

# Hifenização Automática

Leonardo Araujo

UFSJ



# Introdução

- Evolução histórica da hifenização:
  - Textos antigos: Sem marcas de hifenização.
  - Introdução da hifenização para gerenciar quebras de linha e evitar espaços excessivamente largos.
- Importância:
  - Evitar grandes espaços em texto justificado.
  - Manter o fluxo estético.
  - Linhas curtas têm poucos pontos de quebra entre palavras - evitar rios e lagos.
- Desvantagens:
  - Interromper o fluxo de leitura e a compreensão.
  - Pode levar à confusão.
  - Visualmente desagradável.

# Introdução

- Evolução histórica da hifenização:
  - Textos antigos: Sem marcas de hifenização.
  - Introdução da hifenização para gerenciar quebras de linha e evitar espaços excessivamente largos.
- Importância:
  - Evitar grandes espaços em texto justificado.
  - Manter o fluxo estético.
  - Linhas curtas têm poucos pontos de quebra entre palavras - evitar rios e lagos.
- Desvantagens:
  - Interromper o fluxo de leitura e a compreensão.
  - Pode levar à confusão.
  - Visualmente desagradável.

# Introdução

- Evolução histórica da hifenização:
  - Textos antigos: Sem marcas de hifenização.
  - Introdução da hifenização para gerenciar quebras de linha e evitar espaços excessivamente largos.
- Importância:
  - Evitar grandes espaços em texto justificado.
  - Manter o fluxo estético.
  - Linhas curtas têm poucos pontos de quebra entre palavras - evitar rios e lagos.
- Desvantagens:
  - Interromper o fluxo de leitura e a compreensão.
  - Pode levar à confusão.
  - Visualmente desagradável.

During what many archaeologists call the formative period, Amazonian societies were deeply involved in the emergence of South America's highland agrarian systems, and possibly contributed directly to the social and religious fabric constitutive of the Andean civilizational orders.

In 1500, Vicente Yáñez Pinzón was the first European to sail into the river. Pinzón called the river flow Río Santa María de la Mar Dulce, later shortened to Mar Dulce (literally, sweet sea, because of its fresh-water pushing out into the ocean). For 350 years after the first European encounter of the Amazon by Pinzón, the Portuguese portion of the basin remained an untended former food gathering and planned agricultural landscape occupied by the indigenous peoples who survived the arrival of European diseases. There is ample evidence for complex large-scale, pre-Columbian social formations, including chiefdoms, in many areas of Amazonia (particularly the inter-fluvial regions) and even large towns and cities. For instance the pre-Columbian culture on the island of Marajo may have developed social stratification and supported a population of 100,000 people. The Native Americans of the Amazon rain forest may have used Terra preta to make the land suitable for the large scale agriculture needed to support large populations and complex social formations such as chiefdoms. One of Gonzalo Pizarro's lieutenants, Francisco de Orellana, during his 1541 expedition, east of Quito into the South American interior in search of El Dorado and the Country of the Cinnamon was ordered to explore the Coca River and return when the river ended. When they arrived to the confluence to the Napo River, his men menaced to mutiny if they did not continue. On 26 December 1541, he accepted to be elected chief of the new expedition and to conquest new lands in name of the king. The 49 men began to build a bigger ship for riverine navigation. During their navigation on Napo River they were threatened consistently by the Omaguas. They reached Negro River on 3 June 1542 and there I finally arrived to the Amazon River, that was so named because they were attacked by fierce female warriors like the mythological Amazons. The icamiabas Indians dominated the area close to the Amazon River, rich in gold. When Orellana went down the river in search of gold, descends Andes (in 1541), the river was still called Grande Río, Mar Dulce or Río da Canela (Cinnamon).

Fathom holystone broadside run an  
shot across the bow yo smart hands  
deadlighting Plate Fleet pirate. Reef an lubber  
just lubber leads ye hearties. Laying  
hogsheading gabion pinnace Brethren of the  
Coast loot to be account broadside  
highjacking Tar sheet. Run a shot across the  
Shiver me timbers fire  
shipping in the Caribbean hole hearties keelhaul  
wench brigantine sutler mutiny.

Figura 1: Rios e Lagos.



Figura 2: Rios, Lagos, Órfãos e Viúvas.

## O papel do TeX

- Hifenização automática desde 1977.
  - Donald Ervin Knuth (1977).
  - Franklin Mark Liang (1983).
- Usa `patgen` para a geração de regras.
- Hunspell (LibreOffice, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher).
- `tex-hyphen` - repositório central para todos os padrões de hifenização - regras para 66 idiomas e outros recursos.
- Variáveis do TeX `\lefthyphenmin` e `\righthyphenmin`.

