

Differentiating Brazilian and European Portuguese multiword adverbs

Izabela Müller¹, Nuno Mamede², Jorge Baptista¹

¹Universidade do Algarve, INESC-ID Lisboa.

²Universidade de Lisboa, INESC-ID Lisboa.

Abstract

The identification of Portuguese multiword adverbs in texts such as *de fininho* ‘sneakily’ or *em plena luz do dia* ‘in broad daylight’ is an essential part of determining the meaning units forming a text. This paper explores the boundaries of adverbial multiword expressions in European and Brazilian Portuguese varieties. The primary purpose is to determine the frequency and distribution of the expressions in the *pt TenTen18 corpus* from Sketch Engine. Furthermore, we aim to analyze lexical and grammatical differences, as well as some cases of ambiguity and *faux-amis* (false-friends) between the two varieties. Finally, we seek to establish the boundaries of the variety-specific lexicon (European and Brazilian), which is part of our ongoing investigation on identifying, collecting, and providing a syntactic-semantic description of multiword adverbs in (Brazilian) Portuguese based on similar studies for European Portuguese, French, and Spanish. We adopt and adapt the classification proposed by Molinier and Levrier (2000), based on the Lexicon-Grammar theoretical-methodological framework.

Keywords: Brazilian Portuguese, European Portuguese, language varieties, compound adverb, multiword expressions.

Resumo

A identificação de advérbios compostos em português em textos, como *de fininho* e *em plena luz do dia*, é uma parte essencial para determinar as unidades de significado que formam um texto. Este artigo explora os limites das expressões adverbiais compostas nas variedades do português europeu e brasileiro. O principal objetivo é verificar a frequência e distribuição das expressões nos corpora *ptTenTen18* do Sketch Engine. Além disso, pretendemos analisar diferenças lexicais e gramaticais, bem como alguns casos de ambiguidade e *faux amis* (falsos amigos) entre as duas variedades. Por fim, buscamos estabelecer a extensão do léxico específico de cada variedade (europeia e brasileira), o que faz parte da nossa investigação em curso sobre identificação, recenseamento e descrição sintático-semântica de advérbios compostos em português (brasileiro) com base em estudos semelhantes para o português europeu, francês e espanhol. Adotamos a classificação proposta por Molinier & Levrier (2000) com base no quadro teórico-metodológico Léxico-Gramática.

Palavras-chave: Português brasileiro, português europeu, variedades da língua, advérbios compostos, expressões multipalavra.

1. Introduction

This paper delineates differences and similarities in adverbial constructions, specifically multiword or compound adverbs, across Brazilian and European Portuguese. The analyses presented here are part of a broader ongoing study focusing on identifying, classifying, and describing the lexical and syntactic features of multiword adverbs in Brazilian Portuguese (Müller *et al.*, 2022, 2023). The motivation for this study is to extend and refine an existing lexicon of multiword adverbs of Portuguese, which already contains expressions from

European Portuguese, initially compiled and analyzed by Palma (2009), by including the adverbial expressions from Brazilian Portuguese. Ultimately, the goal is to create a comprehensive lexicon with linguistic descriptions that can be used in Natural Language Processing (NLP) and for several applications, to facilitate the identification of these lexical units in texts.

The distinction between two language varieties can be established on other different linguistic dimensions, including orthography, lexicon, phonetics, morphology, syntax, semantics, or pragmatics (Zampieri *et al.*, 2012). This paper examines the syntactic and lexical differences between Brazilian and European Portuguese multiword adverbs based on their occurrence in written texts, as can be observed in the large-sized (over 1 billion tokens) Portuguese *ptTenTen18* corpus (Kilgarrif *et al.*, 2014) available on the Sketch Engine¹ platform. The (normalized) frequency in each corpus partition (one from BP and the other from EP), corresponding to texts in a given variety, will indicate the usage trend in each variety. The information will be later integrated into the STRING² system, a natural language processing (NLP) pipeline developed for Portuguese (Mamede *et al.*, 2012). The linguistic resources and outcomes from this study will be made available to the scientific community.

We start this work by describing its objectives, the concept of compound/multiword adverbs, and the theoretical framework employed in our approach to studying compound, multiword adverbs in Portuguese (Section 1). Section 2 describes the study's methodology, particularly the criteria for gathering multiword/compound adverbs in Brazilian Portuguese and the procedures for sampling our data, and contrasting Brazilian and European Portuguese, using different corpora of smaller dimensions against the large corpus *ptTenTen18* in Sketch Engine. Section 3 discusses our findings and results and provides a detailed analysis of interesting grammatical, idiomatic, and lexical variation cases. We present our final remarks and future work in Section 4.

1.1 Compound adverbs

Compound adverbs or *multiword adverbial expressions*, referred to as *adverbial locutions* in Portuguese grammatical terminology, are expressions constituted of two or more words in specific combinations, that operate as a single lexical unit. The structure of these expressions is usually frozen, meaning that their elements occur in a constrained sequence that, in general, does not allow permutation, insertion, transposition, or reduction (ellipsis) of their components (Guimier, 1996; M. Gross, 1982, 1986; Mel'čuk, 2023). A defining characteristic of multiword adverbs is their semantic non-compositionality; that is, the meaning of the expression cannot often be derived from the individual meanings of its components, therefore producing their idiomatic nature³:

- (1) [As vacinas foram criadas] *a toque de caixa* [na tentativa de controlar a pandemia]
'Vaccines were created *on the fly/hastily* in an attempt to control the pandemic'
lit.: 'by the touch of drums/boxes'

The example above (1) illustrates how the fixed expression *a toque de caixa* is used. This idiom demonstrates limited or no flexibility in terms of lexical variation on (i) the substitution of their elements: **a toque de caixote/pandeiro* 'by touch of box/tambourines': (ii) the insertion of any lexical element **a toque de uma grande Caixa* 'by touch of a big box': (iii) the deletion of one of the elements, **a toque* 'by touch': and

¹ <https://www.sketchengine.eu/pttnten-portuguese-corpus/> [last access: 28/03/2024] (all URLs were checked on this date).

² https://string.hlt.inesc-id.pt/wiki/Main_Page

³ We displayed the compound adverb in *italics* separately from the base sentence, shown in squared brackets. A tentative translation is also provided. If an equivalent idiomatic expression exists, this might have been used. In some cases, the literal (word-by-word) translation is also provided, highlighting the expression's idiomatic meaning. Most examples in this paper were retrieved from the *ptTenTen18*, accessible through the Sketch Engine platform, and eventually adapted, though only minimally.

lastly, (iv) the transposition of their elements, /**a caixa de toque* ‘the box of touch’. Besides the fact that this string is rigidly frozen, its overall meaning is non-compositional, as denoted by the translation equivalent ‘hastily/on the fly’.

1.2. Formal classification (M. Gross, 1986, 1996a)

In this study, we adopt the theoretical-methodological framework of Lexicon-Grammar, conceptualized by Maurice Gross (1975, 1981, 1996b), and based on the Transformational Operator Grammar of Zellig S. Harris (1991). This framework considers that the minimal unit of linguistic analysis is the elementary sentence consisting in the expression of a minimal semantic predicate. This principle states that only in an elementary sentence can one adequately determine the syntactic properties and, therefore, the precise meaning of linguistic expressions. For this reason, the classification of compound adverbs is established on the syntactic properties of lexical units (both simple and compound) within the framework of an elementary sentence (M. Gross, 1984, p. 275), particularly within the elementary sentence determined by the nuclear predicative element that defines the sentence. This predicative element can be a predicative noun, a full verb, or a predicative adjective. When the adverb modifies one of these components, it acts as an *internal modifier* of that sentence. *External adverbial modifiers* modify the entire sentence and can often appear detached from it by commas or other formal devices.

In order to deal with the adverbial modifiers as a single construct, M. Gross (1986, p. 12) proposed the concept of *generalized adverb* (in French: *adverbe généralisé*), which integrates both compositional, syntactically analyzable, expressions and frozen, non-compositional, and idiomatic expressions (2)–(5). These constructions include:

(i) *simple* adverbs, both *non-derived* adverbs,

- (2) [O Leo estuda] *lá / muito / assim*
‘Leo studies there / a lot / in this way’

and *derived* adverbs, mainly with the suffix *-mente*, ‘ly,’ originated from adjectives:

- (3) [O Leo estuda] *afincadamente = de forma afincada*
‘Leo studies diligently’ = ‘in a diligent way’

(ii) circumstantial complement phrases, typically with the morphosyntactic structure of a prepositional phrase:

- (4) [O Leo estuda] *todos os dias / nos feriados / em casa / com afinco*
‘Leo studies every day / on holidays / at home / with diligence’

(iii) circumstantial subordinate clauses, both free and frozen:

- (5) [O Leo estuda o caso] *enquanto eu descanso / enquanto o diabo esfrega um olho*
‘Leo studies the case while I rest / quickly’ lit.: ‘while the devil rubs one eye’

M. Gross (1986) proposed a taxonomic method to label and organize French compound adverbs based on their internal sequence of grammatical categories. The basic structure for these adverbs is a prepositional phrase of the form *Prep Det N Modif*, where *Prep* stands for a preposition, *Det* for a determiner, *N* for a nominal element, and *Modif* for a modifier. The presence, absence, or variation in the arrangement of these components results in various formal categories/classes. Based on this structure, M. Gross (1986) has sub-classified the French adverbial expressions into 16 classes, of which this study selects ten for the classification of compound adverbs in Portuguese. Table 1 displays the current breakdown of the lexicon and the classes of compound adverbs considered at this stage. It includes the structural description of each class, which is determined by their different internal structure (sequence of grammatical categories). For instance, the class PAC is made of adverbial expressions containing a *Preposition*, an *Adjective* and a frozen nominal *Component*. The other

classes follow suit. Table 1 also includes an example of each compound adverb of the assigned class, and denotes the adverb's language variety (EP: European Portuguese; BP: Brazilian Portuguese; or EPBP: common to both varieties). At this stage, the lexicon compiles approximately 3,500 compound adverbs, with nearly half of the entries shared by both varieties. The numbers represent the amount of compound adverbs there are per class, per variety and common to both varieties. For instance, the total of the PAC class is 269 compound adverbs, being 25 exclusive of EP, 129 exclusive of BP and 115 compound adverbs shared by both. Notice that these numbers may vary as the study advances.

