
Hyphenation Exception Log

Barbara Beeton

This is the periodic update of the list of words that \TeX fails to hyphenate properly. The full list last appeared in *TUGboat* 16, no. 1, starting on page 12, with an update in *TUGboat* 22, no. 1/2, pages 31–32. The present list contains exceptions found since the last update.

This list is specific to the hyphenation patterns that appear in the original `hyphen.tex`, that is, the patterns for US English. It is been subdivided into two parts: English words, and names and non-English words that occur in English texts. The first listed column gives results from \TeX 's `\showhyphens{...}`; entries in the second column are suitable for inclusion in a `\hyphenation{...}` list. In most instances, inflected forms are not shown for nouns and verbs; note that all forms must be specified in a `\hyphenation{...}` list if they occur in your document.

Thanks to all who have submitted entries to the list. Since some suggestions demonstrated a lack of familiarity with the rules of the hyphenation algorithm, here is a short reminder of the relevant idiosyncrasies. Hyphens will not be inserted before the number of letters specified by `\lefthyphenmin`, nor after the number of letters specified by `\righthyphenmin`. For US English, `\lefthyphenmin=2` and `\righthyphenmin=3`; thus no word shorter than five letters will be hyphenated. (For the details, see *The \TeX book*, page 454. For a digression on other views of hyphenation rules, see below under “English hyphenation”.) This particular rule is violated in some of the words listed; however, if a word is hyphenated correctly by \TeX except for “missing” hyphens at the beginning or end, it has not been included here.

Some other permissible hyphens have been omitted for reasons of style or clarity. While this is at least partly a matter of personal taste, an author should think of the reader when deciding whether or not to permit just one more break-point in some obscure or confusing word. There really are times when a bit of rewriting is preferable.

One other warning: Some words can be more than one part of speech, depending on context, and have different hyphenations; for example, ‘analyses’ can be either a verb or a plural noun. If such a word appears in this list, hyphens are shown only for the portions of the word that would be hyphenated the same regardless of usage. These words are marked with a ‘*’; additional hyphenation points, if

needed in your document, should be inserted with discretionary hyphens.

The reference used to check these hyphenations is *Webster's Third New International Dictionary*, Unabridged.

English hyphenation

It has been pointed out to me that the hyphenation rules of British English are based on the etymology of the words being hyphenated as opposed to the “syllabic” principles used in the U.S. Furthermore, in the U.K., it is considered bad style to hyphenate a word after only two letters. In order to make T_EX defer hyphenation until after three initial letters, set `\lefthyphenmin=3`.

Of course, British hyphenation patterns should be used as well. A set of patterns for UK English has been created by Dominik Wujastyk and Graham Toal, using Frank Liang's PATGEN and based on a file of 114925 British-hyphenated words generously made available to Dominik by Oxford University Press. (This list of words and the hyphenation break points in the words are copyright to the OUP and may not be redistributed.) The file of hyphenation patterns may be freely distributed; it is posted on CTAN in the file `tex-archive/language/hyphenation/ukhyph.tex` and can be retrieved by anonymous FTP or using a Web browser.

Hyphenation for languages other than English

Patterns now exist for many languages other than English, including languages using accented alphabets. CTAN holds an extensive collection of patterns in `tex-archive/language/hyphenation` and its subdirectories.

Converting this list into a list of hyphenation exceptions

Werner Lemberg has created a script that will convert this article into a real `\hyphenation` block that can be incorporated into a document either directly or by inputting a file. His work has necessitated some changes to the macros used to format the list, but the appearance of the list will not change. Many inflected forms will be included

automatically, some evident in the printed version, but many included silently.

The script, `hyphenex.sh`, runs under Unix and is posted on CTAN, in `tex-archive/info/digests/tugboat/hyphenex.sh`

The List — English words

<code>an-ti-holo-mor-ptic</code>	anti-holo-mor-ptic
<code>au-toround-ing</code>	auto-round-ing
<code>biomed-i-cal</code>	bio-med-i-cal
<code>book-seller</code>	book-sell-er
<code>caller</code>	call-er
<code>coas-so-cia-tive</code>	co-asso-cia-tive
<code>dam-selfly(ies)</code>	dam-sel-fly(ies)
<code>falling</code>	fall-ing
<code>geostrophic</code>	ge-o-strophic
<code>halfway</code>	half-way
<code>hotbed</code>	hot-bed
<code>id-iosyn-cratic(ally)</code>	idio-syn-cratic(-i-cal-ly)
<code>in-stallers</code>	in-stall-er
<code>lovestruck</code>	love-struck
<code>metabo-lite</code>	me-tab-o-lite
<code>nephews</code>	neph-ews
<code>nephrite(ic)</code>	neph-rite(ic)
<code>pheny-lala-nine</code>	phenyl-ala-nine
<code>pho-tooff-set</code>	pho-to-off-set
<code>pro-peller</code>	pro-pel-ler
<code>pro-pelling</code>	pro-pel-ling
<code>quasiequiv-a-lent</code>	qua-si-equiv-a-lent
<code>rect-an-gu-lar</code>	rec-tan-gu-lar
<code>seller</code>	sell-er
<code>speller</code>	spell-er
<code>spelling</code>	spell-ing
<code>teller</code>	tell-er
<code>wingspan</code>	wing-span

Names and non-English words used in English text

<code>Florid-ian</code>	Flor-i-d-ian
<code>Ghostscript</code>	Ghost-script
<code>GhostView</code>	Ghost-View
<code>Grass-man-nian</code>	Grass-mann-ian
<code>Lo-jban</code>	Loj-ban
<code>Maxwell</code>	Max-well
<code>Mi-crosoft</code>	Mi-cro-soft
<code>Netscape</code>	Net-scape
<code>Palatino</code>	Pala-tino
<code>Pfaf-fian</code>	Pfaff-ian
<code>philosophis-che</code>	phil-o-soph-i-sche
<code>Rad-hakr-ish-nan</code>	Ra-dha-krish-nan

◇ Barbara Beeton
American Mathematical Society
`bnb@ams.org`