

Improving the portuguese hyphenation rules

Leonardo Araújo

Aline de Lima Benevides

June 6, 2024

Abstract

Portuguese hyphenation rules are available for more than 35 years and have done a good job. Nonetheless they still make mistakes and leave some hyphenation points unmarked. Although most undetected hyphenation points are located near word boundaries, what will be irrelevant for T_EX typographic purposes, they are still useful to hyphenate proper nouns, new words or pseudoword, and for usage in other applications, such as text-to-speech conversion. A list of 85 638 hyphenated words acquired from online dictionaries was used along *patgen* to create improved rules, leading to better hyphenation of Portuguese words.

1 Introduction

Hyphenation in text wrapping was not used for a long time. Words should fit entirely in a line, or they would be broken in arbitrary places. Initially, no markers were used to indicate word wrapping, leading to potential confusion and unintended interpretations. As a result, orthographers advocated for the introduction of a sign to indicate such breaks. Portuguese faced the same gradual introduction of a hyphenation sign to mark words wrapping across lines. Even though the usage of a hyphenation sign (=) was advocated by orthographers (Magalhães Gândavo 1574), few documents used such sign until the end of the 18th century (Araújo and Maruyama 2015).

In some cases, hyphenation hinders smooth reading and should be avoided in child literature. In opposition, both excessive spacing and insufficient spacing between words also impose difficulty in the reading process, making hyphenation fundamental when texts use short line lengths. As a line gets shorter, the number of breaking candidates between words decreases, leading to awkward spaces between words and among letters. For that matter, automatic hyphenation plays an important role in good typesetting.

T_EX is a typesetting system which carefully deals with these issues, automatically arranging text on a page to create a good reading experience. Automatic hyphenation is an important part on this process, promoting an even-tempered distribution of elements on the page. Line height and line length, paragraph length, font size and typeface, letter-spacing, and word-spacing are some factors which influence the text legibility and readability. Space between words should not be too long creating lakes and rivers in the text, nor too tight, impairing the legibility and readability.

Another important matter to consider is ambiguities that might be created when a word is partitioned during the hyphenation process. In English we should avoid hyphenations such as *re-cover*, *re-form*, *re-sign*, *the-rapist*, *depart-mental*, and *mans-laughter* (in Portuguese, some examples are: *de-putada*, *fede-ração*, *acumula*, *após-tolo*, *cú-bico*). Hyphenations that might lead the reader to pronounce a word incorrectly should also be prevented. That is the case of *considera-tion*, in English (and *pe-rigo* in Portuguese).

In some situations, hyphenation is also a matter of style. Some partitioning choices sound better than others. These conflicting alternatives typically arises when a words has many possible hyphenation points. Consider *ar-chae-ol-o-gist* (or *ar-che-ol-o-gist*), which is preferably partitioned as *archae-ologist* (or *arche-ologist*) in opposition to *archaeol-ogist* (or *archeol-ogist*) or *archaeolo-gist* (or *archeolo-gist*). It is preferable to keep whole morphemes together. In the previous example: *archae* (or *arche*, meaning “ancient”, “primitive”) and *-ologist* (“one who studies the topic”). In Portuguese, it is also more elegant to avoid splits between double consonants or vowels, even if an hyphenation point do exists between those letters. For example, *pressu-rizar* is preferable over *pres-surizar* and *empreen-dedor* is preferable over *empre-endedor*. Even so, exceptions exists, it is preferable to partition *micro-organismo* rather than *microor-ganismo* (keeping morphemes together is favored over splitting a double vowel). Numerous factors come into play when choosing the optimal hyphenation point for a word.

The general rule for Portuguese hyphenation is to split a word into its syllables. A syllable is made of a mandatory nucleus, filled by a vowel, and optional peripheral consonants (before or after the nucleus). In some situations, the syllabic division does not respect the ethnologic constituents. The usage of the prefixes *bis-* and *in-* are examples of this circumstance. The correct syllabifications are *bi-sa-vô*, *i-nobs-tan-te*¹, and *i-na-ti-vo*, where the prefixes are split into two syllables. But, as pointed out previously, it is preferable to

¹The rule of syllable division could lead to two possible partitions: *i-nobs-tan-te* and *in-obs-tan-te*, but the first is preferable.

keep morphemes together rather than splitting them apart. Therefore we should favor the hyphenation *bisa-vô* over *bi-savô*, *inobs-tante* over *i-nobstante*, and *ina-tivo* over *i-nativo*. This last example could also lead to a misunderstanding since the word *nativo* (*native* in English) emerges from this word break.

Each language has its hyphenation rules, which can be categorized into two groups: those driven by morphology (etymology) and those driven by pronunciation. An algorithmic approach employs a logical system to analyze words and apply the hyphenation rules of a specific language. Since hyphenation rules vary significantly between languages, an algorithm must be developed for each one. While a logic-based system can be efficient and compact, it still needs to address exceptions through hard-coded rules.

The automation of this process can employ various approaches, including:

Dictionary-Based Approach: This method restricts hyphenation possibilities to entries found in the dictionary.

Algorithmic-Based Approach: This approach can be applied to any sequence encountered in a text.

Rule-Based Model: This model recognizes prefixes, suffixes, morphemes, or specific sequences suitable for hyphenation.

Pattern Matching: By using a corpus of hyphenation examples in a language, this approach identifies letter sequences that determine suitable hyphenation points. Patterns encompass prefixes, suffixes, exceptions, and special hyphenation rules of the language.

Mixed Approach: This approach combines two or more of the previously described methods to enhance hyphenation accuracy and flexibility.

Analyzing only the immediate surroundings may not always suffice to determine a potential hyphenation point. For instance, consider *de-moc-ra-cy* and *dem-o-crat*, where the immediate surrounding of the letter *e* does not provide a clear indication of the hyphenation point. Even when facing chosen patterns that typically make hyphenation straightforward, exceptions can arise. Take the sequence *tion*, for example. It may seem logical to place a break before this pattern, but the word *cation* is hyphenated as *cat-ion*, highlighting the influence of etymology on the way a word is split.

A word with multiple meanings may have distinct hyphenations based on its intended meaning. ‘For instance, the Swedish word form *glassko* has three different meanings, and can be hyphenated as *glas-sko* (glass shoe), *glass-ko* (ice cream cow) and in the non-standard way, *glass-sko* (ice cream shoe)’ (Németh 2006). In Portuguese, the word *sublinha* might be hyphenated in two different ways: *su-bli-nha* when representing the inflected form of the verb *sublinhar* (to underline) or *sub-li-nha* when referring to the line under (underline as a noun).

Machine learning techniques were employed to hyphenate Norwegian text (Kristensen and Langmyhr 2001). The study revealed that, overall, the \TeX approach outperformed a neural network. Both methods demonstrated similar performance in identifying correct hyphenation points and minimizing incorrect hyphenations when tested on a small word list (with the neural network slightly outperforming in correctly recognizing hyphenation points). However, when tested on a larger word list, the \TeX approach proved superior in avoiding incorrect hyphenations compared to the neural network.

The original \TeX hyphenation algorithm, introduced by Knuth (1977), primarily focused on the English language and employed three main rules: (1) suffix removal; (2) prefix removal; and (3) the Vowel-Consonant-Consonant-Vowel (VCCV) breaking rule². Tests demonstrated that it could identify 40% of allowable hyphen locations (F. M. Liang 1983). The \TeX hyphenation algorithm later adopted the approach proposed by Frank M. Liang, which involves competing patterns. The \TeX 82 algorithm employs five alternating levels of hyphenating and inhibiting patterns. The program for pattern generation, known as PATGEN, was created by (F. Liang and Breitenlohner 1991) and has been utilized to generate hyphenating patterns for numerous languages (P. Sojka 1995b; P. Sojka 1995a; P. Sojka et al. 2005; P. Sojka and Antoš 2003; Scannell 2003). It involves sweeping a database of hyphenated words in a language to identify both hyphenating and inhibiting patterns, ultimately creating a list of competing patterns for that specific language.

The effective hyphenation of words by \TeX will actually depends on the following factors: (1) document language, which will determine which set of patterns to apply; (2) characters used, since some might block hyphenation at their edges; (3) the value of the internal variables `lefthyphenmin` and `righthyphenmin`³, which defines the minimum sequence length of characters at the left and right borders before any hyphenation is allowed.

²The VCCV rule in hyphenation patterns, places the syllable boundary between two consecutive consonants when they appear between two vowels. For example, in the word *sudden*, the syllable break occurs between the *d* and the second *d*, making it *sud-den*. The VCCV pattern is a relatively common syllable division pattern in English. This rule ensures accurate hyphenation, maintaining word readability and pronunciation when words are split at the end of a line in printed text.

³The variables `lefthyphenmin` and `righthyphenmin` are language dependant and are defined in *tlpobj* files (`/usr/local/texlive/20XX/tlpkg/tlpobj/hyphen-xxxxxx.tlpobj`). Default values varies in the range from 1 to 3. English and Portuguese, for example, use `lefthyphenmin=2` and `righthyphenmin=3`.

Despite the fact that T_EX hyphenation algorithm and rules are old, they are, to these days, the most frequently used approach, even outside the T_EX's world. The grounds for this is Hunspell, a spell checker and morphological analyzer that is adopted in many softwares (e.g. LibreOffice, OpenOffice.org, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher, among others (Hunspell's Team 2023a)). Hunspell uses T_EX hyphenation rules (Hunspell's Team 2023b; Levien 1998), making T_EX hyphenation widespread in the computer world. That is a result of T_EX approach simplicity and versatility. The algorithm works effectively, as it already supports rules for 66 languages (T_EX pattern authors 2023), and offers the flexibility to create rules for any currently unsupported languages.

Unfortunately, certain hyphenation rules cannot be implemented using the T_EX hyphenation algorithm. For example, in German, hyphenation can lead to letter change or insertion. Additionally, compound words lack hyphens, resulting in extended letter sequences without visible separation and even repetitions of the same letter, as seen in examples like *Wasserrinne* and *Schiffahrt*. Furthermore, the German spelling reform made some changes, making it necessary to create a different set of rules for German hyphenation. For example, the word *Schiffahrt* should be hyphenated as *Schiff-fahrt*, preserving the *fs* from each word that makes this compound. The hyphenation should insert an *f* that is not part of the written form. That was not a problem for the old written form of the word: *Schiffahrt*. Also, the old hyphenating rules of German grammar stated the hyphenation *Bäk-ker* for the word *Bäcker*, *Zuk-ker* for the word *Zucker* and *pak-ken* for the word *packen*. Nowadays, those words are hyphenated as *Bä-cker*, *Zu-cker* and *pa-cken*, respectively. Some words have also changed their hyphenation point after the spelling reform. For example, *Fen-ster* became *Fens-ter* and *mei-stens* became *meis-tens*. Some problems in compound word hyphenation in T_EX are discussed in P. Sojka (1995a).

Moreover, as mentioned earlier, certain hyphenations may be preferred for stylistic reasons, or to avoid ambiguity, or for better reading experience. Some words might have multiple hyphenations, depending on the intended meaning. In such cases, T_EX may not efficiently address the hyphenation challenge, as it would necessitate case-by-case handling.

2 Patterns for T_EX hyphenation

To simplify, if we consider only the Latin alphabet, with no diacritics, the patterns used in T_EX hyphenation are of the form: `^\.?[0-9]?([a-z]+[0-9]?)+\.?$,` where we have described it using a regular expression⁴. One example of such pattern is `4z1z2`, which is composed of a sequence of letters and numbers. In general, we use characters/symbols from the language alphabet, along with Hindu-Arabic numbers, to express either hyphenation facilitation or inhibition. Odd numbers indicate a good hyphenation point, whereas even numbers indicate a bad place to break. The given example states that the sequence has a good breaking point between the first and the second *z* and an hyphenation should be inhibited before the first *z* and after the second *z*. For example, the hyphenation of the word *piz-za*, *fiz-zle* and *mez-zanine* use this rule, where we see the hyphen placed between the two *z*'s and no hyphen before, nor after the *z*'s. Patterns may also use period symbol (`.`) to indicate word boundaries. The pattern `.sh2` applies to beginning of words, implying that the *s* and the *h* should stick together in beginning of a word and an hyphenation should also be inhibited after the *h*. For example, this pattern is used in *Sher-lock*.

Hyphenation rules are organized in levels, from 1 to 9, where odd numbers represent hyphenating levels and even numbers represent inhibiting levels. Each level works as an exception level of its predecessor. For example, the rule `sh1er` indicates a good hyphenation point between the *h* and the *e* in the sequence *sher*. A rule at a higher level, as `.sh2`, implies an exception to the lower level rule. When we see *sher*, in the beginning of a word, the rule `.sh2` applies and hyphenation proposed by the lower level rule `sh1er` should be hindered. That is the case in the hyphenation of the word *Sher-lock*. The full example is provided in Listing 1, where we might see all pertinent English rules taking place in the hyphenation of *Sher-lock*.

⁴Regular expressions (regex) are powerful search patterns used in text processing to find, match, and manipulate strings of text. They are a fundamental tool in programming and are supported in many programming languages. The regex given here uses the Perl syntax and might be broken in the following parts: `^` and `$` mark the start and end of the string; `\.?` specifies an optional period; `[0-9]?` allows for an optional single digit; `([a-z]+[0-9]?)+` matches sequences of lowercase letters interleaved with optional digits.

.	s	h	e	r	l	o	c	k	.
	0	0	2						.sh2
	0	2	0						s2h
	0	0	1	0	0				sh1er
				0	1	0			r1l
				0	3	0	4		r3lo4
							0	0	1
max:	0	2	2	0	3	0	4	0	1
final:	s	h	e	r	l	o	c	k	-

3 Patgen

Patgen works on glyph⁶ indices rather than character codes. Each glyph is represented by a single byte. That amount to 256 indices, where 13 of them are reserved for the digits 0-9 and the characters ‘, ‘-’, and ‘*’. The remaining 243 are used to represent symbols of a given language. To run *patgen*, a translation file is necessary. This file defines the values of certain language-specific parameters (in the first line) and enumerates the various forms in which language symbols may appear (all subsequent lines). In the first line, positions 1 and 2 are used to set the value of **lefthyphenmin**, and positions 3 and 4 are used to set the value of **rightthyphenmin**. These values determine the minimum length of a string that may be generated by a hyphenation procedure. To set a single-digit value, leave the first position blank, i.e., place a space in position 1 and 3 for **lefthyphenmin** and **rightthyphenmin**, respectively. Positions 5, 6 and 7 are used to define alternative values for the special characters ‘, ‘-’ and ‘*’.

1	2	3	4	5	6	7
lefthyphenmin		rightthyphenmin		·	–	*

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
x	e	X	E	x	é	X	ê	x	É	X	Ê	x	\' {e}	X	\^ {e}	x	\' {E}	X	\^ {E}
delimiter	letter e	delimiter	letter E	delimiter	letter é	delimiter	letter ê	delimiter	letter É	delimiter	letter Ê	delimiter	TeX sequence for ê	delimiter	TeX sequence for ê	delimiter	TeX sequence for Ê	delimiter	TeX sequence for Ê

⁶ *Glyph* is commonly used in linguistics, typography, and computer graphics to refer to a specific graphical representation of a character or symbol, which can be the entire symbol or a distinct visual element within it.

we have assumed that the many forms in which we may find the character e will be equivalent for hyphenation (pattern matching) purpose. As another example, see the next line which defines the character π (taken from Haralambous (2021)): `#p#P#\varpi ##`.

Patgen also needs a dictionary file, which is a list of pre-hyphenated words from which *Patgen* extracts patterns to create hyphenation rules. To ensure *Patgen*'s proper functionality, the translation file and the dictionary file must utilize the same encoding, even if it is a multi-byte encoding. The translation file describes how to handle byte sequences representing a glyph, and *Patgen* will work seamlessly when there are at most 243 symbols in the given language.

The syntax to run *patgen* is described in Listing 2

Listing 2: Syntax to run *patgen*.

```
patgen dictionary_file initial_pattern_set output_file translation_file
```

The dictionary file is a list of correctly hyphenated words, one per line; the initial pattern file is a set of hyphenation rules to be used as a starting point; the output file is the set of final rules created by *patgen*; and the translation file which maps the many forms each symbol in a language might appear in T_EX documents. If it is desired to run *patgen* from scratch, starting from an empty set of rules, just use an empty file as the initial pattern set.

Patgen uses a few parameters along its executions:

hyph_start, hyph_finish : Two numbers between 1 and 9 (separated by a space), representing the desired pattern levels in the final set of rules. Odd pattern levels are hyphenating levels and even pattern levels are inhibiting levels. Higher level numbers prevail over lower ones, creating exceptions, exceptions over exceptions, and so on... **hyph_start** and **hyph_finish** specify the first and last levels, respectively, to be considered during the rule creation process.

pat_start, pat_finish : Patterns at each level are chosen in order of increasing pattern length (usually starting with length 2). This is controlled by the parameters **pat_start** and **pat_finish** specified at the beginning of each level. These are the minimum and maximum lengths of patterns we are interesting in. Their values range between 1 and 15.

good weight, bad weight, threshold : Each level of patterns is tested over all words in the dictionary. A pattern is only if it satisfies the following formula: $\alpha \times \# \text{good matches} - \beta \times \# \text{bad matches} \geq \eta$, where α is the **good weight**, β is the **bad weight** and η is the **threshold**.

4 T_EX hyphenation rules for portuguese

Rezende (1987) was the first to create patterns for Portuguese hyphenation in T_EX. The pattern set was last updated in 2015, incorporating contributions from Rezende and Almeida (2015), resulting in a total of 307 rules⁷. These rules effectively hyphenate the majority of Portuguese words.

However, in light of certain cases, we propose an analysis of the default rules to identify areas for improvement. By addressing specific issues and considering non-typical patterns, we aim to enhance hyphenation accuracy. The methodology used for this analysis is described in Section 5.

Out of the default T_EX rules, 252 (82%) follow the pattern 1CV, which represents the recurring CV syllables in Portuguese. As for the consonants, there are typically 18 considered (b, c, ç, d, f, g, j, k, l, m, n, p, r, s, t, v, x, and z) that can combine with 14 vowels (a, e, i, o, u, â, ã, é, í, ó, ú, ê, and ô), resulting in these CV patterns that indicate favorable hyphenation points before the consonants⁸. It is worth noting that there might be other rules or exceptions in the hyphenation patterns not covered by these default rules. To accommodate exceptions, Rezende (1987) proposes the following rules:

- 20 rules created for cases involving consonants b, c, d, f, g, k, p, t, v, or w followed by l or r⁹;

⁷The set of rules for Portuguese hyphenation is short when compared to rules in other languages. For example, English currently has 4938 rules, Russian has 7023 rules and German has 34011 rules.

⁸The complete list of these 252 patterns is given here: 1ba, 1be, 1bi, 1bo, 1bu, 1bá, 1bâ, 1bã, 1bé, 1bí, 1bó, 1bú, 1bê, 1bô, 1ca, 1ce, 1ci, 1co, 1cu, 1cá, 1câ, 1cã, 1cé, 1cí, 1cô, 1cú, 1cê, 1cô, 1ça, 1çe, 1çi, 1ço, 1çu, 1cá, 1câ, 1cã, 1cé, 1cí, 1cô, 1cú, 1cê, 1cô, 1da, 1de, 1di, 1do, 1du, 1dâ, 1dã, 1dê, 1dí, 1dó, 1dú, 1dê, 1dô, 1fa, 1fe, 1fi, 1fo, 1fu, 1fá, 1fâ, 1fã, 1fé, 1fí, 1fó, 1fú, 1fê, 1fô, 1ga, 1ge, 1gi, 1go, 1gu, 1gá, 1gâ, 1gã, 1gé, 1gí, 1gô, 1gú, 1gê, 1gô, 1ja, 1je, 1ji, 1jo, 1ju, 1já, 1jâ, 1jã, 1jé, 1jí, 1jó, 1jú, 1jê, 1jô, 1ka, 1ke, 1ki, 1ko, 1ku, 1ká, 1kâ, 1kã, 1ké, 1kí, 1kó, 1kú, 1kê, 1kô, 1la, 1le, 1li, 1lo, 1lu, 1lá, 1lâ, 1lã, 1lé, 1lí, 1ló, 1lú, 1lê, 1lô, 1ma, 1me, 1mi, 1mo, 1mu, 1má, 1mâ, 1mã, 1mé, 1mí, 1mó, 1mú, 1mê, 1mô, 1na, 1ne, 1ni, 1no, 1nu, 1ná, 1nâ, 1nã, 1né, 1ní, 1nó, 1nú, 1nê, 1nô, 1pa, 1pe, 1pi, 1po, 1pu, 1pá, 1pâ, 1pã, 1pé, 1pí, 1pó, 1pú, 1pê, 1pô, 1ra, 1re, 1ri, 1ro, 1ru, 1rá, 1râ, 1rã, 1ré, 1rí, 1ró, 1rú, 1rê, 1rô, 1sa, 1se, 1si, 1so, 1su, 1sá, 1sâ, 1sã, 1sé, 1sí, 1só, 1sú, 1sê, 1sô, 1ta, 1te, 1ti, 1to, 1tu, 1tá, 1tâ, 1tã, 1té, 1tí, 1tó, 1tú, 1tê, 1tô, 1va, 1ve, 1vi, 1vo, 1vu, 1vâ, 1vã, 1vê, 1ví, 1vó, 1vú, 1vê, 1vô, 1xa, 1xe, 1xi, 1xo, 1xu, 1xá, 1xâ, 1xã, 1xé, 1xí, 1xó, 1xú, 1xê, 1xô, 1za, 1ze, 1zi, 1zo, 1zu, 1zá, 1zâ, 1zã, 1zé, 1zí, 1zó, 1zú, 1zê, 1zô.

⁹The 20 additional rules are: 1b2l, 1b2r, 1c2l, 1c2r, 1d2l, 1d2r, 1f2l, 1f2r, 1g2l, 1g2r, 1k2l, 1k2r, 1p2l, 1p2r, 1t2l, 1t2r, 1v2l, 1v2r, 1w2l, 1w2r.

Table 1: Considering the six dictionaries used, the first line of this table presents the number of words that a given number of hyphenations were found in the dictionaries. The second line presents the number of words that have a given number of hyphenation as its most frequent form found in the dictionaries.

hyphenations	6	5	4	3	2	1	0
hyphenations found	16 036	15 820	10 325	7 783	8 520	27 840	333 253
same hyphenations	15 841	15 642	10 299	7 745	8 435	28 368	n/a

- 3 rules for **c**, **l** or **n** followed by **h**¹⁰;
- 23 distinct patterns were introduced to indicate hyphenation points between vowels or between **c**'s, **r**'s, and **s**'s¹¹;
- 8 patterns adhere to the **1[**gq**]u4V** pattern, signaling beneficial hyphenation points before **g** or **q**, followed by a sequence of **u** and a vowel, with an inhibiting point between the **u** and the subsequent vowel¹²;
- 1 pattern represented as **1-**, denoting that a hyphen indeed serves as a beneficial hyphenation point.

5 Methodology

To assess the performance of each set of rules, we require a collection of correctly hyphenated Portuguese words. We employed word frequency data to ensure we selected representative, commonly used words, avoiding excessively rare ones. Our comprehensive word list was curated using online dictionaries, Portuguese corpora, and verified against Portuguese grammars to validate hyphenation rules. We also performed manual corrections where necessary. We should not expect an entire corpus to follow a set of conventions for a written language nor uniformity among a variety of corpora, especially those harvested from the Internet. "It is notoriously difficult to prescribe rules governing the use of a written language; it is even more difficult to get people to 'follow the rules.' This is in large part due to the nature of written language, in which the conventions are not always in line with actual usage and are subject to frequent change" (Palmer 2010, p. 14).

5.1 Data collection

Initially, we selected CETENFolha as the source corpus, but this approach presented two issues: many words were still missing, and the corpus was based on text from 1994, prior to the implementation of the Orthographic Agreement. The Orthographic Agreement of the Portuguese Language was conducted in 1990, and its transition period started in 2009, becoming mandatory in 2016. However, even after the agreement, some words continue to have different spellings in the participating countries. To address these idiosyncrasies, we aim to develop hyphenation rules that accommodate variations in spelling. For instance, the word *reception* might be written as *receção* (Portugal) or *recepção* (Brazil); the word *action* might be written as *acção* (Portugal) or *ação* (Brazil); and the word *project* might be written as *projecto* (Portugal) or *projeto* (Brazil).

Due to the peculiarities mentioned, we made the decision to integrate the word list from Palavras NET, and subsequently, we augmented it with words sourced from the Portuguese Wikipedia dump. From the Wikipedia data, we narrowed down the selection to a subset of 50 721 words, accounting for 95% of occurrences in the corpus. This threshold was set to filter out typos and infrequent words. The initial word list became quite extensive, comprising 419 578 words. However, we refined the list by retaining only those words for which we could find hyphenation data in at least one of the following online dictionaries: Michaelis, Priberam, Wikcionário, Aulete, Portal da Língua Portuguesa, and Dicio. Consequently, this curation process resulted in a final dictionary containing 85 638 words.