Notably, this study excludes certain classes initially proposed by M. Gross (1986). Please see Section 2.1 for a detailed description of the types of multiword adverbs that were included in or excluded from this study.

Table 1. Breakdown of the lexicon of compound adverbs in Portuguese.

Class	Structure	Example	EP	%	BP	%	EP-BP	%	Total	% Class
PC	<i>Prep C</i>	<i>em vão</i> 'in vain'	77	8%	497	49%	444	44%	1018	29%
PDETC	<i>Prep Det C</i>	<i>pelo menos</i> 'at least'	104	14%	271	36%	386	51%	761	21%
PAC	<i>Prep Adj C</i>	<i>de má vontade</i> 'unwillingly'	25	9%	129	48%	115	43%	269	8%
PCA	<i>Prep C Adj</i>	<i>por maioria absoluta</i> 'by an absolute majority'	45	14%	119	36%	166	50%	330	9%
PCDC	<i>Prep C1 de C2</i>	<i>por conta da casa</i> 'on the house'	43	15%	136	47%	109	38%	288	8%
PCPC	<i>Prep C1 Prep C2</i>	<i>da cabeça aos pés</i> 'from head to toe'	60	16%	160	41%	166	45%	386	11%
PCONJ	<i>Prep C1 Conj C2</i>	<i>com unhas e dentes</i> 'tooth and nail'	16	6%	102	40%	139	54%	257	7%
PF	<i>Frozen Clause</i>	<i>dito isso</i> 'having said that'	5	5%	72	69%	27	26%	104	3%
PV	<i>Prep V W</i>	<i>até dizer chega</i> 'until we have had enough'	1	4%	11	42%	14	54%	26	7%
PJC	<i>Conj C</i>	<i>e por aí vai</i> 'and so on'	2	3%	52	81%	10	16%	64	18%
Total			378	11%	1,546	44%	1,576	45%	3,503	

1.3 Syntactic-semantic classification

Our analysis considered the syntactic-semantic description and classification of adverbs proposed by Molinier and Levrier (2000) to examine the functions and different relations multiword adverbs establish within and with sentences. These authors' work focuses on the classification of French, single-word, derived adverbs ending in *-ment*. It lays the foundation for our study. For the most part, we consider that this classification framework largely applies not only to the corresponding simple-word adverbs ending in *-mente* in Portuguese, as identified by Fernandes (2011), but can also be extended or adapted to the description of Portuguese compound adverbs. In essence, the authors proposed a set of criteria to distinguish and classify adverbial constructions based on the syntactic relations the adverbs establish with the sentence as a whole (P) or with the elements within the sentence on which they operate (M). They propose two main types of adverbial constructions:

- (P) the adverbs that function as external modifiers of sentences; and
- (M) the adverbs that modify internal elements within sentences.

Take the following examples (6) and (7) as base sentences:

- (6) *Apesar de tudo*, [eu acredito nela]
'After all, [I believe her]'
- (7) [O Leo sabe o poema] *de cor e salteado*
'[Leo knows the poem] by heart' (lit.: 'by heart and in random order')

Two main properties, (A) and (B), are used to verify whether the compound adverb operates as an internal or as an external modifier:

- (A) Topicalization and negation of the main predicate:
 - (6a) *Apesar de tudo*, [eu (**não**) acredito nela]
'After all, I do not believe her'
 - (7a) **De cor e salteado*, [o Leo **não** sabe o poema]
'By heart, [Leo does not know the poem]'
- (B) Clefting (or extraction using *ser...que* 'it be...that')
 - (6b) **É apesar de tudo* [que eu acredito nela]
'It is after all that [I believe her]'
 - (7b) *É de cor e salteado* que [o Leo sabe o poema]
'It is by heart and spelling that [Leo knows the poem]'

In general terms, *sentence external modifiers*, such as *apesar de tudo* 'after all' in (6a), can undergo topicalization, which means the adverb can be moved to the beginning of the sentence while the sentence can be under the scope of negation. Since external modifiers operate on the entire sentence, they lie outside the scope of the negation of the predicate they operate on. On the other hand, *sentence external modifiers* cannot be subjected to *clefting* (or the extraction *ser...que* 'it be...that'), as this operation focuses on the internal constituents of a sentence only, as seen in (6b). On the contrary, *sentence-internal modifiers*, such as *de cor e salteado* 'by heart', typically cannot be topicalized when the sentence is under negation, as seen in (7a). Their role is to modify specific elements within the sentence; in this case, this is the *manner* modifier of the verb. Thus, these *internal modifiers* can undergo *clefting* (7b).

Molinier and Levrier (2000) describe the class of external proposition-modifying adverbs based on the simultaneous conjunction of the topicalization property of the adverb from the negated sentence (A) and the impossibility of its clefting (B). Other combinations of these properties are relegated to the various classes of adverbial modifiers internal to the proposition. The adverbs that function as external modifiers to the proposition are further divided into three classes: (i) conjunctive adverbs (PC), (ii) disjunctive adverbs of style (PS), and (iii) disjunctive adverbs of attitude (PA). The latter one is subdivided into four subclasses: (a) adverbs of habit (PAh), (b) evaluative (PAe), (c) modal (PAm), and (d) subject-oriented adverbs (PAs). Sentence-internal modifier adverbs are divided into six subclasses: (iv) manner adverbs (MV), (v) subject-oriented manner adverbs (MS), (vi) adverbs of time (MT), (vii) point-of-view adverbs (MP), (viii) quantification adverbs (MQ), and (ix) focusing adverbs (MF). Finally, we added a new class (x) locative adverbs (ML). Locatives are now new in the realm of adverbs; however, they were not considered in Molinier and Levrier's (2000) classification.

All 3,510 multiword adverbs collected for this study were classified into syntactic-semantic classes as proposed by Molinier and Levrier (2000). Table 2 presents the distribution of the lexicon of multiword adverbs by the classes described above and provides diatopic information regarding European Portuguese (EP) and Brazilian Portuguese (BP). For instance, the MV class (manner adverbs), which has the highest number of multiword adverbs in our study, includes 274 adverbs exclusive to EP, 963 exclusives to BP, and 927 common to both varieties, totaling 2,164 multiword manner adverbs in the lexicon. The data for the remaining classes is shown in a similar way.

Table 2. Distribution of multiword adverbs according to their syntactic-semantic classification

Class	Example	EP	BP	EP-BP	Total	%
PC (conjunctive)	<i>afinal de contas</i> ‘after all’	15	93	122	230	0.07%
PS (disjunctive of style)	<i>com toda a franqueza</i> ‘in all honesty’	4	27	27	58	0.02%
PA (disjunctive of attitude)	<i>em geral</i> ‘in general’	2	28	35	65	0.02%
MV (manner)	<i>por amor à pátria</i> ‘for love of country’	274	963	927	2,164	0.62%
MS (subject-oriented manner)	<i>de boa fé</i> ‘in good faith’	9	36	63	108	0.03%
MT (time)	<i>ao romper do dia</i> ‘at the break of day’	55	206	251	512	0.15%
MP (point of view)	<i>na prática</i> ‘in practice’	0	2	2	4	0.00%
MQ (quantification)	<i>aos montes</i> ‘in abundance’	13	90	66	169	0.05%
MF (focalizer)	<i>em especial</i> ‘especially’ <i>nos confins do mundo</i> ‘at the ends of the earth’	1	7	11	19	0.01%
ML (locative)		7	102	72	181	0.05%
Total		382	1.556	1.576	3.510	

1.4. Related work

With growing interest in studies dealing with multiword expressions, specifically multiword adverbs, across other languages, parallel studies have contributed to a better understanding of compound adverbs' syntactic and semantic features. Palma (2009) built the first lexicon of compound adverbs in European Portuguese with approximately 1,600 entries, along with a description of their internal structure and syntactic properties. This work was later compared and contrasted to their equivalent constructions in Spanish (Català *et al.*, 2020) to highlight the challenges concerning the automatic translation between Spanish and European Portuguese compound adverbs. Fernandes (2011), in turn, applied the syntactic-semantic taxonomy proposed by Molinier and Levrier (2000) to the 1,000 most frequently occurring derived adverbs in Brazilian Portuguese, in view of their disambiguation using machine learning techniques. These data were later added to the lexicon of the Portuguese NLP system STRING (Mamede *et al.*, 2012).

Regarding research on other languages, an electronic dictionary containing 6,800 multiword adverbs with syntactic and semantic descriptions was initially developed for the French language (M. Gross, 1986) and later integrated into NLP systems (Laporte & Voyatzi, 2008). Similarly, De Gioia's (2001) work on Italian described around 6,000 compound adverbs. For the Japanese language, the Dictionary of Multiword Expressions-JDMWE (Shudo *et al.*, 2011) catalogs 6,000 adverbial expressions, complete with their syntactic and semantic descriptions. In Czech, a study by Žižková (2018) added 103 new compound adverbs to an existing morphological MWE dictionary, extending it to 407 entries. Català (2003) created a comprehensive dictionary of approximately 6,000 frozen adverbial expressions for Spanish. This work follows M. Gross' (1996a)

syntactic framework, classifying these expressions into 11 classes. Moreover, the SentiText lexicon, developed in Spanish (Moreno-Ortiz *et al.*, 2013), specifically for Sentiment Analysis, includes 2,255 MWE adverbs.

2. Methods

In this Section, we present the methods used in this paper. Section 2.1. describes the steps and criteria for building the lexicon of multiword adverbs in Brazilian Portuguese, bootstrapping it from the lexicon originally built by Palma (2009) for the European Portuguese. Section 2.2., explains the selection of 300 most frequent multiword adverbs. Section 2.3. describes how these adverbs were used for determining their distribution in a large-sized corpora, namely, *ptTenTen18* corpus in both partitions (EP and BP).

