Verb-framed and satellite-framed patterns in Brazilian and European Portuguese

Manú Castro Nilsson

Stockholms universitet

Verb-framed and satellite-framed patterns in Brazilian and European Portuguese

Manú Castro Nilsson

Abstract

The goal of this work is to contribute to the understanding of the framing typological domain of motion events in Brazilian Portuguese (BP) and European Portuguese (EP) through corpus analysis of intransitive pleonastic double Path constructions, which are constructions that express 'the same' trajectory in both the verb and its satellite. Data has been retrieved from the ptTenTen11 corpus.

Traditionally Portuguese is described as a verb-framed language, meaning that it expresses the semantic trajectory component (Path) in the verb, rather than in the satellite.
The double Path constructions are interesting since they do not have a 'clear place' within the framing typological approach to motion. The Boundary crossing constraint, which is well-known and studied constraint in Spanish, but that has not been found in studies on Portuguese is also analyzed here. With these two entry points an attempt is made to relate BP and EP to the framing typological classification into language types.

The double Path constructions appear to be more frequent in EP than in BP and no challenges to the Boundary crossing constraint are identified in the data for either of the Portuguese varieties.

Keywords

Verb-framed and satellite-framed patterns in Brazilian and European Portuguese

Manú Castro Nilsson

Sammanfattning

Syftet med den här studien är att bidra med förståelse kring hur brasiliansk portugisiska (BP) och europeisk portugisiska (EP) passar in i domänen av uttryck för rörelse från en plats till en annan enligt *framing*-typologin. Datan är hämtad från korpusen ptTenTen11 och uttryck med upprepad semantisk information om en rörelses tillryggalagda sträcka undersöks.

Inom ramverket för *framing*-typologin är portugisiska ett så kallat verb-inramat språk som tenderar att uttrycka en rörelses sträcka i verbet och inte i en så kallad satellitkonstruktion. Då upprepningen av denna semantiska information gällande sträckan inte har en självklar plats inom *framing*-typologin är dessa typer av konstruktioner intressanta att analysera närmre. En inom *framing*-typologin omtalad begränsning på hur verb-inramade språk kan beskriva situationer när en gräns korsas undersöks också. Begränsningen är välstuderad inom till exempel spanskan, men verkar inte ha tagits upp i någon större utsträckning i litteratur om portugisiska.

De semansiskt upprepande konstruktionerna visar sig vara vanligare i EP än i BP och begränsningen gällande hur konstruktioner tenderar att se ut när en gräns korsas motsägs inte av datan.

Nyckelord

Portugisiska, rörelseverb, framing-typologi, pleonasm

Table of contents

List of abbreviations

BP   Brazilian Portuguese

CME   COSTA-motion event

EP   European Portuguese

VBP   Vernacular Brazilian Portuguese

1 Introduction

This study investigates motion event constructions in Brazilian Portuguese (BP) and European Portuguese (EP) from a framing typological point of view. According to Talmy (2000), motion events consist of the four basic building blocks Figure, Ground, Path, and Motion. Figure can be described as the 'highlighted' object that moves in relation to some backgrounded object, or Ground, and Path is the trajectory that the Figure travels along. As for Motion, which denotes motion itself, a specific type of motion is of special relevance for this study: translational motion. Translational motion is described as "motion in which the location of the Figure changes in the time period under consideration" (Talmy 2000: 25). The Spanish sentence in (1) below is illustrative of how a translational motion event can be portrayed in the framing typological literature.

(1)   Spanish expression of Motion conflated with Path in the verb [glossing and translation from source text] (Talmy 2000: 49)

   La botella entró   a la   cueva   (flotando)
   The bottle   MOVED-in to the   cave   (floating)
   "The bottle floated into the cave"

Considering (1) and the just mentioned motion components, the Path element can be identified in *entró* 'MOVED-in/enter'. As for the English translation however, *floating* is used as the main verb, while the 'MOVED-in' sense is expressed with *into.* Framing typologists refer to adverbials such as *into* as satellites, which is a broad category of almost any form of closed class element that act as a dependent with the motion verb as its head (Talmy 2000: 102). Based on general tendencies of where the Path component is expressed, framing typological language types emerge, according to this approach. Spanish, which expresses Path in the verb is a verb-framed language, whereas English, which expresses Path in the satellite, is a satellite-framed language (Talmy 2000: 49, 131).

Portuguese is not as well-studied as Spanish when it comes to framing types, and there are opposing claims as to what lexicalization pattern is the most prevalent one for BP and EP. (i) Queriquelli & Moura (2021) suggest that BP is increasingly showing signs of becoming a satellite framed variety, while (ii) Batoréo & Ferrari (2016) claim that EP is showing signs of turning into a satellite framed variety, while BP is more stable in its verb-framed patterns (p. 6126; p. 59). Amongst the examples provided by Queriquelli & Moura (2021) there are Vernacular Brazilian Portuguese (VBP) examples of what appears to be double Path constructions, exemplified in (2) below. The double Path motion event construction is underlined in VBP, the gloss, as well as the English translation.

(2)   Vernacular Brazilian Portuguese [own gloss and translation] (Queriquelli & Moura 2021:6127)

   Ele pegou   minha   mochila   e   <u>saiu</u>   <u>pra</u>   <u>fora</u>   da   sala
   He picked.up   my   backpack   and <u>exited</u>   <u>to</u>   <u>outside</u> of.the   room
   'he picked up my backpack and <u>exited (to the outside)</u> of the room'

It is not entirely clear what to make of double Path constructions such as (2) from a framing typological perspective. Further, not only is there a Path component in both the verb and the satellite, but they also express 'the same' semantic information – *saiu* 'exit' and *pra fora* '(to) outside'. One part of the study is to examine to what extent intransitive pleonastic motion event constructions, such as the example in (2) are distributed across BP and EP. Also, considering that Portuguese is an understudied language from a framing typological perspective, a more prototypical characteristic of verb-framed languages known as the Boundary crossing constraint will be tested for in the Portuguese varieties as well. The constraint is claimed to hold for verb-framed languages and entails that Manner verbs are not allowed in contexts where the crossing of a boundary is expressed (Cifuentes-Férez & Molés-Cases 2020: 89). A Manner verb is a verb such as the English *floated* in (1) above, which describes *how* a particular Figure is moving, rather than expressing its trajectory (Talmy 2000: 26).

2 Background

The background section consists of four parts. The first one (2.1) provides a general background of Portuguese, along with contrastive particularities of Brazilian Portuguese and European Portuguese – henceforth BP and EP respectively. Following the introduction to Portuguese, the second part (2.2) introduces Motion events, first from the theoretical perspective of Leonard Talmy's (2000) framing typology, after which Sinha & Kuteva's (1995) Distributed spatial semantics is presented as an alternative descriptive approach. The second part (2.2) continues by contextualizing the framing typological perspective in two more recent studies on Portuguese. First, a study by Queriquelli & Moura (2021) is described, where it is claimed that BP is behaving increasingly as a satellite-framed language, after which an investigation by Batoréo & Ferrari (2016) is presented, where they state that EP is displaying satellite-framed tendencies in comparison with BP. Considering that the investigation concerns pleonastic constructions (constructions of semantically repeated elements), the second part (2.2) proceeds with treating the notion of 'pleonasm' and presents a working definition of it. The second section (2.2) is concluded with an overview of selected examples from the corpus data, for the reader to get familiarized with the types of contexts that the BP and EP pleonastic constructions appear in. Part three (2.3) summarizes the preceding sections, which feeds into the final part (2.4) where the aims and research questions of the thesis are specified.

2.1 Portuguese: an introduction

Considering that Portuguese is the language treated, the first section of this part (2.1.1) consists in an overall introduction to this highly globalized and pluricentric language. In the second section (2.1.2), focus is shifted to describing and exemplifying distinctive characteristics of Brazilian Portuguese (BP) and European Portuguese (EP), as well as highlighting points of differentiation between them.

2.1.1 Overall genealogy and sociolinguistic background

By 2014, Portuguese was estimated to be spoken as a first or second language by 221-245 million people in the world (Müller de Oliveira 2014: 28). Portuguese holds the status of 'official language' in, or in parts of, ten countries across four continents: Portugal, Brazil, Angola, Cape Verde, Guinea-Bissau, Mozambique, São Tomé e Príncipe, Equatorial Guinea, East Timor and the Special Autonomous Region of Macau, China. Overall, the number of Portuguese speakers is considered to be growing at a moderate level, with Southern Africa as its main center of growth – estimated to reach 160 million speakers by the year 2060 (Müller de Oliveira 2014: 28-30). When accounting for the 'Portuguese language world', in terms of how Portuguese language policies are organized, Müller de Oliveira (2014) describes a decentralized and *dual* standard, with Portugal actively engaging in external language actions in East Timor and the Portuguese speaking African countries (known by the acronym PALOP: *Paises Africanos de Lingua Oficial Portuguesa*), while Brazil concentrates on 'its own' territory on the American continent (p. 34-5). Following Brazilian independence from Portuguese colonial rule, and in particular since the proclamation of the Brazilian Republic in 1889, political relations between Brazil and Portugal have initiated separate standardization processes – e.g. two orthographies, two academies, two dictionaries (although there are most likely more Portuguese dictionaries than just these two), and two approaches to proficiency certification (Müller de Oliveira 2014: 36). Beyond the Portuguese speaking regions, Portuguese (either one of the two standardized variants) is also an official language in the European Union (EU), the Africa Caribbean Pacific Secretariat (ACP), the African Union (AU), the Southern African Development Community (SADC), the Organization of American States (OAS), the Southern Common Market (MERCOSUL), and Partners of the Americas (Hutchinson et al. 2003: 240). Regardless of one's perspective on the divergence of standard varieties, it is undoubtedly safe to say that Portuguese holds a global and officially recognized status in several spheres of political and economic power on the continents where it is used.

Genealogically, Portuguese grew out of (Vulgar) Latin, which was brought to the Iberian Peninsula by Roman soldiers. The settlement of Germanic groups as the Roman Empire faded, and the Arab conquest of the peninsula in the year 711, are further recognized as major episodes of influence on the shaping of what we today call Portuguese. For instance, most place names beginning with 'al' are of Arabic origin, which is also the case with several terms for denoting agricultural products and technological features that were brought to the Iberian region from North Africa (Hutchinson et al. 2003: 241).

2.1.2 Brazilian and European Portuguese: some indicators of difference

The following section compares and contrasts the Brazilian (BP) and European Portuguese (EP) varieties, on the basis of a set of illustrative points of lexical differentiation, retrieved quantitatively by Kilgarfiff et al. (2014). Subsequently, selected examples of differing 'dialectal' preferences in the area of denominal verbs and corresponding periphrastic constructions will be presented.

In their paper *PtTenTen: A Corpus for Portuguese Lexicography*, Kilgarriff et al. (2014) describe, discuss and exemplify steps in the process of using corpus technology when preparing material for the Oxford Portuguese Dictionary – a Portuguese-English, English-

Portuguese dictionary (p. 111). Since ptTenTen is the main data source and methodological tool for this thesis as well, the corpus will be properly introduced in the Method and Data section 3.1 further down. However, before proceeding with the current presentation of BP and EP, it must be mentioned that ptTenTen is constituted by data crawled off of websites (mainly) with the domains .pt and br, which as such represent EP and BP respectively (Kilgarriff et al. 2014: 115). The previously presented description of BP and EP as the two main regional variants of Portuguese is thus reflected in the ptTenTen corpus' classification of them as subcorpora, according to the website domains.

Keywords, which are words (or single-token items) that are more frequent in a subcorpus than in the whole corpus, can be used in order to look at what is distinctively a feature of BP, and what is of EP (Kilgarriff et al. 2014: 119). While it comes as no surprise that a geographical adjective such as *paulista* (meaning, from São Paulo state) is a BP keyword, since São Paulo is a state in Brazil, keywords within the category Grammatical words indicate divergencies at a 'deeper', or structural/grammatical level (Kilgarriff et al. 2014: 120). The Grammatical keywords are listed in Table 1 below.

Table 1 Keywords: Grammatical words (Kilgariff et al. 2014:120)

Brazilian ptTenTen	European ptTenTen
diante 'in front of, in view of'	*perante* 'in front of, in view of'
você 'you, 2nd p. sing.'	*teu* 'yours, 2nd p. sing.'
porém 'but'	*vosso* 'yours, 2nd p. pl.'
	vós 'you, 2nd p. pl.'
	este 'this'
	isto 'this'
	quer 'whether, want [verb]'
	aquando 'when'

Grammatical 'keywords' in the ptTenTen11 corpus, with grammatical tokens that are more common for the Brazilian variety listed on the left, and typical grammatical token of the European variety listen on the right.

As can be seen in Table 1, the expression of second person pronoun, in both its singular and plural forms differ between the varieties. *Portuguese: An Essential Grammar*, which is an EP grammar with a chapter dedicated to BP, states that *Tu* is the informal second person pronoun, to be used only for friends, relatives and children. *Você* is described as slightly more formal, but not formal enough for addressing those that you have never met before, or a superior. In such cases *o senhor* 'sir' or *a senhora* 'madam' should be used (Hutchinson et al. 2003: 43). Concerning the EP/BP difference, the grammar also claims that "**Tu** and **vós** are hardly ever used in Brazil. They have been superseded by **você**, **vocês**, **o senhor** [sir] and its variants a **senhora** [madam], **os senhores** [sirs], **as senhoras** [madams]" (Hutchinson et al. 2003: 214). Kilgarriff et al. (2014) also point out that this is a known 'dialectal' difference, which (on a side note) interestingly has a parallel in the Peninsular/Latin American Spanish divide as well (p. 120).

If we turn to the morphology of verbs, Batoréo & Ferrari (2016) – whose work will be more thoroughly presented below (in section 2.2.5) when discussing their view on motion events in Portuguese – suggest that there is a higher degree of acceptability of denominal verbs in BP, such as saying *parabenizar* 'to congratulate', (from Portuguese *parabéns*, meaning 'congratulations'), whereas EP speakers only use the periphrastic expression *dar os parabéns* 'to give congratulations' (p. 64). The same phenomena of denominal verbs in BP can be seen in meteorological terms. There are some core similarities with regards to describing weather conditions in both varieties: periphrastic constructions are used when it is cold (*faz frio,* lit. '(it) makes cold'), denominalized expression are used when it is raining (*chove,* '(it) rains') or snowing (*neva*, '(it) snows'). However, in the case of wind, BP is claimed to allow the denominalized construction *ventar* 'to wind/to be windy', which is not the case in EP contexts, where only the periphrastic *faz vento* (lit. '(it) makes wind') would be accepted (Batoréo & Ferrari 2016: 64).

While the differences between BP and EP presented above might be perceived as peripheral or superficial, it is precisely these types of lexical and pragmatic (forms of address) differences that usually appear to be brought up when comparisons are made, apart from phonological divergencies, which are not to be mentioned here. As will become clearer in the upcoming section (2.2) on motion events, the denominalization versus periphrastic constructions is an issue that goes beyond meteorological terms, which were just a specific example of one such instantiation, which happens to be more widely known.

