invasion. Thus, the contact with English in Brazil has been less extreme than for example in India (British colony), or Hawaii (American annexation).

For reasons of space the Brazilian diaspora and Brazilian migration to English-speaking countries such as the USA, the UK and Australia is not discussed, although this undoubtedly is an important part of the contact between PDE and BP. What is taken into consideration, however, is the global spread of English through media, ICT, academia, and the increasing need for English proficiency due to the status of English as a global lingua franca (cf. Crystal, 2003). In the current era of globalization, contact with English within any given nation, is practically inevitable, and Kachru (2006) has made a model for the global spread of English.

Kachru (2006) divides global Englishes and English proficiency within nations in three circles:

- i. the inner circle, typically nations such as the US and the UK,
- ii. the expanding circle, e.g. India and other former British colonies
- iii. the outer circle, typically nations with little direct contact with L1 (first language) English-speakers and low English language proficiency.

Brazil is promptly placed in the outer circle (see Kachru, 2006), which implies that there is limited English proficiency within the Brazilian population.

Montes (2016) explains that the initial contact between Portuguese and English in Brazil began with the slave trade in 1530 and was later sustained by English interests in Brazilian timber. Nonetheless, the only foreign languages that were taught in Brazil until 1809 were classic Latin and Greek (see Montes, 2016, p. 6), and for a brief period after Brazilian independence in 1889, all European languages, except Portuguese, were excluded from the curriculum (Montes, 2016, p. 7). For most of the 1900s, French was considered the language of prestige and therefore the Brazilian elite prioritised learning French rather than English until the aftermath of World War II.

After World War II, English has gained a different status in the Brazilian society. English proficiency is a prerequisite for higher education, but both higher education and English proficiency is only accessible to those who have completed secondary education and English tutoring in expensive private schools. In addition, the use of English in Brazil is restricted through national policies in order to preserve and protect BP from English loanwords and other influence (cf. Tsuda, 2010) and it is severely deprioritised in public schools (Montes, 2016).

Despite English proficiency being regarded as a commodity only available to the elite, English has had a considerable impact on the BP lexicon (Montes, 2016). Words, such as *shopping, kit, love, list, design, short, tablet, online, site* are frequent in colloquial BP (Montes, 2016, p. 11). Many English loanwords, and most of the ones mentioned here, have gone through both semantical changes and morphological changes. In addition, “Brazilian English” is developing its own phonological and morpho-syntactic characteristics (cf. Montes, 2016, p. 14).

English loanwords are so excessively used and modified that several Brazilian songs use irony to criticise the use of English by the Brazilian middle class (Montes, 2016, p. 27). By referring to Kachru, Montes emphasise that English in Brazil have instrumental, interpersonal, regulative, and innovative functions (2016, p. 17). The innovative functions of English in Brazil could have implications for the thesis, in the sense that English linguistic features, such as lexemes and syntax that make BP resemble English, have been internalised as a way of creating innovation and change in BP.

English has become the dominant foreign language (henceforth: FL), both in schools and in the Brazilian society in general. Nonetheless, FLs in general are still severely deprioritised in the public-school system. The argument for deprioritising English is that literacy is more important

than FL learning (Montes, 2016, p.9). Brazil has also experienced a severe shortage in qualified FL teachers, and this makes English competence a commodity only available through expensive private schools, and thus English language skills are associated with the rich elite.

To summarise, BP is influenced by English, especially through media and other forms of indirect communication in the era after World War II. *Estar indo* could be a grammatical loan from English, but this thesis has not investigated how contact with other languages may have influenced the grammaticalization of *-ndo* progressives and the use of *estar indo*. It is possible that other languages use periphrastic progressives even more frequently than English. Crystal (2003, amongst others) have emphasised that *-ing* progressives are used more frequently in Indian English than in other Englishes, and that could be attributed to multilingualism influencing English in India. In Brazil, however, Portuguese is the superstrate and African and American indigenous languages are substrates that, together with English, have influenced the BP lexicon and the grammar.

6.4. Innovation versus language contact in the grammaticalization of English *be going to* and BP *estar indo*.

Jakobi argues that the French grammaticalization of GO into a future marker is comparable with the *be going to* construction, and he writes that “(...) GO is the heart of the *be going to* construction and the component that extended its meaning from a purely lexical (...) one to grammatical meaning (...)” (Jakobi 2004, pp.8-9). This thesis emphasises that although, GO is the only lexical word in the *be going to* phrase it is important to acknowledge that, contrary to French ALLER, BP and Spanish IR, it was the entire English *-ing* periphrasis that grammaticalized into an auxiliary.

Increased frequency and grammaticalization of Spanish and BP progressive *-ndo* phrases have been attributed to language contact with English (cf. Rocha 2012, Cacoullos 2000). However, there is no indication that Spanish *estar yendo* (i.e. *be going to*) is in the process of grammaticalizing, and the Spanish-speaking population of more than 50 million vastly outnumber the BP population in the USA⁴⁷, thus the hypothesis about language contact between

⁴⁷ US now has more Spanish-speakers than Spain – only Mexico has more. (30 June 2015). *The Guardian*. Retrieved from: <https://www.theguardian.com/us-news/2015/jun/29/us-second-biggest-spanish-speaking-country> (03.10.2017)

BP and English as the main cause of grammaticalization of *-ndo* periphrases and *estar indo* requires further investigation. As mentioned in section 3.2., one explanation could be that the BP *estar indo* is more prone to grammaticalization because *-ndo* progressives are more frequent in BP than in Spanish and EP (cf. Rocha 2012). As seen from the comparison of EMdE data and BP data in chapter 5, the development of BP *estar indo* and EMdE *be going to* share some similarities, as e.g. occurring in the same ambiguous linguistic contexts and with similar collocates, but this could be due to innovation in the two languages.