## O papel do TeX

- Hifenização automática desde 1977.
  - Donald Ervin Knuth (1977).
  - Franklin Mark Liang (1983).
- Usa `patgen` para a geração de regras.
- Hunspell (LibreOffice, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher).
- `tex-hyphen` - repositório central para todos os padrões de hifenização - regras para 66 idiomas e outros recursos.
- Variáveis do TeX `\lefthyphenmin` e `\righthyphenmin`.



## O papel do TeX

- Hifenização automática desde 1977.
  - Donald Ervin Knuth (1977).
  - Franklin Mark Liang (1983).
- Usa `patgen` para a geração de regras.
- Hunspell (LibreOffice, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher).
- `tex-hyphen` - repositório central para todos os padrões de hifenização - regras para 66 idiomas e outros recursos.
- Variáveis do TeX `\lefthyphenmin` e `\righthyphenmin`.

## O papel do TeX

- Hifenização automática desde 1977.
  - Donald Ervin Knuth (1977).
  - Franklin Mark Liang (1983).
- Usa `patgen` para a geração de regras.
- Hunspell (LibreOffice, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher).
- **tex-hyphen - repositório central para todos os padrões de hifenização** - regras para 66 idiomas e outros recursos.
- Variáveis do TeX `\lefthyphenmin` e `\righthyphenmin`.

## O papel do TeX

- Hifenização automática desde 1977.
  - Donald Ervin Knuth (1977).
  - Franklin Mark Liang (1983).
- Usa `patgen` para a geração de regras.
- Hunspell (LibreOffice, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher).
- **tex-hyphen - repositório central para todos os padrões de hifenização** - regras para 66 idiomas e outros recursos.
- Variáveis do TeX `\lefthyphenmin` e `\righthyphenmin`.

## Critérios para Hifenização

- **Fonológico:** Baseado na divisão de sílabas faladas.
  - Exemplo: “com-pu-ta-dor”, mostrando a divisão silábica.
- **Morfológico (ou Etimológico):** Focado nas partes da palavra que carregam significado (prefixos, raízes, sufixos).
  - Exemplo: “in-feliz”, mostrando o prefixo e a raiz.
- **Ortográfico:** Segue as convenções padrão de escrita.
  - Exemplo: “ad-vo-ga-do”, as sequências consonantais constituiriam sílabas próprias, entretanto a ortografia impede que sejam separadas deixando uma sequência não permitida pela fonotática da língua (“a-d-vo-ga-do”).
- **Semântico:** Considera o contexto para evitar quebras ambíguas ou inadequadas.
  - Evitar “fede-ração” para evitar uma interpretação inadequada.
- **Estilo:** A hifenização também pode ser uma questão de estilo, onde algumas divisões soam melhor ou são mais elegantes.
  - A separação “entres-safra” é preferível à “en-tressafra” ou “entressa-fra”.

## Critérios para Hifenização

- **Fonológico:** Baseado na divisão de sílabas faladas.
  - Exemplo: “com-pu-ta-dor”, mostrando a divisão silábica.
- **Morfológico (ou Etimológico):** Focado nas partes da palavra que carregam significado (prefixos, raízes, sufixos).
  - Exemplo: “in-feliz”, mostrando o prefixo e a raiz.
- **Ortográfico:** Segue as convenções padrão de escrita.
  - Exemplo: “ad-vo-ga-do”, as sequências consonantais constituiriam sílabas próprias, entretanto a ortografia impede que sejam separadas deixando uma sequência não permitida pela fonotática da língua (“a-d-vo-ga-do”).
- **Semântico:** Considera o contexto para evitar quebras ambíguas ou inadequadas.
  - Evitar “fede-ração” para evitar uma interpretação inadequada.
- **Estilo:** A hifenização também pode ser uma questão de estilo, onde algumas divisões soam melhor ou são mais elegantes.
  - A separação “entres-safra” é preferível à “en-tressafra” ou “entressa-fra”.

## Critérios para Hifenização

- **Fonológico:** Baseado na divisão de sílabas faladas.
  - Exemplo: “com-pu-ta-dor”, mostrando a divisão silábica.
- **Morfológico (ou Etimológico):** Focado nas partes da palavra que carregam significado (prefixos, raízes, sufixos).
  - Exemplo: “in-feliz”, mostrando o prefixo e a raiz.
- **Ortográfico:** Segue as convenções padrão de escrita.
  - Exemplo: “ad-vo-ga-do”, as sequências consonantais constituiriam sílabas próprias, entretanto a ortografia impede que sejam separadas deixando uma sequência não permitida pela fonotática da língua (“a-d-vo-ga-do”).
- **Semântico:** Considera o contexto para evitar quebras ambíguas ou inadequadas.
  - Evitar “fede-ração” para evitar uma interpretação inadequada.
- **Estilo:** A hifenização também pode ser uma questão de estilo, onde algumas divisões soam melhor ou são mais elegantes.
  - A separação “entres-safra” é preferível à “en-tressafra” ou “entressa-fra”.