2.1. Computational lexicon of Portuguese MW adverbs

The ongoing work consists of collecting, providing a formal classification, and describing the syntactic properties of Brazilian Portuguese multiword adverbs. To build a large-scale lexicon of compound adverbs in Portuguese, we started examining an initial lexicon built for European Portuguese, Palma (2009), which comprises approximately 1,600 multiword adverbs, and to identify the adverbs that were common to Brazilian Portuguese, indicating whether they occur in BP or not. Secondly, we perused various Brazilian idioms and multiword expression-specific dictionaries (Rocha, 2011; Schwab, 1985; Silva, 2013) to collect multiword adverbs in that variety of Portuguese. To systematically include or exclude expressions, we established the following criteria:

1 – Idiomatic Expressions: we include in the lexicon idiomatic, i.e., semantically non-compositional adverbial expressions that feature some degree of fixedness; therefore, these expressions restrict the operations of substitution, insertion, deletion, or transposition of its elements (8), as the example below:

- (8) [O êxito não acontece] *do pé para a mão*
lit.: ‘Success does not happen *unexpectedly*’ lit.: ‘from foot to hand’
**da mão para o pé* lit.: ‘from hand to foot’
**do pé esquerdo para a mão direita* lit.: ‘from the left foot to the right hand’
**do pé para o braço* lit.: ‘from foot to arm’

Other examples of multiword adverbs include idiomatic temporal expressions (9)–(10), e.g.:

- (9) *No tempo do Onça*, [ele já era singular]
lit.: ‘*In the old days*, he was already unique’ lit.: ‘in the time of the Jaguar’ (BP)
- (10) [Por que chamar uma cidade por um apelido dado pelos portugueses] *no tempo dos Afonsinhos?*
lit.: ‘Why call a city by a nickname given by the Portuguese *in the old days*’ lit.: ‘in the time of the Afonsinhos?’ (EP)

2 – Single-word equivalents: we include adverbial constructions that directly correspond or are equivalent to a single word adverb such as *devagar* ‘slowly’ (*a conta-gotas* ‘drop by drop’), and derived adverbs ending in *-mente*, such as (11), *em geral* ‘in general’ = *geralmente* ‘generally,’ e.g.:

- (11) *Em geral* = *geralmente*, [o problema que se verifica nos estabelecimentos comerciais]
‘In general = generally, this is a problem in commercial establishments’

3 – Expressions presenting limited lexical variation: we admit adverbial constructions that allow some degree of variation in their components, e.g.:

- (12) *A certa/dada/esta/essa/*aquela/?alguma altura*, [o ministro resolveu definir a questão]
 ‘At a certain/a given/this/that point, the minister decided to define the issue’

These types of expressions presented in example (12), exhibit some variation of elements, though several idiosyncratic combinatorial constraints can be found. For example, in Brazilian Portuguese, temporal adverb formed with *altura* ‘point, level’ can display variation of the preposition *a / em certa altura* ‘at/in a certain point/level’ (10), and the indefinite determiner *certa* ‘certain’ can be replaced by *dada* ‘given’ and the demonstratives *esta/essa/?aquela*, (with varying degrees of acceptability). For conciseness, these expressions are represented by a *local grammar*, and only one entry is referred to in the multiword adverbs’ lexicon. However, notice that a detailed description of these somewhat ‘productive’ expressions is necessary. For example, with preposition *a* ‘at’ and the determiner *mesma* only the **locative** interpretation exists: *Ele sentado e eu de pé estávamos quase à mesma altura* ‘He sitting and me standing, we were almost the same height’. In European Portuguese, both **locative** and **temporal** values exist, even if the temporal use is rarer: *A inscrição e o balão estão à mesma altura* ‘The inscription and the balloon are at the same height’; *Outro nome de relevo, chegado à mesma altura [...] foi Carlos* ‘Another prominent name, arriving at the same time [...], was Carlos’. (An idiomatic, frozen use of the same string, though rarer, is also attested: [...] *eu pude responder à mesma altura* ‘I was able to respond at the same level.’ Conversely, the expression *na mesma altura* ‘at the same time’, with the preposition *em* ‘in’, seems only to have a **temporal** value in European Portuguese, and predominantly a **locative** sense in Brazilian Portuguese, e.g., [...] *com máquinas modernas e precisas, sempre mantidas na mesma altura para um perfeito nivelamento*. ‘with modern and precise machines, always kept at the same height for perfect leveling’, though some rare **temporal** use can be attested: *Na mesma altura, os jesuítas [...] abandonaram as aldeias [...]* ‘At the same time, the [...] Jesuits [...] abandoned the villages’.

4 – Frozen comparative constructions: we include only those that are exclusive to Brazilian Portuguese (13)–(14), in order to complement the lexicon already represented in the STRING system for the European Portuguese and developed from the data of Ranchhod (1991), e.g.:

- (13) [Tudo está caminhando] *como o diabo gosta*, [não é?]
 ‘Everything is going the way of the devil likes it, isn't it?’
- (14) [Tem gente que] *fuma como uma caipora* [e não tem câncer!]
 ‘There are people who smoke a lot and do not get cancer!’ lit.: ‘like a Caipora’

Thus, in the process of assembling the lexicon of compound adverbs in Portuguese, we followed well-defined criteria, which led us to exclude some constructions, to ensure consistency in our work.

1 – Prepositional/conjunctive constructions as seen in the examples (15)–(16) below:

- (15) [Ele fez isso] *a custa de [muito suor/muito esforço]*
 ‘He did it at the cost of a lot of sweat/effort’
- (16) [Queria saber, o que vai fazer] *em relação a [este assunto/esse problema?]*
 ‘I would like to know what you're going to do about this issue/this problem’

These types of expressions select a distributional free element, and for the purpose of this study, we considered only completely frozen adverbial expressions, or with limited variation. These expressions belong to the classes PCDN (Prep + C + *de* + Noun) and PCPN (Prep + C + Prep + Noun), proposed by M. Gross (1986), included in the work of Palma (2009) for the European Portuguese.

2 – Predicative, idiomatic constructions associated with the support verb *estar* (17) ‘to be’ which have already been described by Ranchhod (1990), e.g.:

- (17) [João está] *com a corda no pescoço*
‘João is *in a difficult situation*’ lit.: ‘has the rope around his neck’

3 – Adverbial constructions associated with the support verb *ter* (18) ‘to have’ e.g.:

- (18) [Sou Português] *com muita honra*
Tenho muita honra em ser Português.
‘I am Portuguese with great honor’ (= I have great honor to be Portuguese)

often associated with support verb *estar com*, (19)–(20a/b) lit.: ‘be with’, and that can be reduced to adverbial phrases:

- (19) [O Pedro] *tem fome*
lit.: ‘Pedro has hunger’ (=Pedro is hungry)
- (20a) [O Pedro] *está com fome*
lit.: ‘Pedro is with hunger’ (=Pedro is hungry)
- (20b) [O Pedro fez isso] (estando) *com fome*
‘Pedro did this (being) with hunger’

4 – Temporal compositional, productive expressions as seen in the examples (21)–(22):

- (21) *De noite*, [performances de música e luzes tomam conta da região]
‘*At night*, performances of music and lights take over the area’
- (22) [Os resultados devem ser conhecidos] *na próxima semana*
‘The results should be known *next week*’

These types of constructions allow a relatively wide range of formal and lexical variation, especially when involving modifiers such as *próximo* ‘next’, *passado* ‘past’ and many others (Baptista, 1999). Most of these productive expressions are already listed as *local grammars* (LG) in the STRING system.

5 – Comparative adverbial constructions such as the following examples (23)–(24):

- (23) [Ela fuma] *como uma chaminé* [e se recusa a comer verduras]
‘She smokes *like a chimney* and refuses to eat vegetables’
- (24) [Meu cunhado é] *surdo como uma porta*
‘My brother-in-law is *deaf as a doornail*’

While these expressions are frozen and idiomatic, they typically associate with a specific verb or adjective. As mentioned above, these expressions are the subject of a separate and future work.

So far, our study has compiled approximately 1,900 new expressions in Brazilian Portuguese, expanding the lexicon to a total of 3,500 compound adverbs. While many expressions are common to both varieties and may seem similar, several aspects merit consideration. These include determining whether an expression in European Portuguese carries the same meaning as its Brazilian counterpart and vice versa, as well as identifying cases of false cognates, grammatical disparities, and lexical differences. The validation process for each

multiword adverb follows a two-step approach: firstly, the proposed entry undergoes review by two native speakers, one from each variety of Portuguese; secondly, its precise meaning and syntax are then confirmed through corpus-based verification.

2.2. Sampling seeds

We aimed to use large-sized corpora to determine the frequency and compare the distribution of compound adverbs in European and Brazilian Portuguese. We have elected the *PortugueseTenTen18* corpus (Kilgarrif *et al.*, 2014), developed under the WaC (Web-as-Corpus) framework (Wagner Filho *et al.*, 2018), recently crawled from the web (data downloaded in February–June 2018), and accessible through the Sketch Engine⁴ platform. The corpus contains 8.7 billion tokens (7.4 billion words). It mainly consists of 3 sub-corpora, retrieved, respectively, from the Brazilian national top domain .br (+5.5 billion tokens), the Portuguese national top domain .pt (206.8 million tokens), and Wikipedia 2018 (240.5 million tokens). It is possible to query each sub-corpus (.br or .pt) separately, which was particularly suitable for our purposes in this paper.

In the absence, to the best of our knowledge, of an API that would enable us to retrieve from the corpus the entire list of adverbial MWE collected so far in the lexicon of Portuguese compound adverbs, it was necessary to retrieve these adverbs manually. Due to the current size of the lexicon (+3,500 entries), a selection of its entries had to be established.