BP has been and still is influenced by the contact with a vast variety of languages from America, Europe, Africa and Asia. The same can be said about English. Both English and BP could be prone to grammaticalization due to their history of contact with other languages. Scholars have even hypothesised that Middle English could have been a creole emerging from contact between Old Norse, Old English and French (see Dawson 2003, among others). However, there are also innovative forces that shape English and BP. Trudgill argues that some dialects may be more prone to language change phenomena, such as phonetic erosion, than others, and in a more elaborate research project it would be interesting to study whether BP possess some of those qualities mentioned by Trudgill.

6.5. A brief summary of some differences between BP *estar indo* and English *be going to*

Since PDE *be going to* has undergone obligatorification and paradigmatisation it fits a narrow definition of grammaticalization, but BP *estar indo* is neither obligatory nor has it paradigmaticalized, thus some could argue that but BP *estar indo* is not grammaticalizing. However, as suggested by Diewald (2010, 2011) and Lehmann (2015) there are varying degrees of obligatoriness and paradigmatisation. Lehmann (2015) stresses that obligatoriness is not necessarily a criterion for grammatical items (thus, neither an absolute criterion for grammaticalization) (cf. Diewald 2011, and Hopper and Traugott 2003). Whether a grammatical element is obligatory or not depends entirely upon the context (Diewald 2011, p. 367).

For instance, as illustrated in (44), omitting personal subject pronouns (pro-drop) is obligatory in Spanish, in European Portuguese it is optional (depending on the context), and in Brazilian Portuguese pro-drop of certain pronouns, e.g. first-person singular, would be obligatory in some settings, e.g. formal academic writing, but in other contexts personal subject pronouns are

obligatory, but object pronouns are dropped (see e.g. Farrell 1990). Observe the following examples:

(44)a. Spanish: *¿Lo hiciste?*

[It (you-encoded through inflectional morphology of the verb) did]

‘Did [you] do it?’

b. European Portuguese: *Fizeste-lo?*

[(you-encoded through inflectional morphology of the verb) did-it]

‘Did [you] do it’

c. Brazilian Portuguese: *Voce fez?*

[You did]

‘Did you do [it]?’

In recent years, there have been several instances of grammaticalization in BP, and the recent development of *estar indo V* is a fragment of the processes of change in BP. As illustrated in the examples above, the paradigms for pronouns and verb inflection has taken a different direction in BP than in European Portuguese and Spanish (cf. Zilles, 2007). For instance, the nouns ‘vossa mercê’ (‘your mercy’) and ‘a gente’ (literally: ‘the people’) have been grammaticalized into pronouns (Zilles, 2007), and this has completely altered both the pronoun paradigms and the paradigms of verb inflection in BP.

To conclude, some issues that may impact the grammaticalization of BP *estar indo* are:

- Brazil’s history of multilingualism and language contact which could make BP more prone to grammaticalization than e.g. European Portuguese.
- The increased frequency of *-ndo* progressives in BP (cf. Rocha, 2012)
- The high degree of secondary grammaticalization of *-ndo* progressives in BP (e.g. the FPP discussed in Rocha 2012)
- General tendency of phonetic erosion and deletion of clitics in BP

Chapter 7: Conclusions

This thesis has discussed whether the BP phrase *estar indo* is in the process of grammaticalization. After acquiring data from Twitter and the BYU corpora this study concludes that BP *estar indo* occurs in critical context, i.e. the type of linguistic conditions that enable grammaticalization (see Traugott, 2012). In addition, BP *estar indo* could possibly be emerging in switch contexts, which would indicate that it has undergone reanalysis but in order to verify this theory a quantitative study would be necessary.

In the present thesis, however, a quantitative study was not possible, because BP *estar indo* is still rare in written language. Nor does it occur with inanimate subjects, or in passive voice constructions, which according to Wu et.al. (2016) are important traits of full grammaticalization. The BP *estar indo* is influenced by a complex set of changes in pronoun paradigms and the emergence of new future verb phrases, and the English *be going to* developed in a very different linguistic environment, thus this study concludes that it is possible that BP *estar indo* could be an instance of grammaticalization that stays *in progress* instead of undergoing complete grammaticalization, but still, more studies on this topic are required.

The increased frequency and grammaticalization of *-ing* and *-ndo* periphrases have influenced and continue to influence both PDE *be going to* and BP *estar indo*. In other words, the data that has been investigated here, and the research by Traugott (2012) and Rocha (2012), amongst others, give evidence to the conclusion that the secondary grammaticalization of English *-ing* and BP *-ndo* progressives influences the grammaticalization of BP *estar indo* and English *be going to*. In other words, AUX *be going to* and AUX *estar indo* are examples of primary grammaticalization because they used to be unambiguous lexical items that developed grammatical properties, and this primary grammaticalization is part of a bigger process in which secondary grammaticalization of *-ing* and *-ndo* progressives play an important part.

In short, The Twitter corpus contains more than the double amount of collocates compared to the amount of collocates in Traugott's research, but since both Traugott and the study of the Twitter corpus are qualitative studies, it is impossible to compare the frequency of EMdE *be going to* and BP *estar indo*. Studies of data in the BYU corpora show that the frequency of *estar indo* is very low compared to the frequency of PDE *be going to*. Increased frequency is an important indicator of grammaticalization (cf. Hopper and Traugott, 2003). Thus, this thesis

concludes that if the frequency of *estar indo* does not increase it is very unlikely that it will grammaticalize completely in the same manner as PDE *be going to* has.