2.2 Motion events

This section familiarizes the reader with the framing typology, as presented by Talmy (2000), in the first part (2.2.1). Subsequently, in the second part (2.2.2) an alternative approach to understanding motion, known as Distributional spatial semantics, is outlined. Part three (2.2.3) provides a short overview of framing typological literature on Portuguese. In sections four (2.2.4) and five (2.2.5), respectively, two papers of arguably opposing views are presented: first Queriquelli & Moura (2021), who suggest that BP is increasingly behaving in a satellite-framed manner (while making no claims regarding the state of EP), and then Batoréo & Ferrari (2016), who claim that EP is behaving increasingly in a satellite-framed manner, and that BP is more stable in its verb-framed patterns. Pleonasms are defined and contextualized in part six (2.2.6), and the final section (2.2.7) presents a selection of pleonastic BP and EP constructions from the corpus, which provide an overview of the data in the study.

2.2.1 Framing typology

In the second volume of his seminal work *Toward a Cognitive Semantics*, Talmy (2000) defines a Motion event as encompassing either a situation that contains motion, or a situation of prolonged stationary location. Perhaps confusingly, Motion (with capital M) is itself further presented as one of the four internal components that make up a Motion event – all together, the four internal components are: Figure, Ground, Path, and Motion. Figure is the object that moves in relation to another reference object, or Ground. Path is the trajectory followed, or

site occupied by the Figure, whereas Motion is the presence of motion itself, or locatedness in the event. This study focuses on the former (motion itself), which is differentiated from the latter (locatedness, or stationary location), by containing the occurrence (rather than the non-occurrence) of translational motion, which is presented as "motion in which the location of the Figure changes in the time period under consideration" (Talmy 2000: 25). Translational motion thus excludes such phenomena as rotation, oscillation and other forms of self-contained motion. Adding to the four internal components of the Motion Event (Figure, Ground, Path, and Motion), there is the *external* Co-event, which is argued to be linked to either Manner or Cause, in most cases (Talmy 2000: 26). Example (3) below illustrates two cases in English where the Motion component is conflated with an external Co-event – Manner in (3a) and Cause in (3b).

(3)   Conflation of Motion and Manner (a) and Motion and Cause (b) in English (Talmy 2000: 26)

    a) *Manner*: The pencil rolled off the table

    b) *Cause*: The pencil blew off the table

Based on the English examples in (3) above, Talmy (2000) presents Manner as being expressed in *rolled,* and Cause in *blew*, while *off* expresses Path in both (3a) and (3b) (p. 26). This pattern of Motion and Co-event being conflated in the verb root is argued to be characteristic for English, where characteristic is defined according to Figure 1 below.

Figure 1: Listed features that define 'characteristic' according to Talmy (2000: 27).

    (i)   *colloquial* in style, instead of literary or "stilted"
    (ii)   *frequent* rather than occasional in occurrence in speech
    (iii)   *pervasive*, in the sense of "a wide range of semantic notions [being] expressed in this [lexicalization] type"

Judging from the definition of 'characteristic' in Figure 1 above, Talmy (2000) is concerned with *general* typological patterns, that are both wide spread and casual in their form of expression, rather than highly local or literary forms of expression. In contrast to the English examples in (3), another typological pattern of the expression of motion is the type that conflates Motion and *Path* in the verb root. Such a conflation pattern is typical of the Romance language branch (of which Portuguese is a part of) and is illustrated in the Spanish examples in (4) below (Talmy 2000: 49).

(4)   Spanish expressions of Motion conflated with Path in the verb [glossing and translation from source text] (Talmy 2000: 49)

    a) La botella entró   a la   cueva   (flotando)
        The bottle   MOVED-in   to the   cave   (floating)
        "The bottle floated into the cave"

    b) La botella salió   de   la   cueva   (flotando)

11

The bottle   MOVED-out from the   cave   (floating)
"The bottle floated out of the cave"

The difference between the two patterns is noticeable in the Spanish examples contrasted with their respective English translations, where the parenthesized and optional gerund expression of Manner in Spanish - *flotando* 'floating' – is expressed as the obligatory main verb in English, *floated*. The optional Spanish expression of Manner is what Talmy (2000) calls a *satellite*. Satellites are understood to be any closed class elements "that relate to the verb root as a dependent to a head" (p. 102). While the Manner expression *flotando* 'floating' is the satellite in the Spanish expression in (4), the Path *off* (in *rolled off* and *blew off*) is the satellite in the English examples in (3). With their comparatively inverted encoding of semantic components across constituents in Motion event constructions, English and Spanish are thus claimed to belong to typologically separate framing types. Languages that belong to the type exemplified in the English example (3) – with Path expressed in the satellite – are known as satellite-framed, whereas languages that belong to the type exemplified in the Spanish example (4) – with Path expressed in the verb – are known as verb-framed (Talmy 2000: 131).

As a final comment for this section it is worth mentioning that Talmy's (2000) focus on the verb root rather than the verb is not considered relevant in this study. While Talmy (2000) argues that focusing on the verb root allows for the possibility of comparing typologically distinct languages, the analysis and comparison of semantic components in BP and EP motion event constructions in this study analyses the frequency of entire sequences/constructions and does not go down to the morphological level (see Method section 3.2 for further description) (p. 27). The next section (2.2.1.1) will describe the Boundary crossing constraint, which is argued to be a defining characteristic of verb-framed languages, such as Portuguese.

2.2.1.1   Boundary crossing constraint

Within the framing typological classification of languages, there are certain features considered characteristic, and/or indeed definable for the proposed framing types. The Boundary crossing constraint is one such well studied feature, whereby verb-framed languages have been claimed to not permit the use of Manner verbs in motion event constructions where a boundary is crossed (Cifuentes-Férez & Molés-Cases 2020: 89). Example (5) below illustrates how the 'same' boundary crossing event is encoded differently in English, German, and Spanish.

(5)   Typological contrast of boundary crossing events in English, German and Spanish (Cifuentes-Férez & Molés-Cases 2020: 90)

    a) English
        The bird flew out of the cage

    b) German
        Der   *Vogel*   *flog*   *aus* *dem*   *Käfig*
        The.M   bird   fly.PST.3SG from the.M.DAT cage

    c) Spanish
        El   *pájaro* *salió*   *de*   *la*   *jaula (volando)*

The.M   bird   exit.PST.3SG   from the.F cage (fly.GER)

As can be seen in (5), in the verb slot, where the satellite-framed languages English and German express Manner of motion (*flew*/*flog*), Spanish expresses Path (*salió* 'exited'), and reserves the encoding of Manner for the optional satellite gerund *volando* 'flying'. While Spanish is one of the more well-studied languages with regard to the Boundary crossing constraint, within a framing typological framework, no literature that specifically mentions Portuguese in that regard was found.

Before moving on to the framing typological studies that are specifically centered on BP and EP, another (non-framing typological) take on how to understand and analyze expressions of motion in languages of the world will be accounted for. The next section (2.2.2) describes Distributed spatial semantics, which contributes to the possibility of a more nuanced discussion with its distinct approach to semantics.

2.2.2 Another perspective on the semantics of motion

While Motion Events, from a framing typological point of view, is the central perspective in the current comparison between BP and EP (and the respective literature included on the two varieties), it is far from the only framework on the semantics of linguistic constructions involving motion and spatial relations. Distributed spatial semantics, as advocated for by Sinha & Kuteva (1995), provides another angle on how to understand and describe the semantics of the spatial relationship between Figure and Ground. Even though they use Langacker's terminology of *trajector* and *landmark* in their stead, they do comment upon them as being alternative terms to Talmy's Figure and Ground. For the sake of cohesion with the rest of the investigation Talmy's terminology is adopted here (Sinha & Kuteva 1995: 168-9). The following section briefly outlines Sinha & Kuteva's (1995) description of what they denote as the 'localist approach' in studies on the semantics of locative particles (such as prepositions in English), and subsequently describe, with the use of cross-linguistic examples, how their distributed spatial semantics stands as an alternative typological candidate to said local approach to meaning (p. 167).

The local form of semantics, that Sinha & Kuteva (1995) problematize, is one where closed class spatial grams are construed and analyzed as the sole carriers of semantic locative information. Although such a position, they claim, is rarely argued for explicitly, the consequences of its assumptions are widespread in the body of linguistic literature on locatives (Sinha & Kuteva 1995: 167-68). At the same time, there is an "often-remarked strong dependence of the contextual meaning of the locative particle upon the meanings of syntagmatically co-present items." (Sinha & Kuteva 1995: 168). Thus, instead of paying attention exclusively to the individual meaning of individual spatial grams, Sinha & Kuteva (1995) rather have in mind a categorization of languages based on whether the expression of spatial relational meaning (i.e., the spatial relation between Figure and Ground) is *overtly* or *covertly* distributed across simultaneous selections from both closed and open form classes (p. 195). A language with an *overtly* distributed spatial semantics is Ewe – as illustrated in (6) below –, and a language with a *covertly* distributed spatial semantics is English, exemplified in (7) (Sinha & Kuteva 1995: 181).

(6)   Ewe (Sinha & Kuteva 1995: 189)

*Agálă **do   le   do   me***
Crab **exit   be.at** hole **inside**
Lit: The crab exited at the inside of the hole
'The crab has got out of the hole'

(7)   English (Sinha & Kuteva 1995: 181-2)

The boy jumped OVER the fence
The girl swam ACROSS the river
The plane broke THROUGH the sound-barrier

The bold segments in (6) are explicitly (overtly) specified spatial grams whereas the capitalized adverbials in (7) represent optional specifications. The optionality of the explicitly spatial elements in (7) indicate a distribution of the spatial relational meaning across the surrounding Ground and Figure NPs, which are open class items rather than closed class 'spatial grams'.

Although no specific threshold is presented for determining whether a language is overt or covert in its distribution of spatial relational information, there are indeed notable differences between Ewe and English. While Ewe explicitly encodes the spatial relational information across several (obligatory) grammatical elements – *do* 'exit', *le* 'be.at', *me* 'inside' – the English sequences contrastingly presents optional adverbials. In the latter case of English then, the Path-profile of the motion can be recovered from the verb and the ground NP combined and can thus be said to be covertly distributed across the sequence (Sinha & Kuteva 1995: 182).

The framing typological understanding of how semantic components are encoded in motion event constructions appears to be predetermined to if the semantic components are mapped onto either open class lexical items (e.g., verbs) or closed class grammatical items (e.g., satellites). As illustrated above, the Distributed spatial semantics approach bypasses such restrictions, through scrutiny of *across which components* (obligatory or optional) *is the motion event actually expressed*. While the two stances will be contrasted further in the Discussion section (5), the next section (2.3.4) consists in a short overview of literature of framing typology and Portuguese.

2.2.3 Previous studies on framing typology and Portuguese

This short section serves as a brief account to represent in what sorts of framing typological literature that Portuguese has been studied or included in. There indeed appears to be more accounts on BP than EP. Indeed, apart from the study by Batoréo & Ferrari (2016), which will be described in section 2.2.5 below, all investigations that were found concerned BP. A noted similarity across the works is a critical stance to the dichotomous classification of languages as either verb-framed or satellite-framed. Santos Filho (2017) discusses the difficulty of classifying languages into any of the two categories verb-framed and satellite-framed, due to the within language variation of lexicalization patterns. As an illustration of this variation, he presents examples of satellite framed patterns in BP that make use of Path prefixes, such as *ex-* in *expulsos* 'expelled' (Santos Filho 2017: 113-15). The same author devotes attention to

similar BP satellite framed patterns in his doctoral thesis *Padrão tipológico do português: Um estudo dos vestigios de satélites na expressão do movimento e do trajeto* (Santos Filho 2013). In a similar vein, Meirelles (2019) mentions that BP cannot be considered to "present a definite typological pattern" (p. 1101). She claims that the forms of expressing trajectories in BP has to do with wider syntactic restrictions rather than the semantic component of movement (Meirelles 2019: 1103). A thesis by Rodrigues da Silva Júnior (2015) approaches BP motion events from a generativist framework. While it is beyond my competence to fully apprehend such a perspective within the scope of this thesis, it suggests a third lexicalization pattern type that conflates both Manner and Path. The next section (2.2.4) presents Queriquelli & Moura's (2021) claim that BP is showing signs of turning into a satellite-framed language variety.

2.2.4 Brazilian Portuguese as satellite-framed: Queriquelli & Moura

Basing their work on the framing typology of Talmy – described in section 2.2.1 above – Queriquelli & Moura (2021) suggest that "[t]he contemporary Brazilian vernacular […] is in the process of becoming a satellite-framed language" (p. 6126). Although they do not present any systematic results that back up their claims, their study aims at investigating *why* Brazilian Portuguese (BP) is changing, and whether it is typologically returning to the satellite-framed standard of Latin, or if it is the case of a new satellite-framed standard. Further, there is an expressed intent to "in an essayistic way" explore the encoding of Path and Manner in BP, Old Portuguese, and Latin, as well as to look at possible alternatives that could explain the supposed pattern change in BP (Queriquelli & Moura 2021: 6126). From the onset, the authors refrain from glossing any examples presented and only sporadically provide translations of the non-English excerpts in the text. The BP and English sentences in (8) below (from a section titled *starting from consensus*) are provided as illustrative of the framing typology divide between verb-framed and satellite-framed languages.

(8)　Brazilian Portuguese [own gloss] (Queriquelli & Moura 2021:6126)

 a) *Mas a janela abaixou deslizando e eu fiquei preso para fora*
  But the window went.down sliding and I left prisoner to outside

 b) 'But the window <u>slid down</u>, and I was trapped outside'

While the English sentence in (8b) is presented as having been retrieved from the Corpus of Contemporary American English, there is no source given for the BP sentence in (8a). As a Portuguese speaker of a 'moderate level of proficiency', the BP sentence is perceived as a Portuguese – almost word by word – mirroring of (8b). Origin aside, the English example is described as expressing manner of movement by means of the verb root *slid*, and direction in the satellite *down*, while the Portuguese equivalent marks direction at the verb root *abaixou* 'go.down', and manner of movement in the satellite *deslizando* 'sliding'. Through these two examples and a claim about general consensus on the matter, English is presented as a typical satellite-framed language, and Portuguese as a typical verb-framed language (Queriquelli & Moura 2021: 6126).

The authors then proceed to mention that Latin also was a satellite-framing language. In line with the presentation of the previous examples, excerpts in Latin are then listed alongside Portuguese and English translations. The Portuguese translations are here claimed to be from translations of Julius Caesar and Cicero, while the source of the English sentences is unspecified (Queriquelli & Moura 2021: 6126). The Latin examples, and their corresponding translations, are presented as in (9) and (10) below:

(9)   Latin [own gloss based on https://en.wiktionary.org/wiki/Latin] (Queriquelli & moura 2021:6126)

Ipse   omnes   copias   castris   e-   duxit
Himself   all   troops   military.camp out/away- lead

(10)   Portuguese [own gloss] (Queriquelli & Moura 2021:6126)

Ele mesno   retirou   todas as   tropas do   campo da batalha
He himself removed all   the troops of.the field   of battle

'He himself removed [or moved away] all the troops from the battlefield'

Claiming that the motion events in the Portuguese translations increasingly vary with other expressions (and these alternative expressions are simply listed in the text, without making any explicit connections between the translations and the listed expressions), it is subsequently abruptly stated that the contemporary Brazilian vernacular is in the process of turning into a satellite-framed language "just as Latin was" (Queriquelli & Moura 2021: 6126). The use of *just as* is especially interesting here, since it is later stated in the text that the encoding of Path and Manner in the satellite-framing structures of BP and Latin are indeed different (Queriquelli & Moura 2021: 6134). Another phenomenon claimed to be indicative of the supposedly emergent satellite-framed behavior of BP is constructions such as *descer pra baixo,* and *subir pra cima* – translated as 'descend down' and 'ascend up' respectively. A contextualized example of this type of expression, which is further described to be considered 'redundancy addictions' by prescriptive grammars, is shown in (11) below (Queriquelli & Moura 2021: 6127).

