# Automatic Hyphenation

Leonardo Araujo

UFSJ

# Introduction

- Historical evolution of hyphenation:
    - Early texts: No hyphenation marks.
    - Introduction of hyphenation to manage line breaks and prevent overly wide gaps.
- Importance:
    - Avoid large gaps in justified text.
    - Maintain aesthetic flow.
    - Short lines have few breakpoints between words - avoid rivers and lakes.
- Drawbacks:
    - Disrupt the reading flow and comprehension.
    - May lead to confusion.
    - Visually unappealing.

# Introduction

- Historical evolution of hyphenation:
    - Early texts: No hyphenation marks.
    - Introduction of hyphenation to manage line breaks and prevent overly wide gaps.
- Importance:
    - Avoid large gaps in justified text.
    - Maintain aesthetic flow.
    - Short lines have few breakpoints between words - avoid rivers and lakes.
- Drawbacks:
    - Disrupt the reading flow and comprehension.
    - May lead to confusion.
    - Visually unappealing.

# Introduction

- Historical evolution of hyphenation:
    - Early texts: No hyphenation marks.
    - Introduction of hyphenation to manage line breaks and prevent overly wide gaps.
- Importance:
    - Avoid large gaps in justified text.
    - Maintain aesthetic flow.
    - Short lines have few breakpoints between words - avoid rivers and lakes.
- Drawbacks:
    - Disrupt the reading flow and comprehension.
    - May lead to confusion.
    - Visually unappealing.

During what many archaeologists call the formative period, Amazonian societies were deep- ly involved in the emergence of South America's highland agrarian sys- tems, and possi- bly contributed directly to the social and religious fabric constitutive of the Andean civilizational orders.

In 1500, Vicente Ya- ñez Pinzón was the first European to sail into the river. Pinzón called the river flow Río Santa María de la Mar Duke, 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 Ama- zon by Pinzón, the Portuguese portion of the basin remained an untended former food gath- ering and planned agricultural landscape occupied by the indigenous peo- ples who survived the arrival of European diseases. There is ample evi- dence for complex large-scale, pre-Columbian social formations, includ- ing chiefdoms, in many areas of Amazonia (particu- larly the inter-fluvial regions) and even large towns and cities.For instance the pre-Columbian culture on the island of Marajo may have developed so- cial stratification and supported a population of 100,000 people.The Na- tive 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 com- plex social formations such as chiefdoms. One of Gon- zalo Pizarro's lieuten- ants, Francisco de Orellana, during his 1541 expedi- tion, east of Quito into the South American interi- or 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 big- ger ship for riverine navigation. Dur-ing their navigation on Napo Riv- er 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 i c a m i a b a s 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 Rio, Mar Dulce or Rio da Canela ( C i n n a -

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 Breth n of the Coast loot to be account broadside highjacking Tar sheet. Run a shot across the Shiver me timbers fire shippingintheCaribbean hole hearties keelhaul wench brigantine sutler mutiny.
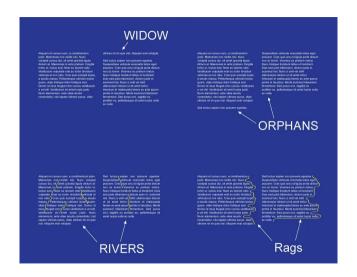
Figure 1: Rivers and Lakes.

Figure 2: Rivers, Lakes, Orphans and Widows.

## TeX's role

- Automatic hyphenation since 1977.
  - Donald Ervin Knuth (1977).
  - Franklin Mark Liang (1983).
- Uses patgen for rule generation.
- Hunspell (LibreOffice, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher).
- tex-hyphen - central repository for all hyphenation patterns - rules for 66 languages and other resources.
- TeX variables \lefthyphenmin and \righthyphenmin.

## TeX's role

- Automatic hyphenation since 1977.
  - Donald Ervin Knuth (1977).
  - Franklin Mark Liang (1983).
- Uses `patgen` for rule generation.
- Hunspell (LibreOffice, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher).
- tex-hyphen - central repository for all hyphenation patterns - rules for 66 languages and other resources.
- TeX variables \lefthyphenmin and \righthyphenmin.

## TeX's role

- Automatic hyphenation since 1977.
  - Donald Ervin Knuth (1977).
  - Franklin Mark Liang (1983).
- Uses `patgen` for rule generation.
- Hunspell (LibreOffice, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher).
- tex-hyphen - central repository for all hyphenation patterns - rules for 66 languages and other resources.
- TeX variables `\lefthyphenmin` and `\righthyphenmin`.

## TeX's role

- Automatic hyphenation since 1977.
    - Donald Ervin Knuth (1977).
    - Franklin Mark Liang (1983).
- Uses `patgen` for rule generation.
- Hunspell (LibreOffice, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher).
- tex-hyphen - central repository for all hyphenation patterns - rules for 66 languages and other resources.
- TeX variables `\lefthyphenmin` and `\righthyphenmin`.