Table 1 presents the number of distinct words with a given number of hyphenations found in the dictionaries. Additionally, it displays the number of words that received the highest agreed-upon hyphenation. For instance, the word *como* was consistently hyphenated as *co-mo* in all six dictionaries, therefore, it contributes to both the counts in the column representing six hyphenations (first column). On the other hand, the word *sua* received different hyphenations: *su-a* in four dictionaries, *sua* in one dictionary, and one dictionary did not provide any hyphenation result. This leads to an increment in the first line, second column (since hyphenations were found in five dictionaries) and the second line, third column (since the most frequent hyphenation was found in four dictionaries). It is evident now that the number of identical hyphenations can exceed the total number of hyphenations found, especially for smaller values.

¹⁰Those 3 rules are: 1c2h, 1l2h, 1n2h.

¹¹Those patterns are: a3a, a3e, a3o, c3c, e3a, e3e, e3o, i3a, i3e, i3i, i3o, i3â, i3ê, i3ô, o3a, o3e, o3o, r3r, s3s, u3a, u3e, u3o, u3u.

¹²They are the following: 1gu4a, 1gu4e, 1gu4i, 1gu4o, 1qu4a, 1qu4e, 1qu4i, 1qu4o.

In summary, our data compilation involved *CETENFolha*, *Palavras NET*, and *Wikipedia*, and hyphenated the words using the original T_EX hyphenation rules. Subsequently, we compared the hyphenated results with those obtained from the dictionary. By conducting a thorough performance appraisal, we systematized the errors and delved into the relevant literature to identify potential rules that could complement and enhance the existing hyphenation system.

6 Writing systems

There is a diverse collection of writing systems, categorized into logographic, syllabic, and alphabetic systems. Although distinct, these systems can be built on an interplay of these categories (Coulmas 2003; Palmer 2010). The principles guiding an alphabetic or phonemic writing system vary significantly based on the language and its history. Here are some key points worth highlighting:

- 1) The most common principle involves representing the sounds of a language with written symbols, using one or a combination of symbols to represent each sound (e.g., the orthographic system of Portuguese uses an alphabet of 27 letters to represent different sounds of the language).
- 2) Etymology also plays a crucial role, where the spelling of a word reflects its origins and historical development (e.g., the French language often reflects the Latin roots of words in its spelling).
- 3) Morphology serves as a guide in structuring many languages, using specific letters or symbols to indicate word endings, prefixes, or suffixes (e.g., Russian uses different forms of the Cyrillic alphabet to indicate gender and case in its nouns).
- 4) The evolution of a language over time also contributes to its written form (e.g., the spelling of many English words has changed over time to reflect changes in pronunciation).

Orthographic systems for languages with logographic writing systems, such as Chinese ideograms, cuneiform writing, and Egyptian hieroglyphs, differ significantly from alphabetic writing systems. In logographic systems, individual symbols or characters represent entire words or ideas, rather than phonemes or sounds. As a result, the principles guiding their orthographic systems are based more on semantic and visual principles than on phonemic principles. Hyphenation is a feature of alphabetic scripts where words are composed of letters, and spaces or hyphens are used to separate words, syllables, or parts of words. Logographic writing systems, such as the Chinese, function differently and do not employ hyphenation (Honorof and Feldman 2006).

7 Portuguese spelling system

Portuguese employs an alphabetical writing system, which means its spelling is guided by phonological principles (Cagliari 2015). The correlation between spelling and pronunciation influences word hyphenation, as words are divided into syllables based on the phonemic system. It is important to note that different languages follow diverse principles for word division. For example, English is primarily guided by morphological principles, evident in words like *walk-ing*, *un-happy*, *work-s*, and *ear-ly*. Additionally, other factors also impact word hyphenation in English, such as the distinction between long and short vowels, which function within the context of open or closed syllable, respectively; and the presence of doubled consonants and digraphs (Lin 2011; Yavas 2020). Although each language has its driving principles for the hyphenation process, multiple factors come into play, leading to various solutions in certain scenarios. For instance, in Portuguese, the phonological principle would lead to *hipe-rativo*, while the morphological principle would lead to *hiper-ativo*. Both approaches seem valid and are indeed found in online dictionaries¹³. Some rather rare words, such as *hiperalgesia*, may not be subject to morphological influences. Since it has a low frequency of occurrence, the individual may not be aware of its morphological components and therefore hyphenate it as *hipe-ralgisia*. The form *hiper-algesia* could also be accepted, emphasizing its morphological constituents. Furthermore, there exist numerous exceptions that may be categorized into rules.

Merely stating that an orthographic system is guided by phonological issues does not necessarily mean that its hyphenation rules directly mirror the phonetic counterpart. This is notably apparent in Portuguese, where a strict one-to-one correspondence between letters and sounds is not always observed. The orthographic system operates according to its specific rules. For example, consider consonant clusters that create a single sound (digraphs) in words like *achado*, *ilha*, *sushi*, *carro*, and *massa*. While these digraphs are pronounced as one sound within a single syllable, their representation in writing determines how they are divided. Specifically,

¹³Looking at words with prefixes, it seems that even on those words, the phonological approach was predominant on hyphenation, but we cannot tell if it is a byproduct of automatic hyphenation that might be used to hyphenate words on online dictionaries.

different consonants within a digraph must remain together, whereas identical consonants are separated. As a result, we observe hyphenations like *a-cha-do*, *i-lha*, and *su-shi*, but *car-ro* and *mas-sa*.

In Portuguese, hyphenations are allowed on syllables boundaries and, in general, follow phonological principles. According to the Grammar (Cunha and Cintra 2016; Bergström and Reis 2011; Cegalla 2020), some rules might still apply:

Non-Splitting Rules

1. diphthong or triphthong should not be split (e.g. *mui-to*, *Pa-ra-guai*);
2. the sequences *ia*, *ie*, *io*, *oa*, *ua*, *ue* and *uo*, when in final unstressed position, should not be split (e.g. *gló-ria*, *vi-tó-ria*, *cá-rie*, *es-pé-cie*, *Má-rio*, *má-goa*, *ré-gua*, *tê-nue*, *con-tí-guo*, *am-bí-guo*);
3. consonant clusters starting a syllable should not be split (e.g. *pneu-má-ti-co*, *psi-có-lo-go*, *mne-mô-ni-co*);
4. the digraphs *ch*, *lh*, *nh* should not be split (e.g. *ra-char*, *a-bro-lhos*, *ma-nhã*);
5. bigrams like *gu* and *qu* whose vowel *u* is not pronounced are never separated from the vowel or diphthong that follows it (e.g. *U-ru-guai*, *pe-que*);
6. since they are digraphs, a vowel and its following nasalization marker (a graphic nasal consonant) should not be split (e.g. *am-bi-ção*, *man-cha*);
7. decreasing diphthongs should not be split (e.g. *ai-ro-so*, *ca-dei-ra*, *o-ra-ção*);
8. rising diphthong should not be split (e.g. *a-bai-xo*, *cau-te-la*, *pa-péis*, *cha-péu*, *pre-fei-to*, *reu-nir*, *noi-te*, *ca-la-bou-ço*, *as-te-rói-de*, *re-tri-bui*);
9. disyllables whose syllable has a single vowel should not be split (e.g. *ato*, *rua*, *ódio*, *unha*);
10. words with more than two syllables, when divided, cannot isolate a syllable composed of a single vowel (e.g. *agos-to*, *la-goa*, *ida-de*);

Splitting Rules

11. hiatus vowels and those vowel sequences where each vowel belongs to a different syllable should be split (e.g. *sa-ú-de*, *ra-i-nha*, *do-er*, *vo-os*), the same procedure is used splitting diphthongs in different syllables (e.g. *cai-aís*) or diphthong and vowel in different syllables (e.g. *en-sai-os*);
12. consonant sequences, when in different syllables, should be split (e.g. *af-ta*, *ab-di-car*, *res-ci-são*, *ab-so-lu-to*);
13. the following consonant digraphs should be split: *rr*, *ss*, *mm*, *nn*, *sc*, *sç* and *xc* (e.g. *ter-ra*, *pro-fes-sor*, *co-mum-men-te*, *con-nos-co*, *des-cer*, *cres-ça*, *ex-ce-der*).

Rules 9 and 10 are primarily aimed at ensuring proper readability of the text, aligning with T_EX approach to deal with widows and orphans. As mentioned in Section 1, the variables `lefthyphenmin` and `riquiryphenmin` control the minimum length for fragments of hyphenated words. When those variables are set to values greater than one, Rules 9 and 10 become fiddling rules in T_EX hyphenation, and they could be disregarded. Notwithstanding, the full hyphenation of words is useful, particularly in text-to-speech applications (Libossek and Schiel 2000; Trokanis and Elkan 2010).

Additionally, it is advisable to refrain from splitting disyllables consisting of four letters (e.g., *para*, *como*, *cede*). This considerations also lead to more aesthetically pleasing and intelligible text, and T_EX's control of isolated fragments, through the variables `\lefthyphenmin` and `\riquiryphenmin`, already address this issue.

In some situations a hyphen should be repeated at the start of the following line. They are:

1. cases where compound words using a hyphen are split across lines (e.g., *couve-/flor*, *ex-/presidente*); and
2. cases where splitting a pronoun could result in a different meaning (e.g., *prazer de ver-/me*¹⁴).

Systematizing the rules that guide syllable boundaries and hyphenation in Portuguese is a fundamental step to understand and improve the T_EX hyphenation rules. To ensure the accuracy of hyphenation rules and effectively compare their results, a hyphenation spelling dictionary serves as an indispensable reference. This dictionary provides a comprehensive listing of correctly hyphenated words, enabling the computation of the correctness of each set of hyphenation rules. By consulting the hyphenation spelling dictionary, one can verify whether the hyphenation patterns generated by a set of rules align with established standards. This process involves analyzing how well the hyphenation rules adhere to the accepted conventions of the language, ensuring that they effectively segment words without compromising readability or consistency. Moreover, by comparing the results obtained from different sets of hyphenation rules against the entries in the spelling dictionary, we can assess the efficacy and reliability of each approach. In section 8, we will define the specifications of this reference dictionary, outlining its role in subsequent rule updates and rules generation.

¹⁴Possible conveyed meanings: *pleasure in seeing me* or *worm's pleasure*.

8 Creating the dictionary

A bash script, named `gethyphenations.sh`, was developed to crawl six online dictionaries and extract word hyphenations whenever available. The process involves launching parallel threads for each word in the list, querying a specific online dictionary, and retrieving the corresponding hyphenation. For this purpose, individual scripts were created for each dictionary: `getmichaelishyphenation.sh` for Michaelis, `getpriberamhyphenation.sh` for Priberam, `getwiktionaryhyphenation.sh` for Wikcionário, `getauletehyphenation.sh` for Aulete, `getportalhyphenation.sh` for Portal da Língua Portuguesa, and `getdiciohyphenation.sh` for Dicio. This systematic approach ensures efficient and accurate extraction of hyphenations from various online sources. The resulting data is stored in a CSV¹⁵ file named `hyphenations.csv`.

8.1 Establishing a gold standard

To establish our gold standard for hyphenation, we will compare the hyphenations offered by all six dictionaries. First, we will determine the total count of unique hyphenations across these dictionaries using the script `hyphenationagreements.sh`. It's worth noting that certain dictionaries annotate words with a colon (':') to indicate potential variations in pronunciation or production. Particularly, 'apparent proparoxytones' might appear marked by this colon sign, indicating the possibility of a diphthongs or hiatuses in the word ending. In Portuguese, words ending in *ea*, *eo*, *ia*, *ie*, *io*, *ua*, *ue*, or *uo*, with the stressed syllable before these endings, can be considered either paroxytones ending in a rising diphthong or proparoxytones ending in a hiatus. For instance, words like *his-tó-ri:a* and *on-du-la-tó-ri:o* are examples of words marked as apparent proparoxytones, suggesting they could be regarded as paroxytones (*his-tó-ria* and *on-du-la-tó-rio*) or proparoxytones (*his-tó-ri-a* and *on-du-la-tó-ri-o*). The script `hyphenationagreements.sh` offers four potential approaches to handle these apparent proparoxytones: retaining the hyphenations with the colon mark to indicate an apparent proparoxytone; selecting only the paroxytone or only the proparoxytone counterpart; or counting both paroxytones and proparoxytones. Given the lack of consensus in the literature, our initial approach involves counting both hyphenations, and we will detail below how we resolve this issue. In our hyphenation dictionary, we have identified 1848 words marked as apparent proparoxytones.

As our initial hyphenation reference, we will consider those cases where all six dictionaries are in agreement. This comprehensive list comprises 15 841 words. Their corresponding hyphenations are meticulously compiled and stored in the file `hyphenations6.dic`. This curated list serves as a benchmark for evaluating the accuracy and effectiveness of hyphenation rules generated through computational methods. It will establish a reliable baseline for our hyphenation analysis, enabling us to assess and refine our methodologies effectively.

In case of disagreement, we must exercise caution and examine each instance. We might be tempted to choose the majority vote approach, simply selecting those hyphenations that have the most supporters, but it is unclear how the hyphenations on those online dictionaries were curated; many might use algorithmic approaches, leading to potential flaws shared among them. For example, 5 dictionaries use the hyphenation *quart-zo* while the dictionary Aulete (originally from Portugal) favors the hyphenation *quar-tzo*, which can be explained by the lack of an epenthetic vowel in European Portuguese¹⁶ (Mateus n.d.). Similarly, *ap-nei-a* or *a-pnei-a* and *disp-nei-a* or *dis-pnei-a* should be accepted considering, since there could be an epenthetic vowel (Brazilian Portuguese) or not (European Portuguese). The same happens with *hiperalgesia*, already mentioned in Section 7. Among other examples, we might also highlight *neerlandês*, which appears hyphenated as *ne-er-lan-dês* in five dictionaries and as *neer-lan-dês* in one. If we consider its pronunciation, we would prefer the former hyphenation, but the first one conforms better to regular written syllables in Portuguese. Some strings might accept more than one hyphenation, since they represent different words. For example, *sub-li-nha* and *su-bli-nha*, where the former is the substantive *underline* and the latter is a flexed form of the verb *to underline*.

The procedure to create the second list of hyphenated words involved several steps: 1) select those hyphenations agreed by five dictionaries, that have no alternative hyphenation proposal (5 to 0 votes); 2) from those with alternative proposals, exam whether they constituted apparent proparoxytones (these cases could arise from instances with 5 to 1 votes or 6 to 1 votes, as explained earlier), a) if the word had a diacritic, select the paroxytone option (e.g. *po-lí-cia* instead of *po-lí-ci-a* and *pe-rí-neo* instead of *pe-rí-ne-o*), b) if not, choose the proparoxytone option (e.g. *au-to-ri-a* instead of *au-to-ri-a* and *mar-ce-nei-ro* instead of *mar-ce-neiro*)¹⁷. 3) if the word did not fall into the apparent proparoxytone category, select the hyphenation with five votes. This procedure was implemented in the script `get51.sh`. Only a few hyphenations required manual correction, including the following that were added: *quar-tzo*, *quar-tzi-to*, *su-bli-nha*, *a-pnei-a*, *dis-pnei-a*, *hi-per-al-ge-si-a*,

¹⁵Comma-separated values (CSV) is a text file format defined in RFC 4180. In CSV, commas are used to separate values, and each line in the file represents a new record. Records consist of fields separated by delimiters, typically a single comma.

¹⁶The online dictionary Aulete has its origins on the printed version known as *Dicionário Caldas Aulete*. It is a dictionary from Portugal, reflecting the European Portuguese dialect.

¹⁷We just have to make sure we are not selecting those cases where there is a sequence of the form `[gq]u-[aeio]` (using regex notation), as in *a-pa-zi-gu-ar* and *en-xa-gu-ar*.

and *neer-lan-dês*. The final list, containing 15 642 additional hyphenations, is stored in `hyphenations5.dic`, bringing the total number of hyphenations across both lists to 31 483.

For those hyphenations with four votes, we adopted a similar approach to the one used previously. Many words garnered only four votes, indicating a lack of alternative hyphenations provided by any dictionary, and were thus initially hyphenated as proparoxytones. These instances were typically distinguished by diacritics in the third-to-last syllable and a *V-V* sequence at the end. To address this, we converted these words into paroxytones by eliminating the final hyphen. In cases where alternative hyphenations were available, we assessed the presence of diacritics. If diacritics were detected, we chose the paroxytone option; otherwise, we defaulted to the proparoxytone option. In instances with multiple alternative options, we selected the option with the highest frequency of votes. This procedure is implemented in the script `get4.sh`.

After completing the aforementioned procedures to create our golden hyphenation dictionary, manual corrections may still be necessary to ensure its accuracy and reliability. These manual adjustments, which encompass the corrections described earlier, are documented in the file `replacements.txt`. Utilizing the script `manualcorrections.sh`, we meticulously apply these corrections to refine our hyphenation dictionary and enhance its effectiveness.

9 Results for the default rules

In this section we will apprise the performance of default T_EX hyphenation rules on the list of hyphenated words, which creation was described in Section 8. Table 2 summarizes the results, showing the number of words correctly hyphenated words by the default rules, the number of words wrongly hyphenated and the number of words where a hyphenation point was missed.

Table 2: Results of T_EX default hyphenation rules on `hyphenations6`, `hyphenations5` and `hyphenations4` dictionaries.

word list	# correct	# wrong	# missing	# entries
hyphenations6	15536	30	277	15841
hyphenations5	13976	1151	602	15642
hyphenations4	9000	886	516	10299
total	38512	2067	1395	41782

For the list comprised in `hyphenations6`, below is the complete list of 30 words that were incorrectly hyphenated:

<i>p.si*co*lo*gi*a</i>	<i>p.si*co*se</i>	<i>p.neu*má*ti*ca</i>
<i>p.si*co*ló*gi*co</i>	<i>p.si*co*te*ra*pi*a</i>	<i>p.neu*má*ti*co</i>
<i>p.si*qui*a*tra</i>	<i>p.si*ca*na*li*ti*co</i>	<i>p.si*cos*so*má*ti*co</i>
<i>p.si*ca*na*lis*ta</i>	<i>p.si*co*te*ra*peu*ta</i>	<i>p.si*co*tró*pi*co</i>
<i>p.si*có*lo*go</i>	<i>p.si*có*ti*co</i>	<i>g.nos*ti*cis*mo</i>
<i>p.si*qui-á*tri*co</i>	<i>g.no*mo</i>	<i>g.nós*ti*co</i>
<i>t.che*co</i>	<i>p.si*co*lo*gis*mo</i>	<i>p.si*co*mo*tri*ci*da*de</i>
<i>p.neu*mo*ni*a</i>	<i>p.si*quis*mo</i>	<i>p.si*co*pa*to*ló*gi*co</i>
<i>p.si*qui*co</i>	<i>p.si*co*mo*tor</i>	<i>p.to*se</i>
<i>p.si*co*pa*ta</i>	<i>t.me*se</i>	<i>su.b-lu*nar</i>

In each instance, an erroneous hyphenation point was placed at the beginning of a word, indicating a potential issue in recognizing certain prefixing morphemes in the language. Among those, 20 occurred in words containing the *psi* morpheme, where the algorithm erroneously hyphenated between the initial *p* and the following *s*. Additionally, 3 cases involved the starting *gno* sequence, while the remaining errors were observed in the sequences *pneu* (3), *tch* (1), *tme* (1), *pto* (1), and *sub* (1).

Those erroneous hyphenation points might be corrected by the introduction of a few rules: `.p2si`, `.p2sı` (see rule 1), `.g2no`, `.g2nó` (see rule 4), `p2neu` (see rule 3), `t2c` (see rule 2), `.t2m` (see rule 6), `.p2t` (see rule 5) and `su2b3r`, `su2b3l` (see rule 12). They are deeply discussed in Section 15.

The first 20 examples of missing hyphenations are displayed in the following list:

<i>a-in*da</i>	<i>a-ı</i>	<i>pa*ra-ı*ba</i>
<i>pa-ıs</i>	<i>sa-ır</i>	<i>pre*ju-ı*zo</i>
<i>re*gi-ão</i>	<i>da-ı</i>	<i>cons*tru-ır</i>
<i>sa-ú*de</i>	<i>ca-ır</i>	<i>ve-ı*cu*lo</i>

*co*ca-í*na*
pa-ul
*a*tra-ir*

*subs*ti*tu-ir*
*in*clu-ir*
*ga-ú*cho*

*ra-i*nhã*
*re*li*gi-ão*
 ...

It is a long list with 277 entries, therefore it is unavoidable to analyse it through clusters of certain patterns. It is presented below the counts of immediate context in which those missing hyphenations were found:

38 u-i	6 i-ú	2 o-á	1 s-q	1 e-ê
22 e-í	5 r-q	2 i-é	1 o-w	1 e-é
20 a-i	5 e-ó	2 e-ú	1 o-l	1 e-â
19 a-í	5 e-i	2 é-o	1 o-i	1 b-l
17 u-í	5 a-q	2 e-l	1 o-g	1 a-u
14 í-a	4 a-v	2 e-c	1 o-é	1 a-t
12 i-á	3 o-ó	2 e-á	1 o-b	1 a-r
11 i-ó	3 i-u	2 a-é	1 i-v	1 a-p
11 a-ú	3 e-q	1 u-q	1 i-q	1 a-l
10 i-ã	3 e-ã	1 ú-o	1 í-o	
9 o-í	3 a-ó	1 u-l	1 i-í	
6 o-q	2 u-á	1 u-ã	1 i-c	

On the top of the list we see the missing hyphen in *u-i*, accounting for 38 cases. Among those, 34 were originated from the pattern *u-ir* in the end of word and 13 from *a-ir* in the end of word. To fix them, We will include the rules *u1ir.* and *a1ir.* (see rule 21). The cases that are not covered by this rule are: *tu-im*, *je*su-i*tis*mo*, *ma*lau-i*a*no*, and *con*tri*bu-in*te*. On Section 15 we see how to deal with them.

The default rules already include 17 separating rules for vowel sequences:

a3a	e3e	i3i	i3ô	u3a
a3e	e3o	i3o	o3a	u3e
a3o	i3a	i3â	o3e	u3o
e3a	i3e	i3ê	o3o	u3u

but there are just 3 rules to hyphenate between vowels when one has a diacritic: *i3â*, *i3ê*, *i3ô*. We could then add more 29 rules to account for the missing hyphenations between vowels with diacritics and also the missing rule *i1u*:

a1ê	e1ã	e1ú	i1ô	o1í
a1í	e1é	i1â	í1o	o1ô
a1ô	e1ê	i1ã	i1u	u1â
a1ú	e1í	í1a	i1ú	u1ã
e1â	e1ó	i1é	o1â	u1í
e1ã	é1o	i1í	o1é	ú1o

See more about this on rule 19. These rules will require exceptions for sequences with *[qg]uV'* (where we used *V'* to represent a vowel with a diacritic). See more on rules 19 and 20.

From the list of missing hyphenations above, we find another recurring pattern, a missing hyphen preceding *q*:

a-q	i-q	r-q	u-q
e-q	o-q	s-q	

That issue might be easily solved by adding a hyphenation rule: *1qu*, but as the *q* is always followed by a sequence *uV*, and there are already four hyphenation rules involving this sort of sequence in the default rules (*1qu4a*, *1qu4e*, *1qu4i* and *1qu4o*), the rules *1qu2â*, *1qu2é*, *1qu2ê*, *1qu2i*, *que2i*, *1qu4ã*, *1qu4ó*, *1qu4â* could be added (see rules 19, 20 and 32) but they have not proven to be necessary.

The missing hyphens in *a-v* and *i-v* might be easily solved by introducing the rule *1vô* (see rule 35) which complements the rule *1vo* already included in the default set of rules. Similarly, the missing hyphens *e-l*, *u-l*, *o-l*, and *a-l* are corrected by the introduction of the rule *1lô* rule (see rule 35), complementing the already included rule *1lo*. In the same way, the rules *1cô*, *1gô*, *1bô*, *1tô*, *1rô*, and *1pô* (see rule 35) should be added to fix the missing hyphens *e-c*, *i-c*, *o-g*, *o-b*, *a-t*, *a-r*, and *a-p*, respectively. The missing hyphen *b-l* happened in *su.b-lu*nar*, which also has a wrong hyphenation. Those errors were caused by the default rule *1b21*. The introduction of rule *su2b31* will help solve this matter (see details in rule 12).

The remaining missing hyphen is *o-w*, which comes from the foreign *kilowatt*. Those cases of words borrowed from other languages are dealt in the Foreignness: topic ahead.