In addition, PDE *be going to* occurs with mental verbs, inanimate subjects and in passive constructions, and although *estar indo* has been observed with mental verbs in the Twitter corpus, there is no indication of *estar indo* occurring with inanimate subjects, nor in passive constructions.

About whether the grammaticalization of *estar indo* is borrowed from English, the thesis concludes that the sociolinguistic circumstances within Brazil could have influenced the construction more than English, but the sociolinguistic circumstances may also have made BP more susceptible to borrowings from English. It is possible to argue that grammatical borrowings are typical features of bilingualism, in the same manner as code-switching, and according to Poplack fluent code-mixing is a bilingual skill (cf Poplack, 1988, Poplack and Tagliamonte, 1999). However, English proficiency and bilingualism is low in the Brazilian population and deprioritised by the public administration (Montes, 2016).

Nevertheless, English influence on BP is evident from the large quantity of frequently used English loanwords in colloquial BP, thus it is probable that English influence could have reinforced the grammaticalization. In addition, Brazil's history of extreme language contact that took place during slavery and colonisation, could be an indication that BP is especially prone to grammaticalization. In conclusion, BP *estar indo* could have been influenced by indigenous Tupi-Guarani languages, Bantu and Yoruba languages, as well as English, and further studies are required to determine whether language contact has triggered the initial grammaticalization of BP *estar indo*.

7.1. Implications for future research

Despite its narrow scope, this thesis encompasses theory and research from a range of linguistic fields. It has meshed comparative linguistics together with synchronic and diachronic studies on language change. Both comparative linguistics and grammaticalization are fields that can contribute to cognitive linguistics and cognitive research in general. Recent studies show that sociolinguistic circumstances can be crucial for understanding language change, and thus also for understanding grammaticalization (cf. Trudgill 2016b, 2016b, Zilles, 2007, Bohn, 2003), and for future research on grammaticalization in English and/or BP it would be interesting to use a more holistic approach, including aspects from sociolinguistic research, to further investigate the background and development of *be going to/estar indo*.

The grammaticalization of English *be going to* is continuously revised and debated (cf. Wu et al., 2016, amongst others), and future comparative linguistic studies may shed light on aspects of the grammaticalization process of English *be going to* that have been underestimated earlier. There are still many questions and discussions on the grammaticalization of both English *be going to* and BP *estar indo*. Future studies could aim to investigate the constructions through more quantitative corpora studies, or to study the representation of high-frequency and newly grammaticalized items in foreign language textbooks (as Orozco, 2014, did), or to take sociolinguistic circumstances more into consideration in grammaticalization studies.

Bibliography

- Bastos, A. C. P. (2004). Progressive constructions in Brazilian Portuguese and English. *Revista Letras*, 63.
- Bohn, H. I. (2003). The educational role and status of English in Brazil. *World Englishes*, 22: 159–172. doi:10.1111/1467-971X.00285 Retrieved from <http://onlinelibrary.wiley.com/doi/10.1111/1467-971X.00285/full> (9.2.2018)
- Bybee, J., & Pagliuca, W. (1987). The evolution of future meaning. In: A. G. Ramat, O. Carruba & G. Bernini (Eds), *Papers from the 7th international Conference on Historical Linguistics* (pp. 109–122). Amsterdam: John Benjamins.
- Crystal, D. (2003). *English as a Global Language*, Cambridge, UK: Cambridge University Press
- Cacoullos, R. T. (2000). Grammaticalization, synchronic variation, and language contact: a study of Spanish progressive-ndo constructions, John Benjamins Publishing.
- Dawson, H. C. (2003). Defining the outcome of language contact: Old English and Old Norse. *Ohio State University Working Papers in Linguistics*, 57, 40-57.
- Diewald, G. (2010). On some problem areas in grammaticalization studies. Grammaticalization: Current Views and Issues. K. Stathi, Gehweiler, E., König, . Amsterdam/Philadelphia, John Benjamins: 17-50.
- Diewald, G. (2011). Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics*, 49(2), 365-390.
- Farrell, P. (1990). "Null objects in Brazilian Portuguese." Natural Language & Linguistic Theory 8(3): 325-346.
- Filppula, M., Klemola, J., & Paulasto, H. (2008). *English and Celtic in contact*. Routledge.
- Garcia A.S. (2010). Verbos incompatíveis com o progressivo no inglês. *SOLETRAS (UERJ)*, v. 20: p. 146-164.
- Guimarães MR. (2002). "Alguns problemas na interpretação da progressividade no português do Brasil." *Revista Letras*, Curitiba, n. 58, p. 137-145. jul./dez. Editora UFPR
- Heine, B., & Kuteva, T. (2002). *World lexicon of grammaticalization*. Cambridge University Press.
- Heine, B. (2003). Grammaticalization. The Handbook of Historical Linguistics. B. D. J. a. R. D. Janda, Blackwell Publishing: 575-601.
- Hillman, R. S., & D'Agostino, T. J. (Eds.). (2013). *Understanding Contemporary Latin America*. L. Rienner Publishers. <https://onlinelibrary.wiley.com/doi/abs/10.1111/blar.12144> (28.04.18)

Hopper, P. J. and E. C. Traugott (2003). Grammaticalization. Cambridge, Cambridge University Press.

Jakobi, T. (2004). The English Be Going To Construction and Its Grammaticalization Process.