(11)   Vernacular Brazilian Portuguese [own gloss] (Queriquelli & Moura 2021:6127)

O   fogo subiu   pra   cima da   panela
The   fire ascend.PST.3SG   to   up   the   pan
'the fire ascended up to [above?] the pan'

The translation provided in the text is: 'the fire "*ascend up*" to the pan', but I consider it to not capture the sense of the Portuguese source text.

Without any further explanations by the authors, (11) and the similar expressions that are listed with it, seem to be relevant in being expressions that are deemed redundant by prescriptive grammars, and therefore signal that it is mainly in the vernacular varieties of BP that the change to satellite-framed structures is happening (Queriquelli & Moura 2021: 6127). For clarification, in the motion event construction in example (11) both verb root and satellite express 'the same' Path information. Both the verb *subiu* 'ascend' and the satellite *pra cima*

'(to) up' denote translational movement in an upwards direction, although this is not explicitly mentioned in the text, nor any comments whatsoever concerning what seems to be *double Path marking* are provided.

It is not completely clear to what degree the authors consider the satellite-framed patterns to be of relevance *only* for vernacular varieties of BP. On the one hand they use 'vernacular Brazilian Portuguese' and 'Brazilian Portuguese' quite interchangeably throughout the work, as well as – as mentioned above – stating that it is *mainly* in the vernacular varieties that the framing change is occurring, thus not excluding the same, or similar motion event framing tendencies in other varieties or registers (Queriquelli & Moura 2021: 6127).

2.2.4.1   A comment on vernacular Brazilian Portuguese

Since Queriquelli & Moura (2021) suggest that it is mainly in the vernacular varieties of BP where the shift from verb-framed to satellite-framed motion event patterns is taking place, this section provides an added perspective on what is, or could be meant by vernacular BP (VBP) (p. 6126). Guy & Zilles (2008) present VBP (although they call it PBP, standing for 'popular Brazilian Portuguese') in a context of language loss and endangered varieties (p. 53). Since the distinction between language and dialect is not clear cut, and "any language variety that constitutes the verbal heritage of some speech community and experiences assimilatory contact with a dominant or standard language is potentially at risk", VBP is put forth as a language variety of low status, regionality, exclusion and ethnic minorities, when contrasted with the Portuguese standard (Guy & Zilles 2008: 53). In terms of linguistic characteristics, variable agreement marking is described as a distinctive feature of VBP, meaning that the categorical agreement of for example number and person between subject and verb is not obligatory, as it is in standard Portuguese, which is illustrated in (12) below (Guy & Zilles 2008: 55).

(12)   Vernacular Brazilian Portuguese (Guy & Zilles 2008: 56)

Os	***pai***	*dela*	*foru*	*escravo, que*	*eles*	*eram*	***os***	***negros.***
The.PL	parent.Ø	of.her	were.PL	slave.Ø, for	they	were.PL	<u>the</u>	<u>blacks.PL</u>.

Os	***preto***	*na época*	*de cativeiro*	*era*	*escravo.* ***Os***	***branco***	*que*
The.PL	black.Ø	in time	of slavery	was.SG	slave.Ø. The.PL	white.Ø	who
era	*senhor*						
was.SG	master.Ø						

'Her parents were slaves, because they were the blacks. The blacks in slavery-days were slaves, it was whites who were the masters.'

In the text segment in (12) there are four two-word noun phrases, out of which one shows standard agreement: *os negros,* the blacks (underlined in the example). In the remaining three two-word noun phrases all definite articles *os* 'the' express plurality, while the head nouns lack the final 's', which expresses plurality in standard Portuguese – thus, there is no number agreement (Guy & Zilles 2008: 56). Although this study does not explicitly investigate VBP, but rather the standard varieties Brazilian Portuguese (BP) and European Portuguese (EP), it is important that the reader is aware of the existence of inter-variational differences such as

those mentioned in this section. After all, no measures are taken to methodically exclude VBP (or other Portuguese non-standard varieties) from the investigated data. See Section 3 for description of method and data. Moving on from Brazil to Europe, the proceeding section (2.2.5), outlines Batoréo & Ferrari's (2016), claim that EP is behaving increasingly in a satellite-framed manner.

2.2.5 European Portuguese as satellite-framed: Batoréo & Ferrari

Quite to the contrary of the suggestions made in the previous section, Batoréo & Ferrari (2016) propose that European Portuguese (EP) appears to be "moving towards satellite-framing", while Brazilian Portuguese (BP) is described as having more stable verb-framed patterns (p. 59). Their study focuses on the distribution of the Path component in what they denominate as COSTA-motion event (henceforth CME) constructions, which describe movement along or towards the coast or a basin of water. The verb-framed versions of CME constructions, that one would expect from a prototypically verb-framed Romance languages, such as Portuguese, are exemplified in (13) below (Batoréo & Ferrari 2016: 66).

(13)   Portuguese [own glosses] (Batoréo & Ferrari 2016: 66)

margear / marginar
a. *As   casas   que   margeavam   o   rio   foram   inundadas*
    The.PL houses   that line.IMPF   the   river were   flooded
    'the houses that lined the river were flooded'
b. *As   casas   marginavam o   rio   até   à   foz*
    The.PL houses   border.IMPF   the river until to.the (river.)mouth
    'the houses bordered the river to its mouth'

The underlined verbs in the verb-framed example are denominalizations from COSTA nouns – *margem* 'border' > *margear/marginar* 'to line/border' – in which Motion and Path are conflated. While the verb-framed pattern might be the only expected one, from a strictly Talmyan point of view, satellite-framed constructions are also attested in the data. The satellite-framed constructions conform to the periphrastic structure of '[Verb + Preposition + COSTA noun]' and are exemplified in (14) below (Batoréo & Ferrari 2016: 70).

(14)   Portuguese [own gloss and encoding of motion event components] (Batoréo & Ferrari 2016: 70)

O   barco   aproximou-se   [da   costa.]
The   boat   approach=REFL of.the   coast
                VERB   [PREP   COSTA.NOUN]
'The boat approached/got closer to the coast'

Underlined = motion event construction
Within square brackets [] = satellite

As indicated by the underlined and bracketed motion event construction elements in (14), the Path component is expressed in the periphrastic satellite (Batoréo & Ferrari 2016: 70). There is no mention of the fact that all satellite-framed example sentences also seem to conflate Motion and Path at the verb root, which would make them structurally similar to the double

Path constructions, such as *entrar pra dentro* 'enter inside'*,* and *subir pra cima* 'go.up up' in Queriquelli & Moura (2021), described in the previous section. While both verb-framed and satellite-framed patterns are attested in BP and EP CME constructions, there appears to be a significant difference in how the patterns are distributed across the varieties in the Linguateca corpus, which constitutes the data of the study. In the case of CMEs in Linguateca, BP shows a preference for verb-framed constructions. Each of the top three most common COSTA terms investigated (*costear*/*costa, margear*/*margem,* and *beirar*/*beira*), appear over the 50% mark in the denominal form – 53% for *costear,* 58% for *margear*, and 59% for *beirar*. In EP, the denominal version of the construction respectively occurs 38%, 42% and 6%. The numbers are further contextualized in Figure 2 below, which contains two result graphs from the text, with frequencies for both denominal as well as periphrastic instances (Batoréo & Ferrari 2016: 74).

Figure 2: Contrastive frequencies between BP and EP of (a) most common verb-framed costa-motion event patterns and (b) most common satellite-framed costa-motion event patterns (data from Batoréo & Ferrari 2016: 74).
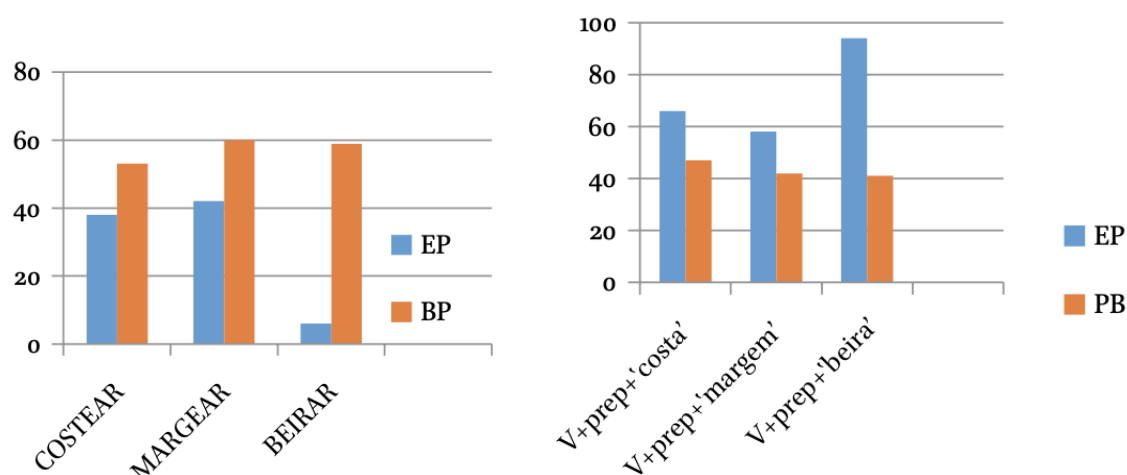
Since there are two identified pattern options for the CMEs: what EP lacks in denominalization occurrences, it 'makes up' for in periphrastic constructions. The case of V + Prep + *beira* stands out, by accounting for 94% of all EP *beirar/beira* instances, while it is 66% for the periphrastic alternative of 'costa', and 58% for 'margem' (Batoréo & Ferrari 2016: 75). In conclusion, Batoréo & Ferrari's (2016) study finds that verb-framed and satellite-framed strategies are used in both varieties, that the verb-framed strategy is 'privileged' in BP, and that the satellite-framed strategy is significantly more prevalent in EP (75)

Considering that BP is the variety spoken by considerably more people, it strikes me as odd that the data for the study is: *CETEMPúblico*, with 191 million words, for EP, and *NILC/São Carlos*, with 32 million words, for BP (Batoréo & Ferrari 2016: 60). Regardless of the *level* of contrast between BP and EP, and the perhaps questionably narrow scope of CMEs, the study by Batoréo & Ferrari (2016) undoubtedly indicates that both varieties make use of verb-framed and satellite-framed patterns, and that there indeed appears to be some sort of difference in how some motion events are encoded. It remains rather unclear, however, how relevant these particular motion events are for a more general picture.

20

2.2.6 Pleonasms and redundancy

When mentioning the attitude of prescriptive grammars with respect to the double Path expressions, that they claim to be indicators of vernacular varieties of Brazilian Portuguese (BP), Queriquelli & Moura (2021) bring up 'redundancy' (p. 6127). Although semantically close to 'redundancy', 'pleonasm' is however considered a more appropriate term for the task at hand, since it lacks the information-theoretic judgement borne by the former, implying that a message contains elements that contribute with *no new information*, which is not already expressed in the rest of the message (Lehmann 2005: 2). According to Lehmann (2005), a pleonastic expression – generally speaking – typically contains two constituents, and one of the constituents is entailed by the other, in the sense that the meaning of one constituent is part of (and thus entailed by) the other (p. 3). Formally, this relationship is given in the following definition: "an expression $E_1 + E_2 \ldots E_n$, is **pleonastic** iff it contains a meaning component F that is included in the meaning of more than one E_i." (Lehmann 2005: 4). In (15) below, *back* is included in the meaning of *return*, and this shared semantic component is referred to by Lehmann (2005) as the *focal component*, which is underlined in the example (p. 3).

(15)   Illustration of a focal component by itself (i) and in a pleonastic expression (ii). The focal component is underlined [own adaptation of formulations in the text] (Lehmann 2005:3)

(i)   *Return* = 'go back'
(ii)   *Return back* = 'go back back'

Lehmann (2005) proceeds to state that there are three generic syntactic relations in the "higher levels of grammar" between the focal component and the unit that shares its semantic expression, namely: (i) *sociation*, in which the two elements are typically synonyms, (ii) *government*, where the dependent is selected by the head (such as in *dream a dream*) and (iii) *modification*, which consists of a syntactically optional modifier that is semantically redundant. Modificative constructions make up the core of pleonastic constructions (Lehmann 2005: 5-6). Example sentence (ii) in (15) above is an example of a modification, which as a notion also appears to apply quite well to the 'redundant' (vernacular) BP expressions listed in Queriquelli & Moura (2021). Example (16) illustrates a Portuguese motion event modification, where the focal component is in bolded text and the entire pleonastic expression is underlined.

(16)   Vernacular Brazilian Portuguese [own gloss and translation] (Queriquelli & Moura 2021:6127)

*Ele   pegou   minha   mochila   e   **saiu**   pra   fora   da   sala*
He   picked.up   my   backpack   and   **exited**   to   outside   of.the   room
'he picked up my backpack and exited (to the outside) of the room'

As this section hopefully has made clear to the reader, approaching the Portuguese double Path constructions as *pleonasms* rather than *redundancies* is to acknowledge a tight semantic relationship between the verb and the satellite – *saiu* 'exited' and *pra fora* 'to out' respectively – without making any claims regarding to what degree any of the constituents

contribute with meaningful information. In order to get more familiar with some of the contexts that BP and EP pleonastic double Path constructions are used in, the next section (2.2.) is an overview of selected examples from the ptTenTen11 corpus.

2.2.7 Selected examples from the data: an overview of Portuguese double Path constructions

The upcoming section presents a brief and qualitative overview, with illustrative examples, of what type of data the study deals with. While quantitative calculations, and explicitly numerical comparisons are reserved for the results section, these selected examples constitute snippets of the semantic continuum, and range of contexts of use of the intransitive pleonastic motion event constructions in BP and EP. The examples were selected from randomized selections from the data through the Sketch Engine interface – a process that is further elaborated on in the Method section (3).

Several of the retrieved results proved to be from football commentary contexts. These sports related contexts further constitute rather prototypical examples of the Figure (in most cases a ball) being described *while in translational motion*. (17) and (18) below, illustrate football commentary contexts from BP and EP respectively, where the football is described as either going out of the field, or going out of a specific zone on the field.