10 Updating the rules

We now begin incorporating the rules proposed in the previous section and appraise the results of each one. During this process, we will consider the necessity of updating those rules, including exceptions, or even discarding the proposed rules if needed.

After introducing the rules: `.p2si`, `.p2s1`, `.g2no`, `.g2nô`, `t2c`, `p2neu`, `.t2m`, `.p2t`, `su2b3r`, and `su2b3l`, the number of incorrect hyphenations decreased to 3, and the number of missing hyphenations decreased to 278. At first sight, we expected the number of incorrect hyphenations to drop to zero. However, the introduction of rule `su2b3l` introduced hyphenation errors in *su-b.li*me*, *su-b.li*ma*çã*o and *su-b.li*ma*do*.

Listing 3: Hyphenation of the word *sublime* after the introduction of rule `su2b3l`. This rule introduces a wrong hyphenation point, demanding an exception rule to fix it.

.	s	u	b	l	i	m	e	.
1	0	0						1su
0	0	2	3	0				su2b3l
		1	2	0				1b2l
			1	0	0			1li
					1	0	0	1me
max:	1	0	2	3	0	1	0	0
final:	s	u	b	—	l	i	—	m e

By inserting the rule `.su3b4li`, we fixed the hyphenation errors caused by the introduction of the rule `su2b3l`, as illustrated in Listing 3. This adjustment resulted in zero hyphenation errors and 276 missing hyphens.

The next step is to introduce the rules to address the missing hyphenation points. These rules are: `a1ir.`, `u1ir.`, `i3ã`, `i3ê`, `i3ô`, `1qu`, `1vô`, `1lô`, `1cô`, `1gô`, `1bô`, `1tô`, `1rô`, `1pô`, `a1é`, `a1í`, `a1ô`, `a1ú`, `e1ã`, `e1â`, `e1ã`, `e1é`, `e1ê`, `e1í`, `e1ô`, `e1o`, `e1ú`, `i1ã`, `i1â`, `i1a`, `i1é`, `i1í`, `i1ô`, `i1o`, `i1u`, `i1ú`, `o1ã`, `o1é`, `o1í`, `o1ô`, `u1ã`, `u1â`, `u1í`, `ú1o`. After inserting these rules, the number of incorrect hyphenations increases to 18 due to the inclusion of hyphens in *qu.í*, *gu.í* and *qu.á*. To overcome these errors, we need to add exception rules: `1qu2ã`, `1qu2í`, and `1gu2í`. See more on this on rules 19 and 20.

We have reached a point where there is no wrong hyphenation but still there are 19 missing hyphenation. Here is the list of the missing ones:

<i>a-in*da</i>	<i>re-in*te*gra*çã</i> o	<i>re-in*ci*dir</i>	<i>tu-im</i>	<i>re-im*pri*mir</i>
<i>ra-i*nha</i>	<i>cam*pa-i*nha</i>	<i>pa-in*ço</i>	<i>je*su-i*tis*mo</i>	
<i>con*tri*bu-in*te</i>	<i>re-in*te*grar</i>	<i>pi*xa-im</i>	<i>ben*jo-im</i>	
<i>ra-iz</i>	<i>ta-i*nha</i>	<i>re-im*pres*sã</i> o	<i>ma*lau-i*a*no</i>	

Observing the instances above, 17 out of 19 involve a missing hyphen before the vowel *i*. The following rules will be used to solve those issues: `a1ind`, `a1i1nh` (see rule 24), `e1imp` (see rule 18), `e1inc`, `e1inf`, `e1ing`, `e1ins`, `e1int`, `e1inv`, (see rule 15), `u1int`, `u1ind` (see rule 16), and `u1iz.`, `a1iz.` (see rule 22).

After incorporating these rules, the number of missing hyphenations decrease to only 8:

<i>pa-ul</i>	<i>pa-in*ço</i>	<i>tu-im</i>	<i>ben*jo-im</i>
<i>qui*lo-watt</i>	<i>pi*xa-im</i>	<i>je*su-i*tis*mo</i>	<i>ma*lau-i*a*no</i>

Since these remaining cases comprises only a few rare words, we now proceed to integrate the dictionary `hyphenations5` and evaluate the results of the current rules against this set. In this new dataset, the number of incorrect hyphenations increases to 1121, with the majority stemming from apparent proparoxytones. These cases can be easily rectified by avoiding the final hyphen which leaves the vowels *a*, *e* or *o* alone in the last syllable. To address this, we introduce rules `4a.`, `4e.`, `4o.` (see ??). While a more comprehensive rule set could be developed to account for all scenarios¹⁸, we opted for these more concise rules despite their limitations. The addition of these rules reduces the number of wrong hyphenations to 35; however, it increases the number of missing hyphenations from 133 to 900. Among these, 770 additional missing hyphenations occur before the final vowel of a word. Considering the smaller number of cases, the perceived severity (a missing hyphen is deemed less problematic than an incorrect one), and the position of the word's end, we have chosen to retain these rules.

The resulting errors are shown in the following list:

¹⁸A better rule would be to restrain the final hyphenation, leaving the vowel alone, if the preceding syllable has a vowel with diacritic. That would require an exhaustive list of all scenarios.

<i>sa*gu.ão</i>	<i>c.za*ris*ta</i>	<i>p.so*ri*a*se</i>	<i>núp*ci.as</i>	<i>flâ*mu.la</i>
<i>p.seu-dô*ni*mo</i>	<i>bre.ch*ti*a*no</i>	<i>t.za*ris*ta</i>	<i>su.b-li*nha</i>	<i>a-p.nei-a</i>
<i>s.ta*li*nis*mo</i>	<i>c.za*ris*mo</i>	<i>g.nais*se</i>	<i>a*rá*bi.as</i>	<i>c.ni*dá*rio</i>
<i>a.b-rup*to</i>	<i>su.b-li*te*ra*tu*ra</i>	<i>m.ne-mô*ni*co</i>	<i>quar-t.zo</i>	<i>dis-p.nei-a</i>
<i>su.b-li*mi*nar</i>	<i>ca*ra*min*gu.ás</i>	<i>p.so*as</i>	<i>e*fe*mé*ri.de</i>	<i>gli*có*li.se</i>
<i>quar-t.zi*to</i>	<i>d.ze*ta</i>	<i>su.b-lin*gual</i>	<i>cri*sân*te.mo</i>	<i>hi*pe.r-al*ge*si-a</i>
<i>c.za*ri*na</i>	<i>m.ne-mô*ni*ca</i>	<i>es*tô*ma.go</i>	<i>e*xé*qui.as</i>	<i>ne.er*lan*dês</i>

From these, we have proposed the rules: **1gu4á**, **1gu4ã**, **1qu4ã** (see rule 20), **.m2n** (see rule 10), **c2za** (see rule 7), **.s2** (see rule 8), and **1p2seu1d** (see rule 14).

The inclusion of these rules leave us with 21 incorrect hyphenations and 899 missing ones. The list of wrong hyphenations follows below:

<i>a.b-rup*to</i>	<i>d.ze*ta</i>	<i>su.b-lin*gual</i>	<i>e*xé*qui.as</i>	<i>ne.er*lan*dês</i>
<i>su.b-li*mi*nar</i>	<i>p.so*ri*a*se</i>	<i>núp*ci.as</i>	<i>a-p.nei-a</i>	
<i>quar-t.zi*to</i>	<i>t.za*ris*ta</i>	<i>su.b-li*nha</i>	<i>c.ni*dá*rio</i>	
<i>bre.ch*ti*a*no</i>	<i>g.nais*se</i>	<i>a*rá*bi.as</i>	<i>dis-p.nei-a</i>	
<i>su.b-li*te*ra*tu*ra</i>	<i>p.so*as</i>	<i>quar-t.zo</i>	<i>hi*pe.r-al*ge*si-a</i>	

Most of these cases occur in rare words or loanword. We decided not to address them. Analyzing the missing hyphenation points, as mentioned earlier, most result from the rules (4a., 4e., 4o.) we proposed to handle apparent proparoxytones. This leave us with 125 missing hyphenations to address. We observe that a few additional rules with diacritic counterpart need to be introduced: **1dô**, **1fô**, **1mô**, **1nô**, **1sô**, **1zô** (see ??), **u3é**, **1gu4é**, **1qu4é** (see ??). After incorporating these rules, we now have 21 incorrect and 856 missing hyphenations, 86 of which do not involve the final vowel.

To account for a few more missing cases, we propose the inclusion of the following rules: **tu1i**, **bu1i**, **co1in**, **nu1i**, **cu1i**, **o1in**, **u1in**, **su1i**, **i1e**, **ra1is**, **ju1i**, **fu1i**, **du1i**, **do1im**, **au1i**, **u1i1ç**. With these additions, the number of missing hyphenations drops to 818, with 772 cases involving final vowels, as mentioned earlier. Among these, only two cases involve a final vowel with a diacritic: *fu*zu-ê* and *ba*ba*la-ô*). The list of the remaining 46 follows:

<i>ru-im</i>	<i>re-ur*ba*ni*zar</i>	<i>a*bra-â*mi*co</i>	<i>in*flu-ên*cia</i>	<i>quar-t.zo</i>
<i>flu-i*dez</i>	<i>voy-eu*ris*ta</i>	<i>ca*far*na-um</i>	<i>de*po-i*men*to</i>	<i>es*pon*sa-is</i>
<i>a*da-il</i>	<i>ba-ha*men*se</i>	<i>ku-wai*ti*a*no</i>	<i>re-u*nir</i>	<i>ba*la-us*tra*da</i>
<i>voy-eu*ris*mo</i>	<i>dar-wi*nis*ta</i>	<i>ma*ru-im</i>	<i>co-i*bir</i>	<i>con*gru-ên*cia</i>
<i>a.b-rup*to</i>	<i>ma-çô*ni*co</i>	<i>su.b-lin*gual</i>	<i>flu-ên*cia</i>	<i>tran*se-un*te</i>
<i>me*ga-watt</i>	<i>pat-chu*li</i>	<i>te-ô*ni*mo</i>	<i>con*flu-ên*cia</i>	<i>di*flu-ên*cia</i>
<i>ca-im</i>	<i>su.b-li*te*ra*tu*ra</i>	<i>a*lu-i*men*to</i>	<i>a*nu-ên*cia</i>	<i>in*de-is*cen*te</i>
<i>su.b-li*mi*nar</i>	<i>fa-im</i>	<i>mal-</i>	<i>su.b-li*nha</i>	
<i>dar-wi*nis*mo</i>	<i>voy-eu*ris*ti*co</i>	<i>thu*si*a*nis*mo</i>	<i>a*flu-ên*cia</i>	
<i>quar-t.zi*to</i>	<i>tai-wa*nês</i>	<i>mal-thu*si*a*no</i>	<i>in*con*gru-ên*cia</i>	

From this list, we may create a new rule: **u1ê**, which would also require two exception rules: **gu2ê** and **qu2ê**. Incorporating them, we solve the nine cases of missing hyphen *u-ê*, dropping the number of missing hyphenations to 809.

We now move to the last dictionary file: **hyphenations4**. Using the rules we have developed so far, we achieve 0 incorrect and 29 missing hyphenations in this dictionary. Of these, 26 involve final vowels, and the remaining 3 are: *re-in*de*xa*ção*, *pa*ra-i*ba*no* and *ru-ther*ford*)

Considering some additional words that are not on the list, we gather an additional set of rules to be incorporated: **1çô** (e.g., *ma-çô-ni-co*), **1xô** (e.g., *a-xô-nio*, *sa-xô-ni-co*), **a1â** (e.g., *a-bra-â-mi-co*), **a1ã** (e.g., *a-bra-ão*, *ca-na-ã*), **a1ô** (e.g., *fa-ra-ô-ni-co*), **e1ô** (e.g., *le-ô-ni-das*, *na-po-le-ô-ni-co*), **o1ã** (e.g., *jo-ão*, *la-go-ão*), **o1ê** (e.g., *bo-ê-mi-o*), **cu2i** (e.g., *cui-da-do*, *cir-cui-to*; this rule requires the exception rule **cu3in** to avoid unintended errors in words like *pi-cu-i-nha*, *i-mis-cu-in-do*), **.ne4o** (e.g., *neo-li-be-ral*, *neo-na-zis-ta*), **o1im** (e.g., *co-im-bra*; this rule requires the exception rule **co2ima** to avoid unintended errors in words like *coi-ma*), **o2i1na** (e.g., *boi-na*, *coi-na*), and **pro1i1b** (e.g., *pro-i-bir*). After incorporating those rules, the number of incorrect and missing hyphenations, respectively, will be: 0, 538 (**hyphenations6**); 21, 818 (**hyphenations5**); and 0, 33 (**hyphenations4**).

Table 3: Hyphenation in the dictionaries of words having conflict between morphological and syllabic information.

word	hyphenation	Michaelis	Priberam	dictionary Wikitionary	Aulete	Portal	Dicio
sublinhar	su-bli-nhar sub-li-nhar	x	x	x	x	x	
reiniciar	rei-ni-ci-ar re-i-ni-ci-ar	x	x	x	x	x	x
ciberespaço	ci-be-res-pa-ço ci-ber-es-pa-ço	x	x	x	x	x	x
hiperalgesia	hi-pe-ral-ge-si-a hi-per-al-ge-si-a	x	x	x	x	x	x
autoimagem	au-toi-ma-gem au-to-i-ma-gem	x			x	x	x

11 Limitations

In general, although the erroneous words share some common characteristics that might allow for a reduction in hyphenation errors to some extent, there is a limitation inherent in the way \TeX 's rules are conceived. Below, we present the systematics found in these data that could be encompassed by a different rule structure:

Morphological determination: Prefixes such as *re-*, *sub-*, *ciber-*, *hiper-*, and *auto-*, among others, require separating the prefix from its stem, which can lead to phonological issues. For example, words like *reiniciar*, *sublinhar*, *ciberespaço*, *hiperalgesia*, and *autoimagem* contain prefixes and could be hyphenated as *re-i-ni-ci-ar*, *sub-li-nhar*, *ci-ber-es-pa-ço*, *hi-per-al-ge-si-a*, and *au-to-i-ma-gem*, respectively, to respect their morphological formation. However, considering that Portuguese hyphenates its words based on syllabic phonological correlates, words that require morphological information, such as these, might not have their hyphenation performed correctly. These examples are presented in Table 3.

Foreignness: There is a group of words in the corpus that are terminologies or words incorporated into the Portuguese language without full phonological adaptation, such as *darwinismo*, *quilowatt*, and *esfiha*. The lack of adaptation makes the phonological pattern very specific to the word, making it impossible to incorporate their cases \TeX 's rules alongside the other rules. The solution is to add them to the exception word list.

Word-initial consonant clusters: Portuguese has few cases of consonant clusters at the beginning of a word. They are, in general, etymological remnants and are currently unproductive in the language, since there are no neologisms with this pattern. Encounters like *ps-* and *pn-* are more frequent, as they are present in words like *psicologia* and *pneu*, which have moderate frequency in the language. These can be predicted by specific rules that would cover 49 and 13 words in the corpus, respectively. However, there are consonant clusters that are found in very specific and low-frequency words. Although possible, it is not worth adding very specific rules for clusters found in words like *dzeta*, *gnu*, *cnidário*, *ftálico*, and *gnaisse* – which amount to only five words.

Abbreviations, Acronyms, or Initialisms: Whether for efficiency, convenience, clarity, or specialized jargon, it is common to use shortened versions of words or phrases. *Abbreviation* is a method employed to achieve this shortening. In our Portuguese corpus, we find examples such as *etc.*¹⁹, *Dr.*²⁰, *Exmo.*²¹, *cap.*²², *Univ.*²³, *ed.*²⁴, *s.n.*²⁵. Another shortened form is an *initialism*, which consists of using the initial letters of words to create a shortened version. However, initialisms may not always conform to the hyphenation rules described in this work, as they do not necessarily follow the orthographic or phonotactic standards of the language. Some abbreviations found in the corpus include *SESC*²⁶, *INSS*²⁷, *PCdoB*²⁸, *PM*²⁹, and *UFRJ*³⁰. *Acronyms* are a specific type of shortening where the first letters (or groups of letters) of each

¹⁹Latin expression *et cetera*, meaning ‘and other similar things’.

²⁰*doutor* (doctor, person with PhD title, but popularly used to designate an erudite individual)

²¹*Excelentíssimo* (honourable)

²²*capítulo* (chapter)

²³Universidade (university)

²⁴*edição* (edition)

²⁵sine nomine, Latin expression meaning ‘without a name’, mostly used in the context of publishing.

²⁶Serviço Social do Comércio

²⁷*Instituto Nacional do Seguro Social* (National Institute of Social Security)

²⁸*Partido Comunista do Brasil* (Communist Party of Brazil)

²⁹*Polícia Militar* (military police)

³⁰Universidade Federal do Rio de Janeiro

word are combined to form a new pronounceable word. In the corpus, we encounter examples like *Ana-tel*³¹, *Ovni*³², *Sida*³³ (in Portugal) and *Mercosul*³⁴. These various shortened forms play an important role in written language, providing concise ways to represent longer words or phrases. It is important to note that abbreviations, acronyms, and initialisms are generally treated as single units and are not hyphenated.

12 Creating a new set of rules using Patgen

In this section, we describe the process of creating a new set of hyphenation rules using *patgen*. This involves defining specific parameters, choosing a reference dictionary with correct hyphenations, and using a translation file tailored for *patgen*. In the previous Section 3 we have detailed how *patgen* works.

A partial of the content of the translation file is shown in Listing 4.

Listing 4: Portuguese translation file.

```

1  1 1
2  %% This file portuguese.tra defines the letters used for generating
3  %% Portuguese hyphenation patterns with patgen.
4  a A
5  á Á
6  â Â
7  ã Ã
8  b B
9  c C
10 c C
...
39 x X
40 y Y
41 z Z

```

The parameters mentioned in Section 3 to run *patgen* are provided interactively during the execution of *patgen*. We might automatize this process using a script. The Listing 5 presents the code block responsible for passing these parameters to *patgen*.

Listing 5: Portuguese translation file.

```

{
  printf "${hyph_start}_${hyph_finish}\n"
  for ((i=hyph_start; i<=hyph_finish; i++)); do
    printf "${pat_start}_${pat_finish}\n"
    printf "${weights}\n"
  done
  printf "y"
} | patgen ../data/portuguese.dic "$filename" "output_${identifier}" ../data/portuguese.
tra | tail -n 2;

```

As an example of execution, when choosing `hyph_start=1`, `hyph_finish=9`, `pat_start=2`, `pat_finish=5`, `good weight=1`, `bad weight=1`, and `threshold=1`, we achieve 99.97%(86799) correct hyphenations, 0.01%(12) incorrect hyphenations, and 0.03%(25) missing hyphenations. It produced a set of 799 rules, where just 123 are present in the handcrafted set of rules.

We have also used the approach used by P. Sojka and O. Sojka (2019) in creating the script `make-full-pattern`. In this approach, the parameters are loaded from a file and might be distinct for each level. We start from an empty set and create progressive higher order rules, always using the rules from the previous step as the starting point at the next step. Using this approach with the 8 level parameters they used to create hyphenation rules to German, we achieved 99.98%(86820) correct hyphenations, 0.01%(9) incorrect hyphenations, and 0.00%(4) missed hyphenations. The result was a set of 593 rules, with only 69 in common with those handcrafted rules proposed earlier.

The results from the set of rules crafted by *patgen* seems outstanding, but we must be skeptical of its generalization power. To inquire on this matter, we decided to retrieve the list of words where just three hyphenations were found in the dictionaries. We have also dealt with apparent proparoxytones, to be consistent in our approach. In this new set, the number of correct, incorrect and missing hyphenations for each set of rules is given Table 4. It is important to highlight that, among those 385 missing hyphenations in the handcrafted

³¹Agência Nacional de Telecomunicações (National Telecommunications Agency)

³²Objeto voador não identificado (unidentified flying object - UFO)

³³Síndrome da Imunodeficiência Adquirida (acquired immunodeficiency syndrome - AIDS)

³⁴Mercado Comum do Sul (Southern Common Market).

Table 4: Compared results between two use cases of *patgen* and the handcrafted rule set.

rule set	correct	incorrect	missing
fixed parameters	7465	150	183
make-full-pattern	647	47	7089
handcrafted	7335	42	385

rules, 278 cases correspond to the last hyphen, leaving a single vowel as a final syllable. Comparing to the other set of rules, we have 7 (in 183, for fixed parameters case), and 257 (in 7089, for the **make-full-pattern** script).

Establishing this comparison leave us no doubt that our handcrafted rules achieved the best performance, also retaining a small number of useful rules, what leads to good generalization.

13 Comparing the set of rules

Let us remember that the standard T_EX hyphenation rules, proposed by Rezende and Almeida (2015), amount to XXX errors, in a set of XXXX words. Among those, XXX correspond to places where the hyphenation was not carried out and XXX to mismarkings. Faced with the errors made by the standard hyphenator, a set of 135 rules were established, which reduced the errors to XXXX, with XXXX unmarked hyphenation places and XXX mismarkings. Such rules are presented and exemplified below, as well as the necessary exception rules and an example in which the application of the rule is demonstrated.

14 Universal syllabic pattern generation

Universal syllabic pattern generation by Jakub Máca, Petr Sojka, Ondřej Sojka (TUG 2023).

15 Hyphenation rule patches

OLD TEST

OLD TEST

Foi criado um conjunto de 135 regras para aprimorar as hifenizações no português. Analisando um conjunto de 37 798 palavras, as regras padrão apresentam 1368 erros (1273 pontos de hifenização não marcados e 125 pontos marcados erroneamente)³⁵, enquanto as novas regras apresentaram 132 erros (118 pontos de hifenização não marcados e 35 pontos marcados erroneamente)³⁶.

Foi criado um conjunto de 37798 palavras obtidas do corpus CETENFolha, juntamente com palavras do Palavras NET e algumas palavras do corpus baseado na Wikipedia. Foram mantidas apenas as palavras para as quais encontrou-se a hifenização nos seguintes dicionários online: Michaelis, Priberam, Wikcionário, Aulete, Portal da Língua Portuguesa e Dicio. Adotamos como hifenização ‘correta’ aquela que ao menos 4 dicionários concordavam. Algumas hifenizações foram realizadas manualmente pelo autor. Considerando a frequência de ocorrência das palavras nos corpus, o conjunto de palavras escolhido representa 55% do total das ocorrências relativo ao Wikipedia e 58% em relação ao CETENFolha. Este percentual é relativamente baixo, mas deve-se ao fato de que não obtivemos sucesso em encontrar a hifenização de muitas palavras nos dicionários analisados e também por não observarmos um consenso de ao menos 4 dicionários com uma mesma hifenização.

Analisando os erros resultantes da hifenização utilizando as regras padrão do TeX, foram estabelecidas algumas regras como forma de abarcar um conjunto maior de palavras. Tais regras são elencadas a seguir, expondo também as regras de exceção necessárias e um exemplo onde a regra será utilizada.

1 regra: **.p2si, .p2s1** – psicologia, psíquico

2 regra: **t2c** – tchau

3 regra: **p2neu** – pneumonia

4 regra: **.g2no, .g2nó** – gnomo, gnóstico

5 regra: **.p2t** – ptose

6 regra: **.t2m** – tmese

7 regra: **c2za** – czar

³⁵A soma não é 1368 pois temos 30 casos em que os dois tipos de erros ocorrem.

³⁶35 palavras apresentam ambos tipos de erro.