We defined a subset of this lexicon's entries, which will work as seeds for retrieving those compound adverbs from *PortugueseTenTen18*. The selection was based on the frequency of the entire lexicon of multiword adverbs that could be observed in two smaller corpora:

1. A fragment of the European Portuguese CETEMPúblico corpus (Rocha & Santos, 2000),⁵ with approximately 10 million words, built from the online edition of the *Público* daily newspaper (2004–2009); and
2. A corpus built from several years of the online edition of the *Folha de São Paulo* daily newspaper, from 1994–1998, with approximately 24 million words (Folha de São Paulo, 1999).⁶

Both corpora consist of journalistic texts and cover a wide variety of topics, so it was deemed appropriate for this exploratory phase.

The first step involved verifying how many of the 3,500 expressions in the multiword adverbs' lexicon were identified in both corpora. We encountered a total of 2,080 expressions in the two corpora. Specifically, 1,584 were found in the CETEMPúblico corpus and 1,762 in the Brazilian corpus, with 1,266 shared between both corpora. In the second stage, 300 compound adverbs,⁷ representing just under 10% of the entries in the current lexicon, were manually selected, respecting a set of criteria. These criteria prioritized (i) the longer and often highly idiomatic adverbs such as *com conhecimento de causa* 'with knowledge of the facts'; *contra tudo e contra todos* 'against everything and everyone'; (ii) adverbs that do not constitute sub-expressions of a longer adverb or adverbial expression, e.g., *no tempo* 'BR: in the outdoors' as opposed to *no tempo das vacas frias* 'in the time of cold cows, in bad times'; and (iii), as much as possible, non-ambiguous or non-polysemous expressions, e.g., *a cântaros* 'in buckets'. Additionally, we chose 30 adverbs from each variety (EP and BP) with 0-occurring adverbs in each respective corpus to examine if the asymmetry in frequency on these smaller corpora was consistent across both *ptTenTen18* partitions (EP and BP).

⁴ <https://www.sketchengine.eu/>

⁵ www.linguateca.pt/cetempublico

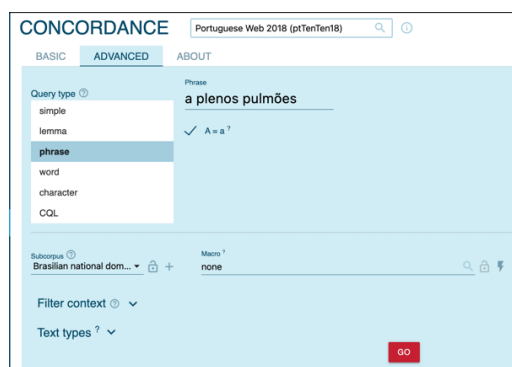
⁶ The authors thank Professor Oto A. Vale (UFSCar, São Carlos, Brasil) for making this resource available.

⁷ See Appendix. Further information, including examples, can be found in: https://string.hlt.inesc-id.pt/wiki/Compound_Adverbs

2.3. Sampling from the Portuguese Web 2018 (ptTenTen18) corpus

Once the seed expressions had been selected, a concordance of each multiword adverb was retrieved from the Portuguese Web 2018 *ptTenTen18* corpus using the Sketch Engine interface (Figure 1). This was done twice, once for the European Portuguese partition and then for the Brazilian Portuguese partition.

Figure 1. Sketch Engine concordance query: interface menu settings for retrieving the compound adverb *a plenos pulmões* lit.: ‘at full lungs, at the top of his lungs’.



The frequency of the adverbial expressions within each partition of the corpus was registered, and the corresponding concordances were downloaded for subsequent inspection. Following this, the frequency of each adverbial expression in each partition of the corpus was calculated. A fragment of the resulting table is shown in Table 3. Note that the frequency figures provided by the *ptTenTen18* corpus for each sub-corpus are calculated from the total value of tokens in both sub-corpus BP and EP combined (8,731,838,327), without differentiation by language variety. To address this, we recalculated the frequencies as the ratio of the number of occurrences of the given phrase by the total number of tokens specific to each language sub-corpus, namely, 5,542,074,775 for BP and 206,869,477 for EP, respectively. Table 3 presents a selection of 11 examples of multiword adverbs that appear in both corpora, CETEMPúblico and *Folha de São Paulo*, as verified within the European and Brazilian partition of the *ptTenTen18*, along with their relative frequency. The table is systematically organized in descending order of the most frequent multiword adverbs in the BP (BP-SE#) corpus.

Table 3. Frequency data from corpora

Adverbs	EP CETEMPúblico	BP FolhaSPaulo	EP-SE#	BP-SE#	EP-SE%	BP-SE%
<i>no entanto</i>	2,114	4,351	49,948	922,648	0.000241447	0.000166481
<i>hoje em dia</i>	135	334	6,828	139,835	0.000033006	0.000025232
<i>por enquanto</i>	169	816	2,341	100,341	0.000011316	0.000018105
<i>o tempo todo</i>	5	281	415	74,205	0.000002006	0.000013389
<i>passo a passo</i>	19	63	658	64,323	0.000003181	0.000011606
<i>caso contrário</i>	55	233	2,282	62,587	0.000011031	0.000011293
<i>de preferência</i>	43	190	224	59,645	0.000001083	0.000010762
<i>a longo prazo</i>	62	203	2,475	47,768	0.000011964	0.000008619
<i>assim sendo</i> ^a	28	24	2,524	45,916	0.000012201	0.000008285
<i>de vez em quando</i>	60	173	2,379	43,422	0.000011500	0.000007835
<i>graças a deus</i>	13	84	853	42,284	0.000004123	0.000007630

^a The combination ‘assim sendo’ appears in similar contexts as *therefore* and other conjunctive adverbs (Molinier & Levrier, 2000: class PC pp. 56–58). We provisionally included it in the lexicon, due to its frequency and the statistical relevance of the construction (Student t-test: EP 48.794 and BP 201.855; Manning & Schütze, 1999)

Of the 3,500 multiword adverbs that currently integrate the lexicon, 496 occurrences were not detected in the EP-CETEMPúblico corpus, and 318 occurrences did not appear in the BP-FolhaSPaulo. In order to verify the relevance of the asymmetry found in the distribution of some adverbs in these two corpora, namely, adverbs in the lexicon that were not found in the corpora CETEMPúblico and FolhaSPaulo, a selection of 30 of the non-occurring adverbs occurring in only one of the corpus but not in the corpus of the other variety of Portuguese were queried in the two Sketch Engine *ptTenTen18* partitions EP and BP, and positive results were obtained in every search, as seen in Table 4 and Table 5. The findings underscore the critical role that large-scale corpora play in a study like this, highlighting the importance of substantial language samples to map out the distribution of some expressions across the varieties of the language.

Table 4. Asymmetry between EP CETEMPublico and Sketch Engine EP

Adverb	EP CETEMPublico	BP FolhaSPaulo	EP-SE#	BP-SE#	EP-SE%	BP-SE%
<i>a todo momento</i>	0	37	42	17,590	0.00000020	0.00000317
<i>às gargalhadas</i>	0	19	69	987	0.00000033	0.00000017
<i>da boca para fora</i>	0	5	39	863	0.00000018	0.00000015
<i>de forma contínua</i>	0	4	186	7,124	0.00000089	0.00000128
<i>sem dó nem piedade</i>	0	5	116	1,501	0.00000056	0.00000027

Table 5. Asymmetry between BP Folha de SaoPaulo and Sketch Engine BP

Adverb	EP CetemPublico	BPFolhaSPaulo	EP-SE#	BP-SE#	EP-SE%	BP-SE%
<i>a abarrotar</i>	14	0	166	84	0.00000080	0.00000001
<i>a cântaros</i>	1	0	15	118	0.00000007	0.00000002
<i>a par e passo</i>	12	0	180	73	0.00000087	0.00000001
<i>a preto e branco</i>	46	0	624	356	0.00000301	0.00000006
<i>ao pôr do sol</i>	4	0	80	1,578	0.00000038	0.00000028

3. Results and discussion

3.1. General remarks

We began by relating the frequency of the list of expressions selected as seeds in the smaller corpora (EP, BP) with their corresponding frequency in the *ptTenTen18* corpus (EP-SE, BP-SE), having obtained a very high Pearson correlation coefficient:

- Pearson EP~EP-SE: 0.976
- Pearson BP~BP-SE: 0.979

This very high correlation indicates that the seed selection method can be considered adequate since the frequency values found in the smaller corpora are not much skewed from the frequency values found in the *ptTenTen18* corpus, irrespective of the difference in the size of each corpus.

Secondly, we compared the correlation between the frequency values of each corpus pairwise, i.e., the smaller corpora (EP, BP) and the values obtained in each partition of the *ptTenTen18* corpus (EP-SE, BP-SE):

- Pearson EP-BP: 0.978
- Pearson EP-SE~BP-SE: 0.967

This high correlation indicates that the frequency values of the list of adverbs in each corpus/language do not differ considerably (and only by a negligible difference of 0.011). This is quite interesting since the relative size of the smaller corpora (about 10 million words in PE and 24 million in BP) contrasts with the relative size of the two partitions of the *ptTenTen18* corpus, which are very unevenly sized (206,8 million tokens for EP and

5,5 billion tokens for BP). We interpret this result as a confirmation of our initial intuition that the most frequently occurring compound adverbs in each language variety show a quite similar frequency rate.

In the following subsections, we investigate several interesting cases when comparing the concordances of the compound adverbs found in the *ptTenTen18* corpus.

3.2. Detailed analysis

Case 1: Reduction

As previously mentioned, our selection process aimed to include only complete expressions, meaning a string of words that constitute a compound adverb but that was not a substring of a longer expression. For instance, in the case of *em flagrante* (26) and *em flagrante delito* (27), the first one can be considered a substring of the second. We retrieved both strings from the two partitions of the *ptTenTen18* corpus and show in Table 6 the results.