(17)   Brazilian Portuguese (own gloss and translation)

A	*bola*	*saiu*	*pra*	*fora,*	*houve*	*uma*	*chute do*	*jogador do*	*Sport*
The	ball	exited	to	out,	COP	a	shot of.the	player of.the	Sport

no	*Saci*
on.the	Saci

'The ball <u>went out</u>, there was a shot from the Sport [football team] player on Saci [football player]'

(18)   European Portuguese (own gloss and translation)

Douglas	*vai*	*buscar*	*a*	*bola*	*à*	*linha,*	*mas*	*o*	*remate*	*sai*	*para*	*fora*
Douglas	goes	get	the	ball	to.the	line,	but	the	shot	exits	to	out

da	*pequena*	*área*	*e*	*permite*	*que*	*Jorginho*	*corte o*	*passe*
of.the	small	area	and	lets/allows	that	Jorginho	cut the	pass

'Douglas goes to get the ball at the line, but the shot <u>goes outside</u> of the small zone/area and lets Jorginho cut the pass'

Both examples include the most frequent double Path constructions in the corpus, *sair para fora* 'exit (to) out' and are thus Boundary crossing events – the ball is described as *crossing a physical boundary* on the field. Apart from the prototypical, and concrete, Boundary crossing events in (17) and (18), *sair para fora* 'exit (to) outside' was also identified as occurring in metaphorical contexts, where *imaginary boundaries* are crossed, such as the ones in (19) and (20), both from EP.

(19) European Portuguese (own gloss and translation)

Amor: deixe o   seu lado mais louco sair para fora e   divirta -se
Love: let   the your side most crazy exit to out and have.fun=REFL.3SG

'Love: let your most crazy side <u>come out/forth</u> and have fun'

(20) European Portuguese (own gloss and translation)

Aos   18 anos,   decidiu elaborar um projeto dirigido aos jovens dos
at.the 18 years,   decided elaborate a   project directed to.the young   of.the
bairros   sociais   – "dar -lhes   algo   que os fizesse
neighborhoods social.PL   – "give =them something that them make.IMPF.3SG
<u>*sair para fora* da</u>   *sua realidade e   acreditar que era possível*
<u>exit to   out</u> of.the   their reality   and believe   that was possible
mudar   de vida"
change of life"

'at age 18, (he) decided to elaborate a project directed to the young/youth of the social (socially vulnerable?) neighborhoods – "give them something that would make them <u>come out of</u> their reality and believe that it was possible to change (way of) life."'

In (19) the double path construction is part of an imperative statement, where one person encourages another to let their 'crazy side come out' of/from their usual self, whereas (20) concerns coming out of a problematic/socially vulnerable reality. Although a motion verb construction, *sair para fora* 'exit (to) outside' was also identified in expressing a figure being in one locative state, as contrasted with another state, or perhaps more accurately, going from one state to another – from in/small-state to out/enlarged-state. The recipe excerpt in (21) is an example of such a context, where the batter/dough/that which is supposed to be poured into the molds is described as *saem para fora* '[they/the molds] exit (to) outside' in conjunction with the verb growing: 'they will grow and stick/extend out (of the mold?)'.

(21) European Portuguese (own gloss and translation)

Coloca -se   em forminhas de papel (enche -se   as fominhas
Placed =REFL.3SG in   molds.DIM of paper (fill   =REFL.3SG the molds.DIM
só   até meio porque depois crescem e   <u>saem para fora</u>).
only until half because later   grow   and <u>exit to   out</u>).

'(They are) placed/put in small paper molds (the molds are only to be half filled/filled until half because later they (the mold content) grow and <u>stick/extend out (of the mold?)</u>).'

It is not completely clear to what extent there is a boundary being crossed in (21). After all, an augmentative interpretation of *sair para fora* 'exit (to) outside' as reinforcing *crescem* or raising caution concerning the possibility of the content becoming too large, is indeed conceivable. Lastly, for the identified contexts of *sair para fora* 'exit (to) outside', a contrastive (with in/inside) use was noted as well. (22) is an illustration of such as context, where 'exit outside of the church' in order to 'bring people inside' is the contrastive case, which has been highlighted in the example.

(22) Brazilian Portuguese (own gloss and translation)

É que, independente da realidade, a resposta do discípulo consiste em
Is that, independently of.the reality, the answer of.the disciple consists in
ser missionário, isto é, <u>sair</u> <u>para</u> <u>fora</u> da Igreja para trazer as
being missionary, that is, <u>exit</u> <u>to</u> <u>out</u> of.the church to bring the
pessoas <u>para</u> <u>dentro</u>
people <u>to</u> <u>inside</u>

'It is such that, independently of reality, the answer of the disciple consists in being a missionary, that is, <u>go outside of the church (in order to) bring people inside of the church</u>'

While more time could be dedicated to the examples with *sair* 'exit', we now turn to the second most common construction, *entrar para dentro* 'enter (to) inside'. For this construction, all 10 randomly sampled contexts for EP (section 3.4 below describes the sampling procedure) were syntactically structured with *de* 'of' directly following the double PATH construction, in the following way: *entrar para dentro de X* 'enter (to the) inside of X'. Thus, these contexts all explicitly denominate the GROUND enclosure that the FIGURE enters, inevitably lending for a boundary crossing semantic interpretation, as illustrated in (23) and (24) below. While not all *entrar* 'enter' double PATH constructions in BP were of the explicit boundary crossing kind, 7 out of 10 in the sample indeed were.

(23) Brazilian Portuguese (own gloss and translation)

quando não há quase mais ninguem. Então entr -o para
when no COP almost more no.one. Then enter -1SG.PRS to
dentro de mim e adormeç -o
in of me and fall.asleep -1SG.PRS

'when there is almost no one left. Then I <u>go inside</u> of mine/my place and fall asleep'

(24) European Portuguese (own gloss and translation)

E se cada aluno pudesse entrar para dentro do livro de história e
And if each pupil could enter to in of.the book of history and
encarnar uma personagem?
embody a character?

'what if each pupil could enter (to the inside) of the history book and embody/become a character/person?'

As (24) shows, there were attested imaginary boundaries amongst the double PATH constructions with *entrar* 'enter' as well. Although my Portuguese language intuitions are not those of a native speaker, I do read *dentro do* 'inside of' as a fundamental part in the interpretation of *actually entering the story* of the book and become/embody a character, as contrasted with *being in the boo*k in the sense of being mesmerized by it, completely caught up with it or lending one's full attention to it. Further, interestingly, one of the contexts that does not contain *de* (of) and a subsequent GROUND specification directly after the double PATH construction is a metalinguistic comment on not only *entrar pra dentro* 'enter (to) inside', but also on *subir pra cima* 'go.up (to) up', *descer pra baixo* 'descend (to) down', and *sair pra fora* 'exit (to) outside'. In the comment, which is displayed in (25) below, the just listed

expressions are used as reference examples of pleonasms, when (or so it seems) condemning the 'pleonastic' expression *interagem junto* '(they) interact together'.

(25) Brazilian Portuguese (own gloss and translation)

As	*paredes*	*interagem*	*junto?*	*Isso*	*é*	*assombração*	*pleonástica!*	*Só*
The	walls	interact	together?	That	is	haunting	pleonastic!	Only

pode	*ser!*	*Pleonasmo,*	*se*	*você*	*não*	*está*	*ligando*	*o*	*nome*	*à*	*pessoa,*	*é*
can	be!	Pleonasm,	if	you	NEG	is	attaching	the	name	to.the	person,	is

aquela	*figura*	*de linguagem*	*que*	*soube*	*pra*	*cima,*	*desce*	*pra*
that	figure	of language/speech	that	go.up	to	up,	descend	to

baixo,	*entra*	*pra*	*dentro*	*e*	*sai*	*pra*	*fora.*
down,	enter	to	in,	and	exit	to	out.

'The walls interact together? That is a pleonastic haunting/ghost(?)! It can only be! Pleonasm, if you aren't calling/attaching the name to the person, it's that/the same figure of speech as <u>go up (to) up, descend (to) down, enter (to) in, and exit (to) out</u>.'

When it comes to lemmas of *subir para cima* 'go.up (to) up', it is the construction that displays the largest divide between the two regional Portuguese varieties, amongst the investigated double path constructions (which is shown in section 4.1 further down). In EP all contexts, except for one, in the randomly retrieved 10 context sample had an explicit ground specification, with *de* 'of'. (26) below, is an example of one such instance. In contrast to EP, only one of the instances in the 10 context sample appears with the preposition *de* 'of' after the *subir para cima* 'go.up (to) up' construction in BP.

(26) European Portuguese (own gloss and translation)

Resson	*-o*	*um*	*bocad*	*-inho*	*=)*	*e*	*já*	*arranj*	*-ei*	*uma*	*nova*
Snore	-1SG.PRS	a	bit	-DIM	=)	and	already	arrange	-1SG.PST	a	new

Tecnica	*de*	*subir*	*para*	*cima*	*do*	*sofa*	*atraves*	*das*	*pernas*	*dos*
Technique	of	go.up	to	up	of.the	sofa	through	of.the	legs	of.the

don	*-inho*	*-s…*
owner	-DIM	-PL…

'I snore a little/itty-bitty bit =) and already found a new technique of going/getting up to the sofa through the legs of (my) cute little owners…'

The overview with 10 randomly retrieved samples per double Path construction for each Portuguese variety – some of which were presented in this section –, illustrate that there is a wide range of contexts that the double Path constructions are used in. The Boundary crossing contexts are of special relevance in this section, since their expression in contexts with Manner verbs and Path satellites are considered a classificatory characteristic for verb-framed languages, as was described in section 2.2.1.1 above. While the examples so far have all been double Path constructions, one part of the study consists of investigating the presence of Path satellites in conjunction with Manner verbs in BP and EP, which is a task that will be presented in the upcoming section (2.4) along with the other aims and research questions of the study. However, before the aims and research questions of the study are presented, this entire section will first be summarized (2.3).

2.3 Summary

Having two standards of what is recognized as one and the same language, such as BP and EP, implies that (some form of linguistic or cultural) differences undoubtedly are claimed to exist by the institutions and/or speech communities involved. While there indeed are recognizable lexical contrasts – such as second person pronoun choices – between BP and EP, the works by Queriqiuelli & Moura (2021) and Batoréo & Ferrari (2016) propose divergencies at the grammatical level, within the domain of motion event constructions, both based on the framing typological work by Talmy (2000). As outlined above, Queriquelli & Moura (2021) claim that there are indications of BP turning into a satellite-framed variety, while Batoréo & Ferrari (2016) suggest the opposite: EP is behaving in a more satellite-framed manner, while BP is more stable in its verb-framed structure. Since the literature suggest that prototypical Romance languages – Portuguese included – are verb-framed, and that Queriquelli & Moura (2021) mention redundancy when accounting for prescriptive perspectives on their exemplified double Path constructions, it is inevitable not to provide a perspective on said redundancy. The notion of *pleonasm,* rather than *redundancy,* is here argued to be more suitable as a working concept in formulating the aspirations of this project. According to Lehmann (2005), the former (redundancy) is rooted in an idea of an added element which contributes with no new information, while the latter (pleonasm) simply denotes such contexts where the meaning of one element is entailed by another element. A selection of examples from the ptTenTen11 corpus data clearly shows that the double Path constructions are used in a wide range of contexts, such as metaphorical, metalinguistic, and boundary crossing events. The latter (boundary crossing events) is central in the formulation of the Boundary crossing constraint, which has been claimed to hold for verb-framed languages. It states that Manner verbs 'aren't permitted' in contexts where the crossing of a boundary is expressed (Cifuentes-Férez & Molés-Cases 2020). Contrasted with the framing typological studies summarized above, Sinha & Kuteva (1995) present a different view on Motion and spatial relations. Their Distributed spatial semantics, problematizes and goes beyond the 'localist approach', which they claim is a widely held assumption that spatial grams carry all the locative information. Instead, they suggest that meaning is rather distributed across selections from *both* open and closed form classes, rather than being restricted to the latter.

Interspersed in the reviewed literature are both evident points of conflict, such as the different framing typological categorizations of BP and EP, and more suggestive differences in how to approach the domain of motion and spatiality in language, such as the indicated disagreements between Framing typology and Distributed spatial semantics. The proceeding section (2.4) embeds these considerations into the aims and research questions of this project.

2.4 Aims and research questions

While the central aim of this study is to compare Brazilian Portuguese (BP) and European Portuguese (EP) in terms of their respective frequency of intransitive pleonastic motion event constructions in the *ptTenTen11* corpus, a broader aim is to relate the Portuguese varieties to the framing typological framework. In the literature, there are opposing claims concerning the morphosyntactic and semantic systematicity of BP and EP motion event constructions, but a lack of systematic and quantitative data on the matter. Queriquelli & Moura (2021) propose

that there are indications of BP turning into a satellite-framed variety – indications that they reserve for future studies. Included in their illustrative examples of claimed satellite-framed tendencies within BP is a set of double Path constructions, that appear to be expressing 'the same' Path information in both the verb and the satellite, such as (27) below.

(27) Vernacular Brazilian Portuguese (own gloss and translation) (Queriquelli & Moura 2021:6127)

> *o fogo subiu*   *pra cima da   panela*
> the fire   ascend.PST.3SG   to   up   the   pan
> 'the fire ascended up to (above?) the pan'

> The translation provided in the text is: 'the fire "*ascend up*" to the pan', but I consider it to not capture the sense of the Portuguese source text.

The underlined elements in the Vernacular Brazilian Portuguese example in (23) are constituted by a verb (*subiu* 'ascend'), and a satellite (*cima* 'up') – interconnected with a preposition (*pra* 'to'). Semantically, and individually, both these elements express an upwards Path movement.

Batoréo & Ferrari (2016) suggest that BP is more stable than EP in its verb-framed encoding of motion event constructions, while EP increasingly makes use of satellite-framed strategies (p. 59). Their body of evidence is a rather restricted type of motion event construction, that they denominate CMEs, which describe movement along, or towards the coast or a basin of water (Batoréo & Ferrari 2016: 66). Two morphosyntactic strategies are described for the semantic expression of Portuguese CMEs: (i) verb-framed costa noun denominalization, as exemplified in (28), and (ii) a satellite-framed periphrastic construction, conforming to the structure 'Verb + Preposition + costa noun', as exemplified in (29) (Batoréo & Ferrari 2016: 70).

(28) Portuguese [own glosses] (Batoréo & Ferrari 2016: 66)

> *margear / marginar*
>   a. *As   casas   que   margeavam   o   rio   foram   inundadas*
>   The.PL   houses   that   line.IMPF   the   river were   flooded
>   'the houses that lined the river were flooded'
>   b. *As   casas   marginavam o   rio   até   à   foz*
>   The.PL   houses   border.IMPF   the   river until to.the   (river.)mouth
>   'the houses bordered the river to its mouth'

(29) Portuguese [own gloss and encoding of motion event components] (Batoréo & Ferrari 2016: 70)

> *O   barco   aproximou -se*   [*da   costa.*]
> The   boat   approach   =REFL of.the   coast
>   VERB   [PREP   COSTA.NOUN]
> 'The boat approached/got closer to the coast'

> Underlined = motion event construction
> Within square brackets [] = satellite

As for relating BP and EP (and Portuguese in general) to the framing typology of motion event constructions, there appears to be a lack of cohesive overview. While Queriquelli & Moura (2021) appear to be speculating, rather than systematically analyzing, and Batoréo & Ferrari (2016) scrutinize the rather small domain of CMEs, their respective works diverge in terms of which variety of Portuguese they claim to conform to what pattern type within the framing typological literature.