# TeX's role

- Automatic hyphenation since 1977.
  - Donald Ervin Knuth (1977).
  - Franklin Mark Liang (1983).
- Uses `patgen` for rule generation.
- Hunspell (LibreOffice, Mozilla Firefox, Mozilla Thunderbird, Google Chrome, macOS, InDesign, memoQ, Opera, Affinity Publisher).
- tex-hyphen - central repository for all hyphenation patterns - rules for 66 languages and other resources.
- TeX variables `\lefthyphenmin` and `\righthyphenmin`.

# Criteria for Hyphenation

- Phonological: Based on spoken syllable division.
  - Example: "com-pu-ter", showing syllable breaks.
- Morphological (or Etymological): Focused on word parts that carry meaning (prefixes, roots, suffixes).
  - Example: "un-happy", showing the prefix and root.
- Orthographic: Follows standard writing conventions.
  - "week-end" in English, following writing conventions.
- Semantic: Considers context to prevent ambiguous or inappropriate breaks.
  - Avoiding "re-sign" when "resign" is intended to keep the meaning clear.
- Style: Hyphenation can also be a matter of style, where some divisions sound better or are more elegant.
  - "coop-eration" is preferable over "co-operation".

# Criteria for Hyphenation

- Phonological: Based on spoken syllable division.
  - Example: "com-pu-ter", showing syllable breaks.
- Morphological (or Etymological): Focused on word parts that carry meaning (prefixes, roots, suffixes).
  - Example: "un-happy", showing the prefix and root.
- Orthographic: Follows standard writing conventions.
  - "week-end" in English, following writing conventions.
- Semantic: Considers context to prevent ambiguous or inappropriate breaks.
  - Avoiding "re-sign" when "resign" is intended to keep the meaning clear.
- Style: Hyphenation can also be a matter of style, where some divisions sound better or are more elegant.
  - "coop-eration" is preferable over "co-operation".

# Criteria for Hyphenation

- Phonological: Based on spoken syllable division.
    - Example: "com-pu-ter", showing syllable breaks.
- Morphological (or Etymological): Focused on word parts that carry meaning (prefixes, roots, suffixes).
    - Example: "un-happy", showing the prefix and root.
- Orthographic: Follows standard writing conventions.
    - "week-end" in English, following writing conventions.
- Semantic: Considers context to prevent ambiguous or inappropriate breaks.
    - Avoiding "re-sign" when "resign" is intended to keep the meaning clear.
- Style: Hyphenation can also be a matter of style, where some divisions sound better or are more elegant.
    - "coop-eration" is preferable over "co-operation".

# Criteria for Hyphenation

- Phonological: Based on spoken syllable division.
    - Example: "com-pu-ter", showing syllable breaks.
- Morphological (or Etymological): Focused on word parts that carry meaning (prefixes, roots, suffixes).
    - Example: "un-happy", showing the prefix and root.
- Orthographic: Follows standard writing conventions.
    - "week-end" in English, following writing conventions.
- Semantic: Considers context to prevent ambiguous or inappropriate breaks.
    - Avoiding "re-sign" when "resign" is intended to keep the meaning clear.
- Style: Hyphenation can also be a matter of style, where some divisions sound better or are more elegant.
    - "coop-eration" is preferable over "co-operation".

# Criteria for Hyphenation

- Phonological: Based on spoken syllable division.
    - Example: "com-pu-ter", showing syllable breaks.
- Morphological (or Etymological): Focused on word parts that carry meaning (prefixes, roots, suffixes).
    - Example: "un-happy", showing the prefix and root.
- Orthographic: Follows standard writing conventions.
    - "week-end" in English, following writing conventions.
- Semantic: Considers context to prevent ambiguous or inappropriate breaks.
    - Avoiding "re-sign" when "resign" is intended to keep the meaning clear.
- Style: Hyphenation can also be a matter of style, where some divisions sound better or are more elegant.
    - "coop-eration" is preferable over "co-operation".

These criteria might differ or be weighted differently across languages:

- Some languages might prioritize phonological rules over others, like Spanish and Portuguese.
- In languages with complex compounding, like German, morphological considerations might be more prominent.

These criteria have evolved or changed with language reforms or in specific linguistic contexts.

It is complex to apply these criteria, especially in multilingual contexts or with non-standard dialects.

These criteria might differ or be weighted differently across languages:

- Some languages might prioritize phonological rules over others, like Spanish and Portuguese.
- In languages with complex compounding, like German, morphological considerations might be more prominent.

These criteria have evolved or changed with language reforms or in specific linguistic contexts.

It is complex to apply these criteria, especially in multilingual contexts or with non-standard dialects.

# Hyphenation rules in TeX

Some English rules:

```
.ach4
.ad4der
.af1t
.al3t
.am5at
.an5c
.ang4
.ani5m
.ant4
.an3te
.anti5s
.ar5s
.ar4tie
```

## Example

```
      .   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       ck1
max:  0   2   2   0   3   0   4   0   1
final: s   h   e   r - l   o   c   k -
```

# TeX hyphenation rules for Portuguese

- 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)