

Automatic Hyphenation

Leonardo Araujo

UFSJ



Introduction

- Historical evolution of hyphenation:
 - Early texts: No hyphenation marks.
 - Introduction of hyphenation for readability.
- 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.

TeX's role

- Automatic hyphenation since 1977.
- 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”.

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)