The Boundary crossing constraint is considered a prototypical feature of the verb-framed language neighbor of Portuguese, Spanish (Cifuentes-Férez & Molés-Cases 2020: 89). However, how this constraint applies to the situation of Portuguese itself seems to be an understudied area of investigation.

The aim of this study is to contribute to an overarching understanding of possible divergencies between BP and EP, within the domain of motion events, by quantitative corpus analysis of data crawled from Brazilian domain and Portuguese domain websites in the ptTenTen11 corpus. More specifically, the research questions are:

1. To what extent are verbs that conflate Motion and Path expressed with satellites that express 'the same' Path information in BP and EP? Is there a significant difference between the two varieties in this regard?

2. Does the boundary crossing constraint - which states that verb-framed languages do not allow for Manner verbs to occur in constructions where a (conceptual) boundary is crossed – apply to BP and/or EP?

3. If the results from double Path constructions and the Boundary Crossing Constraint are taken together, what do they, if anything at all, tell us about the relationship between BP and EP, and the verb- and satellite-framed dichotomy/continuum?

4. Taking a step further back, how central or peripheral are the double Path constructions? Are they widespread enough to be taken into consideration when discussing lexicalization patterns at a more general level or are they rather highly locally restricted phenomena?

Although all 4 research questions to some extent rely on quantitative data and calculations, research question 1 is the one that most directly is associated with quantitative answers. The main purpose of research question 2 is to test for a well-known and considered typical feature of verb-framed languages, in order to support with additional data, beyond the 'unclear' categorization of the double Path constructions. Both research questions 3 and 4 serve as 'contextualizers', relating the frequencies and considerations of the investigated constructions to the framing typological theory, as well as to the language studying endeavor. The next section presents the method and data used to try to reach the aims, as well as answer the research questions of the investigation.

3 Method and data

In this section, the data and the methodological steps taken will be accounted for in four main parts. In the first section (3.1) the corpus data source ptTenTen11 is introduced. The second section (3.2) describes the actual corpus search design and analysis. The third part (3.3) is dedicated to the statistical measures used in the investigation, and the fourth part (3.4) accounts for the used sampling process.

3.1 Data: PtTenTen11

The *Portuguese Web 2011 (ptTenTen11)* (www.sketchengine.eu) consists of over 4.5 billion tokens of data, mainly crawled from websites with Portuguese (.pt) and Brazilian (.br) domains. Thus, it is not restricted to a particular genre or register, but contains both formal and informal texts, ranging from blogs to government agencies. Given the aim of this thesis comparing BP and EP, it is suitable to use a corpus with the same architecture and corpus design for both varieties, since it most likely reduces the likelihood of including possibly 'skewing' factors that have to do with data collection and processing, and that might be difficult to identify.

3.2 Corpus analysis

This section outlines how the Sketch Engine corpus interface was navigated when searching for the desired data. First (3.2.1), a short introduction to CQL (Corpus Query Language) will be given, as well as how it is implemented in the search option of choice for this specific project. In the next section (3.2.2) the procedure of filtering out metalinguistic contexts is described. The third part (3.2.3) is an attempt at establishing a 'measuring stick' for identifying when BP and EP are behaving differently directly in the Sketch Engine interface. In part four (3.2.4) the design of the search strings used (along with motivations for them) are presented. Readers interested in a more in-depth account of the inner mechanisms of the corpus, and other corpora developed within the same framework are referred to Kilgarriff et al. (2014), where the team involved in the ptTenTen11 project describe the background to how it came to be and the methods applied when assembling it.

3.2.1 Sketch Engine and CQL

The Concordance search option is presented by Sketch Engine themselves as their most powerful search tool. It is in particular one feature that makes the Concordance search tool relevant for this study: the possibility it offers of working with CQL (Corpus Query Language). CQL designed search strings can be more complex by allowing for searches to be both less specified, and also have added optional criteria. An example of a 'less specified' search is the option of retrieving results of any token annotated with a certain part of speech tag (see 28b below). Included in a search string of several units, such a less specified element can, for instance, make it easier to identify patterns of what type of verbs occur in a particular

position with regards to another construction (which indeed is relevant when scrutinizing different parameters at play in a given syntactic assemblage, such as motion event constructions).

Beyond the Sketch Engine website itself (www.sketchengine.eu), the document *Cambridge Sketch Engine: Using Corpus Query Language* (2012) was a useful source when constructing search strings that matched aims and research questions of this project. The two search string examples in (30) below illustrate (in 30a) a search string matching research question 1, which asks about the frequency of double Path constructions in BP and EP, and (in 30b) one which is an integral part of research question 2, which asks if there are any boundary crossing events expressed with Manner verbs in BP and EP. Searching for the satellites alone retrieves a list of all verbs that occur with them, which allows for the possible identification of Manner verbs in boundary crossing events. Both strings in (30) include the satellite *para fora* 'to outside'.

(30) CQL search strings for (a.) intransitive pleonastic motion event construction *sair para fora* 'exit (to) outside', and (b.) any verb occurring with the satellite *para fora* '(to) outside'.

    a.   [lemma="sair"] [word="para|pra|prá|pa|pá"] [word="fora"]
    b.   [tag="V.*"] [word="para|pra|prá|pa|pá"] [word="fora"]

All search queries were designed according to the CQL structure displayed in (30), and thus consist of three bracketed/separate items: (i) verb slot, followed by (ii) any of the identified variants of the preposition *para* 'to', and finally (iii) an adverbial slot. (30b) Table 1 below illustrates how the three slots for each search string design represent the Portuguese framing typological componens of 'verb' and 'satellite'.

Table 1: The Framing typological components 'path' and 'satellite' as represented in the CQL search design of this project

	Verb		Satellite			
A) double Path construction	(i)	Lemma of Path verb, e.g., *descer* 'descend'	(ii)	Preposition 'to' in any of the forms identified in the data	(iii)	Path adverbial, e.g., *baixo* 'down'
B) 'bare' adverbial construction	(i)	Lemma of *any* verb, e.g., *correr* 'run'	(ii)	Preposition 'to' in any of the forms identified in the data	(iii)	Path adverbial, e.g., *dentro* 'inside'

Row A) represents each of the three specified search slots included in the search for double Path constructions, whereas row B) represents the three specified search slots for the search for the Path adverbials occurring with any verb.

Section 3.2.4 describes the *search procedure* that includes the presented search string designs in more detail. In the next section (3.2.1.1) the inclusion of various forms for the same preposition *para* 'to' will be explained, and after that (3.2.2) the issue of identifying and filtering out metalinguistic comments is accounted for.

3.2.1.1   A comment on *para* 'to'

It did not require much familiarization with the data at hand to notice the diverse forms used for the preposition traditionally known as *para* 'to'. Although *pra* is a well-known colloquial contraction of *para* in the Lusophone community, *prá*, *pa*, and *pá* were also identified to occur, in varying degrees of frequency. Including all five types of the preposition was considered necessary for not skewing the data in any significant way. After all, in conjunction with the surrounding verb and adverbial, the middle slot is heavily syntactically restricted to not 'permit' other elements than *para* 'to' – at least no such contexts surfaced in the analyzed corpus data. While the orthographic dimensions of *para* 'to' in the Portuguese intransitive pleonastic motion event constructions is not a focal point in this investigation, their distribution across BP and EP might serve to indicate possible differences in register, or context of use of the scrutinized motion event forms. A particular area of interest, considering variations of *para* 'to', is the earlier mentioned suggestion by Queriquelli & Moura (2021), that double Path constructions appear to be a more common feature in vernacular forms of Brazilian Portuguese (p. 6126).

3.2.2 Metalinguistic comments and filtering

Without actively setting out to test for it, metalinguistic comment contexts appeared in the data, where the motion event constructions are commented upon, or used as examples of 'incorrect' language use, rather than being used to express a motion event. Although constituting a "task to be solved", the identified metalinguistic comments followed a structurally consistent pattern of listing at least two of the double Path constructions included in the study, such as the Brazilian Portuguese example in (31) below, where three double Path constructions (all included in this study) are commented upon/used as frequent examples of pleonasms.

(31)   Brazilian Portuguese (own gloss and translation)

As	*paredes*	*interagem*	*junto?*	*Isso*	*é*	*assombração*	*pleonástica!*	*Só*
The	walls	interact	together?	That	is	haunting	pleonastic!	Only
pode	*ser!*	*Pleonasmo,*	*se você não*	*está*	*ligando*	*o*	*nome à*	*pessoa, é*
canbe!		Pleonasm,	if you NEG is	attaching		the	name to.the	person, is
aquela	*figura*	*de linguagem*		*que*	*soube*	*pra*	*cima,*	*desce*	*pra*
that	figure	of language/speech		that	go.up	to	up,	descend	to
baixo,	*entra*	*pra*	*dentro*	*e*	*sai*	*pra*	*fora.*
down,	enter	to	in,	and	exit	to	out.

'The walls interact together? That is a pleonastic haunting/ghost(?)! It can only be! Pleonasm, if you aren't calling/attaching the name to the person, it's that/the same figure of speech as <u>go up (to) up, descend (to) down, enter (to) inside, and exit (to) outside</u>.'

The fact that the double Path constructions are listed together, one immediately after the other, allows for a rather straightforward procedure of filtering out the metalinguistic comment contexts in the Sketch Engine interface. While there is most likely several ways of achieving the same filtering results, the procedure as it was done in this project will hereafter be described step by step. After the first (main) step of searching for a specific motion event construction, according to the CQL search strings presented in the previous section, (i) the *Filter* option was selected amongst the horizontally listed options/functions at the top of the page, above the retrieved results. As a next step, (ii) the *Advanced* tab, which offers a variety of filtering alternatives, was selected. When on the *Advanced Filter* page, (iii) the *Keep lines* drop down menu was changed from *containing* to *not containing*, in order for the contexts matching the subsequently specified CQL search string to be left out of the final results. If the main CQL search query was for *sair para fora* 'exit (to) out', the filter CQL search query would be specified as the bracket square elements in (32b) below, matching all other possible motion event constructions included in the study.

(32)   Parts and steps of filtering out metalinguistic comments in Sketch Engine

   a.   [lemma = "sair"] [word = "para|pra|prá|pa|pá"] [word = "fora"]

   Main CQL search query, corresponding to lemmas of *sair para fora* 'exit (to) out'

   b.   [lemma = "entrar|subir|ascender|baixar|descer"] [word = "para|pra|prá|pa|pá"] [word = "dentro|cima|baixo"] (not, -1:s.. 1:s,+KWIC)

   Second step of filtering out the presence of the other motion event constructions, within a distance of one sentence on each side of the KWIC, or main CQL query in step (a) above. The specification of this filtering step is illustrated as formally presented/described in the Sketch Engine interface.

The *Range* of the filtering search query is specified at the bottom of the *Advanced Filter* page, and was set to a threshold of one sentence on each side of the main search query (or KWIC), as specified in (32b) above – '(not, -1:s.. 1:s, +KWIC)'. While most instances of metalinguistic comments appeared to be in line with the comma separated structure of (31), the range was set to cross sentence boundaries due to contexts such as (33) below, where each pleonastic expression is interjected with a period.

(33)   Brazilian Portuguese (own gloss and translation)

   Subir la   em   cima. *Descer   la   em   baixo.   Sair   para   fora. Entrar*
   Go.up there in   up.   Descend   there in   down. Exit   to   out.   Enter
   para dentro. Terçol   no olho. (Terçol   aparece em   mais algum   lugar?)
   to   in.   Eyesore in eye. (Eyesore   appears in   more some   place?)

   'Ascend up there. Descend down there. Exit (to) outside. Enter (to) inside. Eyesore in the eye. (Do eyesores appear someplace else?)'

The metalinguistic comment filtering step was included for all search queries in the study, as a measure to safeguard against data analysis distortion that would result from including query results that entail a listing and commenting on the motion event constructions of interest in this investigation, rather than investigating their actual *language use*. However, while no engaged attention is directed towards the metalinguistic comments here, the filtering step was applied to BP and EP data separately, to allow for indications of possible differences in frequency of metalinguistic commentary within the speech communities.

3.2.3 Establishing a 'measuring stick' of robustness

Given the comparative nature of this investigation it is crucial with a 'measuring stick' for determining the degree of difference between BP and EP. As a means of establishing such a 'measuring stick' of difference, two illustrative cases are here put forward as suggested representations of each extreme of a bipolar spectrum, spanning from undeniable systematic difference between BP and EP to seemingly no systematic difference between BP and EP. Although perhaps perceived as working in reverse, the case of undeniable regional difference will be presented first. Considering that we are dealing with a comparison, it was decided on as more pedagogical to go from the more clearly defined situation of difference and contrast to where boundaries become increasingly fuzzy and less contrastive. Although both reference cases are here exclusively based in the Sketch Engine interface, section 3.3 describes the other statistical measures included in the significance testing of the project results.

3.2.3.1   Illustrating regional difference: determiner + noun

Kilgarriff et al. (2014) suggest that, when juxtaposing keywords from the two regional varieties in the corpus, the determiner/demonstrative pronoun *este* 'this' is more common in EP than in BP. Traditionally *este* 'this' is used to express closeness to the speaker, as opposed to *esse* 'that', which denotes referents that are close to the interlocutor. However, this distinction is argued to be disappearing in BP, where *esse* 'that' has 'taken over' (Kilgarriff et al. 2014: 120-121). When searched for in *ptTenTen11*, '[lemma = "este"] [tag = "N.* "]' – retrieving all instances of *este* 'this' (including *esta*, depending on the grammatical gender of the 'noun') immediately followed by a noun – resulted in a total of 6,338,313 contexts, distributed across the language varieties according to Figure 3 below.

Figure 3: Distribution of noun phrases consisting of *este* 'this' followed by any noun

(3 items, 6,338,313 total frequency)

		Language variety	Frequency	Relative in text type ?	Relative density ?		
1	☐	Brazil	3,943,825	1,069.63	65.75 %		···
2	☐	Portugal	2,394,486	2,559.11	157.32 %		···
3	☐		2	20.00	1.23 %		···

It is not entirely clear to which variety the blank bottom row is supposed to be ascribed to. However, its 2 instances are of minor significance from a large scale comparative point of view.

As Figure 3 illustrates, both the relative text type frequency (per million), and the relative density is greater in the Portugal language variety, i.e. EP. Although occurring more than a thousand times per one million tokens in both varieties, EP stands for more than twice as many occurrences as BP. Adding to that, the relative density, which represents how typical the search string is for the respective language variety (by dividing the relative frequency of the search string by the relative size of the text type, which in this case is 'language variety'), is below 100% for 'Brazil', or BP – 65% – while it is over 100% for 'Portugal', or EP – 159%. Although the relative density score is calculated automatically for each search, it can formally be represented as in (34) below. Crucially for the relative density score is that if it is below 100% it is less frequent in the text type than in the whole corpus, and thus considered not typical or specific of that text type. On the other hand, if the relative density score is above 100% it is more frequent in the text type than in the whole corpus and can be considered typical or specific of that text type (www.sketchengine.eu/guide/glossary).