Junhui Wu, Qingshun He & Guangwu Feng (2016) Rethinking the Grammaticalization of Future *be going to*: A Corpus-based Approach, *Journal of Quantitative Linguistics*, 23:4, 317-341, DOI: [10.1080/09296174.2016.1226427](https://doi.org/10.1080/09296174.2016.1226427)

Kachru, B. B. (2006). *The English language in the outer circle*. *World Englishes*, 3, 241-255.

Killie, K. (2004). "Subjectivity and the English progressive." *English Language and Linguistics* 8(1): 25-46.

Killie, K. (2007). "On the source(s) and grammaticalization of the Germanic -lik suffix." *Neuphilologische Mitteilungen* 108: 659-682.

Killie, K. (2014a). "The development of the English BE + V-ende/V-ing periphrasis : from emphatic to progressive marker?".

Author's version of an article in the journal: *English Language and Linguistics*. Also available from the publisher at: <http://dx.doi.org/10.1017/S1360674314000148>

Killie, K. (2014b). "Secondary grammaticalization and the English adverbial -ly suffix." *Language Sciences*.

Kranich, S. (2010). *The Progressive in Modern English*.

Lehmann, Christian. (2015). *Thoughts on grammaticalization*. 3rd edition 2015 (Classics in Linguistics1). Berlin: Language Science Press.

Massini-Cagliari, G. (2004). Language policy in Brazil: Monolingualism and linguistic prejudice. *Language Policy*, 3(1), 3-23.

Matthews, P. H. (2007). *The concise Oxford dictionary of linguistics*. Oxford, Oxford University Press.

McArthur, T.(Ed.), (1998). *Superstrate*. *Concise Oxford Companion to the English Language*. : Oxford University Press. Retrieved 26 Jan. 2018, from <http://www.oxfordreference.com/view/10.1093/acref/9780192800619.001.0001/acref-9780192800619-e-1181>.

McWhorter, John (2013) , "Txtng is killing language. JK!!!" Ted Conference. Retrieved from: https://www.ted.com/talks/john_mcwhorter_txtng_is_killing_language_jk#t-362455 (28.11.17).

Mendes, R. B. (2008). Contraction, Grammaticalization and the Phonology/Syntax interface: Progressives in Brazilian Portuguese. In *Canadian Linguistics Association Conference*(pp. 1-10).

Montes, Amanda Lira Gordenstein (2016) "English in Brazil: A Sociolinguistic Profile," *Journal of Pedagogy, Pluralism and Practice*: Vol. 8 : Iss. 1 , Article 4. Available at: <https://digitalcommons.lesley.edu/jppp/vol8/iss1/4>

Nesselhauf, N. (2007). "The spread of the progressive and its future use." *English Language and Linguistics* **11**(1): 191-207.

Nesselhauf, N. (2010). "The development of future time expressions in Late Modern English: redistribution of forms or change in discourse?" *English Language and Linguistics* **14**(2): 163-186.

Norde, Muriel. (2010). "Degrammaticalization" University of Groningen. In: Stathi, K. Gehweiler, E. König, E. (ed.). (2010). *Grammaticalization: Current views and issues*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Orozco, R. (2005). *Distribution of future time forms in Northern Colombian Spanish*. Selected proceedings of the 7th Hispanic linguistics symposium.

Orozco, R. (2007). "The impact of linguistic constraints on the expression of futurity in the Spanish of New York Colombians." *IMPACT-AMSTERDAM AND PHILADELPHIA*- **22**: 311.

Orozco, R., & Thoms, J. J. (2014). The future tense in Spanish L2 textbooks. *Spanish in Context*, *11*(1), 27-49.

Poplack S., & Tagliamonte, S. (1999). The grammaticalization of going to in (African American) English. *Language Variation and Change*, *11*(3), 315-342.

Poplack, S. (1988). Contrasting patterns of code-switching in two communities. *Codeswitching: Anthropological and sociolinguistic perspectives*, 215, 44. https://books.google.no/books?hl=no&lr=&id=ZChqa7p4drcC&oi=fnd&pg=PA215&dq=poplack+code+switching&ots=OdNPI_JYPJ&sig=MDqQH-5FIDS7xXc3MVf0ryIPLPc&redir_esc=y#v=onepage&q=poplack%20code%20switching&f=false (30.04.18)

Priberam, O *Dicionário Priberam da Língua Portuguesa* (DPLP).(2018).: <http://www.priberam.pt/dlpo/> (accessed 20.04.18)

Tagliamonte, S. A., Durham, M., & Smith, J. (2014). Grammaticalization at an early stage: Future be going to in conservative British dialects. *English Language and Linguistics*, *18*(1), 75–108.

The guardian: "US know has more Spanish-speakers than Spain – only Mexico has more." (30 June 2015). *The Guardian*. Retrieved from: <https://www.theguardian.com/us-news/2015/jun/29/us-second-biggest-spanish-speaking-country> (03.10.2017)

Traugott, Elizabeth C., 2012. 'On the persistence of ambiguous linguistic contexts over time: implications for corpus research on micro-changes', in Joybrato Mukherjee & Magnus Huber (eds.), *Corpus Linguistics and Variation in English: Theory and Description*, Amsterdam: Rodopi, 231–46.

Trudgill, Peter. (2011). *A tale of two copulas: language-contact speculations on first-millennium England*. University Press of Southern Denmark : Volda University College

Trudgill, P. (2016)a. I'll git the milk time you bile the kittle do you oon't get no tea yit no coffee more oon't I: phonetic erosion and grammaticalisation in East

Anglian conjunction-formation. *Dialect Matters: respecting vernacular language*. Cambridge : Cambridge University Press.