- 8 regra: .s2 – stalinismo
- 9 regra: .t2 – tsunami
- 10 regra: .m2n – mnemônico
- 11 regra: z1z, p1p, z2z. – pizza, shopping, jazz
- 12 regra: su2b3r, su2b3l – sublunar, subrotina
exceção: .su3b4li
- 13 regra: 1sô – subsônico
- 14 regra: 1p2seu1d – pseudônimo
- 15 regra: e1inc, e1inf, e1ing, e1ins, e1int, e1inv – reincidência, reinfecção, reingressa, reinserção, reintegração, reinventar
- 16 regra: u1int, u1ind – contribuinte, excluindo
- 17 regra: o1in – coincidência, agroindustrial, bioinformática, amendoins, gastrointestinal exceção: o1ina – boina
- 18 regra: e1imp – reimpresso
- 19 regra: a1â, a1ã, a1é, a1í, a1ó, a1ô, a1ú, e1á, e1â, e1ã, e1é, e1ê, e1í, e1ó, e1ô, e1ú, é1o, i1á, i1ã, i1é, i1í, i1ó, i1u, i1ú, í1a, o1á, o1ã, o1é, o1ê, o1í, o1ó, u1á, u1ã, u1é, u1ê, u1í, ú1o – abraâmico, abraão, aéreo, país, caótico, faraônico, saúde, balneário, oceânico, campeã, feérico, veêm, veículo, teórico, napoleônico, conteúdo, néon, diário, região, soviético, iídiche, periódico, feiura, viúva, maníaco, razoável, João, poético, boêmia, heroísmo, alcoólico, usuário, itapuã, suécia, cauê, suíça, flúor
exceção: 1gu2é, 1gu2ê, 1gu2í, 1qu2á, 1qu2é, 1qu2ê, 1qu2í – alguém, português, linguística, aquático, inquerito, sequência, química
- 20 regra: 1gu4á, 1gu4ã, 1qu4ã – jaraguá, saguão, quão
- 21 regra: a1ir., u1ir. – sair, diminuir
- 22 regra: u1iz., a1iz. – juiz, raiz
- 23 regra: í1a – baía
- 24 regra: a1ind, a1i1nh – ainda, rainha
- 25 regra: o1i1nh – moinho
- 26 regra: au1i1c, du1i1c, u1i1ç, u1i1d, cu2i, dru2i, flu2id, bu1i1n, cu1i1n, fu1i1n, nu1i1n, ru1i1na, ru1i1no, su1i1ti, tu1i1ti, u1iz – cacaucultor, sanduicheira, constituição, continuidade, cuidador, druida, fluido, tabuinha, picuinha, fuinha, genuinamente, arruinado, arruinou, jesuitismo, intuitivo, juizado
exceção: cu3i1da1de, bu1i2n1d, bu1i2n1t – acuidade, contribuindo, contribuinte
- 27 regra: i1ur – diurno
- 28 regra: í1o – íon
- 29 regra: ú1o – flúor
- 30 regra: pro1i1b – proibição
- 31 regra: i1un – triunfar
- 32 regra: 1qu4ó, 1qu4â – quórum, equânime
- 33 regra: é3o – alvéolo
- 34 regra: i4a., i4e., i4o., o4a., u4a. – economia, espécie, vazio, destoa, institua
- 35 regra: 1bô, 1cô, 1dô, 1fô, 1gô, 1lô, 1pô, 1mô, 1nô, 1rô, 1tô, 1vô, 1xô, 1zô – robô, recôncavo, judô, telefônica, xangô, camelô, capô, sumô, econômico, tarô, chatô, vovô, saxônia, amazônia

The 135 rules were grouped above in a list of 3 types of rules. They may be further organized into four large groups. The first, which comprises rules 1 to 11, includes consonant clusters such as *czar*, *ptose* and *gnomo*. They, unlike the examples that will be exposed in section ?? (exceptions - item 9 currently), present a set of derived words, which makes their marking advantageous in view of the number of cases that are included in this marking rule. The second group, comprising rules 12 to 18, delimits the morphological boundary between prefixes and radicals. As noted, although phonological issues guide the separation of numerous words in Portuguese, there are also those that are guided by morphology. This is the case of words that have the prefixes *sub-* and *re-*, such as *sublunar* and *reinserção*. The third group, comprising rules 19 to 34, seeks to understand a set of words that have vowel combinations that do not follow the general rules. This is because the Portuguese language has vowel encounters with the second vowel graphically marked that can be separated, forming hiatuses, such as *caótico*, *balneário* and *razoável*, while there are also words with a similar structure that constitute a diphthong, such as *português*, *alguém* and *linguística*. It is remarkable, of course, that the latter are formed by the digraphs *qu-* and *gu-*, while the former by vowels other than *i* and *u*. The fourth and last group, in turn, which comprises rules ?? to 36, which are counterparts of rules that were already in the default rules, but did not contemplate the cases with certain accents. They were then added to encompass words such as *camelô*, *recôncavo*, *amazônia*, and *maçônico*.

It is important to highlight that the words included in these rules are, in general, of low frequency and were incorporated into the Portuguese language without a phonotactic adaptation, which causes these idiosyncrasies and exceptions to the language. They are characterized as marked cases, since it is not possible to defend that they reproduce a phonological pattern of the language because they are not productive, that is, they are not taken as an example and/or derive new words.

O conjunto de 135 regras é apresentado a seguir:

p2si	a1é	o1ã	1cô	1vô	ruli1no
p2si	a1í	o1ê	e1in1c	1xô	suli1ti
p3sia	a1ó	o1ê	e1in1f	1zô	tuli1ti
p3sin	a1ô	o1í	e1in1g	e1imp	u1iz
t2c	a1ú	o1ó	e1in1s	te2im	.m2n
p2neu	e1á	u1á	e1in2st	que2i	1gu4á
.g2no	e1ã	u1ã	e1in2sc	o1im	1gu4ã
.g2nó	e1ã	u1ê	e1in1t	.ne4o	1qu4ã
.g2nô	e1é	u1ê	e1in1v	au1i1c	1çô
.p2t	e1ê	u1í	a1ind	du1i1c	i1ur
su2b3r	e1í	ú1o	a1i1nh	u1i1ç	i1o
su2b3l	e1ó	1gu2é	o1in	u1i1d	ú1o
su3b4li1nh	e1ô	1gu2ê	o2i1na	cu2i	pro1i1b
su3b4li1ma	e1ú	1gu2í	o1i1nh	cu3i1da1de	i1un
su3b4li1me	é1o	1qu2á	1bô	dru2i	1qu4ó
su3b4li1mid	i1á	1qu2é	1dô	flu2id	1qu4ã
1sô	i1ã	1qu2ê	1fô	bu1i1n	é3o
1p2seu1d	i1é	1qu2í	1gô	bu1i2n1d	z1z
c2za	i1i	1flu2id	1mô	bu1i2n1t	p1p
.s2	i1ó	a1ir.	1nô	cu1i1n	z2z.
.t2	i1ú	u1ir.	1pô	fu1i1n	
a1ã	i1a	i1a	1rô	nu1i1n	
a1ã	o1á	1lô	1tô	ru1i1na	

A Tabela 5 apresenta os erros cometidos no conjunto das 37798 palavras analisadas. As palavras omitidas são aquelas que foram hifenizadas corretamente pelos dois conjuntos de regra. Utilizamos aqui a marcação adotada pelo *Patgen*: * para indicar pontos de hifenização corretamente marcados, . para indicar pontos de hifenização marcados erroneamente e - para indicar pontos de hifenização não encontrados. Os pontos marcados erroneamente e aqueles não encontrados são considerados erros. Quando ocorre algum erro de hifenização, marcaremos a palavra com ✗. Quando não houver erro algum, marcaremos a palavra com ✓. Ao final, podemos verificar que as regras padrão apresentaram 1368 erros (3,62%) e as novas regras 132 erros (0,35%).