(26) [Ele foi preso] *em flagrante*

(27) [Ele foi preso] *em flagrante delito*
 ‘He was arrested *in the act/red-handed*’ lit.: *in flagrante delicto*

Table 6. Reduction: *em flagrante*

ADV	BP - SE	%	EP - SE	%
<i>em flagrante</i>	46,560	0.0000084	255	0.0000012
<i>em flagrante delito</i>	3,982	0.0000007	394	0.0000019

We noticed that the full adverbial expression *em flagrante delito*, commonly used in EP, appeared in 394 occurrences, accounting for 61% of the 649 total appearances in the corpus. The reduced form *em flagrante* occurs 255 times, representing 39% of the occurrences. In contrast, in BP, the preference leans significantly towards the reduced form *em flagrante*, which appears 46,560 times making up 93,13% of its 50,542 total occurrences compared to 3,982 instances (7,87%) of the entire expression *em flagrante delito*, as seen in Table 6.

In determining the degree of fixedness of compound nouns (Baptista, 1994; G. Gross, 1988), the reduction of the head of the compound noun is an important indicator of formal fixedness; thus, for example, alongside the combinations *canino / incisivo / molar / pré-molar / siso* ‘canine/incisor/molar/premolar/and wisdom (tooth)’, in which the adjectives are classifiers of *dente* ‘tooth’, one observes the possibility of reducing the noun *dente* ‘tooth’, with the adjective appearing alone while maintaining the value of the combinatorics.

The same seems to happen here with ‘*em flagrante delito*’, where the reduction of the noun *delito* does not affect the interpretation of the combinatorics *em flagrante*. What seems to distinguish the two varieties of Portuguese is the much higher frequency of the reduced form in BP. It might not seem appropriate to consider that these two strings as separate lexical entries, their meaning being, after all, identical. This phenomenon could be treated as a transformational property of the more extended structure (*Prep Adj C = Prep Adj*), possibly signaling a tendency for a shorter form. In this case, we kept both forms, the short and the long to highlight this tendency of use in each language variety.

Other examples of apparent reduction have also been found, but where another solution was adopted. For example, with *em fila* (28) vs. *em fila indiana* (29) lit.: ‘in Indian line’, ‘in a single line’.

(28) [As crianças entraram na sala] *em fila*
 ‘The children entered the room *in line*’

- (29) [As crianças entraram na sala] *em fila indiana*
 ‘The children entered the room *in a single line*, lit: in Indian line’

Table 7. Reduction: *em fila*

ADV	BP - SE	%	EP - SE	%
<i>em fila</i>	3,995	0.00000072	145	0.00000070
<i>em fila indiana</i>	528	0.00000009	39	0.00000010

In this case, it is the adjective that seems to undergo zeroing, as detailed in Table 7. Of the 4,523 occurrences of *em fila* in BP, 88% (3,995 instances) represent a shorter form, compared to 12% (528 instances) that use the entire expression *em fila indiana*. Similarly, in EP, the trend towards the shorter form is evident, with 78% (145 instances) of the 184 total occurrences, and 22% (39 instances) corresponding to the use of the entire multiword adverb. However, we note that the shorter form *em fila* should not be directly associated with the larger one, since the expression *em fila indiana* seems to apply to a narrower distribution (primarily people) and involves predicates with the meaning of **movement** (e.g. *entrar* ‘enter’, *passar* ‘pass’), unlike the more general-purpose combination *em fila*, which can be applied to a wider variety of objects, including non-human nouns and does not necessarily involve the meaning of **movement**. Moreover, there are multiple combinations of *em fila* and different nouns, such as [estacionar/parar] *em fila dupla/em dupla fila* ‘park/stop in double line’, so that considering *em fila* a reduced form would raise the issue of choosing the appropriate adjective that would have been reduced. Because of this, the two forms (*em fila* and *em fila indiana*) as well as other combinatorics involving *fila* ‘line’ were considered different lexical entries.

In examining another example of reduction, we observe that the expression [saber] *de cor e salteado* [know] ‘by heart’ tends to be reduced to *de cor* ‘id.’, (30)–(31):

- (30) [O Leo sabe o poema] *de cor*
 ‘Leo knows the poem *by heart*’
- (31) [O Leo sabe o poema] *de cor e salteado*
 ‘Leo knows the poem *by heart*’

This adverbial expression often occurs as a manner modifier of verbs such as *aprender* ‘learn’, *conhecer* ‘know’, *dizer* ‘say’, and *recitar* ‘recite’, as detailed in Table 8. The following frequency values were found in the corpus:

Table 8. Reduction: *de cor*

ADV	BP - SE	%	EP - SE	%
[saber &c.] <i>de cor</i>	2,192	0,00000039	236	0,00000114
[saber &c.] <i>de cor e salteado</i>	387	0,00000007	24	0,00000012

In terms of frequency within the corpus, the reduced form *de cor* is more prevalent than the full string *de cor e salteado*. Furthermore, this shorter form is utilized more often in EP, appearing 9.5 times as frequently, compared to 5.6 times more in BP.

Case 2: Lexical variation

In a notable number of cases, we found that each language variety features an equivalent multiword adverb with some variation in specific lexical elements:

- (32) [Ela vai ganhar] *com os pés nas costas* (BP)
lit.: 'She will *easily* win' lit.: 'with her feet on the back'
- (33) [Se quiserem faço-a] *com uma perna atrás das costas* (EP)
lit.: 'If you like, I'll *easily* do it' lit.: 'with one leg behind the back'

In the examples (32)–(33) above, *pés* (plural) 'feet' and *perna* 'leg' change according to the language variety. Additionally, we observed examples where the BP variant employs the singular form, *com o pé nas costas* 'with a foot on the back.' In EP, the use of the indefinite article is mandatory; that precludes the number variation of *perna* 'leg' (and there was no case of **com as pernas às costas*).

Other interesting examples include (34)–(37):

- (34) [Vizinhos estão vendendo as casas] *a preço de banana* (BP)
lit.: 'Neighbours are selling their houses *at a bargain price*', lit.: 'at the of a banana'
- (35) [Estas pedras compram-se no garimpo e] *ao preço da chuva* (EP)
lit.: 'These stones are bought in the mines and *at a bargain price*', lit.: 'at the price of rain'
- (36) [Não viso lucro nos festivais que realizo, faço] *por amor à camisa* (BP)
'I don't make a profit from the festivals I organize, I do it *for the team*' lit.: 'for the love of the shirt'
- (37) [O que aqui se faz é sobretudo] *por amor à camisola* (EP)
'What we do here is above all *for the team*' lit.: 'for the love of the shirt'

In these examples, a selection of specific lexical items can be seen as reflex from each language variety's cultural context. For instance, in (34)–(35), the term *banana* 'id.' might resonate more closely with the culture of Brazil than that of Portugal. In another instance, in (36)–(37), the noun *camisola* illustrates how the same word can denote entirely different objects across the two varieties: in EP, it means 'T-shirt' whereas in BP, it refers to a 'feminine nightgown'. Furthermore, other expressions (38)–(41), while equivalent in their overall meaning, exhibit entirely different word choices between the two varieties:

- (38) *De mala e cuia*, [foi em busca de seu sonho] (BP)
'He went in search of his dream *with all his possessions*', lit.: 'with bag and gourd'
- (39) [Pretendemos assentar] *de armas e bagagens* [no mercado nacional] (EP)
'We want to settle down *with all our belongings* in the domestic market' lit.: 'with arms and baggage'
- (40) [A questão da intervenção das forças aéreas aliadas estava] *por um fio de cabelo* (BP)
'The question of the intervention of the Allied air forces was hanging *by a thread*' lit.: 'by a strand of hair'
- (41) [O próprio Primeiro-Ministro escapara] *por uma unha negra* (EP)
'The Prime Minister himself had escaped *by a narrow margin*', lit.: 'by a black fingernail'

Determining equivalent expressions across the two language varieties is quite a challenging task that requires a systematic comparison of both lexicons and the collaboration of native speakers from both varieties.

Case 3: Grammatical variation - determiners

When exploring the grammatical differences between BP and EP, it becomes evident that each variety exhibits particularities in its linguistic constructions. Our analysis of compound adverbs in BP and EP shows many grammatical details that are characteristic of each variety.

Some of these differences appear in expressions involving the determiner *todo* ‘all’ (and its inflected forms *todos*, *toda*, *todas*, corresponding to gender-number variations: masculine, feminine, singular, and plural, respectively), among other grammatical cases. This determiner is often followed by the definite article *o* ‘the_masc.sg.’ (and its inflexions). Consider the compound adverb *em todo o caso* ‘in any case’ (42)–(43), a sentence-modifying adverb, often appearing at the front of a sentence followed by a comma:

(42) *Em todo o caso*, [pode-se dizer que ela está certa] (?BP / EP)

(43) *Em todo caso*, [pode-se dizer que ela está certa] (BP / *EP)
‘In any case, you could say she's right’

This variation in the presence of the definite article *o* ‘the_ms’ seems to be systematic. In EP, *todo* is parsed as a *predeterminer* (*Pred*) in the sense of M. Gross (1977), that is, a special type of quantifier that is applied to a noun phrase, whereas in *em todo caso*, it behaves as an adjectival determiner (*DAdj*, M. Gross 1977). This latter class of determiners is linked directly to the noun it determines. Although both cases (with or without *o*) are found in both varieties (EP and BP), there is a clear asymmetry in the distribution of the forms with the definite article, as can be seen in Table 9.