Trudgill, Peter. (2016)b. Contact-related processes of change in the early history of English. In M. Kytö and P. Pahta (eds.) *The Cambridge handbook of English historical linguistics*. Cambridge: Cambridge University Press, 318-334. ISBN: 9781107039353

Tsuda, Y. (2010). Speaking against the hegemony of English: Problems, ideologies, and solutions. In T. K. Nakayama & R. T. Halualani (Eds.), *The handbook of critical intercultural communication* (pp. 248–269). West Sussex, UK: Wiley-Blackwell.

Zhiming, Bao (1998). Theories of language genesis. In: J. Foley (ed.) *English in new cultural contexts: reflections from Singapore*. Singapore: Singapore Institute of Management pp. 41-72

Zilles, A. M. S. (2007). Grammaticalization of a gente as a cluster of changes: evidence from apparent and real time studies. *Fórum Linguístico*, 4(1), 13-46.

Zagona, K. (2012). Ser and Estar: Phrase structure and aspect. *Building a bridge between linguistic communities of the old and the new world: Current research in tense, aspect, mood and modality*, 14, 303-27.

Corpora

Corpora from Brigham Young University:

<http://corpus.byu.edu/> (02.10.2017)

BYU, created by Davies, Mark. (2010-) *The Corpus of Historical American English: 400 million words, 1810-2009*. Available online at <http://corpus.byu.edu/coha/> (02.10.2017)

BYU, created by Davies, Mark. (2008-) *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/> (02.10.2017)

BYU created by Davies, Mark. (2002-) *Corpus del Español: web/dialect 2 billion*.

Retrieved from <https://www.corpusdelespanol.org/web-dial/> (02.10.2017)

BYU created by Davies, Mark and Ferreira, Michael. (2006-) *Corpus do Português*. Available online at <http://www.corpusdoportugues.org> (02.10.2017)

Correa, Kristine L. (2017), The Twitter Corpus: <https://typecraft.org/tc2/ntceditor.html#3885>

POB, Proceedings of the Old Bailey. Version 8, March 2018: <https://www.oldbaileyonline.org> (09.04.2018)

Rocha, Franciane (2012): <http://typecraft.org/TCEditor/1730/> (02.10.2017)

(28.11.17)

Dictionaries:

Matthews, P. H. (2007). *The concise Oxford dictionary of linguistics*. Oxford, Oxford University Press.

Stevenson, A. (2010). *Oxford dictionary of English*. Oxford, Oxford University Press.

Spanish: Real Academia Española (RAE)

Portuguese: *O Dicionário Priberam da Língua Portuguesa* (DPLP)
<http://www.priberam.pt/dlpo/> (accessed 20.04.18)

Appendix

The Twitter corpus

There is free access to the digital corpus, and it is available at:

<https://typecraft.org/tc2/ntceditor.html#3885>.

1. Queria **tá indo pegar** uma praia, mas infelizmente tô separando roupa pra ir pra escola, triste essa vida
2. depois de 2 açais de 500ml a pessoa aqui **tá indo pegar** comida
3. daqui a pouco pegar ônibus quando na vdd queria **tá indo pegar** vc
4. Bom dia pra quem **ta indo pegar** 16 hrs de trabalho, nem podia
5. Nathally só **ta indo pegar** os meninos e eu to aqui kkkk
6. Qm **ta indo pegar** a amiga no aeroporto? Eu mexxxxxma!!!!
7. mas **daqui a pouco** ela **tá indo pegar** ônibus.
8. **Daqui a pouco** tu **ta indo pegar** um molinete e ir pescar umas traíra no açude pra fazer sushi
9. Vou levantar e bota **a** roupa q **daqui a pouco** tenho q **ta indo pegar** ônibus
10. **Daqui a pouco** o Biyombo **tá indo pegar** os rebotes da série do oeste
11. Bom dia , **daqui a pouco** minha mãe **tá indo pegar** meu boletim... Ossada !!!
12. Minha mae **ta indo pegar** laranja no pe ta caçando as abelha agarra nela **daqui a pouco**
13. estudos apontam que quem shippa alfredisis **daqui a pouco** ja **ta indo pegar a** mamadeira pra dormi
14. **daqui a pouco** minha mãe **ta indo pegar** as minhas notas
15. Minha mãe vai viajar **daqui a pouco** e parece que **tá indo pegar** uma nave e não um avião.
16. Meu pai **daqui a pouco ta indo pegar** o aviao e partindo pra vegas, e eu o trem partindo pra cosmos kkk
17. **Daqui a pouco a** menina **tá indo pegar** o lápis no chão e DON'T LOSE YOUR WAY de tanto que essa OST toca
18. Mae **daqui a pouco ta indo pegar** o resultado :(mdds tô cm mt medo d perder meu celular ahhh
19. meu crush não gosto de mim, ele **já ta indo se casar**, to bem triste
20. será **que** ele **tá indo se casar**
21. se eu não sou **a** melhor namorada do mundo, sou **a** mais trouxa, meu namorado **tá indo ver** Thor Ragnarok sozinho
22. Minha irmã **tá indo ver a** liga da justiça e eu indo estudar, q mundo é esse?
23. eu **tô indo viajar** sexta, mas tenho quase ctz q quando voltar vão me internar
24. amanhã **já tô indo viajar** de novo, AAAAH que blz
25. **To indo viajar** por 2 semanas e minha mãe **já** ta desesperada querendo me ensinar a cozinhar, lavar roupa e tudo mais
26. cheguei hoje em casa e quinta **já tô indo viajar** de novo.. era tudo que queria
27. Só mais quatro semanas e **já tô indo viajar**, amém

