

T_EX and Her Friend: Hyphenation

It's better to break a word with a hyphen¹ than to stretch interword² spaces too much. Therefore T_EX tries to divide words into syllables³ when there's no good alternative⁴ available.

But computers are notoriously⁵ bad at hyphenation. When the typesetting⁶ of newspapers began to be fully automated, jokes about "the-rapists who pre-ached on wee-knights" soon began to spread¹.

It's not hard to understand why machines have behaved poorly at this task, because hyphenation is quite a difficult problem. For example, the word 'record' is supposed to be broken as 'record' when it is a noun, but 're-cord' when it is a verb. The word 'hyphenation' itself is somewhat extraordinary²; if 'hy-phen-a-tion' is compared to similar words like 'con-cat-e-na-tion', it's not immediately clear why the 'n' should be attached to the 'e' in one case but not the other. Examples like 'dem-on-stra-tion' vs. 'de-mon-stra-tive' show that the change³ of two letters can actually affect hyphens that are nine positions away.

A good solution to the problem was discovered by Frank M. Liang during 1980-1982, and T_EX includes⁴ the new method. Liang's algorithm⁷ works quickly and finds nearly all of the reasonable⁵ places to insert hyphens; yet it makes few if any errors, and it takes up comparatively⁸ little space in the computer. Moreover, the method is flexible enough to be adapted to any language, and it can also be used to hyphenate words in two languages at the same time⁶. Liang's Ph.D. thesis⁹, published by Stanford University's Department of Computer Science in 1983, explains how to take a dictionary of hyphenated words and teach it to T_EX; i.e., it explains how to compute tables by which T_EX will be able to reconstruct most of the hyphens in the given dictionary, without error.

*If all problems of hyphenation have not been solved,
at least some progress has been made since that night,
when according to legend, an RCA Marketing Manager received
a phone call from a disturbed customer. His 301 had just hyphenated "God".*

— PAUL E. JUSTUS, *There's More to Typesetting Than Setting Type* (1972)

*The committee skeptically re-
commended more study for a bill
to require warning labels on rec-
ords with subliminal messages re-
corded backward.*

— THE PENINSULA TIMES TRIBUNE (April 28, 1982)

Appendix: The Name of the Game

English words like 'technology' originate⁷ from a Greek root beginning with the letters τεχ...; and this same Greek word means *art* as well as technology. Hence the name T_EX, which is an uppercase form of τεχ.

Insiders¹⁰ pronounce the χ of T_EX as a Greek chi, not as an 'x', so that T_EX rhymes with the word *blecchhh*. It's the 'ch' sound in Scottish words like *loch* or German words like *ach*; it's a Spanish 'j' and a Russian 'kh'. When you say it correctly to your computer, the terminal may become slightly¹¹ moist¹².

The purpose of this pronunciation exercise is to remind you that T_EX is primarily concerned with high-quality technical manuscripts¹³. Its emphasis is on art and technology, as in the underlying Greek word. If you merely want to produce a satisfactorily⁸ good document – something acceptable and basically readable but not really beautiful – a simpler system will usually do⁹. With T_EX the goal is to produce the finest quality; this requires more attention to detail, but you will not find it much harder to go the extra distance, and you'll be able to take special pride in the finished product.

*They do certainly give
very strange and new-frangled names to diseases.*

— PLATO, *The Republic*, Book 3 (c. 375 B.C.)

*Technique! The very word is like the shriek
Of outraged Art. It is the idiot name
Given to effort by those who are too weak,
Too weary, or too dull to play the game.*

— LEONARD BACON, *Sophia Trenton* (1920)

¹ 断词用连字符

² 单词间的[术语]

³ 音节

⁴ 可供替代的

⁵ 众所周知的;臭名昭著的

⁶ 排版;组版

⁷ circulate

² exceptional

³ alteration

⁴ incorporates

⁷ 算法[计]

⁵ legitimate

⁸ 相对地;相较而言地

⁶ simultaneously

⁹ 论文

⁷ stem

¹⁰ 知情之人

¹¹ 略微;稍微

¹² 潮湿而微雾

¹³ 原稿;手稿

⁸ passably

⁹ suffice