(34) Relative density measure (www.sketchengine.eu/guide/glossary)

Relative density score formula

$$Relative\ density = \frac{relative\ frequency\ of\ query\ result}{relative\ size\ of\ particular\ text\ type}$$

The formula is based on the formulation in the definition of relative density in the Sketch Engine glossary. Further, for added nuance, according to Sketch Engine themselves, the relative density measure "can be interpreted as how much more/less frequent is the result of the query in this text type compared to the whole corpus" (www.sketchengine.eu/guide/glossary).

Since the relative density measure is aimed at indicating whether the search string is more common in the selected text type (again, in this case: language variety) than in the whole corpus, the 'este/esta + noun' string is thus less frequent in BP than the whole corpus, while more frequent in EP than in the whole corpus. Text type frequency and relative density taken together, arguably point at a systematic difference between EP and BP. As an additional testing step, Figure 4 below shows the total number of occurrences of '[lemma = "esse"] [tag = "N.*"]', within the two text types 'Brazil' and 'Portugal'.

Figure 4: Distribution of noun phrases consisting of *esse* 'that' followed by any noun

(2 items, 6,254,124 total frequency)

		Language variety	Frequency	Relative in text type ?	Relative density ?	
1	☐	Brazil	5,620,501	1,524.38	94.97 %	⋯
2	☐	Portugal	633,623	677.19	42.19 %	⋯

Against the background of *esse* 'that' having been claimed to increasingly be the determiner of choice in contexts previously expressed with *este* 'this' in BP, the data in Figure 4 does not constitute a direct challenge to it. Rather than focusing on BP, however, the EP differences between Figure 3 and Figure 4 are the most striking. Regardless of possible explanations of this pattern, it is undoubtedly safe to say that there is a systematic difference between BP and EP within the grammatical domain of determiner + noun in terms of determiner choice.

3.2.3.2 Illustrating regional non-difference: psych action want-clauses

The aforementioned listing of keywords that differ between BP and EP in Kilgarriff et al. (2014) does little to aid us in the task of identifying an appropriate systematic *non-difference* – i.e. a grammatical domain where the varieties show no signs of variation in expression –, although it can inform us on within which domains *not* to look at. Identifying non-difference between two varieties of 'the same language' should arguably not be difficult, and scrutinizing considered *core Portuguese features* could perhaps retrieve CQL searchable, and suitable enough, queries for setting up one of the sought after comparison parameter poles: regional non-difference.

A search for the structure of psych-action want-clauses, composed of a lemma of the 'want' verb *querer* followed by any verb (formally allowing any conjugation, although all appear to be in the infinitive form), retrieved results that displayed similar scores of relative in text type (per million) frequencies and relative density for both varieties (see previous section 3.2.2.1 for descriptions of both score types). The total raw frequency of the search string '[lemma = "querer"] [tag = "V.*"]', which can be seen in Figure 5 below, was 2,359,799 – out of which 1, 872, 557 tokens pertained to 'Brazil' (BP) and 487,242 to 'Portugal' (EP).

Figure 5: Distribution of psych action want clauses consisting of lemmas of *querer* 'want' followed by any verb

(2 items, 2,359,799 total frequency)

		Language variety	Frequency	Relative in text type ?	Relative density ?		
1	☐	Brazil	1,872,557	507.87	83.86 %		•••
2	☐	Portugal	487,242	520.74	85.98 %		•••

As can be seen in Figure 5, EP has only slightly higher values than BP, both in terms of relative text type frequency (520.74 and 507.87 respectively), as well as the relative density score (85.98% and 83.86% respectively). Given that the value of both relative density scores are close to each other, and (perhaps even more importantly) that they are under 100% is an indicator of psych action want-clause constructions not being typical for any of the varieties compared to the whole ptTenTen11 corpus.

3.2.3.3   Regional difference and non-difference: in sum

Although perhaps difficult to identify and classify the type and degree of difference between two structurally similar regional varieties of the same language – such as BP and EP – the point of the examples above was to illustrate what the relevant scores can look like in the Sketch Engine interface when BP and EP undoubtedly behave differently with respect to a given query, as well as when the frequencies of a search string construction are so close as to be considered (almost) identical on an overall level in BP and EP, without considering more fine-grained levels of analysis beyond the search string itself. To summarize, the goal is thus for the above examples to constitute opposite poles of a general contextualization, that can assist both detection and description of slight systematic differences between BP and EP when looking at the results concerning the intransitive pleonastic motion event constructions, that are the main area of investigation of this project. While these measures provide some indication with regards to the difference *between* the varieties, other statistical measures, such as the chi-squared test, will also be used, and are described in section 3.3. In the next section we turn to the search procedure itself, describing it one step at a time in the order of how they relate to the research questions.

3.2.4 Search procedure

Data belonging to the two main standard Portuguese varieties, BP and EP, are classified as 'text types' within the ptTenTen11 corpus. Therefore, conducting a search within ptTenTen11 retrieves data from both varieties, and further makes intervariational comparisons readily accessible, such as comparing relative text type frequency (per million) and relative density scores (described in the previous section 3.2.3). Since Sketch Engine as a tool, provides the possibility of automatically retrieving these two just mentioned statistical measures, much of the methodological steps were possible to conduct on the website itself (www.sketchengine.eu).

As for the specific search procedures of the project, two main searches were made. The first search – the method of which will be presented in the directly following sub-section – relates to research question 1, and thus focuses on retrieving data on the double Path constructions in BP and EP. The second search relates to research question 2 and is aimed at testing to what extent Manner verbs occur with the same Path satellites as in the double Path constructions.

3.2.4.1   First search: double Path constructions

Considering that the intransitive pleonastic motion event constructions are characterized by the double marking of 'the same' Path – on both the verb and in the satellite –, there are combinatorial restrictions in terms of the availability of Portuguese verbs and corresponding satellites that are considered pleonastic when expressed together. A total of six strings were considered to fulfill this pleonastic double path criteria, and further to align into three rather straight forward categories: boundary crossing (35a), upwards motion (35b), and downwards motion (35c).

(35) Listing and grouping of double Path search strings

   a. Boundary crossing events
       entrar para dentro – 'enter (to) inside'
       sair para fora – 'exit (to) outside'

   b. Upwards motion
       subir para cima – 'go.up (to) up'
       ascender para cima – 'ascend (to) up'

   c. Downwards motion
       baixar para baixo – 'go.down (to) down'
       descer para baixo – 'descend (to) down'

While double Path boundary crossing events perhaps could have been investigated through other verbs, such as *cruzar* '(to) cross', no obvious corresponding satellites that conformed to the structure 'verb + *para* (to) + adverb' were identified. Further, the six strings and three categories in (35) represent two polarity distinctions: in/out and up/down, which set the scene for 'easily' testable distinctions along an axis.

3.2.4.2   Second search: Manner verb constructions

In order to approach the second research question of whether the Boundary crossing constraint applies to BP and/or EP, the next step consists in a more qualitative scrutiny of selected 'Manner verb + boundary crossing Path satellite' constructions that appear in the corpus. As a first step in this second search round, the 'bare' adverbial search strings were used with the Path satellites *para dentro* '(to) inside' and *para fora* '(to) outside', in order to retrieve all instances of these satellites alongside *any* verb preposed to it, as illustrated in (36) below.

(36)   'Bare' Boundary crossing Path satellite search strings

> *VERB para dentro* – 'VERB (to) inside'
> *VERB para fora* – 'VERB (to) outside'

Upon retrieving the results from the two search strings in (36) in a Sketch Engine concordance search, it was possible to select the 'Frequency' option at the top of the page, and then proceed to choose 'KWIC lemmas', which sorts the retrieved lemmas according to frequency. The 20 most frequent verbs for each construction are listed in Appendix 1 and Appendix 2 respectively. Although numerical values are reserved for the Results section (4), the constructions shown in (37) below were selected as 'Manner verb + boundary crossing Path satellite' constructions to select samples from and examine whether their expression (in any way) can be considered to challenge the idea that the Boundary crossing constraint applies to BP and/or EP.

(37)   Boundary crossing Path satellites with Manner verbs

> *correr para dentro* – '<u>run</u> (to) inside'
> *chutar para fora* – '<u>shoot</u> (to) outside'

Neither *correr para dentro* 'run (to) inside' nor *chutar para fora* 'shoot (to) outside' are unambiguously clear cases of Manner verb constructions. The verb *correr* 'to run' indeed describes *how* someone is moving, but in the prototypical case the activity of running also inherently expresses translational motion (unless one perhaps is at a gym). The same can be said of *chutar* 'to shoot', which – as will be shown in the selected examples in the Results section (4) – to a great extent concerns football commentary contexts. How does one shoot a ball without a Path being traversed in reference to some Ground object? Also, while at the surface level appearing as an intransitive construction, *chutar para fora* 'shoot (to) out' is undoubtedly semantically transitive – it is after all a ball that is being shot –, which perhaps does not need clarification given the context of commenting upon a football game. The next section (3.3) covers the statistical measures used for significance testing.

3.3 Statistical measures

Since the first part of this investigation concerns the quantitative relationship between varieties of Portuguese and intransitive pleonastic motion event constructions (i.e. double Path constructions), presented in section 2, a chi-squared test was performed in R (R Core Team 2017) in order to determine if there is a significant categorical association between the two language varieties and the presence or absence of double Path constructions (Desagulier 2017: 178). The null and alternative hypotheses are presented below.

Figure 6: Null and alternative hypotheses

H0: the choice of double Path constructions and variety of Portuguese are independent;

H1: the choice of double Path constructions and variety of Portuguese are interdependent

While the chi-square test statistic indicates whether there is a significant association between the use of double Path constructions and variety of Portuguese, it tells us nothing with regards to the *intensity* of said association. Following Desagulier (2017), a Cramér's V measure will be used to calculate the association intensity, since it is described as the appropriate procedure when the number of row variables differs from the number of column variables, which indeed is the case for our data (p. 183). Unlike the chi-square test, Cramér's V does not have a base-R function, and was therefore calculated according to the formula in (38) (Desagulier 2017: 183).

(38) Cramér's V formula for calculating the intensity of the association between the use of double Path constructions and variety of Portuguese (Desagulier 2017: 183)

$$Cramér's\ V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

$N = sum\ of\ all\ observations.$

$k = the\ smaller\ number\ of\ the\ two\ possible\ values\ of\ row\ variables\ and\ column\ variables$

While α is set to .05 for the chi-square test, the interpretation of the Crámer's V "ranges from 0 (no association) to 1 (perfect association)" (Desagulier 2017: 182).

3.4 Sampling

Two steps of sampling random sub-parts from the retrieved search query results were made within the scope of this study: (i) a sample of 10 random hits for each double path, in order to get an overlook of possibly relevant semantic categories and contexts to look into more closely, and (ii) a sample of 20 random hits for each double path constructions, that involved a more in-depth scrutiny of metaphoric contexts and boundary crossing contexts. The two

sampling procedures were done for BP and EP separately and will be described in more detail in the two upcoming sub-sections.

3.4.1 Overview of the data

As illustrated in the sub-section (2.2.5) of the background, the Sketch Engine 'Get random sample' function was used in order to retrieve 10 random samples per double path construction for each variety, separately. The purpose of these small scale queries was to get an overview of the types of contexts that the constructions appear in respectively. Context is in this case restricted to the semantic expression of the double path construction itself, as a whole – beyond the individual Path components –, as interpreted along with the specified (or not specified) Figure and Ground elements. The samples were retrieved according to the same search design procedure as illustrated in (3.2) above, only that 10 samples were randomized from the total results, following the following steps. Out of the selectable options above the search query results, the three-dimensional box with a question mark retrieves a random sample of the number that one specifies. According to the website, if one specifies that one wishes to retrieve 10 random samples with configurations x and filter y, the same 10 samples will be retrieved each time, as long as the configurations and filter settings are consistent with the original x and y – this, in order to support replicability.

3.4.2 Selected examples of boundary crossing contexts

The same 'Get random sample' procedure as in 3.4.1 above was performed when retrieving a 20 sample size for each of the boundary crossing contexts that occurred with the Manner verbs *correr* 'run' and *chutar* 'shoot'. Although the retrieval of contexts was similar, the purpose of gathering the two 20 context sample groups was as actual pieces of results rather than – as was the case with the 10 context samples – overviewing and familiarize with the available corpus data. The next section (4) will present the results of the investigation.

4 Results

The aim of this study is to contribute to an overarching understanding of possible divergencies between BP and EP within the domain of motion events, by quantitative corpus analysis of data crawled off of Brazilian domain and Portuguese domain websites in the ptTenTen11 corpus. The research questions are more specifically:

1. To what extent are verbs that conflate Motion and Path expressed with satellites that express 'the same' Path information in BP and EP? Is there any significant difference between the two varieties in this regard?

2. Does the Boundary Crossing Constraint apply to BP and/or EP – which states that verb-framed languages do not allow for Manner verbs to occur in constructions where a (conceptual) boundary is crossed?

3. If the results from double Path constructions and the Boundary Crossing Constraint are taken together, what do they, if anything at all, tell us about the relationship between BP and EP and the verb- and satellite-framed dichotomy/continuum?

4. Taking a step further back, how central or peripheral are the double Path constructions? Are they widespread enough, meaning that they ought to be taken into consideration when discussing lexicalization patterns at a more general level, or are they rather highly locally restricted phenomena?

This section presents the results of the study in the order of the research questions as they are listed above. Thus, the first part (4.1) presents the quantitative results related to the presence/absence of double Path constructions in BP and EP, and the inter-variational differences identified. The second part (4.2) displays the results connected to the Manner verbs alongside the boundary crossing Path satellites *para dentro* '(to) inside' and *para fora* '(to) outside'. Although research questions 3 and 4 are implicitly present throughout the Results section, they are made immediately relevant in the subsequent Discussion (section 5), since they rather concern a wider contextualization of the data presented as results.

4.1 Double Path constructions

This first part of the results section presents and illustrates the results connected to the first research question, repeated below, focusing on the frequencies and statistical measurements of the double Path motion event constructions, as distributed both within and across BP and EP.

1. To what extent are verbs that conflate Motion and Path expressed with satellites that express 'the same' Path information in BP and EP? Is there a significant difference between the two varieties in this regard?

The overall (raw) extent to which double Path constructions are distributed across BP and EP is shown in a contingency table below (Table 3). As mentioned previously, metalinguistic comments concerning the constructions themselves as topics of discussion, were filtered out and the column on the right for each variety (given in boldface) lists the double Path frequencies without such metalinguistic commentary contexts in their count.