28. Sabe, **amanhã**, eu vou pegar um horário horário que todo mundo vai pra praia, e eu vou **tá indo trabalhar**
29. Fui obrigada a chamar minha colega pra saber pq ela n **tá indo trabalhar**, se n a partir de **amanhã** vou ter q começar a fazer o serviço dela
30. A gente tem que votar hoje e **amanhã** MUITO pq segunda a maioria **tá indo trabalhar** e estudar
31. **amanhã** é o dia q eu entro tarde na facul e pego o horário de pico onde todo mundo **tá indo trabalhar** ou seja ônibus lotado, já tô irritada
32. Porra meu pai nem fica em casa mais direito, **amanhã** cedinho ele já **tá indo trabalhar** dnv
33. **amanhã** 5 da manhã tenho que **tá indo trabalhar**
34. tenho médico **amanhã** 8h30 e ainda vou ter que ir de ônibus pq minha irmã **tá indo trabalhar** com meu carro
35. Bom mesmo é saber q teu pai **tá indo trabalhar** agr e só volta **amanhã**
36. Só aparece **amanhã** ela dando bom dia e dizendo que **tá indo trabalhar**
37. **Amanhã** tenho que dar o casaco do Pam tadinho **tá indo trabalhar** com frio diz ele né
38. Falar p vcs, queria eu **tá indo trabalhar amanhã**...então n reclamem
39. **amanhã** é feriado por causa do dia do trabalho e tem gente que **tá indo trabalhar**
40. Não sei se vou **ta indo trabalhar amanhã** nem sexta
41. tô vendo q **ano que vem**, metade do **ano** eu **tô indo estudar** de noite
42. **tô indo estudar** med na cidade dele **ano que vem** Kkk
43. **Tô indo estudar** no Cemeit **ano que vem**
44. Bom dia, logo **to indo conhecer** o lugar onde vou **estudar ano que vem**, vaaai
45. **Amanha** começa as 11, no domingo as 13 ai quando vc vai ver **ta indo comer** sushi
46. hahaha e **amanhã ta indo comer** hamburguer <3
47. **amanhã** queria **ta indo comer** um pastel na feira noturna, mas acho que nem vai rola x:
48. aí **amanhã** já **tá indo comer** burgui, pastel, etc e isso tudo vencendo na geladeira da pobre amiga lia
49. **AMANHA** ESSAS HORAS A GENTE VAI **TA INDO COMER**
LSDKFJLSKDJFLSKDJFLSKDJFLSKDJFLSKDJF QUERO **AMANHA** LOGO
50. @kecampion vamos **ta indo comer** **amanha** parka
51. **amanha** agente **ta indo comer** japa na casa do tofani
52. a gente **ta indo comer** sushi **amanha**
53. mas tb bom dia pra quem **daqui a pouco tá indo comprar** o ingresso do show da aurora <3 e pq moção vai comprar o da daughter hj tb aaaa
54. **daqui a pouco** ela **tá indo comprar** pão de burca, muito apropriadora essa garota
55. Amizade é igual pai, ela tá ali por um tempo mas **daqui a pouco tá indo comprar** cigarro.
56. Nois nem foi dormir e **daqui a pouco** já **tá indo comprar** ingresso pro gordo pqp
57. **Daqui a pouco** meu pai **tá indo comprar** pão e eu ainda não dormi.
58. 5 e 10 já cara, **daqui a pouco** meu avô **ta indo comprar** pão
59. até meu pai tava falando que **daqui a pouco ta indo comprar** remédio pra eu conseguir dormir
60. **Daqui a pouco** ele **tá indo comprar** pra namorada, se ele for hétero né nunca se sabe
61. papi **tá indo comprar** **daqui a pouco**
62. **daqui a pouco** você **ta indo comprar** pão, o caixa ta sem troco e pergunta se você aceita "eu te amo" de troco
63. **Daqui a pouco ta indo comprar** pão de terno e gravata
64. Adivinha só quem vai p praia **daqui a pouco** e arreassem **tá indo comprar** o biquíni????