Table 9. Determiners: *todo* (*o*)

ADV	BP - SE	%	EP - SE	%
<i>em todo o caso</i>	2,178	0.00000039	1,085	0.00000524
<i>em todo caso</i>	8,143	0.00000146	64	0.00000030

According to these figures, the longer expression (with the article *o*) is the preferred form in EP against a small fraction of cases occurring without the article *o*. An inverse distribution can be seen in BP, which shows a clear tendency to omit the article *o*. Yet, the occurrence of both forms (with and without the article *o*) indicates a flexible grammatical structure, where BP seems to accept both variations. One could argue that this phenomenon occurs due to the assimilation of the final vowel of *todo* and the article *o*, which is then reflected in the spelling. However, a similar linguistic phenomenon can be seen in examples (44)–(45), where both forms (presence and absence of *a* ‘the_fs’) are accepted and utilized, as shown in Table 10:

(44) [Estivemos] *em toda a parte*

(45) [Estivemos] *em toda parte*
‘We have been *everywhere*’

Table 10. Determiners: *toda (a)*

ADV	BP - SE	%	EP - SE	%
<i>em toda a parte</i>	2,615	0.00000047	418	0.00000202
<i>em toda parte</i>	8,262	0.00000149	51	0.00000024

In Brazilian Portuguese, there seems to be a noticeable tendency to omit the definite articles *o/a* ‘the_masc./fem.’ in the singular form when they precede *todo/toda*. This pattern is less evident in EP, as the low frequency of the form without the article *o* ‘the’ in EP seems fortuitous and might not indicate a consistent trend. Other examples are *a todo o momento/a todo momento* ‘at all times’, *a todo o custo / a todo custo*, ‘lit: at all cost, at all costs’, *a toda a hora / a toda hora* ‘all the time/at every time’ *com toda a certeza, com toda certeza* ‘lit.: with all certainty, surely’ *a toda a velocidade / a toda velocidade* ‘at full speed’.

Nonetheless, when we analyzed adverbial expressions containing the determiners in their plural forms *todos/todas* all_masc.pl./fem.pl., such as *com todos os detalhes*, ‘with all the details’, *com todas as letras* ‘with all letters’ we observed that this phenomenon of omitting the articles, especially in BP, rarely occurs. This consistency is even more pronounced in EP, where our analysis did not reveal any instances of the plural forms omitting the articles. This indicates a clear preference in both varieties for retaining the articles in pluralized constructions (Table 11). Therefore, the zeroing of definite article after *todo* ‘all’ in singular should be treated as a phonological process rather than a syntactic constraint, and the lexical entries of both varieties are the same (with the articles). For automatic processing, however, we made distinct lexical entries for BP and EP.

Table 11. Determiners: *todos (os) / todas (as)*

ADV	BP-SE	%	EP-SE	%
<i>com todos os detalhes</i>	2,407	0.00000043	37	0.00000017
<i>com todos detalhes</i>	52	0.000000009	0	0
<i>com todas as letras</i>	4,905	0.00000088	87	0.00000042
<i>com todas letras</i>	25	0.000000004	0	0

Notice also that this singular-plural opposition may be semantically relevant: in BP *todos os dias* ‘every day’ is opposed to the singular form, *todo o dia* ‘all day long’, and *todo dia* ‘daily/every day’ in which the article is usually absent; *todas as vezes* ‘all the times’, is opposed to the singular *toda vez*, ‘every time’; *com todas as forças*, ‘with all one’s strength, intensely’ can be opposed to *com toda a força*, ‘with full force, with all the might’, the first seems to represent a mental/psychological strength, and the second has a physical sense. Naturally, there are cases where only one of the forms exists: [*mentir*] *com todos os dentes* ‘lit.: with all the teeth,’ ‘lie through your teeth’ (cp. **com todo o dente* ‘with all the tooth’ and **com todos dentes* ‘with all teeth’).

Case 4: Grammatical variation - prepositions

Of all the differences in adverbial expressions between EP and BP examined so far in this study, many cases are related to the choice of prepositions in each variety. We resorted to SE corpora to attest the distribution of particular cases, such as the following examples (46)–(47):

- (46) [O seu coração batia] *a mil à hora* (EP)
 ‘Her heart was beating *at full speed*’ lit.: ‘a thousand an hour’

- (47) [Minha cabeça girava] *a mil por hora* (BP)
 ‘My head was spinning *at full speed*’ lit.: ‘a thousand per hour’

Table 12. Prepositions: *a/por*

ADV	BP - SE	%	EP - SE	%
<i>a mil à hora</i>	2	0.00000001	9	0.00000004
<i>a mil por hora</i>	1,049	0.00000018	4	0.00000001

The examples above indicate that the use of both prepositions *à* ‘by’ and *por* ‘per’ is possible in EP, though there is a tendency for using *à* rather than *por* (almost 20 times more frequent). In BP, the preposition *por* seems to be prevalent (522 times more frequent). The very same phenomenon occurs for similar expressions that indicate **velocity**, whether in their idiomatic or literal sense (48)–(51):

- (48) [O Leo fala] *a cem à hora* (EP) (idiomatic)
 ‘Leo speaks *very fast*’ lit.: ‘at a hundred an hour’
- (49) [O carro corria] *a duzentos à hora* (EP) (literal)
 ‘The car was running *at two hundred an hour*’
- (50) [O Leo fala] *a cem por hora* (BP) (idiomatic)
 ‘Leo speaks *fast*’ lit.: ‘at a hundred an hour’
- (51) [O carro corria] *a duzentos por hora* (BP) (literal)
 ‘The car was running *at two hundred per hour*’

A systematic case of variation is found with the preposition *até* ‘until, up to’. In EP (52), this word is combined with another preposition, *a* ‘at’, forming a compound preposition (or conjunction) *até a* ‘up to’; this latter can then be contracted with the following article (*até ao* ‘up to_the.masc.sg.’). On the other hand, in BP (53), the preposition is kept as a simple word, without the preposition *a* for the most cases (Table 13):

- (52) [Estão enterrados no pântano] *até ao pescoço* (EP)
 ‘They are buried *up to their necks* in the swamp’
- (53) [O presidente está enrolado] *até o pescoço* (BP)
 ‘The president is *up to his neck* in trouble’

Table 13. Prepositions: *até (a/ao)*

ADV	BP - SE	%	EP - SE	%
<i>até ao pescoço</i>	64	0.00000001	60	0.00000029
<i>até o pescoço</i>	2,395	0.00000043	7	0.00000003

Here, we observe that the cases are entirely contrasting, and their distribution is quite asymmetric. The numbers indicate that, in EP, it is evident that the combination preposition + article is preferred, while in BP the preference is to leave out the preposition, maintaining the article. However, both cases can be found (with and without the preposition), and the numbers indicate an inclination towards dropping the preposition, making

the form *até a* appear only in a residual (normalized) frequency. Yet, this analysis might not fully express the depth of the issue as the presence of the masculine article may be a factor, since the presence of the preposition *a* would produce an obligatory contraction. As we observe in Table 14, when the article is in the feminine form, it raises questions about whether in the cases where the *a* without the proper accent corresponds to a spelling mistake, given the homophony in BP (but not EP) with contraction *à* ‘at_the.fem.sg.’ (54)–(55):

- (54) [As empresas exploram os funcionários] *até à última gota*
 ‘Companies exploit their employees *completely*’ lit.: ‘to the last drop’
- (55) [Tenho que aproveitar isso ao máximo] *até a última gota*
 ‘I have to make the most of it, *to the last drop*’

Table 14. Prepositions: *à / a*

ADV	BP - SE	%	EP - SE	%
<i>até à última gota</i>	2	0.0000000003	34	0.0000001643
<i>até a última gota</i>	1,194	0.0000002154	7	0.0000000338

As mentioned above, the non-accented *a* predominates in BP, being 597 times more frequent than the contraction of the preposition *a* followed by the article *a* = *à*, while in EP, the opposite occurs, the dominant form is with *à*, and the few cases without the accent may be incidental.

Other examples of different uses of prepositions correspond to temporal expressions, such as (56)–(57) below.

- (56) [A companhia oferece voos] *de hora a hora* (EP)
 ‘The company offers *hourly* flights’
- (57) [A previsão do tempo é dada] *de hora em hora* (BP/EP)
 ‘The weather forecast is given *every hour*’

Table 15. Prepositions (*de hora prep hora*)

ADV	BP - SE	%	EP - SE	%
<i>de hora em hora</i>	3,006	0.0000340	26	0.00000012
<i>de hora a hora</i>	60	0.0000007	84	0.00000040

In this case, we notice that the preposition *em* ‘in’ prevails in BP, while in EP, although the preposition *a* is frequent, it seems that the use of the preposition *em* in the same type of expressions should not be discarded. It suggests that the (normalized) frequency can be used as important information in both lexicographic descriptions and when building variety-aware lexicons for the computational processing of Portuguese (Müller *et al.*, 2024).

We looked for similar temporal constructions in both partitions of Sketch Engine (Table 16) and found that the frequencies of other expressions denoting time frequency, e.g., *de minuto em minuto/de minuto a minuto* ‘from minute to minute,’ *de dia em dia/de dia a dia/de dia-a-dia*, daily, *de semana em semana/de semana a semana*, weekly, *de mês em mês/de mês a mês* monthly, *de ano em ano/de ano a ano*, yearly.

Table 16. Prepositions: *em/a* in temporal expressions

ADV	SE - BP	%	SE - EP	%
<i>de minuto em minuto</i>	235	0.00000004	3	0.00000001
<i>de minuto a minuto</i>	161	0.00000002	11	0.00000005
<i>de dia em dia</i>	160	0.00000002	12	0.00000005
<i>de dia a dia*</i>	274	0.00000004	11	0.00000005
<i>de dia-a-dia*</i>	102	0.00000001	21	0.00000010
<i>de semana em semana</i>	101	0.00000001	1	0.00000000
<i>de semana a semana</i>	61	0.00000001	13	0.00000006
<i>de mês em mês</i>	253	0.00000004	3	0.00000001
<i>de mês a mês</i>	124	0.00000002	22	0.00000010
<i>de ano em ano</i>	527	0.00000009	10	0.00000004
<i>de ano a ano</i>	287	0.00000005	37	0.00000017

* Ambiguous expressions

As seen in Table 16, in BP, the frequency of the preposition *em* is higher than the preposition *a* in these compound adverbs (except for the pair *de dia em dia/de dia a dia*, ‘daily’). The figures might indicate that both prepositions *em* and *a* may be used interchangeably or with greater flexibility. In EP, however, the frequency of the preposition *em* in some of these expressions is practically null. As in many other situations, the distribution of prepositions in the examples *de hora a hora/de hora em hora* ‘hourly’, *de dia em dia* ‘daily’, and *de ano em ano* ‘yearly’, is misleading, as the variation found is inconsistent with the remaining family of temporal expressions.