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
ainda	a-in*da ✗	a*in*da ✓
país	pa-ís ✗	pa*ís ✓
região	re*gi-ão ✗	re*gi*ão ✓
saúde	sa-ú*de ✗	sa*ú*de ✓
período	pe*rí-o*do ✗	pe*rí*o*do ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
juiz	ju-iz ✗	ju*iz ✓
união	u*ni-ão ✗	u*ni*ão ✓
econômico	e*co-nô*mi*co ✗	e*co*nô*mi*co ✓
aí	a-í ✗	a*í ✓
sair	sa-ir ✗	sa*ir ✓
opinião	o*pi*ni-ão ✗	o*pi*ni*ão ✓
constituição	cons*ti*tu-i*ção ✗	cons*ti*tu*i*ção ✓
saída	sa-í*da ✗	sa*í*da ✓
luís	lu-ís ✗	lu*ís ✓
campeão	cam*pe-ão ✗	cam*pe*ão ✓
aliás	a*li-ás ✗	a*li*ás ✓
distribuição	dis*tri*bu-i*ção ✗	dis*tri*bu*i*ção ✓
avião	a*vi-ão ✗	a*vi*ão ✓
diário	di-á*ri*o ✗	di*á*ri*o ✓
diária	di-á*ri*a ✗	di*á*ri*a ✓
daí	da-í ✗	da*í ✓
cair	ca-ir ✗	ca*ir ✓
paraíba	pa*ra-í*ba ✗	pa*ra*í*ba ✓
prejuízo	pre*ju-í*zo ✗	pre*ju*í*zo ✓
judiciário	ju*di*ci-á*ri*o ✗	ju*di*ci*á*ri*o ✓
suíça	su-í*ça ✗	su*í*ça ✓
influência	in*flu-ên*ci*a ✗	in*flu*ên*ci*a ✓
depoimento	de*po-i*men*to ✗	de*po-i*men*to ✗
instituição	ins*ti*tu-i*ção ✗	ins*ti*tu*i*ção ✓
consequência	con*se-quên*ci*a ✗	con*se*quên*ci*a ✓
usuário	u*su-á*ri*o ✗	u*su*á*ri*o ✓
contribuição	con*tri*bu-i*ção ✗	con*tri*bu*i*ção ✓
diminuir	di*mi*nu-ir ✗	di*mi*nu*ir ✓
fenômeno	fe-nô*me*no ✗	fe*nô*me*no ✓
ruim	ru-im ✗	ru-im ✗
construir	cons*tru-ir ✗	cons*tru*ir ✓
veículo	ve-í*cu*lo ✗	ve*í*cu*lo ✓
cocaína	co*ca-í*na ✗	co*ca*í*na ✓
paul	pa-ul ✗	pa-ul ✗
reunir	re-u*nir ✗	re-u*nir ✗
continuidade	con*ti*nu-i*da*de ✗	con*ti*nu*i*da*de ✓
diálogo	di-á*lo*go ✗	di*á*lo*go ✓
atrair	a*tra-ir ✗	a*tra*ir ✓
ocasião	o*ca*si-ão ✗	o*ca*si*ão ✓
sebastião	se*bas*ti-ão ✗	se*bas*ti*ão ✓
sequência	se-quên*ci*a ✗	se*quên*ci*a ✓
substituir	subs*ti*tu-ir ✗	subs*ti*tu*ir ✓
incluir	in*clu-ir ✗	in*clu*ir ✓
gaúcho	ga-ú*cho ✗	ga*ú*cho ✓
poluição	po*lu-i*ção ✗	po*lu*i*ção ✓
proibido	pro-i*bi*do ✗	pro*í*bi*do ✓
rainha	ra-i*nha ✗	ra*í*nha ✓
religião	re*li*gi-ão ✗	re*li*gi*ão ✓
paraíso	pa*ra-í*so ✗	pa*ra*í*so ✓
frequência	fre-quên*ci*a ✗	fre*quên*ci*a ✓
distribuir	dis*tri*bu-ir ✗	dis*tri*bu*ir ✓
substituição	subs*ti*tu-i*ção ✗	subs*ti*tu*i*ção ✓
contribuinte	con*tri*bu-in*te ✗	con*tri*bu*in*te ✓
conteúdo	con*te-ú*do ✗	con*te*ú*do ✓
leão	le-ão ✗	le*ão ✓
concluir	con*clu-ir ✗	con*clu*ir ✓
proibição	pro-i*bi*ção ✗	pro*í*bi*ção ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
sociólogo	so*ci-ó*lo*go ✗	so*ci*ó*lo*go ✓
construído	cons*tru-í*do ✗	cons*tru*i*do ✓
maceió	ma*cei-ó ✗	ma*cei*ó ✓
diminuição	di*mi*nu-i*ção ✗	di*mi*nu*i*ção ✓
crediário	cre*di-á*ri*o ✗	cre*di*á*ri*o ✓
pivô	pi*vô ✗	pi*vô ✓
substituído	subs*ti*tu-í*do ✗	subs*ti*tu*i*do ✓
razoável	ra*zo-á*vel ✗	ra*zo*á*vel ✓
juíza	ju-i*za ✗	ju*i*za ✓
coincidência	co-in*ci*dên*ci*a ✗	co*in*ci*dên*ci*a ✓
rebelião	re*be*li-ão ✗	re*be*li*ão ✓
noticiário	no*ti*ci-á*ri*o ✗	no*ti*ci*á*ri*o ✓
olimpíada	o*lim*pí-a*da ✗	o*lim*pí*a*da ✓
imobiliário	i*mo*bi*li-á*ri*o ✗	i*mo*bi*li*á*ri*o ✓
austriaco	aus*trí-a*co ✗	aus*trí*a*co ✓
penitenciária	pe*ni*ten*ci-á*ri*a ✗	pe*ni*ten*ci*á*ri*a ✓
psicologia	p.si*co*lo*gi*a ✗	psi*co*lo*gi*a ✓
contribuir	con*tri*bu-ir ✗	con*tri*bu*ir ✓
aéreo	a-é*re*o ✗	a*é*re*o ✓
psicanálise	p.si*ca*ná*li*se ✗	psi*ca*ná*li*se ✓
vestuário	ves*tu-á*ri*o ✗	ves*tu*á*ri*o ✓
destruição	des*tru-i*ção ✗	des*tru*i*ção ✓
viúva	vi-ú*va ✗	vi*ú*va ✓
distribuído	dis*tri*bu-í*do ✗	dis*tri*bu*i*do ✓
juízo	ju-i*zo ✗	ju*i*zo ✓
constituente	cons*ti*tu-in*te ✗	cons*ti*tu-in*te ✗
presidenciável	pre*si*den*ci-á*vel ✗	pre*si*den*ci*á*vel ✓
viável	vi-á*vel ✗	vi*á*vel ✓
suíço	su-i*ço ✗	su*i*ço ✓
destruir	des*tru-ir ✗	des*tru*ir ✓
saído	sa-i*do ✗	sa*i*do ✓
concluído	con*clu-í*do ✗	con*clu*i*do ✓
distribuidora	dis*tri*bu-i*do*ra ✗	dis*tri*bu*i*do*ra ✓
imobiliária	i*mo*bi*li-á*ri*a ✗	i*mo*bi*li*á*ri*a ✓
sinfônica	sin*fô*ni*ca ✗	sin*fô*ni*ca ✓
uísque	u-ís*que ✗	u*ís*que ✓
poética	po-é*ti*ca ✗	po*é*ti*ca ✓
prostituição	pros*ti*tu-i*ção ✗	pros*ti*tu*i*ção ✓
rodoviário	ro*do*vi-á*ri*o ✗	ro*do*vi*á*ri*o ✓
bicampeão	bi*cam*pe-ão ✗	bi*cam*pe*ão ✓
psicológico	p.si*co*ló*gi*co ✗	psi*co*ló*gi*co ✓
proibir	pro-i*bir ✗	pro*i*bir ✓
triunfo	tri-un*fo ✗	tri*un*fo ✓
variável	va*ri-á*vel ✗	va*ri*á*vel ✓
criciúma	cri*ci-ú*ma ✗	cri*ci*ú*ma ✓
sinônimo	si-nô*ni*mo ✗	si*nô*ni*mo ✓
legião	le*gi-ão ✗	le*gi*ão ✓
caído	ca-i*do ✗	ca*i*do ✓
equívoco	e-quí*vo*co ✗	e*quí*vo*co ✓
baía	ba-í-a ✗	ba*í*a ✓
quilômetro	qui-lô*me*tro ✗	qui*lô*me*tro ✓
raiz	ra-iz ✗	ra-iz ✗
psiquiatra	p.si*qui*a*tra ✗	psi*qui*a*tra ✓
tricampeão	tri*cam*pe-ão ✗	tri*cam*pe*ão ✓
psicanalista	p.si*ca*na*lis*ta ✗	psi*ca*na*lis*ta ✓
ruído	ru-i*do ✗	ru*i*do ✓
possuir	pos*su-ir ✗	pos*su*ir ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
coibir	co-i*bir ✗	co-i*bir ✗
soviético	so*vi-é*ti*co ✗	so*vi*é*ti*co ✓
inviável	in*vi-á*vel ✗	in*vi*á*vel ✓
alô	a-lô ✗	a*lô ✓
heroína	he*ro-í*na ✗	he*ro*i*na ✓
destruído	des*tru-i*do ✗	des*tru*i*do ✓
telefônico	te*le-fô*ni*co ✗	te*le*fô*ni*co ✓
saí	sa-í ✗	sa*í ✓
suíte	su-i*te ✗	su*i*te ✓
asiático	a*si-á*ti*co ✗	a*si*á*ti*co ✓
irônico	i-rô*ni*co ✗	i*rô*ni*co ✓
avô	a-vô ✗	a*vô ✓
restituição	res*ti*tu-i*ção ✗	res*ti*tu*i*ção ✓
atribuir	a*tri*bu-ir ✗	a*tri*bu*ir ✓
ingenuidade	in*ge*nu-i*da*de ✗	in*ge*nu*i*da*de ✓
psicólogo	p.si*có*lo*go ✗	psi*có*lo*go ✓
pneu	p.neu ✗	pneu ✓
tetracampeão	te*tra*cam*pe-ão ✗	te*tra*cam*pe*ão ✓
vestiário	ves*ti-á*ri*o ✗	ves*ti*á*ri*o ✓
proteína	pro*te-í*na ✗	pro*te*i*na ✓
espião	es*pi-ão ✗	es*pi*ão ✓
ciúme	ci-ú*me ✗	ci*ú*me ✓
napoleão	na*po*le-ão ✗	na*po*le*ão ✓
teórica	te-ó*ri*ca ✗	te*ó*ri*ca ✓
intermediário	in*ter*me*di-á*ri*o ✗	in*ter*me*di*á*ri*o ✓
constituir	cons*ti*tu-ir ✗	cons*ti*tu*ir ✓
embrião	em*bri-ão ✗	em*bri*ão ✓
autônomo	au*tô*no*mo ✗	au*tô*no*mo ✓
confiável	con*fi-á*vel ✗	con*fi*á*vel ✓
teórico	te-ó*ri*co ✗	te*ó*ri*co ✓
estômago	es*tô*ma*go ✗	es*tô*ma*go ✓
termômetro	ter*mô*me*tro ✗	ter*mô*me*tro ✓
homônimo	ho*mô*ni*mo ✗	ho*mô*ni*mo ✓
excluído	ex*clu-í*do ✗	ex*clu*i*do ✓
penitenciário	pe*ni*ten*ci-á*ri*o ✗	pe*ni*ten*ci*á*ri*o ✓
cardíaco	car*dí-a*co ✗	car*dí*a*co ✓
mutuário	mu*tu-á*ri*o ✗	mu*tu*á*ri*o ✓
aquém	a-quém ✗	a*quém ✓
viário	vi-á*ri*o ✗	vi*á*ri*o ✓
despoluição	des*po*lu-i*ção ✗	des*po*lu*i*ção ✓
medíocre	me*dí-o*cre ✗	me*dí*o*cre ✓
reintegração	re-in*te*gra*ção ✗	re*in*te*gra*ção ✓
marquês	mar-quês ✗	mar*quês ✓
beneficiário	be*ne*fi*ci-á*ri*o ✗	be*ne*fi*ci*á*ri*o ✓
atribuição	a*tri*bu-i*ção ✗	a*tri*bu*i*ção ✓
balneário	bal*ne-á*ri*o ✗	bal*ne*á*ri*o ✓
coreógrafo	co*re-ó*gra*fo ✗	co*re*ó*gra*fo ✓
incômodo	in-cô*mo*do ✗	in*cô*mo*do ✓
sanduíche	san*du-í*che ✗	san*du*i*che ✓
pecuária	pe*cu-á*ri*a ✗	pe*cu*á*ri*a ✓
cirurgião	ci*rur*gi-ão ✗	ci*rur*gi*ão ✓
jaú	ja-ú ✗	ja*ú ✓
reconstruir	re*cons*tru-ir ✗	re*cons*tru*ir ✓
influir	in*flu-ir ✗	in*flu*ir ✓
santuário	san*tu-á*ri*o ✗	san*tu*á*ri*o ✓
sutiã	su*ti-ã ✗	su*ti*ã ✓
poético	po-é*ti*co ✗	po*é*ti*co ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
ensaísta	en*sa-ís*ta ✗	en*sa*ís*ta ✓
porquê	por-quê ✗	por*quê ✓
biólogo	bi-ó*lo*go ✗	bi*ó*lo*go ✓
extrair	ex*tra-ir ✗	ex*tra*ir ✓
reconstituição	re*cons*ti*tu-i*ção ✗	re*cons*ti*tu*i*ção ✓
anfitrião	an*fi*tri-ão ✗	an*fi*tri*ão ✓
anônimo	a-nô*ni*mo ✗	a*nô*ni*mo ✓
aquático	a-quá*ti*co ✗	a*quá*ti*co ✓
excluir	ex*clu-ir ✗	ex*clu*ir ✓
filarmônica	fi*lar-mô*ni*ca ✗	fi*lar*mô*ni*ca ✓
juizado	ju-i*za*do ✗	ju*i*za*do ✓
peão	pe-ão ✗	pe*ão ✓
previdenciário	pre*vi*den*ci-á*ri*o ✗	pre*vi*den*ci*á*ri*o ✓
constituído	cons*ti*tu-í*do ✗	cons*ti*tu*í*do ✓
pornô	por-nô ✗	por*nô ✓
psiquiátrico	p.si*qui-á*tri*co ✗	psi*qui*á*tri*co ✓
moinho	mo-i*nho ✗	mo*i*nho ✓
obstruir	obs*tru-ir ✗	obs*tru*ir ✓
psiquiatria	p.si*qui*a*tri*a ✗	psi*qui*a*tri*a ✓
reindexação	re-in*de*xa*ção ✗	re-in*de*xa*ção ✗
aquário	a-quá*ri*o ✗	a*quá*ri*o ✓
amendoim	a*men*do-im ✗	a*men*do-im ✗
baú	ba-ú ✗	ba*ú ✓
distribuidor	dis*tri*bu-i*dor ✗	dis*tri*bu*i*dor ✓
trair	tra-ir ✗	tra*ir ✓
intuição	in*tu-i*ção ✗	in*tu*i*ção ✓
petroquímica	pe*tro-qui*mi*ca ✗	pe*tro*qui*mi*ca ✓
evoluir	e*vo*lu-ir ✗	e*vo*lu*ir ✓
petroquímico	pe*tro-qui*mi*co ✗	pe*tro*qui*mi*co ✓
asiática	a*si-á*ti*ca ✗	a*si*á*ti*ca ✓
dinamarquês	di*na*mar-quês ✗	di*na*mar*quês ✓
tcheco	t.che*co ✗	tche*co ✓
diminuído	di*mi*nu-í*do ✗	di*mi*nu*í*do ✓
instituir	ins*ti*tu-ir ✗	ins*ti*tu*ir ✓
lâmpião	lam*pi-ão ✗	lam*pi*ão ✓
arquitetônico	ar*qui*te-tô*ni*co ✗	ar*qui*te*tô*ni*co ✓
atraído	a*tra-í*do ✗	a*tra*í*do ✓
gastronômico	gas*tro-nô*mi*co ✗	gas*tro*nô*mi*co ✓
vibrião	vi*bri-ão ✗	vi*bri*ão ✓
caraíba	ca*ra-í*ba ✗	ca*ra*í*ba ✓
ferroviário	fer*ro*vi-á*ri*o ✗	fer*ro*vi*á*ri*o ✓
mobiliário	mo*bi*li-á*ri*o ✗	mo*bi*li*á*ri*o ✓
robô	ro-bô ✗	ro*bô ✓
traído	tra-í*do ✗	tra*í*do ✓
chatô	cha-tô ✗	cha*tô ✓
viés	vi-és ✗	vi*és ✓
biquíni	bi-quí*ni ✗	bi*quí*ni ✓
pneumonia	p.neu*mo*ni*a ✗	pneu*mo*ni*a ✓
galeão	ga*le-ão ✗	ga*le*ão ✓
pseudônimo	p.seu-dô*ni*mo ✗	pseu*dô*ni*mo ✓
psicologicamente	p.si*co*lo*gi*ca*men*te ✗	psi*co*lo*gi*ca*men*te ✓
tatuí	ta*tu-í ✗	ta*tu*í ✓
anuário	a*nu-á*ri*o ✗	a*nu*á*ri*o ✓
contrair	con*tra-ir ✗	con*tra*ir ✓
curió	cu*ri-ó ✗	cu*ri*ó ✓
egoísmo	e*go-ís*mo ✗	e*go*ís*mo ✓
poliéster	po*li-és*ter ✗	po*li*és*ter ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
caótico	ca-ó*ti*co ✗	ca*ó*ti*co ✓
descontraído	des*con*tra-í*do ✗	des*con*tra*í*do ✓
hanseníase	han*se*ní-a*se ✗	han*se*ní*a*se ✓
macedônia	ma*ce-dô*ni*a ✗	ma*ce*dô*ni*a ✓
heroísmo	he*ro-ís*mo ✗	he*ro*ís*mo ✓
instituído	ins*ti*tu-í*do ✗	ins*ti*tu*í*do ✓
periódico	pe*ri-ó*di*co ✗	pe*ri*ó*di*co ✓
abdômen	ab-dô*men ✗	ab*dô*men ✓
paraibano	pa*ra-i*ba*no ✗	pa*ra-i*ba*no ✗
psíquico	p.sí*qui*co ✗	psí*qui*co ✓
reconstituir	re*cons*ti*tu-ir ✗	re*cons*ti*tu*ir ✓
oásis	o-á*sis ✗	o*á*sis ✓
parnaíba	par*na-í*ba ✗	par*na*í*ba ✓
usufruir	u*su*fru-ir ✗	u*su*fru*ir ✓
camelô	ca*me-lô ✗	ca*me*lô ✓
maníaco	ma*ní-a*co ✗	ma*ní*a*co ✓
sabiá	sa*bi-á ✗	sa*bi*á ✓
agronômico	a*gro-nô*mi*co ✗	a*gro*nô*mi*co ✓
astronômico	as*tro-nô*mi*co ✗	as*tro*nô*mi*co ✓
psicopata	p.si*co*pa*ta ✗	psi*co*pa*ta ✓
geógrafo	ge-ó*gra*fo ✗	ge*ó*gra*fo ✓
ideário	i*de-á*ri*o ✗	i*de*á*ri*o ✓
piá	pi-á ✗	pi*á ✓
alcoólatra	al*co-ó*la*tra ✗	al*co*ó*la*tra ✓
biógrafo	bi-ó*gra*fo ✗	bi*ó*gra*fo ✓
coincidir	co-in*ci*dir ✗	co*in*ci*dir ✓
escorpião	es*cor*pi-ão ✗	es*cor*pi*ão ✓
fluência	flu-ên*ci*a ✗	flu*ên*ci*a ✓
ruína	ru-í*na ✗	ru*í*na ✓
delinquência	de*lin-quên*ci*a ✗	de*lin*quên*ci*a ✓
contraído	con*tra-í*do ✗	con*tra*í*do ✓
diluir	di*lu-ir ✗	di*lu*ir ✓
psicose	p.si*co*se ✗	psi*co*se ✓
bioquímica	bi*o-quí*mi*ca ✗	bi*o*quí*mi*ca ✓
fluidez	flu-i*dez ✗	flu*i*dez ✗
futevôlei	fu*te-vô*lei ✗	fu*te*vô*lei ✓
judaísmo	ju*da-ís*mo ✗	ju*da*ís*mo ✓
viúvo	vi-ú*vo ✗	vi*ú*vo ✓
capô	ca-pô ✗	ca*pô ✓
casuísmo	ca*su-ís*mo ✗	ca*su*ís*mo ✓
receituário	re*cei*tu-á*ri*o ✗	re*cei*tu*á*ri*o ✓
agrião	a*gri-ão ✗	a*gri*ão ✓
comerciário	co*mer*ci-á*ri*o ✗	co*mer*ci*á*ri*o ✓
estereótipo	es*te*re-ó*ti*po ✗	es*te*re*ó*ti*po ✓
psicoterapia	p.si*co*te*ra*pi*a ✗	psi*co*te*ra*pi*a ✓
vascaíno	vas*ca-í*no ✗	vas*ca*í*no ✓
arquitetônica	ar*qui*te-tô*ni*ca ✗	ar*qui*te*tô*ni*ca ✓
macroeconômico	ma*cro*e*co-nô*mi*co ✗	ma*cro*e*co*nô*mi*co ✓
restituir	res*ti*tu-ir ✗	res*ti*tu*ir ✓
rutherford	ru-ther*ford ✗	ru-ther*ford ✗
triunfal	tri-un*fal ✗	tri*un*fal ✓
guardião	guar*di-ão ✗	guar*di*ão ✓
oriundo	o*ri-un*do ✗	o*ri*un*do ✓
arqueólogo	ar*que-ó*lo*go ✗	ar*que*ó*lo*go ✓
carbônico	car-bô*ni*co ✗	car*bô*ni*co ✓
stalinista	s.ta*li*nis*ta ✗	sta*li*nis*ta ✓
egoísta	e*go-ís*ta ✗	e*go*ís*ta ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
portuário	por*tu-á*ri*o ✗	por*tu*á*ri*o ✓
recaída	re*ca-í*da ✗	re*ca*í*da ✓
reiniciar	re-i*ni*ci*ar ✗	re-i*ni*ci*ar ✗
reurbanização	re-ur*ba*ni*za*ção ✗	re-ur*ba*ni*za*ção ✗
teólogo	te-ó*lo*go ✗	te*ó*lo*go ✓
baião	bai-ão ✗	bai*ão ✓
genuíno	ge*nu-í*no ✗	ge*nu*i*no ✓
hegemônico	he*ge-mô*ni*co ✗	he*ge*mô*ni*co ✓
resquício	res-quí*ci*o ✗	res*quí*ci*o ✓
retribuição	re*tri*bu-i*ção ✗	re*tri*bu*i*ção ✓
confluência	con*flu-ên*ci*a ✗	con*flu*ên*ci*a ✓
fluir	flu-ir ✗	flu*ir ✓
imperdoável	im*per*do-á*vel ✗	im*per*do*á*vel ✓
distrair	dis*tra-ir ✗	dis*tra*ir ✓
guaíba	gua-í*ba ✗	gua*i*ba ✓
intuitivo	in*tu-i*ti*vo ✗	in*tu*i*ti*vo ✓
orquídea	or-quí*de*a ✗	or*quí*de*a ✓
semiótica	se*mi-ó*ti*ca ✗	se*mi*ó*ti*ca ✓
vovô	vo-vô ✗	vo*vô ✓
campanha	cam*pa-i*nha ✗	cam*pa*i*nha ✓
czar	c.zar ✗	czar ✓
diluição	di*lu-i*ção ✗	di*lu*i*ção ✓
distraído	dis*tra-í*do ✗	dis*tra*í*do ✓
eloquência	e*lo-quên*ci*a ✗	e*lo*quên*ci*a ✓
insubstituível	in*subs*ti*tu-í*vel ✗	in*subs*ti*tu*í*vel ✓
reinventar	re-in*ven*tar ✗	re*in*ven*tar ✓
retribuir	re*tri*bu-ir ✗	re*tri*bu*ir ✓
rubéola	ru*bé-o*la ✗	ru*bé*o*la ✓
abruptamente	a.b-rup*ta*men*te ✗	a.b-rup*ta*men*te ✗
maniqueísmo	ma*ni*que-ís*mo ✗	ma*ni*que*ís*mo ✓
antiquário	an*ti-quá*ri*o ✗	an*ti*quá*ri*o ✓
harmônica	har-mô*ni*ca ✗	har*mô*ni*ca ✓
judô	ju-dô ✗	ju*dô ✓
maniqueísta	ma*ni*que-ís*ta ✗	ma*ni*que*ís*ta ✓
tchau	t.chau ✗	tchau ✓
moído	mo-í*do ✗	mo*i*do ✓
psicanalítico	p.si*ca*na*lí*ti*co ✗	psi*ca*na*lí*ti*co ✓
caíco	ca-í*co ✗	ca*i*co ✓
descontinuidade	des*con*ti*nu-i*da*de ✗	des*con*ti*nu*i*da*de ✓
tarô	ta-rô ✗	ta*rô ✓
absenteísmo	ab*sen*te-ís*mo ✗	ab*sen*te*ís*mo ✓
anuência	a*nu-ên*ci*a ✗	a*nu*ên*ci*a ✓
etíope	e*tí-o*pe ✗	e*tí*o*pe ✓
ofurô	o*fu-rô ✗	o*fu*rô ✓
panteão	pan*te-ão ✗	pan*te*ão ✓
atômico	a*tô-mi*co ✗	a*tô*mi*co ✓
corroído	cor*ro-í*do ✗	cor*ro*i*do ✓
destituído	des*ti*tu-í*do ✗	des*ti*tu*í*do ✓
estagiário	es*ta*gi-á*ri*o ✗	es*ta*gi*á*ri*o ✓
inadiável	i*na*di-á*vel ✗	i*na*di*á*vel ✓
instruído	ins*tru-í*do ✗	ins*tru*i*do ✓
amazônico	a*ma-zô*ni*co ✗	a*ma*zô*ni*co ✓
anuidade	a*nu-i*da*de ✗	a*nu*i*da*de ✓
evoluído	e*vo-lu-í*do ✗	e*vo*lu*i*do ✓
faísca	fa-ís*ca ✗	fa*ís*ca ✓
harmônico	har-mô*ni*co ✗	har*mô*ni*co ✓
patriótico	pa*tri-ó*ti*co ✗	pa*tri*ó*ti*co ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
presidiário	pre*si*di-á*ri*o ✗	pre*si*di*á*ri*o ✓
reconstruído	re*cons*tru-i*do ✗	re*cons*tru*i*do ✓
tricô	tri-cô ✗	tri*cô ✓
antiético	an*ti-é*ti*co ✗	an*ti*é*ti*co ✓
cronômetro	cro-nô*me*tro ✗	cro*nô*me*tro ✓
destituição	des*ti*tu-i*ção ✗	des*ti*tu*i*ção ✓
fiúza	fi-ú*za ✗	fi*ú*za ✓
píer	pí-er ✗	pí-er ✗
psique	p.si*que ✗	psi*que ✓
recair	re*ca-ir ✗	re*ca*ir ✓
sinfônico	sin-fô*ni*co ✗	sin*fô*ni*co ✓
triunfante	tri-un*fan*te ✗	tri*un*fan*te ✓
adail	a*da-il ✗	a*da-il ✗
atônito	a-tô*ni*to ✗	a*tô*ni*to ✓
birô	bi-rô ✗	bi*rô ✓
caféina	ca*fe-í*na ✗	ca*fe*í*na ✓
ideólogo	i*de-ó*lo*go ✗	i*de*ó*lo*go ✓
inequívoco	i*ne-quí*vo*co ✗	i*ne*quí*vo*co ✓
míope	mí-o*pe ✗	mí*o*pe ✓
poá	po-á ✗	po*á ✓
stalinismo	s.ta*li*nis*mo ✗	sta*li*nis*mo ✓
suíno	su-í*no ✗	su*í*no ✓
voyeurismo	voy-eu*ris*mo ✗	voy-eu*ris*mo ✗
camaleão	ca*ma*le-ão ✗	ca*ma*le*ão ✓
dióxido	di-ó*xi*do ✗	di*ó*xi*do ✓
geólogo	ge-ó*lo*go ✗	ge*ó*lo*go ✓
ilíada	i*lí-a*da ✗	i*lí*a*da ✓
impermeável	im*per*me-á*vel ✗	im*per*me*á*vel ✓
jesuíta	je*su-í*ta ✗	je*su*í*ta ✓
suástica	su-ás*ti*ca ✗	su*ás*ti*ca ✓
bioquímico	bi*o-quí*mi*co ✗	bi*o*quí*mi*co ✓
buquê	bu-quê ✗	bu*quê ✓
esôfago	e-sô*fa*go ✗	e*sô*fa*go ✓
folião	fo*li-ão ✗	fo*li*ão ✓
quillowatt	qui*lo-watt ✗	qui*lo-watt ✗
ruir	ru-ir ✗	ru*ir ✓
altruísmo	al*tru-ís*mo ✗	al*tru*ís*mo ✓
equânime	e-quâ*ni*me ✗	e*quâ*ni*me ✓
fruição	fru-i*ção ✗	fru*i*ção ✓
insaciável	in*sa*ci-á*vel ✗	in*sa*ci*á*vel ✓
caída	ca-í*da ✗	ca*í*da ✓
ciático	ci-á*ti*co ✗	ci*á*ti*co ✓
gavião	ga*vi-ão ✗	ga*vi*ão ✓
gratuidade	gra*tu-i*da*de ✗	gra*tu*i*da*de ✓
irremediável	ir*re*me*di-á*vel ✗	ir*re*me*di*á*vel ✓
ladainha	la*da-i*nha ✗	la*da*i*nha ✓
piauiense	pi*au-i*en*se ✗	pi*au-i*en*se ✗
prostituir	pros*ti*tu-ir ✗	pros*ti*tu*ir ✓
triunfar	tri-un*far ✗	tri*un*far ✓
acuidade	a*cu-i*da*de ✗	a*cu*i*da*de ✓
adriático	a*dri-á*ti*co ✗	a*dri*á*ti*co ✓
afrodisíaco	a*fro*di*sí-a*co ✗	a*fro*di*sí*a*co ✓
bioética	bi*o-é*ti*ca ✗	bi*o*é*ti*ca ✓
cocô	co-cô ✗	co*cô ✓
instruir	ins*tru-ir ✗	ins*tru*ir ✓
paradisíaco	pa*ra*di*sí-a*co ✗	pa*ra*di*sí*a*co ✓
poluir	po*lu-ir ✗	po*lu*ir ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
possuído	pos*su-í*do ✗	pos*su*í*do ✓
sublinha	su.b-li*nha ✗	su.b-li*nha ✗
antibiótico	an*ti*bi-ó*ti*co ✗	an*ti*bi*ó*ti*co ✓
apreciável	a*pre*ci-á*vel ✗	a*pre*ci*á*vel ✓
desobstruir	de*sobs*tru-ir ✗	de*sobs*tru*ir ✓
destruidor	des*tru-i*dor ✗	des*tru*i*dor ✓
diáspora	di-ás*po*ra ✗	di*ás*po*ra ✓
diluído	di*lu-í*do ✗	di*lu*í*do ✓
maiúsculo	mai-ús*cu*lo ✗	mai*ús*cu*lo ✓
miúdo	mi-ú*do ✗	mi*ú*do ✓
pentacampeão	pen*ta*cam*pe-ão ✗	pen*ta*cam*pe*ão ✓
retraído	re*tra-i*do ✗	re*tra*i*do ✓
subtrair	sub*tra-ir ✗	sub*tra*ir ✓
suéter	su-é*ter ✗	su*é*ter ✓
açaí	a*ça-í ✗	a*ça*í ✓
alcoólico	al*co-ó*li*co ✗	al*co*ó*li*co ✓
despoluir	des*po*lu-ir ✗	des*po*lu*ir ✓
diurno	di-ur*no ✗	di*ur*no ✓
inexequível	i*ne*xe-quí*vel ✗	i*ne*xe*quí*vel ✓
prontuário	pron*tu-á*ri*o ✗	pron*tu*á*ri*o ✓
psicoterapeuta	p.si*co*te*ra*peu*ta ✗	psi*co*te*ra*peu*ta ✓
saúva	sa-ú*va ✗	sa*ú*va ✓
tabelião	ta*be*li-ão ✗	ta*be*li*ão ✓
abrupto	a.b-rup*to ✗	a.b-rup*to ✗
destituir	des*ti*tu-ir ✗	des*ti*tu*ir ✓
dionisíaco	di*o*ni-sí-a*co ✗	di*o*ni*sí*a*co ✓
lacônico	la-cô*ni*co ✗	la*cô*ni*co ✓
megawatt	me*ga-watt ✗	me*ga-watt ✗
poliuretano	po*li-u*re*ta*no ✗	po*li*u*re*ta*no ✓
protozoário	pro*to*zo-á*ri*o ✗	pro*to*zo*á*ri*o ✓
reincidente	re-in*ci*den*te ✗	re*in*ci*den*te ✓
reintegrar	re-in*te*grar ✗	re*in*te*grar ✓
reinvenção	re-in*ven*ção ✗	re*in*ven*ção ✓
afluência	a*flu-ên*ci*a ✗	a*flu*ên*ci*a ✓
assiduidade	as*si*du-i*da*de ✗	as*si*du*i*da*de ✓
coágulo	co-á*gu*lo ✗	co*á*gu*lo ✓
cuíca	cu-í*ca ✗	cu*í*ca ✓
faraó	fa*ra-ó ✗	fa*ra*ó ✓
fundiário	fun*di-á*ri*o ✗	fun*di*á*ri*o ✓
incendiário	in*cen*di-á*ri*o ✗	in*cen*di*á*ri*o ✓
itapuã	i*ta*pu-ã ✗	i*ta*pu*ã ✓
parcimônia	par*ci-mô*ni*a ✗	par*ci*mô*ni*a ✓
psicótico	p.si*có*ti*co ✗	psi*có*ti*co ✓
redemoinho	re*de*mo-i*nho ✗	re*de*mo*i*nho ✓
redistribuir	re*dis*tri*bu-ir ✗	re*dis*tri*bu*ir ✓
reintrodução	re-in*tro*du*ção ✗	re*in*tro*du*ção ✓
triunfalismo	tri-un*fa*lis*mo ✗	tri*un*fa*lis*mo ✓
zodiaco	zo*dí-a*co ✗	zo*dí*a*co ✓
aminoácido	a*mi*no-á*ci*do ✗	a*mi*no*á*ci*do ✓
arruinar	ar*ru-i*nar ✗	ar*ru*i*nar ✓
ateísmo	a*te-ís*mo ✗	a*te*ís*mo ✓
caim	ca-im ✗	ca-im ✗
exequível	e*xe-quí*vel ✗	e*xe*quí*vel ✓
inconsequência	in*con*se-quên*ci*a ✗	in*con*se*quên*ci*a ✓
laquê	la-quê ✗	la*quê ✓
megalomaniaco	me*ga*lo*ma-ní-a*co ✗	me*ga*lo*ma*ní*a*co ✓
pantaleão	pan*ta*le-ão ✗	pan*ta*le*ão ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
poluidor	po*lu-i*dor ✗	po*lu*i*dor ✓
preâmbulo	pre-âm*bu*lo ✗	pre*âm*bu*lo ✓
arcaísmo	ar*ca-ís*mo ✗	ar*ca*ís*mo ✓
auréola	au*ré-o*la ✗	au*ré*o*la ✓
autodestruição	au*to*des*tru-i*ção ✗	au*to*des*tru*i*ção ✓
casuística	ca*su-ís*ti*ca ✗	ca*su*ís*ti*ca ✓
desunião	de*su*ni-ão ✗	de*su*ni*ão ✓
elogiável	e*lo*gi-á*vel ✗	e*lo*gi*á*vel ✓
gnomo	g.no*mo ✗	gno*mo ✓
neurocirurgia	neu*ro*ci*rur*gi-ão ✗	neu*ro*ci*rur*gi*ão ✓
pediátrico	pe*di-á*tri*co ✗	pe*di*á*tri*co ✓
permeável	per*me-á*vel ✗	per*me*á*vel ✓
pirai	pi*ra-í ✗	pi*ra*í ✓
psicodélico	p.si*co*dé*li*co ✗	psi*co*dé*li*co ✓
psicotécnico	p.si*co*téc*ni*co ✗	psi*co*téc*ni*co ✓
saíra	sa-í*ra ✗	sa*í*ra ✓
sobressair	so*bres*sa-ir ✗	so*bres*sa*ir ✓
subliminar	su.b-li*mi*na-r ✗	sub*li*mi*na-r ✓
ajuizado	a*ju-i*za*do ✗	a*ju*i*za*do ✓
antipoluição	an*ti*po*lu-i*ção ✗	an*ti*po*lu*i*ção ✓
arquétipo	ar-qué*ti*po ✗	ar*qué*ti*po ✓
dadaísmo	da*da-ís*mo ✗	da*da*ís*mo ✓
descontrair	des*con*tra-ir ✗	des*con*tra*ir ✓
gasômetro	ga-sô*me*tro ✗	ga*sô*me*tro ✓
hidroviário	hi*dro*vi-á*ri*o ✗	hi*dro*vi*á*ri*o ✓
inegociável	i*ne*go*ci-á*vel ✗	i*ne*go*ci*á*vel ✓
maiúscula	mai-ús*cu*la ✗	mai*ús*cu*la ✓
miúdas	mi-ú*das ✗	mi*ú*das ✓
perequê	pe*re-quê ✗	pe*re*quê ✓
reconcavo	re-côn*ca*vo ✗	re*côn*ca*vo ✓
saquê	sa-quê ✗	sa*quê ✓
ancião	an*ci-ão ✗	an*ci*ão ✓
bafômetro	ba-fô*me*tro ✗	ba*fô*me*tro ✓
bongô	bon-gô ✗	bon*gô ✓
ciberespaço	ci*be.r-es*pa*ço ✗	ci*be.r-es*pa*ço ✗
condômino	con-dô*mi*no ✗	con*dô*mi*no ✓
dadaísta	da*da-ís*ta ✗	da*da*ís*ta ✓
freático	fre-á*ti*co ✗	fre*á*ti*co ✓
hipocondríaco	hi*po*con*drí-a*co ✗	hi*po*con*drí*a*co ✓
indaíá	in*dai-á ✗	in*dai*á ✓
jacuí	ja*cu-í ✗	ja*cu*í ✓
maleável	ma*le-á*vel ✗	ma*le*á*vel ✓
ministeriável	mi*nis*te*ri-á*vel ✗	mi*nis*te*ri*á*vel ✓
obituário	o*bi*tu-á*ri*o ✗	o*bi*tu*á*ri*o ✓
picuinha	pi*cu-i*nha ✗	pi*cu-i*nha ✗
platô	pla-tô ✗	pla*tô ✓
possuidor	pos*su-i*dor ✗	pos*su*i*dor ✓
psicodelismo	p.si*co*de*lis*mo ✗	psi*co*de*lis*mo ✓
reinserção	re-in*ser*ção ✗	re*in*ser*ção ✓
suingue	su-in*gue ✗	su-in*gue ✗
triunfalista	tri-un*fa*lis*ta ✗	tri*un*fa*lis*ta ✓
variola	va*rí-o*la ✗	va*rí*o*la ✓
apoteótico	a*po*te-ó*ti*co ✗	a*po*te*ó*ti*co ✓
bainha	ba-i*nha ✗	ba*i*nha ✓
bibliófilo	bi*bli-ó*fi*lo ✗	bi*bli*ó*fi*lo ✓
caimento	ca-i*men*to ✗	ca-i*men*to ✗
cambuí	cam*bu-í ✗	cam*bu*í ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
cauim	cau-im ✗	cau-im ✗
darwinismo	dar-wi*nis*mo ✗	dar-wi*nis*mo ✗
eclesiástico	e*cle*si-ás*ti*co ✗	e*cle*si*ás*ti*co ✓
estuário	es*tu-á*ri*o ✗	es*tu*á*ri*o ✓
gigolô	gi*go-lô ✗	gi*go*lô ✓
imbuído	im*bu-i*do ✗	im*bu*í*do ✓
bastião	bas*ti-ão ✗	bas*ti*ão ✓
caíque	ca-i*que ✗	ca*í*que ✓
coincidente	co-in*ci*den*te ✗	co*in*ci*den*te ✓
flúor	flú-or ✗	flú-or ✗
incongruência	in*con*gru-ên*ci*a ✗	in*con*gru*ên*ci*a ✓
indissociável	in*dis*so*ci-á*vel ✗	in*dis*so*ci*á*vel ✓
invariável	in*va*ri-á*vel ✗	in*va*ri*á*vel ✓
mostruário	mos*tru-á*ri*o ✗	mos*tru*á*ri*o ✓
néon	né-on ✗	né-on ✓
oboé	o*bo-é ✗	o*bo*é ✓
oceânico	o*ce-â*ni*co ✗	o*ce*â*ni*co ✓
panteísmo	pan*te-ís*mo ✗	pan*te*ís*mo ✓
platônico	pla*tô*ni*co ✗	pla*tô*ni*co ✓
pneumologia	p.neu*mo*lo*gi*a ✗	pneu*mo*lo*gi*a ✓
proibitivo	pro-i*bi*ti*vo ✗	pro*í*bi*ti*vo ✓
raquítico	ra-quí*ti*co ✗	ra*quí*ti*co ✓
reinterpretação	re-in*ter*pre*ta*ção ✗	re*in*ter*pre*ta*ção ✓
subtraído	sub*tra-i*do ✗	sub*tra*í*do ✓
suinocultura	su-i*no*cul*tu*ra ✗	su*í*no*cul*tu*ra ✓
sumô	su-mô ✗	su*mô ✓
tainha	ta-i*nha ✗	ta*í*nha ✓
acordeão	a*cor*de-ão ✗	a*cor*de*ão ✓
aético	a-é*ti*co ✗	a*é*ti*co ✓
baiuca	bai-u*ca ✗	bai-u*ca ✗
centurião	cen*tu*ri-ão ✗	cen*tu*ri*ão ✓
comediógrafo	co*me*di-ó*gra*fo ✗	co*me*di*ó*gra*fo ✓
conduíte	con*du-i*te ✗	con*du*í*te ✓
coroinha	co*ro-i*nha ✗	co*ro*í*nha ✓
ergonômico	er*go-nô*mi*co ✗	er*go*nô*mi*co ✓
gaélico	ga-é*li*co ✗	ga*é*li*co ✓
grandiloquência	gran*di*lo-quên*ci*a ✗	gran*di*lo*quên*ci*a ✓
impatriótico	im*pa*tri-ó*ti*co ✗	im*pa*tri*ó*ti*co ✓
megalômano	me*ga-lô*ma*no ✗	me*ga*lô*ma*no ✓
miríade	mi*rí-a*de ✗	mi*rí*a*de ✓
neófito	ne-ó*fi*to ✗	ne*ó*fi*to ✓
nipônico	ni-pô*ni*co ✗	ni*pô*ni*co ✓
obstruído	obs*tru-í*do ✗	obs*tru*í*do ✓
pecuniário	pe*cu*ni-á*ri*o ✗	pe*cu*ni*á*ri*o ✓
psicopatologia	p.si*co*pa*to*lo*gi*a ✗	psi*co*pa*to*lo*gi*a ✓
reindexar	re-in*de*xar ✗	re-in*de*xar ✗
reinstalar	re-ins*ta*lar ✗	re*ins*ta*lar ✓
reintroduzir	re-in*tro*du*zir ✗	re*in*tro*du*zir ✓
teleinformática	te*le-in*for*má*ti*ca ✗	te*le*in*for*má*ti*ca ✓
timbaúba	tim*ba-ú*ba ✗	tim*ba*ú*ba ✓
agônico	a*gô*ni*co ✗	a*gô*ni*co ✓
anatômico	a*na-tô*mi*co ✗	a*na*tô*mi*co ✓
anômalo	a*nô*ma*lo ✗	a*nô*ma*lo ✓
ataúde	a*ta-ú*de ✗	a*ta*ú*de ✓
auê	au-ê ✗	au*ê ✓
bestiário	bes*ti-á*ri*o ✗	bes*ti*á*ri*o ✓
boá	bo-á ✗	bo*á ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
copião	co*pi-ão ✗	co*pi*ão ✓
ensaística	en*sa-ís*ti*ca ✗	en*sa*ís*ti*ca ✓
fonaudiólogo	fo*no*au*di-ó*lo*go ✗	fo*no*au*di*ó*lo*go ✓
imiscuir	i*mis*cu-ir ✗	i*mis*cu-ir ✗
jupιά	ju*pi-ά ✗	ju*pi*ά ✓
monoteísmo	mo*no*te-ís*mo ✗	mo*no*te*ís*mo ✓
nagô	na-gô ✗	na*gô ✓
psicofarmacologia	p.si*co*far*ma*co*lo*gi*a ✗	psi*co*far*ma*co*lo*gi*a ✓
psicologismo	p.si*co*lo*gis*mo ✗	psi*co*lo*gis*mo ✓
ruindade	ru-in*da*de ✗	ru-in*da*de ✗
sanduícheria	san*du-i*che*ri*a ✗	san*du*i*che*ri*a ✓
sociável	so*ci-á*vel ✗	so*ci*á*vel ✓
sublevação	su.b-le*va*ção ✗	sub*le*va*ção ✓
abstrair	abs*tra-ir ✗	abs*tra*ir ✓
alaúde	a*la-ú*de ✗	a*la*ú*de ✓
almôndega	al*môn*de*ga ✗	al*môn*de*ga ✓
antiaéreo	an*ti*a-é*re*o ✗	an*ti*a*é*re*o ✓
atrabiliário	a*tra*bi*li-á*ri*o ✗	a*tra*bi*li*á*ri*o ✓
bangalô	ban*ga-lô ✗	ban*ga*lô ✓
bôer	bô-er ✗	bô-er ✗
breviário	bre*vi-á*ri*o ✗	bre*vi*á*ri*o ✓
canônico	ca-nô*ni*co ✗	ca*nô*ni*co ✓
carnaúba	car*na-ú*ba ✗	car*na*ú*ba ✓
doído	do-í*do ✗	do*í*do ✓
ensaísmo	en*sa-ís*mo ✗	en*sa*ís*mo ✓
esfiha	es*fi-ha ✗	es*fi-ha ✗
graúna	gra-ú*na ✗	gra*ú*na ✓
hemodiálise	he*mo*di-á*li*se ✗	he*mo*di*á*li*se ✓
hinduísmo	hin*du-ís*mo ✗	hin*du*ís*mo ✓
influído	in*flu-í*do ✗	in*flu-í*do ✗
intuir	in*tu-ir ✗	in*tu*ir ✓
latifundiário	la*ti*fun*di-á*ri*o ✗	la*ti*fun*di*á*ri*o ✓
mimeógrafo	mi*me-ó*gra*fo ✗	mi*me*ó*gra*fo ✓
pajeú	pa*je-ú ✗	pa*je*ú ✓
prosaísmo	pro*sa-ís*mo ✗	pro*sa*ís*mo ✓
psicodrama	p.si*co*dra*ma ✗	psi*co*dra*ma ✓
psicopedagogia	p.si*co*pe*da*go*gi*a ✗	psi*co*pe*da*go*gi*a ✓
quinquênio	quin-quê*ni*o ✗	quin*quê*ni*o ✓
radiofrequência	ra*di*o*fre-quên*ci*a ✗	ra*di*o*fre*quên*ci*a ✓