65. @jessicapiccagli durmo sim **daqui a pouco** kk e vc q **tá indo comprar** açai essas horas?
66. Minha mãe me liga p falar que **daqui a pouco tá indo comprar** roupa com a minha irmã
67. **Daqui a pouco a gente ta indo comprar** ahauau
68. Então para com isso --' **daqui a pouco tá indo comprar** roupa na sessão G³
69. o show na praça eh só de noite agora a gente **ta indo comprar** as coisas pra festa **daqui a pouco**
70. Ele disse que **daqui a pouco ta indo comprar** o meu *o*
71. pia boy do caralho, usahusahsauh, **daqui a pouco tá indo comprar** um carro @chupadax
72. aposto q ela **ta indo comprar** mais paçoca **daqui a pouco**
73. **Daqui a pouco a gente tá indo visitar** ele na cadeia por agressão de mulher também, é um ogro esse menino
74. E assim mesmo, **daqui a pouco já ta indo visitar** ele
75. é, falou tambem que **daqui a pouco ta indo namorar... vou** ai p ficar conversando com teus parentes
76. meu primo **ta indo namorar, daqui a pouco vou** é caçar meu rumo tb
77. Espero que não chova pq **daqui a pouco a gente ta indo andar** de barco
78. isaaaa **daqui a pouco a gnt ta indo andar** na praia, ar ir com a gnt?
79. Cheguei aqui no hotel :) **daqui a pouco a gente tá indo andar** pela cidade.
80. apareceu aqui do nada, rs, **daqui a pouco agente ta indo andar** de skate..
81. **daqui a pouco agente ta indo andar** de patins"
82. ATAH PENSEI QUE TWITTER TINHA PARADO NOVAMENTE **JA TA INDO CONVERSAR** COM A CORDA DE SUICÍDIO
83. Tamiris ta tão triste q **ja ta indo conversar** na praça
84. **Já ta indo conversar** com o Grêmio pra assinar amanhã e fazer a maior zueira so futebol do Brazil
85. Quer que eu vá la ver pq ta ela e minha tia falando que é melhor e **ja ta indo conversar** com a mulher pra trocar
86. **Já tà indo conversar** com os amigos
87. a iara **já ta indo conversar** com as vaca do sitio
88. **ja ta indo conversar** com o [@roberto stein](#) , aushaush'
89. metade do **ano que vem** quero **ta indo viajar** ja
90. será **que ano que vem** você vai **ta indo viajar** com a gente ?
91. Se fosse junho do **ano que vem** eu não ia ligar, pelo menos eu ia **tá indo viajar** pra bem longe desse lugar
92. **Ano que vem** meu melhor amigo, meu amor, minha vida por fim meu irmão **ta indo viajar** comigo
93. ahhh mas o gui **ano que vem** tbm **ta indo viajar**.. Mas já estou esperando aqui pra gente sair rs
94. Queria muito **ta indo viajar** com minhas amigas esse final de semana.. Mas, chega logo **ano que vem!**
95. Pois é , mais tbm ele **ta indo viajar** amanhã , vai pra escola de aprendiz de marinheiro , e só volta **ano que vem** :(
96. meu melhor amigo **ta indo viajar** semana **que vem**, ficar um **ano** fora, e eu nao to tendo coragem de ver ele :/
97. minha irma **ta indo viajar** semana **que vem** p voltar so em janeiro do **ano que vem**
98. Nossa nas ferias do **ano que vem** eu vo **ta indo viajar** com o meu carro, e sem bico nenhum podendo fala nada! 18 Chega logo seu lindo! Hah
99. pra esse povo **que ta indo viajar** ... da nada **ano que vem** eu vo **ta indo viajar** tbm KK
100. fiquei sabendo **que** você passo no teste e **ano que vem ta indo viajar** ! parabens manooo :D

101. Mano minha mãe **ta indo viajar** daqui a pouco
 102. e ai eu penso q terça ele **ta indo viajar**
 103. Advinha quem acordou e já **tá indo estudar** pra prova
 104. meu ex crush perdeu de ano e **ta indo estudar** na minha escola ano **que vem**
 105. podia **ta indo estudar** mas to indo dormir pois já desisti do ano **que vem**
 106. Ano **que vem** minha nova amiguinha **ta indo estudar** no whallther ...
 107. A pessoa **ta indo estudar** fora mes **que vem** e eu vo fica aqui escutando already home e chorando
 108. E quando a crush avisa **que tá indo estudar** na sua faculdade ano **que vem**
 109. Meu irmão tem futuro, **ta indo estudar** no Pedro II ano **que vem...**
110. semana **que vem** eu vo **ta indo estudar** feito uma corna pras provas finais
 111. É foda ce pensar **que semana que vem ta indo estudar** fora e vai deixar sua mãe sozinha.
 112. O marcus **ta indo estudar** no elite ano q **vem**, vê se agora aparece lá pra ver a gente
 113. é impressão minha ou todo mundo **ta indo estudar** no exathum ano **que vem**
 114. só de pensar **que ano que vem** eu ja podia **tá indo estudar** nos EUA
 115. Muita gente nova **tá indo estudar** no Olimpo ano **que vem**, nossa.
 116. fiquei sabendo no domingo **que** meu brother [@heliobona](#) **tá indo estudar** em buenos aires mês **que vem**, vai fazer falta demais esse mlk :/
 117. espero entra na faculdade **ano que vem** também, **tô indo estudar** infelizmente
 118. **Ano que vem To indo viajar estudar** em outro país só não sei qual ainda
 119. E ai menina tudo bom? **Tô indo viajar ano que vem** pra Irlanda Vamo???
 120. **ano que vem** provavelmente **tô indo estudar** pra esses lados ♡♡
 121. Se tudo der errado **amanhã to indo ver** liga da justiça
 122. ai carol num faz isso cmg **amanha** eu **to indo ver** vcs olha os modos
 123. **Amanhã tô indo** pegar meu combo BK pq BK é melhor ta dito
 124. **amanhã tô indo** arrumar o da amanda, n vai durar um dia, conheço!!! agr o da babi vai n sei kkkk qnd eu for eu arrumo
 125. **amanha to indo** cortar o cabelo com o trindade...
 126. **Amanhã to indo** ver as opções de cursos de inglês.
 127. falei isso pra minha professora de fotografia e **amanhã tô indo** ter uma conversa com meu coordenador sobre ela, vou piorar ainda mais a coisa
 128. amorzíneos, **amanhã tô indo** comprar as luzes pra poder gravar meus vídeos daqui!
 129. hoje to indo fazer o enem e **amanhã to indo** viajar, acho q nao to tao na pior assim
 130. 22:00 Tem Enem **amanhã, tô indo** dormir.
 131. Então faz passar logo que **amanhã to indo** te ver
 132. **Amanhã to indo** trocar, nem ferrando q fico cm essa merda aq
 133. **amanhã to indo** furar o mamilo tb zé
 134. **amanhã to indo** pega bolo
 135. Bah, **amanhã to indo** comprar pra dar pra vocês
 136. **tô indo** assistir o filme agora kkkk
 137. to indo nessa, **amanha to indo** fazer exame
 138. ah, amor? **amanha to indo** fazer exame, então so entrarei aqui na parte da tarde, me espera ta?
 139. **Amanha to indo** fica um semana no meio do nada