Case 5: Grammatical difference: verb forms

One of the most interesting differences between both varieties, EP and BP is the use of the gerund form (ending in *-ndo*) ‘-ing’, as in *abandar/abanando* ‘wave/waiving’ commonly used in BP. This contrasts with the periphrastic use of the *a* + infinitive form, which is prevalent in EP. Both forms are used to indicate progressive/continuous tenses, among other values, as in (58).

- (58) *O Pedro está a ler / lendo o livro*
‘Pedro is reading the book’

Gerund forms can also be found in idiomatic adverbial constructions such as (59)–(60). Besides this variation, we find also an alternation of prepositions *com* ‘with’ and *de* ‘of’. The distribution of these forms is shown in Table 17. Furthermore, in BP, the noun *mão* ‘hand’ can also be used in singular, though rarely.

- (59) [Ele vai sair] *de-com mãos abanando* [só com os dias trabalhados] (BP)
‘He'll come out empty-handed, only with the days worked’
- (60) [Não podia sair de lá] *de-com mãos a abanar* (EP)
‘I couldn't leave there empty-handed’

Table 17. Verb forms: alternation gerund/infinite

ADV	BP-SE	%	EP-SE	%
<i>de mãos abanando</i>	639	0.0000001152	0	0
<i>de mãos a abanar</i>	5	0.0000000009	67	0.0000003238
<i>com as mãos abanando</i>	157	0.0000000283	0	0
<i>com as mãos a abanar</i>	1	0.0000000001	3	0.0000000145
<i>de mão abanando</i>	54	0.0000000097	0	0
<i>com a mão abanando</i>	11	0.0000000019	0	0

The examples in Table 17 show how rigid such combinations are in each variant, with 0 cases ending in *-ndo* ‘ing’ in EP and only 5 and 1 cases of *a* + infinitive form in BP.

Case 6: False friends (*faux amis*) and other complex cases

In some cases, the same adverbial expression might have a different meaning in each variety, the so-called *faux amis*, false friends. For instance, consider the adverbial expression *à francesa* ‘in French style’. In both varieties, EP and BP, the expression *à francesa* is preceded by certain verbs such as *sair*, ‘to leave’ or *despedir-se* ‘to say goodbye’ *à francesa*, ‘discretely, without much fuss’

- (61) [Eles despediram-se e saíram] *à francesa*
‘They said goodbye and left *discretely*’ lit.: ‘in French style’

In EP, exclusively, the same expression may be used as a reduction of a full expression *à grande e à francesa* ‘with excess, with pomp’, which is usually preceded by verbs such as (*comer*, ‘eat’ *divertir-se*, ‘have fun’ etc.).

In BP, instead, there are other meanings associated with *à francesa*; the first means to cut a pizza in small squares/finger-food style (62). Another meaning is to serve food *à francesa*, a particular way of serving dinner ‘in a French style’, typical during weddings and formal dinners.

- (62) [O pizzaiolo cortou a pizza] *à francesa*
‘The pizza chef cut the pizza in French style’

In the following examples (63)–(64), we observe different prepositions in the same expression:

- (63) [Não posso afirmar] *a pés juntos* [porque não é de fonte directa] (EP)
‘I can't say *with all my heart* because it's not from a direct source’ lit.: ‘with feet together’
- (64) [Mãe, eu juro] *de pés juntos* [que não fui eu] (BP)
‘Mum, I swear *with all my heart* it wasn't me’ lit.: ‘with feet together’

The distribution of these two expressions in the corpus is given in Table 18. We considered this case to be complex because, although the construction *de pés juntos* occur in EP and BP, in EP it only used as a literal (free) phrase, and according to our search in the EP partition of the corpus *ptTenTen18*, the mere 7 occurrences of *de pés juntos* were preceded by different verbs, (e.g.: *saltar*, ‘to jump’, *entrar*, ‘to enter’, *colocar*, ‘to put’) indicating a sequence of words, that are non-idiomatic, but compositional. It does not compare to the idiomatic use of *de pés juntos*, common in BP, where the preposition can vary, and the expression can be used both literally and idiomatically.

Table 18. Prepositions and meaning

ADV	BP - SE	%	EP - SE	%
<i>a pés juntos</i>	70	0.0000008	123	0.0000014
<i>de pés juntos</i>	770	0.0000088	7	0.00000008

4. Conclusion and future work

This work presented similarities and differences between multiword/compound adverbs in Brazilian and European Portuguese. Our primary focus involved assessing the degree of similarity and difference of adverbial constructions in these two varieties of Portuguese. After running our lexicon in two smaller corpora, CETEMPúblico and Folha de São Paulo, we selected the 300 most frequently occurring, multiword compound adverbs that were then retrieved from each partition (EP and BP) of the *ptTenTen18* corpus. We verified the distribution, the frequency of these expressions and gathered examples to evaluate our findings. While we observe a significant overlap indicating that both varieties share a great number of adverbial expressions, we detected asymmetries in some respects, primarily in grammatical and lexical aspects. Interestingly, the proportion of the syntactic-semantic classes seems to indicate a consistent distribution when compared to the single-word adverbs ending in *-mente* ‘ly’ (Fernandes, 2011), with adverbs of manner and time predominating, followed by quantification, conjunctive, and locative adverbs. This observation confirms that the syntactic/semantic classification proposed by Molinier and Levrier (2000) can be extended or adapted to multiword (compound) frozen adverbs.

In conclusion, this work makes a contribution to the field by providing an in-depth analysis of multiword adverbs in European and Brazilian Portuguese, which might be resourceful for detecting similar language varieties, translations and in the identification of these expressions as lexical units. Future steps include the implementation of the lexicon of multiword adverbs in the STRING system and the production of annotated corpora for training and evaluating NLP systems.

Acknowledgements

This work was supported by the Foundation for Science and Technology (FCT) (Proj.Ref. UIDB/50021/2020, DOI: 10.54499/UIDB/50021/2020) and by the Doctoral Program in Language Sciences of University of Algarve.

References

- Baptista, Jorge (1994) *Estabelecimento e formalização de classes de nomes compostos*. Masters dissertation, Universidade de Lisboa, Portugal.
- Baptista, Jorge (1999) *Manhã, tarde, noite*. Analysis of temporal adverbs using local grammars. *Seminários de Linguística* 3, pp. 1–27.

- Català, Dolors, Jorge Baptista & Cristina Palma (2020) Problèmes formels concernant la traduction des adverbes composés (espagnol/portugais). *Langue(s) & Parole*, 5, pp. 67–82.
- Català, Dolors (2003) *Les adverbes composés. Approches contrastives en linguistique appliquée*. Doctoral dissertation, Universitat Autònoma de Barcelona, Spain.
- da Silva, João [em linha] (2013) *Dicionário Brasileiro de Fraseologia*. Available at http://www.josepereira.com.br/DBF_2013.pdf
- De Gioia, Michele (2001) *Avverbi idiomatici dell'italiano. Analisi lessico-grammaticale*. l'Harmattan Italia.
- Fernandes, Gaia (2011) *Automatic disambiguation of -mente ending adverbs in Brazilian Portuguese*. Masters thesis, Universidade do Algarve and Universitat Autònoma de Barcelona, Portugal and Spain.
- Folha de São Paulo (1999) CD-ROM Folha - Edição 99. Texto integral da Folha de S. Paulo de 1994 a 1998. Publifolha. Available at <http://bd.folha.uol.com.br/cdrom.html>
- Gross, Gaston (1988) Degrée de figement de noms composés. *Languages* 90, pp. 57–72.
- Gross, Maurice (1975) *Méthodes en syntaxe*. Hermann.
- Gross, Maurice (1977) *Grammaire transformationnelle du français : 2 - Syntaxe du nom* (2nd ed.). Cantilène. [Original publication in 1986]
- Gross, Maurice (1981) Les bases empiriques de la notion de prédicat sémantique. *Langages* 63, pp. 7–52.
- Gross, Maurice (1982) Une classification des phrases figées du français. *Revue québécoise de linguistique* 11 (2), pp. 151–185. <https://doi.org/10.7202/602492ar>
- Gross, Maurice (1984) Lexicon-grammar and the syntactic analysis of French. In *International Conference on Computational Linguistics (COLING)*. Association for Computational Linguistics, pp. 275–282. <https://doi.org/10.3115/980491.980549>
- Gross, Maurice (1986) *Grammaire transformationnelle du français : 3 - Syntaxe de l'adverbe*. ASSTRIL.
- Gross, Maurice (1996a) *Les expressions figées en français : noms composés et autres locutions*. Editions Ophrys.
- Gross, Maurice (1996b) Lexicon-Grammar. In Keith Brown & Jim Miller (eds.), *Concise Encyclopedia of Syntactic Theories*, Pergamon, pp. 244–259.
- Guimier, Claude (1996) *Les adverbes du français: le cas des adverbes en -ment*. Editions Ophrys.
- Harris, Zellig (1991) *Theory of language and information: a mathematical approach*. Clarendon Press.
- Kilgarriff, Adam, Miloš Jakubiček, Jan Pomikálek, Tony Berber Sardinha & Pete Whitelock (2014) PtTenTen: A corpus for Portuguese lexicography. In Tony Berber Sardinha, Telma de Lurdes São Bento Ferreira (eds.), *Working with Portuguese corpora* (1st ed.). Bloomsbury Academic, pp. 280–287.
- Laporte, Éric & Stavroula Voyatzi (2008) An electronic dictionary of French multiword adverbs. In *Language resources and evaluation conference. Workshop towards a shared task for multiword expressions*, pp. 31–34. Available at <https://shs.hal.science/halshs00286546/file/MWEAdverbLexicon.pdf>
- Mamede, Nuno, Jorge Baptista, Cláudio Diniz & Vera Cabarrão (2012) STRING: A hybrid, statistical and rule-based natural language processing chain for Portuguese. In *Proceedings of the 10th International Conference on Computational Processing of Portuguese (PROPOR'12)*. Springer-Verlag, pp. 17–20. Available at <http://www.inesc-id.pt/ficheiros/publicacoes/8578.pdf>
- Manning, Christopher & Hinrich Schütze (1999) *Foundations of statistical natural language processing*. MIT press.
- Mel'čuk, Igor (2023) *General phraseology: theory and practice*. John Benjamins.
- Molinier, Christian & Françoise Levrier (2000) *Grammaire des adverbes : description des formes en -ment*. Droz.
- Moreno-Ortiz, Antonio, Chantal Pérez-Hernández & Maria Del-Olmo (2013) Managing multiword expressions in a lexicon-based sentiment analysis system for Spanish. In *Proceedings of the 9th Workshop on Multiword Expressions*. Association for Computational Linguistics, pp. 1–10. Available at <https://aclanthology.org/W13-1001.pdf>