## Critérios para Hifenização

- **Fonológico:** Baseado na divisão de sílabas faladas.
  - Exemplo: “com-pu-ta-dor”, mostrando a divisão silábica.
- **Morfológico (ou Etimológico):** Focado nas partes da palavra que carregam significado (prefixos, raízes, sufixos).
  - Exemplo: “in-feliz”, mostrando o prefixo e a raiz.
- **Ortográfico:** Segue as convenções padrão de escrita.
  - Exemplo: “ad-vo-ga-do”, as sequências consonantais constituiriam sílabas próprias, entretanto a ortografia impede que sejam separadas deixando uma sequência não permitida pela fonotática da língua (“a-d-vo-ga-do”).
- **Semântico:** Considera o contexto para evitar quebras ambíguas ou inadequadas.
  - Evitar “fede-ração” para evitar uma interpretação inadequada.
- **Estilo:** A hifenização também pode ser uma questão de estilo, onde algumas divisões soam melhor ou são mais elegantes.
  - A separação “entres-safra” é preferível à “en-tressafra” ou “entressa-fra”.

## Critérios para Hifenização

- Fonológico: Baseado na divisão de sílabas faladas.
  - Exemplo: “com-pu-ta-dor”, mostrando a divisão silábica.
- Morfológico (ou Etimológico): Focado nas partes da palavra que carregam significado (prefixos, raízes, sufixos).
  - Exemplo: “in-feliz”, mostrando o prefixo e a raiz.
- Ortográfico: Segue as convenções padrão de escrita.
  - Exemplo: “ad-vo-ga-do”, as sequências consonantais constituiriam sílabas próprias, entretanto a ortografia impede que sejam separadas deixando uma sequência não permitida pela fonotática da língua (“a-d-vo-ga-do”).
- Semântico: Considera o contexto para evitar quebras ambíguas ou inadequadas.
  - Evitar “fede-ração” para evitar uma interpretação inadequada.
- Estilo: A hifenização também pode ser uma questão de estilo, onde algumas divisões soam melhor ou são mais elegantes.
  - A separação “entres-safra” é preferível à “en-tressafra” ou “entressa-fra”.



Esses critérios podem diferir ou ter pesos diferentes entre as línguas:

- Algumas línguas podem priorizar regras fonológicas sobre outras, como o espanhol e o português.
- Em línguas com compostos complexos, como o alemão, as considerações morfológicas podem ser mais proeminentes.

Esses critérios evoluíram ou mudaram com reformas linguísticas ou em contextos linguísticos específicos.

É complexo aplicar esses critérios, especialmente em contextos multilíngues ou com dialetos não padronizados.

Esses critérios podem diferir ou ter pesos diferentes entre as línguas:

- Algumas línguas podem priorizar regras fonológicas sobre outras, como o espanhol e o português.
- Em línguas com compostos complexos, como o alemão, as considerações morfológicas podem ser mais proeminentes.

Esses critérios evoluíram ou mudaram com reformas linguísticas ou em contextos linguísticos específicos.

É complexo aplicar esses critérios, especialmente em contextos multilíngues ou com dialetos não padronizados.

## Regras de Hifenização no TeX

Algumas regras para o português:

1b2l

1b2r

1ba

1be

1bi

1bo

1bu

1bá

1bâ

...

a3a

a3e

a3o

...

## Exemplo

	.	g	u	i	m	a	r	ã	e	s	.	
	1	0	0									1gu
	1	0	4	0								1gu4i
				1	0	0						1ma
						1	0	0				1rã
max:	1	0	4	1	0	1	0	0	0	0		
final:	g	u	i	-	m	a	-	r	ã	e	s	

## Regras de hifenização para o português

- P.J. de Rezende. Portuguese hyphenation table for TeX (1987)
- P.J. de Rezende, J.J.D. Almeida. Hyphenation patterns for Portuguese (1987, 1994, 1996, 2015)
- L.C. Araujo, A. Benevides. Enhancing TeX hyphenation rules for Portuguese (2024)
- L.C. Araujo, A. Benevides. Aperfeiçoando a Hifenização Automática em Português no TeX (2024)