reinstaurar	re-ins*ti*tu-ir ✗	re*ins*ti*tu*ir ✓
supercampeão	su*per*cam*pe-ão ✗	su*per*cam*pe*ão ✓
supersônico	su*per*sô*ni*co ✗	su*per*sô*ni*co ✓
tragicômico	tra*gi-cô*mi*co ✗	tra*gi*cô*mi*co ✓
tríade	trí-a*de ✗	trí*a*de ✓
vacuidade	va*cu-i*da*de ✗	va*cu*i*da*de ✓
agogô	a*go-gô ✗	a*go*gô ✓
altruísta	al*tru-ís*ta ✗	al*tru*ís*ta ✓
antiácido	an*ti-á*ci*do ✗	an*ti*á*ci*do ✓
autobiógrafo	au*to*bi-ó*gra*fo ✗	au*to*bi*ó*gra*fo ✓
beduíno	be*du-í*no ✗	be*du*í*no ✓
bordô	bor-dô ✗	bor*dô ✓
casuístico	ca*su-ís*ti*co ✗	ca*su*ís*ti*co ✓
czarina	c.za*ri*na ✗	cza*ri*na ✓
demoníaco	de*mo*ní-a*co ✗	de*mo*ní*a*co ✓
desconstruir	des*cons*tru-ir ✗	des*cons*tru*ir ✓
ergômetro	er-gô*me*tro ✗	er*gô*me*tro ✓
evoé	e*vo-é ✗	e*vo*é ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
graúdo	gra-ú*do ✗	gra*ú*do ✓
iídiche	i-í*di*che ✗	i*í*di*che ✓
impronunciável	im*pro*nun*ci-á*vel ✗	im*pro*nun*ci*á*vel ✓
influenciável	in*flu*en*ci-á*vel ✗	in*flu*en*ci*á*vel ✓
irrefreável	ir*re*fre-á*vel ✗	ir*re*fre*á*vel ✓
irretorquível	ir*re*tor-quí*vel ✗	ir*re*tor*quí*vel ✓
jataí	ja*ta-í ✗	ja*ta*í ✓
juquiá	ju*qui-á ✗	ju*qui*á ✓
painho	pa-i*nho ✗	pa*i*nho ✓
pindaíba	pin*da-í*ba ✗	pin*da*í*ba ✓
pneumologista	p.neu*mo*lo*gis*ta ✗	pneu*mo*lo*gis*ta ✓
porciúncula	por*ci-ún*cu*la ✗	por*ci*ún*cu*la ✓
prolegômenos	pro*le-gô*me*nos ✗	pro*le*gô*me*nos ✓
psiquismo	p.si*quis*mo ✗	psi*quis*mo ✓
quiproqué	qui*pro-quó ✗	qui*pro*quó ✓
refluir	re*flu-ir ✗	re*flu*ir ✓
reimplantar	re-im*plan*tar ✗	re*im*plan*tar ✓
reurbanizar	re-ur*ba*ni*zar ✗	re-ur*ba*ni*zar ✗
reutilizar	re-u*ti*li*zar ✗	re-u*ti*li*zar ✗
sainte	sa-in*te ✗	sa-in*te ✗
tsunami	t.su*na*mi ✗	tsu*na*mi ✓
usucapião	u*su*ca*pi*ão ✗	u*su*ca*pi*ão ✓
zebuíno	ze*bu-í*no ✗	ze*bu*í*no ✓
aeroporto	a*e*ro*por*tu-á*ri*o ✗	a*e*ro*por*tu*á*ri*o ✓
agroquímica	a*gro-quí*mi*ca ✗	a*gro*quí*mi*ca ✓
ajuizar	a*ju-i*zar ✗	a*ju*i*zar ✓
baraúna	ba*ra-ú*na ✗	ba*ra*ú*na ✓
casuísta	ca*su-ís*ta ✗	ca*su*ís*ta ✓
catatônico	ca*ta-tô*ni*co ✗	ca*ta*tô*ni*co ✓
conteudista	con*te-u*dis*ta ✗	con*te-u*dis*ta ✗
continuista	con*ti*nu-ís*ta ✗	con*ti*nu*ís*ta ✓
czarista	c.za*ris*ta ✗	cza*ris*ta ✓
decair	de*ca-ir ✗	de*ca*ir ✓
deísta	de-ís*ta ✗	de*ís*ta ✓
diácono	di-á*co*no ✗	di*á*co*no ✓
elegíaco	e*le*gí-a*co ✗	e*le*gí*a*co ✓
faraônico	fa*ra-ô*ni*co ✗	fa*ra*ô*ni*co ✓
feérico	fe-é*ri*co ✗	fe*é*ri*co ✓
hidroavião	hi*dro*a*vi-ão ✗	hi*dro*a*vi*ão ✓
hodômetro	ho*dô*me*tro ✗	ho*dô*me*tro ✓
inconciliável	in*con*ci*li-á*vel ✗	in*con*ci*li*á*vel ✓
ioiô	i.o-i.ô ✗	i.o-i.ô ✗
macrobiótica	ma*cro*bi-ó*ti*ca ✗	ma*cro*bi*ó*ti*ca ✓
marruá	mar*ru-á ✗	mar*ru*á ✓
meteórico	me*te-ó*ri*co ✗	me*te*ó*ri*co ✓
panteísta	pan*te-ís*ta ✗	pan*te*is*ta ✓
papaína	pa*pa-í*na ✗	pa*pa*í*na ✓
patuá	pa*tu-á ✗	pa*tu*á ✓
perdoável	per*do-á*vel ✗	per*do*á*vel ✓
prostituído	pros*ti*tu-í*do ✗	pros*ti*tu*í*do ✓
pulôver	pu-lô*ver ✗	pu*lô*ver ✓
ravióli	ra*vi-ó*li ✗	ra*vi*ó*li ✓
reincorporação	re-in*cor*po*ra*ção ✗	re*in*cor*po*ra*ção ✓
reingresso	re-in*gres*so ✗	re*in*gres*so ✓
reinqüirir	re-in*qui*rir ✗	re-in*qui*rir ✗
reinstalação	re-ins*ta*la*ção ✗	re*ins*ta*la*ção ✓
reinterpretar	re-in*ter*pre*tar ✗	re*in*ter*pre*tar ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
saião	sai-ão ✗	sai*ão ✓
samaúma	sa*ma-ú*ma ✗	sa*ma*ú*ma ✓
sublocação	su.b-lo*ca*ção ✗	sub*lo*ca*ção ✓
turboélice	tur*bo-é*li*ce ✗	tur*bo*é*li*ce ✓
ventoinha	ven*to-i*nha ✗	ven*to*i*nha ✓
voyeurista	voy-eu*ris*ta ✗	voy-eu*ris*ta ✗
alagoinha	a*la*go-i*nha ✗	a*la*go*i*nha ✓
aluvião	a*lu*vi-ão ✗	a*lu*vi*ão ✓
amiúde	a*mi-ú*de ✗	a*mi*ú*de ✓
amoníaco	a*mo*ní-a*co ✗	a*mo*ní*a*co ✓
antagônico	an*ta-gô*ni*co ✗	an*ta*gô*ni*co ✓
antiquíssimo	an*ti-quís*si*mo ✗	an*ti*quís*si*mo ✓
babuíno	ba*bu-i*no ✗	ba*bu*i*no ✓
bahamense	ba-ha*men*se ✗	ba-ha*men*se ✗
barômetro	ba-rô*me*tro ✗	ba*rô*me*tro ✓
bibelô	bi*be-lô ✗	bi*be*lô ✓
biótipo	bi-ó*ti*po ✗	bi*ó*ti*po ✓
bisavô	bi*sa-vô ✗	bi*sa*vô ✓
bocaiuva	bo*cai-u*va ✗	bo*cai-u*va ✗
borderô	bor*de-rô ✗	bor*de*rô ✓
brechtiano	bre.ch*ti*a*no ✗	bre.ch*ti*a*no ✗
cariótipo	ca*ri-ó*ti*po ✗	ca*ri*ó*ti*po ✓
chororô	cho*ro-rô ✗	cho*ro*rô ✓
copaíba	co*pa-i*ba ✗	co*pa*i*ba ✓
czarismo	c.za*ris*mo ✗	cza*ris*mo ✓
darwinista	dar-wi*nis*ta ✗	dar-wi*nis*ta ✗
decaída	de*ca-i*da ✗	de*ca*i*da ✓
diálise	di-á*li*se ✗	di*á*li*se ✓
diástole	di-ás*to*le ✗	di*ás*to*le ✓
esquálido	es-quá*li*do ✗	es*quá*li*do ✓
esvair	es*va-ir ✗	es*va*ir ✓
farisaísmo	fa*ri*sa-ís*mo ✗	fa*ri*sa*ís*mo ✓
fiduciário	fi*du*ci-á*ri*o ✗	fi*du*ci*á*ri*o ✓
filarmônico	fi*lar-mô*ni*co ✗	fi*lar*mô*ni*co ✓
financiável	fi*nan-ci-á*vel ✗	fi*nan*ci*á*vel ✓
fruir	fru-ir ✗	fru*ir ✓
fuzuê	fu*zu-ê ✗	fu*zu*ê ✓
geômetra	ge-ô*me*tra ✗	ge*ô*me*tra ✓
gnose	g.no*se ✗	gno*se ✓
jabô	ja-bô ✗	ja*bô ✓
janaúba	ja*na-ú*ba ✗	ja*na*ú*ba ✓
lusiada	lu*sí-a*da ✗	lu*sí*a*da ✓
maçônico	ma-çô*ni*co ✗	ma*çô*ni*co ✓
macrobiótico	ma*cro*bi-ó*ti*co ✗	ma*cro*bi*ó*ti*co ✓
mafuá	ma*fu-á ✗	ma*fu*á ✓
manômetro	ma-nô*me*tro ✗	ma*nô*me*tro ✓
mediúnico	me*di-ú*ni*co ✗	me*di*ú*ni*co ✓
microrregião	mi*cror*re*gi-ão ✗	mi*cror*re*gi*ão ✓
neoplatônico	ne*o*pla-tô*ni*co ✗	ne*o*pla*tô*ni*co ✓
patchuli	pat*chu*li ✓	pat-chu*li ✗
pecuário	pe*cu-á*ri*o ✗	pe*cu*á*ri*o ✓
psicolinguista	p.si*co*lin*guis*ta ✗	psi*co*lin*guis*ta ✓
psicolinguística	p.si*co*lin*guís*ti*ca ✗	psi*co*lin*guís*ti*ca ✓
psicomotor	p.si*co*mo*tor ✗	psi*co*mo*tor ✓
psicomotricista	p.si*co*mo*tri*cis*ta ✗	psi*co*mo*tri*cis*ta ✓
psicopatía	p.si*co*pa*ti*a ✗	psi*co*pa*ti*a ✓
psicopedagogo	p.si*co*pe*da*go*go ✗	psi*co*pe*da*go*go ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
psicossomática	p.si*cos*so*má*ti*ca ✗	psi*cos*so*má*ti*ca ✓
reconstituente	re*cons*ti*tu-in*te ✗	re*cons*ti*tu-in*te ✗
reincidir	re-in*ci*dir ✗	re*in*ci*dir ✓
saruê	sa*ru-ê ✗	sa*ru*ê ✓
subaquático	su*ba-quá*ti*co ✗	su*ba*quá*ti*co ✓
subliteratura	su.b-li*te*ra*tu*ra ✗	sub*li*te*ra*tu*ra ✓
subsidiário	sub*si*di-á*ri*o ✗	sub*si*di*á*ri*o ✓
substituível	subs*ti*tu-í*vel ✗	subs*ti*tu*í*vel ✓
teutônico	teu-tô*ni*co ✗	teu*tô*ni*co ✓
tmese	t.me*se ✗	tme*se ✓
toboágua	to*bo-á*gua ✗	to*bo*á*gua ✓
aldeão	al*de-ão ✗	al*de*ão ✓
amniocentese	am*ni*o*cen*te*se ✓	a.m-ni*o*cen*te*se ✗
anamnese	a*nam*ne*se ✓	a*na.m-ne*se ✗
anauê	a*nau-ê ✗	a*nau*ê ✓
antipatriótico	an*ti*pa*tri-ó*ti*co ✗	an*ti*pa*tri*ó*ti*co ✓
antônimo	an-tô*ni*mo ✗	an*tô*ni*mo ✓
antropônimo	an*tro-pô*ni*mo ✗	an*tro*pô*ni*mo ✓
aviário	a*vi-á*ri*o ✗	a*vi*á*ri*o ✓
babalaô	ba*ba*la-ô ✗	ba*ba*la*ô ✓
braúna	bra-ú*na ✗	bra*ú*na ✓
corruíra	cor*ru-í*ra ✗	cor*ru*í*ra ✓
corrupião	cor*ru*pi-ão ✗	cor*ru*pi*ão ✓
decaído	de*ca-í*do ✗	de*ca*i*do ✓
deificação	de-i*fi*ca*ção ✗	de-i*fi*ca*ção ✗
demiurgo	de*mi-ur*go ✗	de*mi*ur*go ✓
descair	des*ca-ir ✗	des*ca*ir ✓
desembainhar	de*sem*ba-i*nhar ✗	de*sem*ba*i*nhar ✓
diáfano	di-á*fa*no ✗	di*á*fa*no ✓
dicotômico	di*co-tô*mi*co ✗	di*co*tô*mi*co ✓
dodecafônico	do*de*ca-fô*ni*co ✗	do*de*ca*fô*ni*co ✓
ensaístico	en*sa-ís*ti*co ✗	en*sa*ís*ti*co ✓
entusiástico	en*tu*si-ás*ti*co ✗	en*tu*si*ás*ti*co ✓
epônimo	e-pô*ni*mo ✗	e*pô*ni*mo ✓
espectrofotômetro	es*pec*tro*fo-tô*me*tro ✗	es*pec*tro*fo*tô*me*tro ✓
estatuária	es*ta*tu-á*ri*a ✗	es*ta*tu*á*ri*a ✓
faim	fa-im ✗	fa-im ✗
fitoquímica	fi*to-quí*mi*ca ✗	fi*to*quí*mi*ca ✓
formaldeído	for*mal*de-í*do ✗	for*mal*de*i*do ✓
fuinha	fu-i*nha ✗	fu*i*nha ✓
gabião	ga*bi-ão ✗	ga*bi*ão ✓
gauchada	ga-u*cha*da ✗	ga-u*cha*da ✗
gauchismo	ga-u*chis*mo ✗	ga-u*chis*mo ✗
geriátrico	ge*ri-á*tri*co ✗	ge*ri*á*tri*co ✓
gipsita	gip*si*ta ✓	gip-si*ta ✗
guião	gui-ão ✗	gui*ão ✓
ilíaco	i*lí-a*co ✗	i*lí*a*co ✓
iniciático	i*ni*ci-á*ti*co ✗	i*ni*ci*á*ti*co ✓
itapuá	i*ta*pu-á ✗	i*ta*pu*á ✓
japuí	ja*pu-í ✗	ja*pu*í ✓
jesuítico	je*su-í*ti*co ✗	je*su*í*ti*co ✓
lamaísmo	la*ma-ís*mo ✗	la*ma*ís*mo ✓
mantô	man-tô ✗	man*tô ✓
metapsicologia	me*tap*si*co*lo*gi*a ✓	me*tap-si*co*lo*gi*a ✗
mitocôndria	mi*to-côn*dri*a ✗	mi*to*côn*dri*a ✓
miudeza	mi-u*de*za ✗	mi-u*de*za ✗
monomaníaco	mo*no*ma*ní-a*co ✗	mo*no*ma*ní*a*co ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
museólogo	mu*se-ó*lo*go ✗	mu*se*ó*lo*go ✓
ninfomaníaca	nin*fo*ma*ní-a*ca ✗	nin*fo*ma*ní*a*ca ✓
nobiliárquico	no*bi*li-ár*qui*co ✗	no*bi*li*ár*qui*co ✓
odômetro	o-dô*me*tro ✗	o*dô*me*tro ✓
orgiástico	or*gi-ás*ti*co ✗	or*gi*ás*ti*co ✓
painço	pa-in*ço ✗	pa-in*ço ✗
pantagruélico	pan*ta*gru-é*li*co ✗	pan*ta*gru*é*li*co ✓
parquímetro	par-quí*me*tro ✗	par*quí*me*tro ✓
pixaim	pi*xa-im ✗	pi*xa-im ✗
pneumática	p.neu*má*ti*ca ✗	pneu*má*ti*ca ✓
pneumático	p.neu*má*ti*co ✗	pneu*má*ti*co ✓
preá	pre-á ✗	pre*á ✓
priápico	pri-á*pi*co ✗	pri*á*pi*co ✓
psicopedagógico	p.si*co*pe*da*gó*gi*co ✗	psi*co*pe*da*gó*gi*co ✓
psicossomático	p.si*cos*so*má*ti*co ✗	psi*cos*so*má*ti*co ✓
psicotrópico	p.si*co*tró*pi*co ✗	psi*co*tró*pi*co ✓
pterossauro	p.te*ros*sau*ro ✗	pte*ros*sau*ro ✓
quilohertz	qui*lo-hertz ✗	qui*lo-hertz ✗
recondito	re-côn*di*to ✗	re*côn*di*to ✓
reimplantação	re-im*plan*ta*ção ✗	re*im*plan*ta*ção ✓
reincorporar	re-in*cor*po*rar ✗	re*in*cor*po*rar ✓
reingressar	re-in*gres*sar ✗	re*in*gres*sar ✓
reinvestir	re-in*ves*tir ✗	re*in*ves*tir ✓
restituível	res*ti*tu-í*vel ✗	res*ti*tu*í*vel ✓
retrair	re*tra-ir ✗	re*tra*ir ✓
rodamoinho	ro*da*mo-i*nho ✗	ro*da*mo*i*nho ✓
roído	ro-í*do ✗	ro*í*do ✓
ruão	ru-ão ✗	ru*ão ✓
salomônico	sa*lo-mô*ni*co ✗	sa*lo*mô*ni*co ✓
sardônico	sar-dô*ni*co ✗	sar*dô*ni*co ✓
semiólogo	se*mi-ó*lo*go ✗	se*mi*ó*lo*go ✓
solipsismo	so*lip*sis*mo ✓	so*lip*sis*mo ✗
sumaúma	su*ma-ú*ma ✗	su*ma*ú*ma ✓
tataravô	ta*ta*ra-vô ✗	ta*ta*ra*vô ✓
termoquímica	ter*mo-quí*mi*ca ✗	ter*mo*quí*mi*ca ✓
topônimo	to-pô*ni*mo ✗	to*pô*ni*mo ✓
toxicômano	to*xi-cô*ma*no ✗	to*xi*cô*ma*no ✓
trisavô	tri*sa-vô ✗	tri*sa*vô ✓
triunvirato	tri-un*vi*ra*to ✗	tri*un*vi*ra*to ✓
voyeurístico	voy-eu*rís*ti*co ✗	voy-eu*rís*ti*co ✗
xintoísmo	xin*to-ís*mo ✗	xin*to*ís*mo ✓
absenteísta	ab*sen*te-ís*ta ✗	ab*sen*te*ís*ta ✓
aeroviário	a*e*ro*vi-á*ri*o ✗	a*e*ro*vi*á*ri*o ✓
afluir	a*flu-ir ✗	a*flu*ir ✓
agroquímico	a*gro-quí*mi*co ✗	a*gro*quí*mi*co ✓
aguaí	a*gua-í ✗	a*gua*í ✓
alquímico	al-quí*mi*co ✗	al*quí*mi*co ✓
amnesia	am*ne*si*a ✓	a.m-ne*si*a ✗
amnésico	am*né*si*co ✓	a.m-né*si*co ✗
anafrodisíaco	a*na*fro*di*sí-a*co ✗	a*na*fro*di*si*a*co ✓
antialcoólico	an*ti*al*co-ó*li*co ✗	an*ti*al*co*ó*li*co ✓
antinômico	an*ti-nô*mi*co ✗	an*ti*nô*mi*co ✓
anuir	a*nu-ir ✗	a*nu*ir ✓
arapuá	a*ra*pu-á ✗	a*ra*pu*á ✓
aríete	a*rí-e*te ✗	a*rí-e*te ✗
associável	as*so*ci-á*vel ✗	as*so*ci*á*vel ✓
atribuível	a*tri*bu-í*vel ✗	a*tri*bu*í*vel ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
babilônico	ba*bi-lô*ni*co ✗	ba*bi*lô*ni*co ✓
balaustrada	ba*la-us*tra*da ✗	ba*la-us*tra*da ✗
beócio	be-ó*ci*o ✗	be*ó*ci*o ✓
biófilo	bi-ó*fi*lo ✗	bi*ó*fi*lo ✓
biunívoco	bi-u*ní*vo*co ✗	bi*u*ní*vo*co ✓
boião	boi-ão ✗	boi*ão ✓
boiuna	boi-u*na ✗	boi*u*na ✓
butiá	bu*ti-á ✗	bu*ti*á ✓
caingangue	ca-in*gan*gue ✗	ca-in*gan*gue ✗
carnaíba	car*na-í*ba ✗	car*na*í*ba ✓
caúna	ca-ú*na ✗	ca*ú*na ✓
ciática	ci-á*ti*ca ✗	ci*á*ti*ca ✓
clepsidra	clep*si*dra ✓	clep-si*dra ✗
codeína	co*de-í*na ✗	co*de*í*na ✓
conciliábulo	con*ci*li-á*bu*lo ✗	con*ci*li*á*bu*lo ✓
conciliável	con*ci*li-á*vel ✗	con*ci*li*á*vel ✓
condóido	con*do-í*do ✗	con*do*í*do ✓
confluir	con*flu-ir ✗	con*flu*ir ✓
congruência	con*gru-ên*ci*a ✗	con*gru*ên*ci*a ✓
crapô	cra-pô ✗	cra*pô ✓
criptônimo	crip-tô*ni*mo ✗	crip*tô*ni*mo ✓
deão	de-ão ✗	de*ão ✓
decacampeão	de*ca*cam*pe-ão ✗	de*ca*cam*pe*ão ✓
demiúrgico	de*mi-úr*gi*co ✗	de*mi*úr*gi*co ✓
depreciável	de*pre*ci-á*vel ✗	de*pre*ci*á*vel ✓
descaída	des*ca-í*da ✗	des*ca*í*da ✓
dzeta	d.ze*ta ✗	d.ze*ta ✗
elefantíase	e*le*fan*tí-a*se ✗	e*le*fan*tí*a*se ✓
eletroímã	e*le*tro-í*mã ✗	e*le*tro*í*mã ✓
eólico	e-ó*li*co ✗	e*ó*li*co ✓
eólio	e-ó*li*o ✗	e*ó*li*o ✓
epifenômeno	e*pi*fe-nô*me*no ✗	e*pi*fe*nô*me*no ✓
escarabeídeo	es*ca*ra*be-í*de*o ✗	es*ca*ra*be*í*de*o ✓
espadaúdo	es*pa*da-ú*do ✗	es*pa*da*ú*do ✓
espeleólogo	es*pe*le-ó*lo*go ✗	es*pe*le*ó*lo*go ✓
faiscante	fa-is*can*te ✗	fa-is*can*te ✗
feiura	fei-u*ra ✗	fei*u*ra ✓
fotômetro	fo-tô*me*tro ✗	fo*tô*me*tro ✓
franqueável	fran*que-á*vel ✗	fran*que*á*vel ✓
gastrointestinal	gas*tro-in*tes*ti*nal ✗	gas*tro*in*tes*ti*nal ✓
gauchar	ga-u*char ✗	ga-u*char ✗
gnosticismo	g.nos*ti*cis*mo ✗	gnos*ti*cis*mo ✓
gnóstico	g.nós*ti*co ✗	gnós*ti*co ✓
gnu	g.nu ✗	g.nu ✗
grafomaniaco	gra*fo*ma*ní-a*co ✗	gra*fo*ma*ní*a*co ✓
grafômano	gra-fô*ma*no ✗	gra*fô*ma*no ✓
grapiúna	gra*pi-ú*na ✗	gra*pi*ú*na ✓
heterônimo	he*te-rô*ni*mo ✗	he*te*rô*ni*mo ✓
hexacampeão	he*xa*cam*pe-ão ✗	he*xa*cam*pe*ão ✓
hinduísta	hin*du-ís*ta ✗	hin*du*ís*ta ✓
holonômico	ho*lo-nô*mi*co ✗	ho*lo*nô*mi*co ✓
homofônico	ho*mo-fô*ni*co ✗	ho*mo*fô*ni*co ✓
ibijaú	i*bi*ja-ú ✗	i*bi*ja*ú ✓
idiótico	i*di-ó*ti*co ✗	i*di*ó*ti*co ✓
imbuir	im*bu-ir ✗	im*bu*ir ✓
inclinômetro	in*cli-nô*me*tro ✗	in*cli*nô*me*tro ✓
irreconciliável	ir*re*con*ci*li-á*vel ✗	ir*re*con*ci*li*á*vel ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
irreligião	ir*re*li*gi-ão ✗	ir*re*li*gi*ão ✓
irrenunciável	ir*re*nun*ci-á*vel ✗	ir*re*nun*ci*á*vel ✓
isquêmico	is-quê*mi*co ✗	is*quê*mi*co ✓
itapiúna	i*ta*pi-ú*na ✗	i*ta*pi*ú*na ✓
juá	ju-á ✗	ju*á ✓
lamaísta	la*ma-is*ta ✗	la*ma*is*ta ✓
luteína	lu*te-í*na ✗	lu*te*í*na ✓
macrorregião	ma*cror*re*gi-ão ✗	ma*cror*re*gi*ão ✓
mandrião	man*dri-ão ✗	man*dri*ão ✓
masdeísmo	mas*de-is*mo ✗	mas*de*is*mo ✓
maué	mau-é ✗	mau*é ✓
mediático	me*di-á*ti*co ✗	me*di*á*ti*co ✓
melômano	me-lô*ma*no ✗	me*lô*ma*no ✓
mnemônica	m.ne-mô*ni*ca ✗	mne*mô*ni*ca ✓
moinha	mo-i*nha ✗	mo*i*nha ✓
monômero	mo-nô*me*ro ✗	mo*nô*me*ro ✓
monoquíni	mo*no-quí*ni ✗	mo*no*quí*ni ✓
monoteísta	mo*no-te-is*ta ✗	mo*no*te*is*ta ✓
mortuárias	mor*tu-á*ri*as ✗	mor*tu*á*ri*as ✓
mucum	mu*cu-im ✗	mu*cu-im ✗
negligenciável	ne*gli*gen*ci-á*vel ✗	ne*gli*gen*ci*á*vel ✓
neuropsiquiatria	neu*rop*si*qui*a*tri*a ✓	neu*rop-si*qui*a*tri*a ✗
oboísta	o*bo-is*ta ✗	o*bo*is*ta ✓
ofidiário	o*fi*di-á*ri*o ✗	o*fi*di*á*ri*o ✓
oleína	o*le-í*na ✗	o*le*í*na ✓
opiáceo	o*pi-á*ce*o ✗	o*pi*á*ce*o ✓
pancreático	pan*cre-á*ti*co ✗	pan*cre*á*ti*co ✓
panô	pa-nô ✗	pa*nô ✓
parquê	par-quê ✗	par*quê ✓
parvoíce	par*vo-í*ce ✗	par*vo*í*ce ✓
patenteável	pa*ten*te-á*vel ✗	pa*ten*te*á*vel ✓
peã	pe-ã ✗	pe*ã ✓
periósteo	pe*ri-ós*te*o ✗	pe*ri*ós*te*o ✓
picuá	pi*cu-á ✗	pi*cu*á ✓
piquê	pi-quê ✗	pi*quê ✓
piromaníaco	pi*ro*ma*ní-a*co ✗	pi*ro*ma*ní*a*co ✓
piruá	pi*ru-á ✗	pi*ru*á ✓
piúria	pi-ú*ri*a ✗	pi*ú*ri*a ✓
plagiário	pla*gi-á*ri*o ✗	pla*gi*á*ri*o ✓
plebeísmo	ple*be-is*mo ✗	ple*be*is*mo ✓
plenipotenciário	ple*ni*po*ten*ci-á*ri*o ✗	ple*ni*po*ten*ci*á*ri*o ✓
pneumatologia	p.neu*ma*to*lo*gi*a ✗	pneu*ma*to*lo*gi*a ✓
pneumônico	p.neu-mô*ni*co ✗	pneu*mô*ni*co ✓
polifônico	po*li-fô*ni*co ✗	po*li*fô*ni*co ✓
politeísmo	po*li*te-is*mo ✗	po*li*te*is*mo ✓
politeísta	po*li*te-is*ta ✗	po*li*te*is*ta ✓
polítiquês	po*li*ti-quês ✗	po*li*ti*quês ✓
procaína	pro*ca-í*na ✗	pro*ca*í*na ✓
pronunciável	pro*nun*ci-á*vel ✗	pro*nun*ci*á*vel ✓
pseudociência	p.seu*do*ci*ên*ci*a ✗	pseu*do*ci*ên*ci*a ✓
psicofísico	p.si*co*fi*si*co ✗	psi*co*fi*si*co ✓
psicologizar	p.si*co*lo*gi*zar ✗	psi*co*lo*gi*zar ✓
psicometria	p.si*co*me*tri*a ✗	psi*co*me*tri*a ✓
psicomotricidade	p.si*co*mo*tri*ci*da*de ✗	psi*co*mo*tri*ci*da*de ✓
psicopatológico	p.si*co*pa*to*ló*gi*co ✗	psi*co*pa*to*ló*gi*co ✓
psicoprofilaxia	p.si*co*pro*fi*la*xi*a ✗	psi*co*pro*fi*la*xi*a ✓
psicossexual	p.si*cos*se*xu*al ✗	psi*cos*se*xu*al ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
psicoterapêutico	p.si*co*te*ra*pêu*ti*co ✗	psi*co*te*ra*pêu*ti*co ✓
psiquê	p.si-quê ✗	psi*quê ✓
psitacismo	p.si*ta*cis*mo ✗	psi*ta*cis*mo ✓
psoríase	p.so*rí-a*se ✗	p.so*rí*a*se ✗
ptose	p.to*se ✗	pto*se ✓
puído	pu-í*do ✗	pu*í*do ✓
putrião	pu*tri-ão ✗	pu*tri*ão ✓
questiúncula	ques*ti-ún*cu*la ✗	ques*ti*ún*cu*la ✓
reimportar	re-im*por*tar ✗	re*im*por*tar ✓
reimpressão	re-im*pres*são ✗	re*im*pres*são ✓
reinscrever	re-ins*cre*ver ✗	re*ins*cre*ver ✓
reinsserir	re-in*se*rir ✗	re*in*se*rir ✓
remediável	re*me*di-á*vel ✗	re*me*di*á*vel ✓
rubiácea	ru*bi-á*ce*a ✗	ru*bi*á*ce*a ✓
ruinoso	ru-i*no*so ✗	ru*i*no*so ✓
sainha	sa-i*nha ✗	sa*i*nha ✓
sardônica	sar-dô*ni*ca ✗	sar*dô*ni*ca ✓
satiríase	sa*ti*rí-a*se ✗	sa*ti*rí*a*se ✓
setentrião	se*ten*tri-ão ✗	se*ten*tri*ão ✓
sievert	si.e*vert ✗	si.e*vert ✗
sismômetro	sis-mô*me*tro ✗	sis*mô*me*tro ✓
solipsista	so*lip*sis*ta ✓	so*lip-sis*ta ✗
suaíle	su*a-í*le ✗	su*a*í*le ✓
sublevado	su.b-le*va*do ✗	sub*le*va*do ✓
sublocar	su.b-lo*car ✗	sub*lo*car ✓
sublocatário	su.b-lo*ca*tá*ri*o ✗	sub*lo*ca*tá*ri*o ✓
subsônico	sub-sô*ni*co ✗	sub*sô*ni*co ✓
sucuriú	su*cu*ri-ú ✗	su*cu*ri*ú ✓
suingar	su-in*gar ✗	su-in*gar ✗
tacômetro	ta-cô*me*tro ✗	ta*cô*me*tro ✓
taiuiá	tai-ui-á ✗	tai-ui*á ✗
taiwanês	tai-wa*nês ✗	tai-wa*nês ✗
talião	ta*li-ão ✗	ta*li*ão ✓
tchecoslovaco	t.che*cos*lo*va*co ✗	tche*cos*lo*va*co ✓
teísta	te-ís*ta ✗	te*ís*ta ✓
teiú	tei-ú ✗	tei*ú ✓
telômero	te-lô*me*ro ✗	te*lô*me*ro ✓
teníase	te*ní-a*se ✗	te*ní*a*se ✓
terciário	ter*ci-á*ri*o ✗	ter*ci*á*ri*o ✓
torreão	tor*re-ão ✗	tor*re*ão ✓
transeunte	tran*se-un*te ✗	tran*se-un*te ✗
triádico	tri-á*di*co ✗	tri*á*di*co ✓
tripanossomíase	tri*pa*nos*so*mí-a*se ✗	tri*pa*nos*so*mí*a*se ✓
triunfador	tri-un*fa*dor ✗	tri*un*fa*dor ✓
truísmo	tru-ís*mo ✗	tru*ís*mo ✓
tuim	tu-im ✗	tu-im ✗
tuiuiú	tui-ui-ú ✗	tui-ui*ú ✗
tumultuário	tu*mul*tu-á*ri*o ✗	tu*mul*tu*á*ri*o ✓
tzarista	t.za*ris*ta ✗	tza*ris*ta ✓
vietcongue	vi*et*con*gue ✓	vi*et-con*gue ✗
vitriólico	vi*tri-ó*li*co ✗	vi*tri*ó*li*co ✓
vitríolo	vi*trí-o*lo ✗	vi*trí*o*lo ✓
xilocáina	xi*lo*ca-í*na ✗	xi*lo*ca*í*na ✓
abluir	a*blu-ir ✗	a*blu*ir ✓
balaústre	ba*la-ús*tre ✗	ba*la*ús*tre ✓
cabriúva	ca*bri-ú*va ✗	ca*bri*ú*va ✓
daguerreótipo	da*guer*re-ó*ti*po ✗	da*guer*re*ó*ti*po ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
daltônico	dal-tô*ni*co ✗	dal*tô*ni*co ✓
ecônomo	e-cô*no*mo ✗	e*cô*no*mo ✓
egoístico	e*go-ís*ti*co ✗	e*go*ís*ti*co ✓
eletrocardiógrafo	e*le*tro*car*di-ó*gra*fo ✗	e*le*tro*car*di*ó*gra*fo ✓
eletroquímica	e*le*tro-quí*mi*ca ✗	e*le*tro*quí*mi*ca ✓
eletroquímico	e*le*tro-quí*mi*co ✗	e*le*tro*quí*mi*co ✓
embainhar	em*ba-i*nhar ✗	em*ba*i*nhar ✓
faiscar	fa-is*car ✗	fa-is*car ✗
faúlha	fa-ú*lha ✗	fa*ú*lha ✓
galvanômetro	gal*va-nô*me*tro ✗	gal*va*nô*me*tro ✓
hagiógrafo	ha*gi-ó*gra*fo ✗	ha*gi*ó*gra*fo ✓
hebraísmo	he*bra-ís*mo ✗	he*bra*ís*mo ✓
hebraísta	he*bra-ís*ta ✗	he*bra*ís*ta ✓
hipertônico	hi*per-tô*ni*co ✗	hi*per*tô*ni*co ✓
historiógrafo	his*to*ri-ó*gra*fo ✗	his*to*ri*ó*gra*fo ✓
histrião	his*tri-ão ✗	his*tri*ão ✓
jesuitismo	je*su-i*tis*mo ✗	je*su*i*tis*mo ✓
macedônico	ma*ce-dô*ni*co ✗	ma*ce*dô*ni*co ✓
ninfomaniaco	nin*fo*ma*ní-a*co ✗	nin*fo*ma*ní*a*co ✓
paquímetro	pa-quí*me*tro ✗	pa*quí*me*tro ✓
tabuinha	ta*bu-i*nha ✗	ta*bu*i*nha ✓
taquígrafo	ta-quí*gra*fo ✗	ta*quí*gra*fo ✓
taquímetro	ta-quí*me*tro ✗	ta*quí*me*tro ✓
tectônico	tec-tô*ni*co ✗	tec*tô*ni*co ✓
xintoísta	xin*to-ís*ta ✗	xin*to*ís*ta ✓
abiótica	a*bi-ó*ti*ca ✗	a*bi*ó*ti*ca ✓
abiótico	a*bi-ó*ti*co ✗	a*bi*ó*ti*co ✓
abraâmico	a*bra-â*mi*co ✗	a*bra*â*mi*co ✓
abstraído	abs*tra-í*do ✗	abs*tra*i*do ✓
acauã	a*cau-ã ✗	a*cau*ã ✓
acelerômetro	a*ce*le-rô*me*tro ✗	a*ce*le*rô*me*tro ✓
acetaldeído	a*ce*tal*de-í*do ✗	a*ce*tal*de*i*do ✓
acroleína	a*cro*le-í*na ✗	a*cro*le*i*na ✓
adônix	a-dô*nix ✗	a*dô*nix ✓
ainsa	a-in*sa ✗	a-in*sa ✗
aiquidô	ai*qui-dô ✗	ai*quí*dô ✓
alcione	al*ci-o*ne ✗	al*cí*o*ne ✓
aldeído	al*de-í*do ✗	al*de*i*do ✓
aleúte	a*le-ú*te ✗	a*le*ú*te ✓
aliá	a*li-á ✗	a*li*á ✓
almóada	al*mó-a*da ✗	al*mó*a*da ✗
aloé	a*lo-é ✗	a*lo*é ✓
aloés	a*lo-és ✗	a*lo*és ✓
alua	a*lu-á ✗	a*lu*á ✓
alvéolo	al*vé-o*lo ✗	al*vé*o*lo ✓
amebíase	a*me*bí-a*se ✗	a*me*bí*a*se ✓
âmnio	âm*ni*o ✓	â.m-ni*o ✗
ancilostomíase	an*ci*los*to*mí-a*se ✗	an*ci*los*to*mí*a*se ✓
anemômetro	a*ne-mô*me*tro ✗	a*ne*mô*me*tro ✓
anfão	an*fi-ão ✗	an*fi*ão ✓
anfictião	an*fic*ti-ão ✗	an*fic*ti*ão ✓
anião	a*ni-ão ✗	a*ni*ão ✓
aórtico	a-ór*ti*co ✗	a*ór*ti*co ✓
apuí	a*pu-í ✗	a*pu*í ✓
aquífero	a-quí*fe*ro ✗	a*quí*fe*ro ✓
areão	a*re-ão ✗	a*re*ão ✓
areópago	a*re-ó*pa*go ✗	a*re*ó*pa*go ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
arquidiácono	ar*qui*di-á*co*no ✗	ar*qui*di*á*co*no ✓
arteríola	ar*te*rí-o*la ✗	ar*te*rí*o*la ✓
assiriólogo	as*si*ri-ó*lo*go ✗	as*si*ri*ó*lo*go ✓
astroquímica	as*tro-qui*mi*ca ✗	as*tro*qui*mi*ca ✓
ateísta	a*te-ís*ta ✗	a*te*ís*ta ✓
autoimunidade	au*to-i*mu*ni*da*de ✗	au*to-i*mu*ni*da*de ✗
autônimo	au*tô*ni*mo ✗	au*tô*ni*mo ✓
autônômico	au*to-nô*mi*co ✗	au*to*nô*mi*co ✓
averoísmo	a*ver*ro-ís*mo ✗	a*ver*ro*ís*mo ✓
bacteriófago	bac*te*ri-ó*fa*go ✗	bac*te*ri*ó*fa*go ✓
baleárico	ba*le-á*ri*co ✗	ba*le*á*ri*co ✓
banjoísta	ban*jo-ís*ta ✗	ban*jo*ís*ta ✓
batuíra	ba*tu-í*ra ✗	ba*tu*í*ra ✓
behaviorismo	be-ha*vi*o*ris*mo ✗	be-ha*vi*o*ris*mo ✗
benjoim	ben*jo-im ✗	ben*jo-im ✗
bibliógrafo	bi*bli-ó*gra*fo ✗	bi*bli*ó*gra*fo ✓
bilião	bi*li-ão ✗	bi*li*ão ✓
biótico	bi-ó*ti*co ✗	bi*ó*ti*co ✓
biótopo	bi-ó*to*po ✗	bi*ó*to*po ✓
biscainho	bis*ca-i*nho ✗	bis*ca*i*nho ✓
blastômero	blas*tô*me*ro ✗	blas*tô*me*ro ✓
boé	bo-é ✗	bo-é ✓
bosníaco	bos*ní-a*co ✗	bos*ní*a*co ✓
braquiópode	bra*qui-ó*po*de ✗	bra*qui*ó*po*de ✓
bué	bu-é ✗	bu-é ✓
cafarnaum	ca*far*na-um ✗	ca*far*na-um ✗
cajuína	ca*ju-í*na ✗	ca*ju*í*na ✓
camião	ca*mi-ão ✗	ca*mi*ão ✓
candidíase	can*di*dí-a*se ✗	can*di*dí*a*se ✓
canoísta	ca*no-ís*ta ✗	ca*no*ís*ta ✓
carnaubal	car*na-u*bal ✗	car*na-u*bal ✗
caroá	ca*ro-á ✗	ca*ro*á ✓
caseína	ca*se-í*na ✗	ca*se*í*na ✓
catequético	ca*te-qué*ti*co ✗	ca*te*qué*ti*co ✓
catião	ca*ti-ão ✗	ca*ti*ão ✓
cauã	cau-ã ✗	cau*ã ✓
celíaco	ce*lí-a*co ✗	ce*lí*a*co ✓
cnidário	c.ni*dá*ri*o ✗	c.ni*dá*ri*o ✗
coimbrão	co-im*brão ✗	co*im*brão ✓
coleóptero	co*le-óp*te*ro ✗	co*le*óp*te*ro ✓
contraindicação	con*tra-in*dí*ca*ção ✗	con*tra*in*dí*ca*ção ✓
contribuidor	con*tri*bu-i*dor ✗	con*tri*bu*i*dor ✓
coriáceo	co*ri-á*ce*o ✗	co*ri*á*ce*o ✓
curião	cu*ri-ão ✗	cu*ri*ão ✓
curiúva	cu*ri-ú*va ✗	cu*ri*ú*va ✓
cuxiú	cu*xi-ú ✗	cu*xi*ú ✓
decurião	de*cu*ri-ão ✗	de*cu*ri*ão ✓
deiscente	de-is*cen*te ✗	de-is*cen*te ✗
deísmo	de-ís*mo ✗	de*ís*mo ✓
díade	dí-a*de ✗	dí*a*de ✓
diádico	di-á*di*co ✗	di*á*di*co ✓
diádoco	di-á*do*co ✗	di*á*do*co ✓
diáfise	di-á*fí*se ✗	di*á*fí*se ✓
diatômico	di*a-tô*mi*co ✗	di*a*tô*mi*co ✓
diérese	di-é*re*se ✗	di*é*re*se ✓
diferenciável	di*fe*ren*ci-á*vel ✗	di*fe*ren*ci*á*vel ✓
difluência	di*flu-ên*ci*a ✗	di*flu*ên*ci*a ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
díodo	dí-o*do ✗	dí-o*do ✓
domiciliário	do*mi*ci*li-á*ri*o ✗	do*mi*ci*li*á*ri*o ✓
dríade	drí-a*de ✗	drí*a*de ✓
elastômero	e*las-tô*me*ro ✗	e*las*tô*me*ro ✓
eluição	e*lu-i*ção ✗	e*lu*i*ção ✓
éon	é-on ✗	é-on ✓
epicôndilo	e*pi-côn*di*lo ✗	e*pi*côn*di*lo ✓
épsilon	ép*si*lon ✓	ép-si*lon ✗
equidna	e*quid*na ✓	e*qui.d*na ✗
estatuário	es*ta*tu-á*ri*o ✗	es*ta*tu*á*ri*o ✓
esteárico	es*te-á*ri*co ✗	es*te*á*ri*co ✓
estereoquímica	es*te*re*o-quí*mi*ca ✗	es*te*re*o*quí*mi*ca ✓
etnônimo	et-nô*ni*mo ✗	et*nô*ni*mo ✓
eucariótico	eu*ca*ri-ó*ti*co ✗	eu*ca*ri*ó*ti*co ✓
eurasiático	eu*ra*si-á*ti*co ✗	eu*ra*si*á*ti*co ✓
européismo	eu*ro*pe-ís*mo ✗	eu*ro*pe*ís*mo ✓
uropeísta	eu*ro*pe-ís*ta ✗	eu*ro*pe*ís*ta ✓
evangelário	e*van*ge*li-á*ri*o ✗	e*van*ge*li*á*ri*o ✓
feroês	fe*ro-ês ✗	fe*ro*ês ✓
fiável	fi-á*vel ✗	fi*á*vel ✓
fisiólogo	fi*si-ó*lo*go ✗	fi*si*ó*lo*go ✓
fluídico	flu-í*di*co ✗	flu-i*di*co ✗
foliáceo	fo*li-á*ce*o ✗	fo*li*á*ce*o ✓
folíolo	fo*lí-o*lo ✗	fo*lí*o*lo ✓
fotoperíodo	fo*to*pe*ri-o*do ✗	fo*to*pe*ri*o*do ✓
fotoquímica	fo*to-quí*mi*ca ✗	fo*to*quí*mi*ca ✓
fotoquímico	fo*to-quí*mi*co ✗	fo*to*quí*mi*co ✓
friável	fri-á*vel ✗	fri*á*vel ✓
ftálico	f.tá*li*co ✗	f.tá*li*co ✗
fuá	fu-á ✗	fu*á ✓
genômico	ge-nô*mi*co ✗	ge*nô*mi*co ✓
genuinidade	ge*nu-i*ni*da*de ✗	ge*nu*i*ni*da*de ✓
geoquímica	ge*o-quí*mi*ca ✗	ge*o*quí*mi*ca ✓
geoquímico	ge*o-quí*mi*co ✗	ge*o*quí*mi*co ✓
geotectônica	ge*o*tec-tô*ni*ca ✗	ge*o*tec*tô*ni*ca ✓
gimnodesportivo	gi*m-no*des*por*ti*vo ✓	gi.m-no*des*por*ti*vo ✗
glaciário	gla*ci-á*ri*o ✗	gla*ci*á*ri*o ✓
glicoproteína	gli*co*pro*te-í*na ✗	gli*co*pro*te*í*na ✓
gnaisse	g.nais*se ✗	g.nais*se ✗
goês	go-ês ✗	go*ês ✓
graxaim	gra*xa-im ✗	gra*xa-im ✗
helíaco	he*lí-a*co ✗	he*lí*a*co ✓
henoteísmo	he*no*te-ís*mo ✗	he*no*te*ís*mo ✓
heptacampeão	hep*ta*cam*pe-ão ✗	hep*ta*cam*pe*ão ✓
heteroátomo	he*te*ro-á*to*mo ✗	he*te*ro*á*to*mo ✓
higiênico	hi*gi-é*ni*co ✗	hi*gi*é*ni*co ✓
hióideo	hi-ói*de*o ✗	hi*ói*de*o ✓
hipersônico	hi*per-sô*ni*co ✗	hi*per*sô*ni*co ✓
hipocôndrio	hi*po-côn*dri*o ✗	hi*po*côn*dri*o ✓
hipomaníaco	hi*po*ma*ní-a*co ✗	hi*po*ma*ní*a*co ✓
histoquímica	his*to-quí*mi*ca ✗	his*to*quí*mi*ca ✓
ιά	i-á ✗	i*á ✓
ião	i-ão ✗	i*ão ✓
iatroquímica	i*a*tro-quí*mi*ca ✗	i*a*tro*quí*mi*ca ✓
icônico	i-cô*ni*co ✗	i*cô*ni*co ✓
ictiólogo	ic*ti-ó*lo*go ✗	ic*ti*ó*lo*go ✓
indeiscente	in*de-is*cen*te ✗	in*de-is*cen*te ✗