140. vou procurar uma foto do cabelo da Nabrisa q **amanhã to indo** fazer o msmmmm
kk
141. Fala o nome, **amanhã to indo** comprar.
142. Isso deixa tudo bem limpinho que **amanhã tô indo** dormir ai
143. Momentos de tensão: a fiscal não quis aceitar meu RG porque era de 2008. Deu
boa mas **amanhã tô indo** fazer outro.
144. **to indo** fazer hidratação no meu cabelo
145. **Amanhã to indo** tira uns dias em Uruguaiana
146. **Amanhã to indo** ficar com a irmã
147. **amanhã tô indo** fazer minha matricula no aaz
148. **Amanhã to indo** comprar meu COD WW2 em midia fisica. Esse é igual o BO3,
tenho que ter o disco.
149. Inclusive **amanhã to indo** ver com meus filhos, 3 amigos virjão dele e a patroa.
150. que hitoria é essa deque vao trair que **ja** ta traindo que esqueceu de trair que **ta
indo trair**
151. Aí fiquei tipo "**já tá indo trair** a corna de novo ne"
152. Nossa nem fala. Desde sempre foi assim, ele não pode ir ver a familia que **ja** falam
que ele **ta indo trair** a Nanda e tal.
153. O cara mal acaba de casa e **ja ta indo trair** a mina
154. White girls problems: Minha melhor amiga **ta indo visitar** meu melhor amigo
depois de **amanha** no USA e eu ainda não escolhi as coisas que ela vai trazer para mim
155. Quem **logo ta indo visitar** NY??
156. Eu sou do tipo que avisa que **tá indo visitar** o amigo quando **já** está na porta
157. E é cancelado porque **a** clara **ta indo ver** o FIO
158. mas ela **ta indo ver a** kim entao agora vai
159. Bom dia pra quem **tá indo ver a** mamys!
160. quando vc **tá indo roubar** o banco, eu já estou contando o dinheiro bebê
161. "Sr ladrão blablabliblobli", o cara **tá indo roubar** sua casa, mano. Acha que ele vai
ler recadinho na sua geladeira?
162. **Tá indo roubar** cedo em
163. Kkkk ah briza de qnd vc tá perdido na rua e as pessoas acham q vc **tá indo roubar**
já roubo ou ta pensando isso
164. chegou parte das coisas do apê, **amanhã to indo arrumar** <3
165. **to indo arrumar** minhas coisa pra escola **amanhã**
166. acabei de chegar em casa e **tô indo arrumar** mala pra viajar **amanhã** cedo, tô
quase dormindo em cima das roupas
167. **amanhã tô indo arrumar** o da amanda, n vai durar um dia, conheço!!! agr o da
babi vai n sei kkkk qnd eu for eu arrumo
168. **To indo arrumar** a church pra conferência **amanhã**
169. Qq eu **tô indo arrumar** naquela escola **amanhã** @deus kkkkkk
170. **Amanhã tô indo arrumar** as cordas do meu violão, vou tocar umas músicas boas
pra vê se a cultura nasce nessa casa porque
171. sei nem oq q **to indo arrumar** na escola **amanhã**, devia ficar dormindo em casa
172. Bom dia! Depois de fazer uma correria com a minha avó, **tô indo carregar** meu tri
e depois vou trabalhar
173. logo **tô indo carregar** meu bilhete único hoje
174. pela quinta vez eu **tô indo carregar** meu celular
175. Cheguei em casa agora e **já tô indo carregar** água porraaaa

176. **já** carreguei meu celular 4 vezes hj e **tô indo carregar** dnv não aguento maisss
177. Eaii, **tô indo carregar já**
178. **Tô indo carregar** o tri, vou ver se faço uma mão no centro **já** tb
179. Aiai.... **ja to indo carregar** meu cartão.
180. **to indo carregar** um armario e **ja** volto
181. **Ja to indo carregar** o cel so terminar d enviar esse tweet!!
182. ele **tá indo pensar** no quanto ele ama a le hahahahaha
183. adivinha quem **tá indo pensar** na vida durante as aulas
184. tipo agora, deveria **ta indo pensar**, e dormir e não, fico no Twitter falando merda
185. Quando uma pessoa fala que vai deita e ouvir musica, ela **ta indo pensar** em alguém..
186. Deixei Maria das Flores bolada. Ela **tá indo pensar** na vida
187. [@pfvrmatheuso](#) na verdade **ta indo pensar** em mim, chorar por mim, ligar prq mim, nao nao ligar pra ele
188. Meu professor deve **tá indo pensar** em mim, porque tenho sonhado todos os dias com ele.
189. logo ele **ta indo pensar** em mim
190. Nois fala **que ta indo ouvir** mc pepita mas na verdade nois **ta indo ouvir** My immortal do evanescence
191. daí vc escuta pagode em um snap e **já ta indo ouvir**
192. miley cyrus **ja** deu né amiga vamo volta a usa roupa normal nos shows as pessoa **ta indo ouvir** musica nao compra um vibrador
193. your local dc girl **ta indo** ver infinity war assim