**Table 3: Raw frequency and meta comment filtered contexts**

Double Path construction	BP	BP no meta	EP	EP no meta	Total
entrar para dentro	266	**238**	162	**157**	**395**
sair para fora	681	**664**	246	**243**	**907**
ascender para cima	1	**1**	0	**0**	**1**
subir para cima	141	**86**	205	**190**	**276**
descer para baixo	66	**20**	16	**6**	**26**
baixar para baixo	0	**0**	1	**1**	**1**
Total	1155	**1009**	630	**597**	**1606**

EP/BP = number of results retrieved with the double Path CQL search queries

BP/EP No Meta and Total in bold = number of results that were <u>not</u> meta comment contexts (i.e., results that were not filtered out as described in 3.2.2), and thus constitute the final data set for further analysis

In method section 3.2.4.1 the constructions were presented as categorized into the groups 'upward motion', 'downward motion', and 'boundary crossing events'. The results retrieved, however, seem to indicate that it is not possible to talk of such groupings, but rather of individual motion verb behaviors. The difference in frequency between *ascender* 'ascend'/*descer* 'descend' and *subir* 'go.up'/*baixar* 'go.down' is an indication of this. Amongst the double Path constructions both varieties show a relative preference for the verb *sair* 'exit' and in particular disfavor *descer* 'descend' and *ascender* 'ascend'. If any connection is to be made to the semantic categories, it could perhaps be stated that downward motion, with low scores for both *baixar* 'go.down' and *descer* 'descer', is generally dispreferred. There seems to be a general tendency of using double Path constructions in Boundary crossing contexts in BP – *entrar para dentro* 'enter (to) inside' and *sair para fora* 'exit (to) outside' –, with lemmas of *entrar para dentro* 'enter (to) inside' being the second most frequent construction in the Brazilian data, this does not seem to be the case with EP, which has a higher frequency of contexts with *subir para cima* 'go.up to up'. Figure 7 illustrates the variation within language varieties of intransitive pleonastic double Path constructions.

Figure 7: Barplot diagram double Path contexts according to website domain

.br

.pt
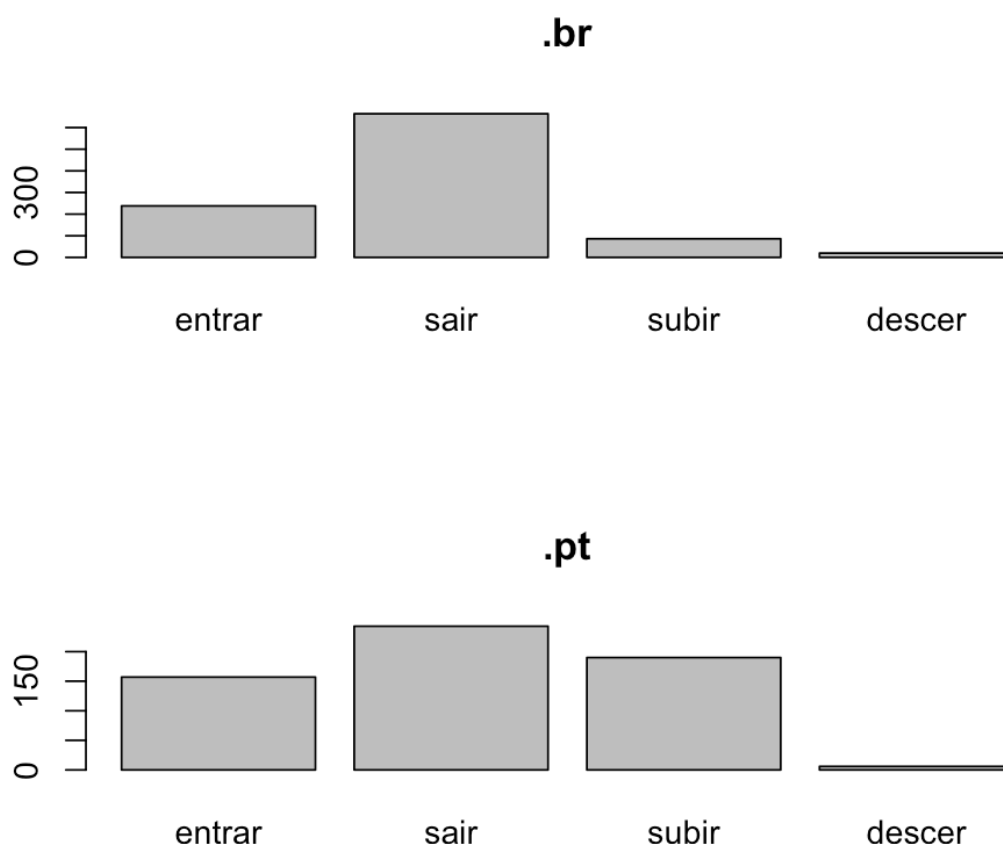
.br represents BP, and .pt represents EP. Baixar para baixo 'go.down (to) down', and ascender para cima 'ascend (to) up' are excluded from this representation due to their lack of presence in the retrieved data. The intention of the diagram is to illustrate the similar intravariational tendencies in usage of the double Path constructions.

If we now turn to the second part of research question 1, concerning if there is a significant difference between BP and EP in terms of how the investigated double Path constructions are distributed across the varieties, Table 4 below displays the *relative in text type frequency*, and *relative density measure* of all constructions (except for *ascender para cima* 'ascend (to) up' and *baixar para baixo* 'go.down (to) down', which both only occurred once in the retrieved results).

Table 4: Relative in text type frequency and relative density measure

Double Path Construction	BP rel type	BP rel dens	EP rel type	EP rel dens
entrar para dentro	0.06	75.54%	0.17	196.37%
sair para fora	0.18	91.79%	0.26	132.37%
subir para cima	0.02	39.07%	0.20	340.46%
descer para baixo	<0.01	96.44%	<0.01	114.01%

Considering the 'measuring stick' introduced in section 3.3.2 the scores presented in Table 4 indicate that the double Path constructions are more prevalent in EP than in BP, since EP has higher scores than BP in both relative text type frequency and relative density across all constructions. While Figure 7 illustrates that *subir para cima* 'go.up (to) up' is used more often in EP, Table 4 emphasizes the same relationship, which can be seen in that EP has a relative in text type frequency which is 10 times as large as the BP frequency – 0.20 versus 0.02 tokens per one million words for *subir para cima*. Note also that the relative density score is 340.46% in EP, and 39.07% in BP. As a measure to contextualize the double Path constructions and relate them to patterns of overall Path satellite use in BP and EP respectively, Table 5 presents the 'bare' Path adverbials as they are distributed across the varieties, regardless of the type of co-text verb – that is, *any* verb and not just the investigated Path verbs.

Table 5: Double Path Constructions as Compared with the 'Bare Path Adverbials'

Verb + Path Satellite Construction	BP freq	BP rel type	BP rel dens	EP freq	EP rel type	EP rel dens	Total freq
V. para dentro*	19,181	5.20	102.66%	4,244	4.54	89.51%	23,425
V. para fora*	28,944	7.85	105.00%	5,617	6.00	80.30%	34,561
V. para cima*	29,333	7.96	106.53%	5,191	5.55	74.29%	34,524
V. para baixo*	14,811	4.02	107.88%	2,402	2.57	68.94%	17,213

Interestingly, Table 5 implies that the overall usage of the four Path satellites is consistently lower in EP, than in BP. In other words, it appears to be the case that the Path satellites are generally more common in BP, while a larger part of the overall Path satellite constructions are of the double Path type – as is indicated by Tables 4 and 5 taken together.

The p-value of the chi-squared test of independence concerning occurrences of double Path constructions and Portuguese variety is close to zero (2.2e-16), which is shown in Figure 8 below. With α set to .05 it is thus significant, meaning that we can reject the null hypothesis that there is no interdependence between occurrences of double Path constructions and variety of Portuguese. Assuming that *no interdependence* between double Path constructions and variety of Portuguese is not a far stretch from *no difference* between BP and EP regarding the extent to which the investigated double Path constructions are attested it is concluded from these results that there is a difference between the varieties. As for the intensity of the

association between double Path constructions and variety of Portuguese, the Cramér's V statistic resulted in .31 (Figure 8), which indicates that there is a moderate intensity of association between Portuguese variety and double Path construction.

Figure 8: Chi-square and Cramér's V result scores for intransitive pleonastic motion event constructions and Portuguese variety

Pearson's Chi-squared test

X-squared = 163.73, df = 3, p-value < 2.2e-16

Cramér's V

0.3194917

The chi-squared p-value of 2.2e-16 is below α, which is set to .05, meaning that it is significant. Regarding Cramér's V, which is measured from 0 (no association) to 1 (perfect association), the statistic of approximately .32 indicates a moderate intensity of association. For further information, such as formulas and reasoning behind the choice of statistical measures, see section 3.3.

Judging from the statistics and the tables in this section, there does seem to be a significant difference between BP and EP in terms of the extent to which the double Path constructions are used. There is overall a moderate association between variety of Portuguese and selection of double Path constructions. Beyond the chi-squared test and the Cramér's V statistic, this association is most evidently seen in that BP most strongly favors the construction *sair para fora* 'exit (to) outside', whereas EP has a slightly more even distribution across the constructions. BP has an overall preference for boundary crossing double Path constructions, whereas *subir para cima* 'go.up (to) up' is markedly more common within EP. A similarity between the two varieties is their strong dispreference of double Path constructions within the semantic domain of motion downwards, which is practically non-existent in the data. An interesting point is that the relative text type frequency and relative density measures of the double Path constructions and bare Path satellites combined indicate that EP favors double Path constructions more than BP. What to make of these differences remains to be discussed in section 5. Before proceeding to the discussion, however, the next section (4.2) will present the results linked to the second research question concerning the Boundary crossing constraint.

4.2 Boundary crossing Manner verb constructions

This section deals with the results connected to the Portuguese Manner verbs and the Boundary crossing constraint, in connection to research question 2, repeated here once more for the sake of clarity.

> 2. Does the Boundary Crossing Constraint apply to BP and/or EP – which states that verb-framed languages do not allow for Manner verbs to occur in constructions where a (conceptual) boundary is crossed?

Following the retrieval of the 'bare' Path satellite/adverbial results that were presented in Table 5 in the previous section (4.1), several Manner verbs appeared to occur alongside them. Not only were Manner verbs overall attested in conjunction with Path satellites, but Manner verbs were identified to occur with the *boundary crossing* Path satellites *para dentro* '(to) inside' and *para fora* '(to) outside'. Since the 'bare' adverbial search retrieved constructions with *any* verb, and not just *motion* verbs (no method for doing such a search could be conceived of), Table 6 below shows the absolute and relative text type frequencies, as well as the relative density scores for the boundary crossing Manner verb constructions, as they are distributed across BP and EP.

Table 6: Boundary crossing constructions with Manner verbs

Manner verb + Path satellite construction	BP freq	BP rel type	BP rel dens	EP freq	EP rel type	EP rel dens	Total freq
correr para dentro	649	0.18	116.58%	49	0.05	34.68%	698
chutar para fora	1,786	0.48	123.71%	24	0.03	6.55%	1,81

Judging from Table 6, both boundary crossing Manner verb constructions are undoubtedly more common in BP than in EP, in all regards counted. For instance, *chutar para fora* 'shoot (to) outside' occurs 1,786 times in BP – and only 24 times in EP –, has a relative text type frequency of 0.48 (meaning that it occurs approximately 1 time per two million words) – EP has a relative text type frequency of 0.03 –, and has a relative density score of 123.71%, which hints at it being more common in the BP part of the corpus than the entire corpus as a whole. The relative density score of EP for *chutar para fora* 'shoot (to) outside' is 6.55%. The just listed numbers are indicative of the 'measuring stick' case in section 3.2.3.1 where the regional varieties can be identified to clearly (appear to) behave differently already in the Sketch Engine interface. Section 4.2.1 lists examples from randomly selected samples of the constructions in both Portuguese varieties.

4.2.1 Selected boundary Crossing Contexts

The following section presents selected illustrative examples from the retrieved results that are linked to the contexts aimed at scrutinizing whether the Boundary crossing constraint applies to BP and/or EP. First (4.2.1.1) selected examples of *correr para dentro* 'run (to) inside' are shown, after which (4.2.1.2) selected contexts of *chutar para fora* 'shoot (to) outside' are listed.

4.2.1.1 *correr para dentro* 'run (to) inside'

This sub-section presents selected examples of the boundary crossing Manner verb construction *correr para dentro* 'run (to) inside'. The small 20 context size randomized samples for each Portuguese variety had a majority of actual physical boundaries being crossed, rather than say metaphorical boundaries (which there are examples of in section 2.2.7). While the example in (39) is from the BP part of the corpus, such expressions where someone runs inside into hiding as a consequence of crying, shame, or anger appeared to be common in both BP and EP.

(39)   Brazilian Portuguese (own gloss and translation)

Terminei a goiaba antes de correr pra dentro da casa
Finish.1SG.PST the guava before of running to in of.the house
chorando
crying

'I finished/ate up the guava before running (to the) inside of the house, crying'

The EP example in (40) is another representation of a physical boundary being crossed with *correr* 'run' as the Manner verb in a construction with the Path satellite *para dentro* '(to) inside'.

(40)   European Portuguese (own gloss and translation)

Como estava atrasada, só tive tempo de correr para dentro do
Since was.1SG late only had.1SG time of running to in of.the
comboio que já estava na plataforma quase a partir
bus which already was.3SG on.the platform almost to leave

'Since I was late, I only had time to run inside of the bus that was already on the platform, almost ready to leave'

While (40) is not a context where emotions are explicitly expressed, such as (37), it does express a situation 'out of the ordinary', where the person retelling their experiences *emphasize* managing to get inside of the bus when it was about to leave. Whether these contexts could be classified as emphatic expressions is unclear, however.

4.2.1.2 *chutar para fora* **'shoot (to) outside'**

This sub-section shows selected examples of lemmas pertaining to the boundary crossing Manner verb construction *chutar para fora* 'shoot (to) outside'. All boundary crossing Manner verb constructions that include the verb *chutar* 'shoot' in the BP sample were from football related contexts except for one, which expressed a metaphorical type of boundary crossing, where salesmen where described as being kicked out of the company, when being fired. A typical football related context is presented in (41) below.

(41) Brazilian Portuguese (own gloss and translation)

Após cruzamento, Adriano, cara a cara com o   goleiro   Juninho, chutou
After crossing,   Adriano,   face to face with the   goalkeeper Juninho, shot
para fora
to   out

'After crossing, Adriano, face to face with the goalkeeper Juninho shot [the ball] outside [of the goal]'

Although much less frequent, the EP constructions involving *chutar* 'shoot' follow a similar pattern of consisting of a large majority of football related boundary crossing events, where a ball and a goal in many cases are only covertly implied.

4.3 Results in sum

Based on the retrieved results there appears to be a significant difference between BP and EP in terms of the extent that intransitive pleonastic double Path constructions are used in the two varieties. According to the Cramér's V statistic the intensity of this difference is at a moderate level. Amongst the investigated double Path constructions there is a shared relative preference for *sair para fora* 'exit (to) outside' for both varieties, and a shared dispreference for double Path constructions that express downward motion – i.e. *descer para baixo* 'descend (to) down' and *baixar para baixo* 'go.down (to) down'. When looking at the relative density score, which indicates if a certain construction is more common in a specific variety than in the whole Portuguese corpus, they show that the Path adverbials as a whole (in combination with any verb) are less typical of EP, and typical of BP. This difference reflects the moderate level implied in the significance statistics scores above. What is interesting is that, at the same time, the relative density measure also indicates that the double Path constructions are markedly more common in EP than in BP. In sum, Path adverbials in general are more common in BP than in EP, while the double Path constructions with the same adverbials are more common in EP than in BP.