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
indemnização	in*dem*ni*za*ção ✓	in*de.m-ni*za*ção ✗
indemnizar	in*dem*ni*zar ✓	in*de.m-ni*zar ✗
instituidor	ins*ti*tu-i*dor ✗	ins*ti*tu*i*dor ✓
intercambiável	in*ter*cam*bi-á*vel ✗	in*ter*cam*bi*á*vel ✓
interferômetro	in*ter*fe-rô*me*tro ✗	in*ter*fe*rô*me*tro ✓
íon	í-on ✗	í*on ✓
ipsilateral	ip*si*la*te*ral ✓	ip-si*la*te*ral ✗
ípsilon	íp*si*lon ✓	íp-si*lon ✗
isômero	i-sô*me*ro ✗	i*sô*me*ro ✓
isotônico	i*so-tô*ni*co ✗	i*so*tô*ni*co ✓
itaúba	i*ta-ú*ba ✗	i*ta*ú*ba ✓
jandaíra	jan*da-í*ra ✗	jan*da*í*ra ✓
jaó	ja-ó ✗	ja*ó ✓
jauá	jau-á ✗	jau*á ✓
jihadista	ji-ha*dis*ta ✗	ji-ha*dis*ta ✗
joinvilense	jo-in*vi*len*se ✗	jo-in*vi*len*se ✗
jundiá	jun*di-á ✗	jun*di*á ✓
jutaí	ju*ta-í ✗	ju*ta*í ✓
kuwaitiano	ku-wai*ti*a*no ✗	ku-wai*ti*a*no ✗
lemná	lem*na ✓	le.m-na ✗
lemniscata	lem*nis*ca*ta ✓	le.m-nis*ca*ta ✗
lemnisco	lem*nis*co ✓	le.m-nis*co ✗
leônico	le-ô*ni*co ✗	le*ô*ni*co ✓
lidocaína	li*do*ca-í*na ✗	li*do*ca*í*na ✓
liechtensteinense	li*e.ch*tens*tei*nen*se ✗	li*e.ch*tens*tei*nen*se ✗
limnologia	lim*no*lo*gi*a ✓	li.m-no*lo*gi*a ✗
litíase	li*tí-a*se ✗	li*tí*a*se ✓
macaúba	ma*ca-ú*ba ✗	ma*ca*ú*ba ✓
magnetômetro	mag*ne-tô*me*tro ✗	mag*ne*tô*me*tro ✓
maís	ma-ís ✗	ma*ís ✓
malauiano	ma*lau-i*a*no ✗	ma*lau-i*a*no ✗
malônico	ma-lô*ni*co ✗	ma*lô*ni*co ✓
maruí	ma*ru-í ✗	ma*ru*í ✓
maruim	ma*ru-im ✗	ma*ru-im ✗
meã	me-ã ✗	me*ã ✓
meão	me-ão ✗	me*ão ✓
meiótico	mei-ó*ti*co ✗	mei*ó*ti*co ✓
microbiólogo	mi*cro*bi-ó*lo*go ✗	mi*cro*bi*ó*lo*go ✓
midríase	mi*drí-a*se ✗	mi*drí*a*se ✓
miíase	mi-í-a*se ✗	mi*í*a*se ✓
miliário	mi*li-á*ri*o ✗	mi*li*á*ri*o ✓
mnemônico	m.ne-mô*ni*co ✗	mne*mô*ni*co ✓
monofônico	mo*no-fô*ni*co ✗	mo*no*fô*ni*co ✓
moquém	mo-quém ✗	mo*quém ✓
morrião	mor*ri-ão ✗	mor*ri*ão ✓
neodarwinismo	ne*o*dar-wi*nis*mo ✗	ne*o*dar-wi*nis*mo ✗
nobiliário	no*bi*li-á*ri*o ✗	no*bi*li*á*ri*o ✓
núcleolo	nu*clé-o*lo ✗	nu*clé*o*lo ✓
octaédrico	oc*ta-é*dri*co ✗	oc*ta*é*dri*co ✓
odeão	o*de-ão ✗	o*de*ão ✓
oídio	o-í*di*o ✗	o*í*di*o ✓
omíada	o*mí-a*da ✗	o*mí*a*da ✓
omnipotente	om*ni*po*ten*te ✓	o.m-ni*po*ten*te ✗
omnisciente	om*nis*ci*en*te ✓	o.m-nis*ci*en*te ✗
omnívoro	om*ní*vo*ro ✓	o.m-ní*vo*ro ✗
oócito	o-ó*ci*to ✗	o*ó*ci*to ✓
orfeão	or*fe-ão ✗	or*fe*ão ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
oriá	o*ri-á ✗	o*ri*á ✓
ortodôntico	or*to-dôn*ti*co ✗	or*to*dôn*ti*co ✓
ossuário	os*su-á*ri*o ✗	os*su*á*ri*o ✓
osteíte	os*te-í*te ✗	os*te*í*te ✓
ostiário	os*ti-á*ri*o ✗	os*ti*á*ri*o ✓
paleógrafo	pa*le-ó*gra*fo ✗	pa*le*ó*gra*fo ✓
paleólogo	pa*le-ó*lo*go ✗	pa*le*ó*lo*go ✓
panenteísmo	pa*nen*te-ís*mo ✗	pa*nen*te*ís*mo ✓
patauí	pa*tau-á ✗	pa*tau*á ✓
patoá	pa*to-á ✗	pa*to*á ✓
peciolo	pe*ci-o*lo ✗	pe*ci*o*lo ✓
pequiá	pe*qui-á ✗	pe*qui*á ✓
periélio	pe*ri-é*li*o ✗	pe*ri*é*li*o ✓
píon	pí-on ✗	pí*on ✓
pitiríase	pi*ti*rí-a*se ✗	pi*ti*ri*a*se ✓
pium	pi-um ✗	pi-um ✗
planctônico	planc*tô*ni*co ✗	planc*tô*ni*co ✓
pneuma	p.neu*ma ✗	pneu*ma ✓
pneumonite	p.neu*mo*ni*te ✗	pneu*mo*ni*te ✓
pneumotórax	p.neu*mo*tó*rax ✗	pneu*mo*tó*rax ✓
poliúria	po*li-ú*ri*a ✗	po*li*ú*ri*a ✓
poraquê	po*ra-quê ✗	po*ra*quê ✓
preênsil	pre-ên*sil ✗	pre*ên*sil ✓
príon	prí-on ✗	prí*on ✓
procariótico	pro*ca*ri-ó*ti*co ✗	pro*ca*ri*ó*ti*co ✓
procônsul	pro-côn*sul ✗	pro*côn*sul ✓
proteólise	pro*te-ó*li*se ✗	pro*te*ó*li*se ✓
pseudocientífico	p.seu*do*ci*en*ti*fi*co ✗	pseu*do*ci*en*ti*fi*co ✓
pseudofruto	p.seu*do*fru*to ✗	pseu*do*fru*to ✓
psicadélico	p.si*ca*dé*li*co ✗	psi*ca*dé*li*co ✓
psicoativo	p.si*co*a*ti*vo ✗	psi*co*a*ti*vo ✓
psicocirurgia	p.si*co*ci*rur*gi*a ✗	psi*co*ci*rur*gi*a ✓
psicofisiologia	p.si*co*fi*si*o*lo*gi*a ✗	psi*co*fi*si*o*lo*gi*a ✓
psicofonia	p.si*co*fo*ni*a ✗	psi*co*fo*ni*a ✓
psicografia	p.si*co*gra*fi*a ✗	psi*co*gra*fi*a ✓
psicopompo	p.si*co*pom*po ✗	psi*co*pom*po ✓
psicoterápico	p.si*co*te*rá*pi*co ✗	psi*co*te*rá*pi*co ✓
psoas	p.so*as ✗	p.so*as ✗
pterodáctilo	p.te*ro*dác*ti*lo ✗	pte*ro*dác*ti*lo ✓
ptialina	p.ti*a*li*na ✗	pti*a*li*na ✓
ptolemaico	p.to*le*mai*co ✗	pto*le*mai*co ✓
radioquímica	ra*dí*o-quí*mi*ca ✗	ra*dí*o*quí*mi*ca ✓
randômico	ran-dô*mi*co ✗	ran*dô*mi*co ✓
reciário	re*ci-á*ri*o ✗	re*ci*á*ri*o ✓
reimpresso	re-im*pres*so ✗	re*im*pres*so ✓
reimprimir	re-im*pri*mir ✗	re*im*pri*mir ✓
sambaíba	sam*ba-í*ba ✗	sam*ba*í*ba ✓
samnita	sam*ni*ta ✓	sa.m-ni*ta ✗
saxônico	sa-xô*ni*co ✗	sa*xô*ni*co ✓
semiárido	se*mi-á*ri*do ✗	se*mi*á*ri*do ✓
semiótico	se*mi-ó*ti*co ✗	se*mi*ó*ti*co ✓
semipermeável	se*mi*per*me-á*vel ✗	se*mi*per*me*á*vel ✓
sépsis	sép*sis ✓	sép-sis ✗
serôdio	se-rô*di*o ✗	se*rô*di*o ✓
serventuário	ser*ven*tu-á*ri*o ✗	ser*ven*tu*á*ri*o ✓
siálico	si-á*li*co ✗	si*á*li*co ✓
siriaco	si*rí-a*co ✗	si*rí*a*co ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
sonômetro	so-nô-me*tro ✗	so*nô-me*tro ✓
suaíli	su*a-i*li ✗	su*a*i*li ✓
subatômico	su*ba-tô*mi*co ✗	su*ba*tô*mi*co ✓
subdiácono	sub*di-á*co*no ✗	sub*di*á*co*no ✓
sublevar	su.b-le*var ✗	sub*le*var ✓
subliminal	su.b-li*mi*nal ✗	sub*li*mi*nal ✓
sublinear	su.b-li*ne*ar ✗	sub*li*ne*ar ✓
sublingual	su.b-lin*gual ✗	sub*lin*gual ✓
sublunar	su.b-lu*nar ✗	sub*lu*nar ✓
subsequência	sub*se-quên*ci*a ✗	sub*se*quên*ci*a ✓
substituente	subs*ti*tu-in*te ✗	subs*ti*tu-in*te ✗
coimbra	co-im*bra ✗	co*im*bra ✓
coimperador	co-im*pe*ra*dor ✗	co*im*pe*ra*dor ✓
reimposição	re-im*po*si*ção ✗	re*im*po*si*ção ✓
bioimpressão	bi*o-im*pres*são ✗	bi*o*im*pres*são ✓
neoimpressionista	ne*o-im*pres*si*o*nis*ta ✗	ne*o*im*pres*si*o*nis*ta ✓
autoimposto	au*to-im*pos*to ✗	au*to*im*pos*to ✓
reimplementação	re-im*ple*men*ta*ção ✗	re*im*ple*men*ta*ção ✓
teleimpressor	te*le-im*pres*sor ✗	te*le*im*pres*sor ✓
reimposta	re-im*pos*ta ✗	re*im*pos*ta ✓
reimplementar	re-im*ple*men*tar ✗	re*im*ple*men*tar ✓
reimpulsionar	re-im*pul*si*o*nar ✗	re*im*pul*si*o*nar ✓
autoimposição	au*to-im*po*si*ção ✗	au*to*im*po*si*ção ✓
cacaucultor	ca*cau-i*cul*tor ✗	ca*cau*i*cul*tor ✓
sanduche	san*du-i*che ✗	san*du*i*che ✓
ruidosamente	ru-i*do*sa*men*te ✗	ru*i*do*sa*men*te ✓
promiscuidade	pro*mis*cu-i*da*de ✗	pro*mis*cu*i*da*de ✓
inocuidade	i*no*cu-i*da*de ✗	i*no*cu*i*da*de ✓
reincidência	re-in*ci*dên*ci.a ✗	re*in*ci*dên*ci.a ✗
reinclu	re-in*clu-ir ✗	re*in*clu*ir ✓
reinfecção	re-in*fec*ção ✗	re*in*fec*ção ✓
reino	re-in*for*çar ✗	re*in*for*çar ✓
reinsurgir	re-in*sur*gir ✗	re*in*sur*gir ✓
reinstalado	re-ins*ta*la*do ✗	re*ins*ta*la*do ✓
restituído	re-ins*ti*tu-í*do ✗	re*ins*ti*tu*í*do ✓
reinvindicar	re-in*vin*di*car ✗	re*in*vin*di*car ✓
bioincrustação	bi*o-in*crus*ta*ção ✗	bi*o*in*crus*ta*ção ✓
agroindustrial	a*gro-in*dus*tri*al ✗	a*gro*in*dus*tri*al ✓
bioindicadores	bi*o-in*di*ca*do*res ✗	bi*o*in*di*ca*do*res ✓
autoindulgência	au*to-in*dul*gên*ci.a ✗	au*to*in*dul*gên*ci.a ✗
autoinduzida	au*to-in*du*zi*da ✗	au*to*in*du*zi*da ✓
microindústria	mi*cro-in*dús*tri*a ✗	mi*cro*in*dús*tri*a ✓
tecnológico	tec*no-in*dus*tri*al ✗	tec*no*in*dus*tri*al ✓
bioinformática	bi*o-in*for*má*ti*ca ✗	bi*o*in*for*má*ti*ca ✓
coinfecção	co-in*fec*ção ✗	co*in*fec*ção ✓
geoinformação	ge.o-in*for*ma*ção ✗	ge.o*in*for*ma*ção ✗
autoinfecção	au*to-in*fec*ção ✗	au*to*in*fec*ção ✓
neuroinformática	neu*ro-in*for*má*ti*ca ✗	neu*ro*in*for*má*ti*ca ✓
amendins	a*men*do-ins ✗	a*men*do*ins ✓
videomontagem	vi*de*o-ins*ta*la*ção ✗	vi*de*o*ins*ta*la*ção ✓
autointitulado	au*to-in*ti*tu*la*do ✗	au*to*in*ti*tu*la*do ✓
saindo	sa-in*do ✗	sa*in*do ✓
caindo	ca-in*do ✗	ca*in*do ✓
contraindicações	con*tra-in*di*ca*ções ✗	con*tra*in*di*ca*ções ✓
abstraindo	abs*tra-in*do ✗	abs*tra*in*do ✓
pratinha	pra-i*nha ✗	pra*i*nha ✓
lagoinha	la*go-i*nha ✗	la*go*i*nha ✓