- Müller, Izabela, Nuno Mamede & Jorge Baptista (2022) Bootstrapping a Lexicon of Multiword Adverbs for Brazilian Portuguese. In *International Conference on Computational and Corpus-Based Phraseology* Springer, pp. 160–174. https://doi.org/10.1007/978-3-031-15925-1_12
- Müller, Izabela, Nuno Mamede & Jorge Baptista (2023) Advérbios compostos do Português do Brasil. *Revista da Associação Portuguesa de Linguística* 1 (10), pp. 230–250. <https://doi.org/10.26334/2183-9077/rapln10ano2023a13>
- Palma, Cristina (2009) *Estudo contrastivo português-espanhol de expressões fixas adverbiais*. Masters thesis, Universidade do Algarve, Portugal.
- Ranchhod, Elisabete (1990) Sintaxe dos predicados nominais com estar. Instituto Nacional de Investigação Científica (INIC).
- Ranchhod, Elisabete (1991) Frozen adverbs – Comparative forms *Como C* in Portuguese. *Linguisticae Investigationes* 15 (1), pp. 141–170. <https://doi.org/10.1075/li.15.1.07ran>
- Rocha, Paulo Alexandre & Diana Santos (2000) CETEMPúblico: Um corpus de grandes dimensões de linguagem jornalística portuguesa. In Maria das Graças Volpe Nunes (ed.), *V Encontro para o processamento computacional da língua portuguesa escrita e falada (PROPOR 2000)*. ICMC/USP.
- Rocha, Carlos Alberto & Carlos Eduardo Rocha (2011) *Dicionário de locuções e expressões da língua portuguesa*. LEXIKON Editora.
- Schwab, Artur (1985) *Locuções adverbiais* (2nd ed.). Fundação da Universidade Federal do Paraná.
- Shudo, Kosho, Akira Kurahone & Toshifumi Tanabe (2011) A comprehensive dictionary of multiword expressions. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (HLT'11)*. Association for Computational Linguistics, pp. 161–170. Available at <https://aclanthology.org/P11-1017.pdf>
- Wagner Filho, Jorge, Rodrigo Wilkens, Marco Idiart & Aline Villavicencio (2018) The brWaC corpus: a new open resource for Brazilian Portuguese. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), pp. 4339–4344. Available at <https://aclanthology.org/L18-1686>
- Zampieri, Marcos & Binyam Gebrekidan Gebre (2012) Automatic identification of language varieties: the case of Portuguese. In *Empirical Methods in Natural Language Processing. Proceedings of the Conference on Natural Language Processing 2012*. Österreichischen Gesellschaft für Artificial Intelligende (ÖGAI), pp. 233–237. Available at <https://www.oegai.at/konvens2012/proceedings.pdf>
- Žižková, Hana (2018) Improving compound adverb tagging. In *Proceedings of the Twelfth Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN 2018*. Tribun EU, pp. 103–109. Available at <https://nlp.fi.muni.cz/raslan/raslan18.pdf>

Appendix

List of the 300 most frequently occurring multiword adverbs in Portuguese.⁸

à antiga	à mão armada	a torto e a direito	às cegas
a baixa altitude	à margem da lei	à última hora	às claras
à beira-mar	a médio ou longo prazo	a um só tempo	assim sendo
à beira-rio	a médio prazo	a uma só voz	até certo ponto
a bem da verdade	à moda antiga	a vida inteira	até dizer chega
a bem dizer	à noitinha	ad libitum	até nova ordem
a cada instante	a olho nu	afinal de contas	até onde a vista alcança
a cada passo	a olhos vistos	ainda por cima	até ordem em contrário
a cem por cento	à paisana	além do mais	até prova em contrário
a céu aberto	a passos largos	altas horas da noite	bem ou mal
a conta-gotas	a perder de vista	alto e bom som	caso a caso
a contragosto	a peso de ouro	anos a fio	caso contrário
a curta distância	a pique	antes de mais nada	cedo ou tarde
a curto e médio prazo	a plenos pulmões	antes pelo contrário	com armas e bagagens
a curto ou médio prazo	a posteriori	ao acaso	com as próprias mãos
a curto prazo	a pouco e pouco	ao amanhecer	com chave de ouro
à direita e à esquerda	a preço de ouro	ao anoitecer	com conhecimento de causa
a duras penas	à primeira vista	ao ar livre	com mão de ferro
à francesa (1)	a qualquer preço	ao arrepio da lei	com o devido respeito
à francesa (2)	à queima-roupa	ao cair da noite	com o pé direito
a fundo	à risca	ao entardecer	com o pé esquerdo
a fundo perdido	à saciedade	ao fim e ao cabo	com todas as forças
a galope	a sangue frio	ao mesmo tempo	com todo o respeito
a granel	a seco	ao pé da letra	com unhas e dentes
à inglesa	a sete chaves	ao que parece	como de costume
à la carte	a seu tempo	ao que tudo indica	como deve ser
a longa distância	a toda a hora	aos berros	contra a corrente
a longo prazo	a toda a velocidade	aos gritos	contra tudo e contra todos
à luz de velas	a todo o custo	aqui e acolá	da cabeça aos pés
à luz do dia	a todo o vapor	aqui e agora	da noite para o dia
à luz do sol	a toque de caixa	aqui e ali	daí a pouco

⁸ Further information can also be found in: https://string.hlt.inesc-id.pt/wiki/Compound_Adverbs

daí em diante	de pai para filho	em cima da hora	na calada da noite
daí para a frente	de pais para filhos	em duas palavras	na medida do possível
daí por diante	de par em par	em fila indiana	na melhor das hipóteses
daqui a pouco	de parte a parte	em flagrante	na pior das hipóteses
daqui em diante	de ponta a ponta	em flagrante delito	na ponta da língua
daqui para a frente	de porta em porta	em grande escala	na surdina
daqui por diante	de preferência	em grande estilo	nas coxas
de A a Z	de propósito	em linhas gerais	nas entrelinhas
de alto a baixo	de quando em quando	em massa	nem mais nem menos
de antemão	de quando em vez	em partes iguais	nem no melhor dos mundos
de bom grado	de sol a sol	em pé de igualdade	nem por isso
de braços abertos	de tempos em tempos	em poucas palavras	nesta altura do campeonato
de cabo a rabo	de um dia para o outro	em primeira mão	no calor do momento
de comum acordo	de um lado para o outro	em princípio	no curto prazo
de cor e salteado	de um momento para o outro	em surdina	no dia anterior
de corpo e alma	de um tiro	em todo caso	no entanto
de dedo em riste	de uma só vez	em todo o caso	no final das contas
de dia e de noite	de uma vez por todas	em última análise	no melhor dos casos
de então para cá	de vento em popa	em última instância	no pior dos casos
de fonte segura	de vez em quando	em vão	no último minuto
de forma alguma	de viva voz	em voz alta	noite e dia
de geração em geração	do dia para a noite	entre aspas	nos dias de hoje
de hoje em diante	do fundo do coração	frente a frente	nos dias que correm
de igual para igual	do nada	graças a Deus	nos nossos dias
de jeito nenhum	do princípio ao fim	hoje em dia	nos tempos que correm
de má vontade	dos pés à cabeça	lado a lado	num determinado momento
de maneira geral	e assim por diante	mais cedo ou mais tarde	num futuro próximo
de maneira nenhuma	e assim sucessivamente	mais dia menos dia	num piscar de olhos
de mão beijada	e etc	mais do que nunca	o tempo todo
de mão em mão	e ponto final	mais tarde ou mais cedo	olhos nos olhos
de modo algum	e vice-versa	mal e porcamente	ombro a ombro
de morte natural	em abundância	mal ou bem	palavra por palavra
de moto próprio	em câmara lenta	mano a mano	palmo a palmo
de olhos fechados	em carne e osso	muito embora	para caramba
de orelha a orelha	em certa medida	mutatis mutandis	para começo de conversa

para dizer a verdade	por enquanto	por um acaso	sem mais nem menos
para inglês ver	por força	por um triz	sem papas na língua
para o que der e vier	por força das circunstâncias	por unanimidade	sem pestanejar
para todo o sempre	por maioria absoluta	por via das dúvidas	sem sombra de dúvida
para todos os efeitos	por maioria simples	por voto secreto	sem tirar nem pôr
passo a passo	por mérito próprio	porta a porta	tintim por tintim
pelo sim pelo não	por milagre	pouco a pouco	total ou parcialmente
ponto por ponto	por motivo de força maior	preto no branco	uma vez ou outra
por A mais B	por nada deste mundo	pura e simplesmente	única e exclusivamente
por assim dizer	por sua conta e risco	sem dúvida alguma	
por conta própria	por toda a parte	sem dúvida nenhuma	