The retrieved results for the boundary crossing events with Manner verbs and Path satellites are undoubtedly more common in BP than in EP. If one is to do a numerical comparison of frequencies and relative scores, doubt is raised when overviewing the 20 randomly sampled contexts – especially for *chutar para fora* 'shoot (to) outside' that almost exclusively includes football related commentary contexts. The prevalence of narrations of football games might

suggest that genre is rather important for the use of double Path constructions. Although the importance of genre was never explicitly ruled out, the ptTenTen11 corpus had no genre filtering system in place. The difference is less extreme between BP and EP for *correr para dentro* 'run (to) inside', but still evident considering the 'measuring stick' formulated in section 3.2.3. Scrutinizing the randomly selected examples for this construction, they seem to occur in a form of emphatic constructions, where the act of actually getting in or hiding away is highlighted as especially important in the situation described, such as getting onto the bus *right before it leaves.* Section 5 discusses the results (5.1) and the method (5.2) of the investigation as well as proposes suggestions for further research (5.3) within the domain of pleonastic motion events and Portuguese.

5 Discussion

This section discusses the results (5.1), the method (5.2) as well as mentions gaps to be filled in future studies on pleonastic expression and motion events in Portuguese.

5.1 Results discussion

Although the results for research question 1 lead to a rejection of the null hypothesis that the double Path constructions, and BP and EP are independent of each other, and that the intensity of the relation between the two categories (*double Path construction* and *Portuguese variety*) is at a moderate level, it is still rather unclear how a motion event with a double expression of the Path component relates to the overall framing typological approach. Trying to connect the results of research question 1 to the papers by Queriquelli & Moura (2021) and Batoréo & Ferrari (2016), that serve as a backdrop for the comparison, the results of this investigation are more in line with the claims in the latter study. There they state that EP is increasingly behaving as a satellite-framed variety, and that BP is more stable in its verb-framed structure (p. 59). One way of approaching their study with the double Path construction results is to only consider the presence or absence of the Path component in the satellite, as Batoréo & Ferrari (2016) themselves indeed seem to have done. Example (42) below is a repetition of example (12), where the verb *aproximar* '(to) approach' is without a question a Path verb, which makes the construction semantically similar to the pleonastic expression investigated in this thesis (p. 70).

(42) Portuguese [own gloss and encoding of motion event components] (Batoréo & Ferrari 2016: 70)

O	*barco*	*aproximou*	*-se*	[*da*	*costa.*]
The	boat	approach	=REFL	of.the	coast
		VERB		[PREP	COSTA.NOUN]

'The boat approached/got closer to the coast'

Underlined = motion event construction
Within square brackets [] = satellite

49

Further, although claimed to be a periphrastic Path satellite based on a COSTA-noun, *da costa* 'of.the coast' in (42) really appears to be a specification of Ground (Batoréo & Ferrari 2016: 70). Compared to periphrastic satellites such as *da costa* 'of.the coast', the Path satellites in this investigation appear as less ambiguous in their expression of Path. At least there is no mistaking of for example *para dentro* '(to) inside' or *para cima* '(to) up' as Ground elements.

As for Queriquelli & Moura (2021), there were several metalinguistic comments present in the data, that commented upon the double Path constructions as 'bad language use'. A link could perhaps be drawn to non-standard varieties of Portuguese, such as Vernacular Brazilian Portuguese (VBP), that Queriquelli & Moura (2021) mention, where the claimed change of BP to a satellite-framed variety is primarily happening (p. 6127). As a consequence of the structure of the metalinguistic comments in the data of this investigation it could have been possible to quantitatively measure the extent of their occurrence in the corpus. However, as a consequence of time management it was not done, but could be of interest for future research.

With regards to the almost non-existent occurrences of downward motion amongst the double Path constructions in the corpus, it might be so that expressions of motion downwards in general is comparatively rare. Judging from the low count of the constructions *descer para baixo* 'descend (to) down' and *baixar para baixo* 'go.down (to) down' and that the bare adverbial *para baixo* '(to) down' was the least frequent out of all bare adverbials for both BP and EP, surely indicates that such could be the case.

Moving on to research question 2 and if the Boundary crossing constraint can be claimed to apply to BP and/or EP, the results are considered too scarce to make such a claim. At the same time, the scarcity of results with Manner verbs and Path satellites in boundary crossing events can also be seen as indicative of that the boundary crossing constraint is applicable to both BP and EP. Further, both Manner verbs that were identified following the 'bare' adverbial search were not unambiguously expressions of Manner. While it can be argued that few verbs (or 'words' at all) are unambiguous, both *correr* '(to) run' and *chutar* '(to) shoot' are difficult to conceptualize without a Path in mind. Also, *chutar* '(to) shoot' might appear to be intransitive in the surface structure of the expression, but it is really a transitive state of affairs of a ball as the Figure of the event, being *caused* – through shooting – to move *para fora* '(to) outside' of some Ground element.

Now we get to research questions 3 and 4 at the same time – combining the results and widening the scope beyond the framing typological approach. The results concerning the double Path constructions and the Boundary crossing constraint taken together, do not clearly place either Portuguese as a whole or BP and EP specifically along the framing typological continuum from verb-framed to satellite-framed language types. Based on the results for the double Path construction one could claim that EP is closer to the satellite-framed type and BP closer to the verb-framed type, like Batoréo & Ferrari (2016) did. However, similarly to the investigation by Batoréo & Ferrari (2016) that scrutinizes CMEs, the intransitive pleonastic motion event constructions do not appear to be characteristic or general in the sense that Talmy (2000) was interested in. While the double Path constructions to the most part appear to be *colloquial* in style, they cannot really be considered *frequent* (judging from the overall low count of all constructions) (p. 27). Thus, for a framing typologist interested in the general lexicalization patterns of a language, the double Path constructions might be too peripheral in order to say anything substantial of such general and characteristic tendencies. In other words,

regardless of the direction they might take, they would most likely not constitute a challenge to someone interested in the core patterns of the language.

This is a good entry point for Sinha & Kuteva's (1995) alternative approach to spatial relational meaning, *Distributed spatial semantics*. It is evident from the Portuguese corpus data that narrowing down the analysis of locative meaning to the mapping of Path and Manner to overtly expressed surface elements in motion event constructions is restricting. Instead of looking for 'the Path component' as a type classifier in a motion event construction, it could enrich the analysis to instead focus on the Figure and Ground components, and how they relate to each other, since these two are necessarily part of any translational motion event – something that moves in relation to something else – even though information about them might be covertly expressed, such as the case of the football commentary contexts. Sinha & Kuteva's (1995) emphasis on *overt* and *covert* distribution of spatial relational information across a given segment could perhaps deepen the analysis of for example the football commentary contexts, where the relation between Figure and Ground in many cases appeared to be covertly expressed.

5.2 Method discussion

The choice of method was based on the aim of investigating the relationship between BP and EP and certain configurations of motion event constructions, as they are understood according to the framing typological approach put forward by Talmy (2000). Since there appeared to be a general lack of systematic framing typological data on Portuguese, but also several claims on how BP and EP are to be typologically classified, a large-scale corpus analysis seemed appropriate. The implementation of CQL in Sketch Engine was convenient for the search string designs mirroring both double Path constructions and boundary crossing events. Beyond the easy access to quantities of data for both BP and EP within the same corpus architecture and design, the Sketch Engine interface also offered the automatically generated statistical measures of relative text type frequency and relative density score, which indicated how often in one million tokens the search string appeared and how typical of the text type (i.e. Portuguese variety) the search string was respectively. In terms of validity, I would claim that using website domain as proxy for language variety is a quite solid measure, that increase effectivization of data collection without comprising on quality, but rather improve quality, by allowing for easily accessible large bits of data.

One identified restriction with the ptTenTen11 corpus was the lack of possibility to differentiate between genres within the crawled data. Although not identified as an immediate concern for the sake of this study, being able to easily limit searches to certain genres and not just variety of Portuguese, could undoubtedly have added more nuance to the results. However, some genre tendencies did surface regardless of this limitation, such as the prevalence of football commentary contexts within the domain of boundary crossing contexts with Manner verbs.

Another problem with regards to the quantitative part of the investigation was the presupposition of equal corpora sizes of the BP and EP data when comparing the results – proper calculations on their respective sizes could have been embedded in the work for more clarification. Size in general is an issue that could have been explicitly addressed earlier, given the substantial difference in both country and speech community size between Brazil

and Portugal. Regardless of size measurements, the method certainly captures variation within the double Path constructions, and how these tendencies vary between BP and EP. In terms of significance, the chi-squared test and Cramér's V allowed for testing the degree of interdependence between Portuguese variety and type of double Path construction, pointing at different double Path 'behaviors' between the two.

5.3 Further research

There are several possibilities for future research within the domain of motion events in Portuguese. For one, Portuguese in general seems to be relatively understudied compared to its Romance language neighbor Spanish. At first it was planned to include comparisons with Spanish constructions in this investigation as well but due to time management it had to be excluded from the project. Since both languages have a shared past in terms of being spoken on the Iberian Peninsula as well as in South America, there must be several possible entry points to interesting comparisons. While the relevance of pleonastic motion event constructions might not be interesting from a framing typological point of view on standard language varieties, it would be interesting to see studies on these constructions in Portuguese varieties such as VBP. Such a task could perhaps make use of the procedure of identifying metalinguistic commentaries.

6 Conclusion

While EP appears to have a higher frequency of pleonastic intransitive double Path constructions than BP in the ptTenTen11 corpus, it remains to be clarified how these constructions fit within a more general picture (through thorough comparisons with other motion event constructions perhaps?). Although not clear how they relate to such a general view, the frequencies seem to indicate that the investigated double Path constructions are rather peripheral phenomena. Also, no evidence that challenges the Boundary crossing constraint in Portuguese was identified. Manner verbs did occur alongside Path adverbials but perhaps not to the general level sought by framing typological classification. Taken together, the data on double Path constructions and boundary crossing events did not coherently present neither BP nor EP as being on a specific lexicalization Path within the verb-framed and satellite-framed dichotomy/continuum, which of course is linked to the issue of perhaps simply not being general enough phenomena for such a conclusion to be drawn. Judging from the quite varied components that were considered Path elements – for example the periphrastic Ground specifying element in Batoréo and Ferrari's (2016) investigation – perhaps another approach to studying motion events in Portuguese (and in general/cross linguistically) could be enriching. Distributed spatial semantics, which focuses its analysis on the spatial relationship between Figure and Ground, that both are necessary components for the expression of translational motion, could be one such nuancing approach.

7 References

Batoréo, Hanna J & Ferrari, Lilian. 2016. Events of Motion and Talmyan Typology: Verb-framed and Satellite-framed Patterns in Portuguese. *Cognitive Semantic*s. 2, 56–79.

Cambridge Sketch Engine: Using Corpus Query Language (CQL). 2012. Cambridge English Corpus. Cambridge University Press. (https://www.cambridge.org/sketch/help/userguides/CQL %20Help%201.3.pdf) (Accessed 2022-01-27)

Cifuentes-Férez, Paula & Molés-Cases, Teresa. 2020. On the translation of boundary-crossing events: Evidence from an experiment with German and Spanish translation students. *Vigo International Journal of Applied Linguistics*. 17, 87-111.

Desagulier, G. 2017. Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics. Springer International Publishing AG.

Hutchinsson, Amélia P. & Lloyd, Janet. 2003. *Portuguese: An Essential Grammar.* 2nd edn. Routledge.

Kilgarriff, A., Jakubíček, M., Pomikalek, J., Sardinha, T. B., & Whitelock, P. 2014. PtTenTen: A Corpus for Portuguese Lexicography. *Working with Portuguese Corpora*. 111-130.

Lehmann, Christian. 2005. Hypercharacterization. In Booij, Gert E. & van Marle, Jaap (eds.), *Yearbook of Morphology*. 119-154. Heidelberg: Springer. http://www.uni-erfurt.de/ sprachwissenschaft/personal/lehmann/CL_Publ/ Hypercharacterization.pdf

Meirelles, Letícia Lucinda. 2019. Verbos de movimento do português brasileriso: evidências contrauma tipologia binária. *Revista de estudos da linguagem.* 27(2), 1101-1124.

Müller de Oliveira, Gilvan. 2015. Language Policy & Globalization: The Portuguese Language in the Twenty-First Century. Moita-Lopes, Luiz Paulo. *Global Portuguese: Linguistic Ideologies in Late Modernity*. 27-46. Routledge.

Queriquelli, Luiz Henrique Milani & Moura, Heronides. 2021. Path Marking on the Satellite in Latin, Old Portuguese, and Vernacular Brazilian Portuguese. Fórum Linguístico 18(2), 6124–6136. http://dx.doi.org/10.5007/1984-8412.2021.e76267

R Core Team. 2017. R: A language and environment for statistical computing . R Foundation for Statistical Computing. Vienna, Austria http://www.R-project.org/. ISBN 3-900051-07-0

Rodrigues da Silva Júnior, Ilson. 2015. Florianópolis: Universidade federal de Santa Catarina.

Santos Filho. 2013. Florianópolis: Universiaded federal de Santa Catarina.

Santos Filho, Dorival Gonçalves. 2017. O padrão de lexicalização do português brasileiro: evento de movimento. *Revista Electrônica do Netlli.* 5(2), 103-123.

Sinha, Chris & Kuteva, Tania. 1995. Distributed Spatial Semantics1. *Nordic journal of linguistics*, 18(2), 167-199.

Talmy, Leonard. 2000. *Toward a Cognitive Semantics Volume II: Typology and Process in Concept Structuring*. Cambridge, Massachusetts: MIT Press.

Appendices

Appendix 1: top 20 results for 'V.* para dentro'

size: 23425

query: Query:[tag="V.*"][word="para|pra|prá|pa|pá"][word="dentro"]

lemma, Frequency

olhar para dentro, 2451

levar para dentro, 1655

voltar para dentro, 1498

trazer para dentro, 1463

ir para dentro, 1096

correr para dentro, 630

empurrar para dentro, 482

virar para dentro, 455

colocar para dentro, 390

puxar para dentro, 344

entrar para dentro, 343

arrastar para dentro, 272

transportar para dentro, 252

olhar pra dentro, 232

jogar para dentro, 220

pular para dentro, 216

saltar para dentro, 208

sugar para dentro, 190

vir para dentro, 182

passar para dentro, 181

Appendix 2: top 20 results for 'V.* para fora'

size: 34561

query: Query:[tag="V.*"][word="para|pra|prá|pa|pá"][word="fora"]

lemma, Frequency

ir para fora, 2384

chutar para fora, 1518

colocar para fora, 1500

levar para fora, 1308

olhar para fora, 1050

jogar para fora, 917

mandar para fora, 838

viajar para fora, 789

sair para fora, 722

pôr para fora, 585

projetar para fora, 584

cabecear para fora, 579

empurrar para fora, 545

saltar para fora, 523

voltar para fora, 486

arremessar para fora, 460

lançar para fora, 452

puxar para fora, 404

colocar pra fora, 398

arrastar para fora, 368