

# T<sub>E</sub>X and Her Friend: Hyphenation

It's b\_\_\_\_\_ to break a word with a hyphen<sup>1</sup> than to stretch interword<sup>2</sup> spaces too much. Therefore T<sub>E</sub>X tries to divide words into syllables<sup>3</sup> when there's no good alternative<sup>4</sup> a\_\_\_\_\_.

But computers are notoriously<sup>5</sup> bad at hyphenation. When the typesetting<sup>6</sup> of newspapers began to be f\_\_\_\_\_ automated, jokes about "the-rapists who pre-ached on wee-knights" soon began to spread<sup>7</sup>.

It's not hard to understand why machines have behaved poorly at this t\_\_\_\_\_, because hyphenation is quite a difficult problem. For example, the word 'record' is s\_\_\_\_\_ to be broken as 'rec-ord' when it is a noun, but 're-cord' when it is a verb. The word 'hyphenation' itself is somewhat extraordinary<sup>8</sup>; if 'hy-phen-a-tion' is compared to s\_\_\_\_\_ words like 'con-cat-e-na-tion', it's not immediately clear why the 'n' should be a\_\_\_\_\_ to the 'e' in one case but not the other. Examples like 'dem-on-s-tra-tion' vs. 'de-mon-s-tra-tive' show that the change<sup>9</sup> of two letters can actually affect hyphens that are nine p\_\_\_\_\_ away.

A good solution to the problem was discovered by Frank M. Liang during 1980-1982, and T<sub>E</sub>X includes<sup>4</sup> the new method. Liang's algorithm<sup>7</sup> works quickly and finds nearly all of the reasonable<sup>5</sup> places to insert hyphens; yet it makes few if any errors, and it takes up comparatively<sup>8</sup> little space in the computer. Moreover, the method is f\_\_\_\_\_ enough to be adapted to any language, and it can also be used to hyphenate words in two languages at the same time<sup>6</sup>. Liang's Ph.D. thesis<sup>9</sup>, published by Stanford University's Department of Computer Science in 1983, explains how to take a dictionary of hyphenated words and t\_\_\_\_\_ it to T<sub>E</sub>X; i.e., it explains how to compute tables by which T<sub>E</sub>X 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' originate<sup>7</sup> from a Greek root beginning with the letters τεχ...; and this same Greek word means *art* as well as technology. Hence the name T<sub>E</sub>X, which is an uppercase f\_\_\_\_\_ of τεχ.

Insiders<sup>10</sup> pronounce the χ of T<sub>E</sub>X as a Greek chi, not as an 'x', so that T<sub>E</sub>X 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 c\_\_\_\_\_ to your computer, the terminal may become slightly<sup>11</sup> moist<sup>12</sup>.

The purpose of this pronunciation e\_\_\_\_\_ is to remind you that T<sub>E</sub>X is primarily concerned with high-quality technical manuscripts<sup>13</sup>: Its emphasis is on art and technology, as in the underlying Greek word. If you merely want to produce a satisfactorily<sup>8</sup> good document – something acceptable and basically readable but not really beautiful – a simpler system will usually do<sup>9</sup>. With T<sub>E</sub>X 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 p\_\_\_\_\_ 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)

<sup>1</sup> 断词用连字符

<sup>2</sup> 单词间的[术语]

<sup>3</sup> 音节

<sup>4</sup> 可供替代的

<sup>5</sup> 众所周知的;臭名昭著的

<sup>6</sup> 排版;组版

<sup>7</sup> circulate

<sup>8</sup> exceptional

<sup>9</sup> alteration

<sup>4</sup> incorporates

<sup>7</sup> 算法[计]

<sup>5</sup> legitimate

<sup>8</sup> 相对地;相较而言地

<sup>6</sup> simultaneously

<sup>9</sup> 论文

<sup>7</sup> stem

<sup>10</sup> 知情之人

<sup>11</sup> 略微;稍微

<sup>12</sup> 潮湿而微雾

<sup>13</sup> 原稿;手稿

<sup>8</sup> passably

<sup>9</sup> suffice