Table 5: Comparativo dos erros cometidos pelos dois conjuntos de regras de hifenização.

palavra	regras padrão	regras adicionais
canoinha	ca*no-i*nha ✗	ca*no*i*nha ✓
reimpresso	re-im*pres*so ✗	re*im*pres*so ✓
genuino	ge*nu-i*no ✗	ge*nu*i*no ✓
ruína	ru-i*na ✗	ru*i*na ✓
arruinado	ar*ru-i*na*do ✗	ar*ru*i*na*do ✓
arruinou	ar*ru-i*nou ✗	ar*ru*i*nou ✓
suíno	su-i*no ✗	su*i*no ✓
suína	su-i*na ✗	su*i*na ✓
ajuizamento	a*ju-i*za*men*to ✗	a*ju*i*za*men*to ✓
luiz	lu-iz ✗	lu*iz ✓
luíza	lu-i*za ✗	lu*i*za ✓
amnésia	a-m.né*si*a ✗	a*mné*si*a ✓
amniota	a-m.ni*o*ta ✗	a*mni*o*ta ✓
amniótico	a-m.ni-ó*ti*co ✗	a*mni*ó*ti*co ✓
diuréticos	di-u*ré*ti*cos ✗	di*u*ré*ti*cos ✓
diurese	di-u*re*se ✗	di*u*re*se ✓
feiurinha	fei-u*ri*nha ✗	fei*u*ri*nha ✓
íodo	í-o*do ✗	í*o*do ✓
íons	í-ons ✗	í*ons ✓
antíope	an*tí-o*pe ✗	an*tí*o*pe ✓
médiuns	mé*di-uns ✗	mé*di*uns ✓
fluór	flu-ór ✗	flu*ór ✓
sequóia	se-quói*a ✗	se*quói*a ✓
subquântico	sub-quân*ti*co ✗	sub*quân*ti*co ✓
37798	1368	132

References

- Araújo, Antonio Martins de and Toru Maruyama (2015). “A Hifenização em Português”. In: *Idioma* 28, pp. 90–107. URL: http://www.institutodeletras.uerj.br/idioma/numeros/28/Idioma28_a08.pdf.
- Bergström, Magnus and Neves Reis (2011). *Prontuário ortográfico e guia da língua portuguesa*. Leya.
- Cagliari, Luiz Carlos (2015). “Aspectos teóricos da ortografia”. In: *Ortografia da língua portuguesa: história, discurso, representações*. Ed. by Maurício Silva. São Paulo: Contexto, pp. 17–52.
- Cegalla, Domingos Paschoal (2020). *Novíssima Gramática Da Língua Portuguesa*. São Paulo: Companhia Editora Nacional.
- Coulmas, Florian (2003). *Writing Systems: An Introduction to Their Linguistic Analysis*. Cambridge Textbooks in Linguistics. Cambridge University Press. ISBN: 9780521787376.
- Cunha, Celso and Lindley Cintra (2016). *Nova gramática do português contemporâneo*. LEXIKON Editora Digital Ltda.
- Haralambous, Yannis (2021). “A Revisited Small Tutorial on Patgen, 28 Years After”. In: *electronic form, available from CTAN as info/patgen2. tutorial*.
- Honorof, Douglas and Laurie Feldman (Jan. 2006). “The Chinese character in psycholinguistic research: Form, structure, and the reader”. In: *The Handbook of East Asian Psycholinguistics*. Ed. by Ping Li et al. Vol. 1. Cambridge University Press. Chap. 17, pp. 195–208. DOI: 10.2277/0521833337.
- Hunspell’s Team (Apr. 13, 2023a). *Hunspell*. URL: <http://hunspell.github.io/>.
- (Apr. 13, 2023b). *Hunspell Hyphen*. URL: <https://github.com/hunspell/hyphen>.
- Knuth, Donald Ervin (May 13, 1977). *Preliminary preliminary description of TEX*. unpublished.
- Kristensen, T. and D. Langmyhr (2001). “Two regimes of computer hyphenation - a comparison”. In: *IJCNN’01. International Joint Conference on Neural Networks. Proceedings (Cat. No.01CH37222)*. IEEE. DOI: 10.1109/ijcnn.2001.939592. URL: <https://doi.org/10.1109/ijcnn.2001.939592>.
- Levien, Raph (1998). *Brief explanation of the hyphenation algorithm herein*. Hunspell. URL: <https://github.com/hunspell/hyphen/blob/master/README.hyphen>.
- Liang, Frank and Peter Breitenlohner (1991). *PATtern GENERation program for the TEX82 hyphenator*. Tech. rep. 2. Electronic documentation of PATGEN.

- Liang, Franklin Mark (1983). “Word Hyphenation by Computer”. PhD thesis. Stanford University.
- Libossek, Marion and Florian Schiel (2000). “Syllable-based text-to-phoneme conversion for German.” In: *INTERSPEECH*. Citeseer, pp. 283–286.
- Lin, Li-chin (2011). “Fundamental generalizations of English syllabification”. In: *Concentric: Studies in Linguistics* 37.2, pp. 179–208.
- Magalhães Gândavo, Pêro de (1574). *Regras que ensinam a maneira de escrever e orthographia da lingua portuguesa, etc.* Lisboa. URL: <https://purl.pt/324>.
- Mateus, Maria Helena Mira (n.d.). “Questões fonológicas do português”. manuscript.
- Németh, László (2006). “Automatic non-standard hyphenation in OpenOffice.org”. In: *TUGboat* 27.1, pp. 32–37.
- Palmer, David D. (2010). “Tokenisation and Sentence Segmentation”. In: *Handbook of Natural Language Processing*. Ed. by Nitin Indurkha and Fred J. Damerau. CRC Press. Chap. 2, pp. 9–30.
- TeX pattern authors (Apr. 13, 2023). *TeX hyphenation patterns*. URL: <https://www.tug.org/tex-hyphen/>.
- Rezende, Pedro Jussieu de (1987). “Portuguese hyphenation table for TEX”. In: *TUGboat* 8.2, pp. 102–102.
- Rezende, Pedro Jussieu de and José Joao Dias Almeida (2015). *Hyphenation patterns for Portuguese*. <http://mirror.ctan.org/language/hyph-utf8/tex/generic/hyph-utf8/patterns/tex/hyph-pt.tex>.
- Scannell, Kevin (2003). “Hyphenation patterns for minority languages”. In: *TUGboat* 24.2, pp. 236–239.
- Sojka, Petr (Aug. 1995a). *Notes on Compound Word Hyphenation in TEX*. Tech. rep. Masaryk University in Brno, Faculty of Informatics.
- (1995b). “Notes on compound word hyphenation in TEX”. In: *TUGboat* 16.3, pp. 290–297.
- Sojka, Petr et al. (2005). “Competing Patterns in Language Engineering and Computer Typesetting”. PhD thesis. Masaryk University, Brno.
- Sojka, Petr and David Antoš (2003). “Context sensitive pattern based segmentation: A Thai challenge”. In: *Proceedings of EACL 2003 Workshop on Computational Linguistics for South Asian Languages—Expanding Synergies with Europe, Budapest*, pp. 65–72.
- Sojka, Petr and Ondřej Sojka (2019). “The unreasonable effectiveness of pattern generation”. In: *TUGboat* 40.2, pp. 187–193.
- Troglanis, Nikolaos and Charles Elkan (2010). “Conditional random fields for word hyphenation”. In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 366–374.
- Yavas, Mehmet (2020). *Applied English Phonology*. John Wiley & Sons.