



**HAL**  
open science

## TeX and Latin Alphabet Languages

Yannis Haralambous

► **To cite this version:**

| Yannis Haralambous. TeX and Latin Alphabet Languages. Tugboat, 1989, 10 (3), pp.342-345. <hal-02095944>

**HAL Id: hal-02095944**

**<https://hal.science/hal-02095944v1>**

Submitted on 23 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

## Fonts

### Erratum:

#### Typesetting Concrete Mathematics TUGboat Vol. 10, No. 1

Donald E. Knuth

Editor's note: Some erroneous parameter values for the Concrete fonts were given in the footnote on page 33. The following are the correct values. *beak\_darkness* = 4/30 for all Concrete fonts except for *cslc9*, which has *beak\_darkness* = 5/30. Also for *cslc9*, *fudge* = 1.

The METAFONT sources and the macros for using the Concrete fonts can be found in the *tex-fonts* archive at Clarkson. A list of the files appears on page 200, *TUGboat* Vol. 10, No. 2.

---

### TeX and latin alphabet languages

Yannis Haralambous

(Ἐκείνη) μιλάει τή γλώσσα τῶν  
Ἑτρούσκων, μιάν ἄγνωστη, παράωνη  
γλώσσα, πού κανένας δέν τήν ξέρει καί  
κανένας ποτέ δέν θά τήν μιλήσει.

—Κωστούλα Μητροπούλου<sup>1</sup>

TeX provides enough accents and special letters to write in (almost) every latin alphabet language. But there are problems; following D. E. Knuth,<sup>2</sup> "... format designers should provide rules for how to obtain accents and special characters in their particular systems. plain TeX works well enough when accents are infrequent, but (its) conventions are by no means recommended for large-scale applications of TeX to other languages." The best solution to this is to create accented and special letters by METAFONT and to put them in the still-empty 128 boxes of the 256-code fonts.\*

---

\* Editor's note: A technique developed long ago by David Fuchs, but never publicized, permits the creation of "virtual fonts" that have instructions for putting pieces together, at the output driver level, from the same or separate fonts, so that detailed character descriptions are not required for such

But to keep the "universal properties" of TeX, this should be done (a) once and for all, and (b) in the most democratic and complete way. It is important that—in an international level—everybody has the same fonts and ligature conventions (to prevent having for example *un TeX français* and *ein deutsches TeX* with no compatibility to each other).

Now the natural question to ask is: is it possible to cover ALL latin alphabet languages? To answer this question I have searched (for some hours) in the library of the Centre Pompidou in Paris and made a list of languages with their respective accented letters and special characters (beside the well-known 26-letter alphabet). I am neither a linguist, nor a geographer, and this makes me pretty sure that many languages and dialects are missing. So I would be very grateful for additional remarks and corrections (just write to [yannis@frcit171.bitnet](mailto:yannis@frcit171.bitnet)). Also I would like to thank my friends Ana Cristina Matos (Portugal), Vladimir Rosenhaus (Estonia, USSR) and Petr Plecháč (Czechoslovakia) for their precious help.

### List of languages

1. Albanian<sup>3</sup>: č Č, ë Ě
2. Basque<sup>4</sup>: ñ Ñ
3. Breton<sup>5</sup>: é É, è È, ê Ê, ñ Ñ, ô Ô, ü Ü
4. Catalan<sup>6</sup>: à À, ç Ç, é É, è È, í Í, ó Ó, ò Ò, ú Ú, ü Ü
5. Corsican<sup>7</sup>: à À, é É, è È, î Î, ô Ô, ù Ù, ú Û
6. Creole<sup>8</sup>: ä Ä, â Â, ã Ã, é É, ë Ě, ê Ê, ē Ē, ě Ě, ĭ Ĭ, ñ Ñ, ò Ò, ö Ö, õ Ö, ø Ø, ü Ü, ũ Û, ú Ú, ŷ Ÿ
7. Croatian<sup>28</sup>: é Ć, č Č, đ Đ, š Š, ž Ž
8. Czech: á Á, č Č, ď Ď, é É, ě Ě, í Í, ñ Ñ, ó Ó, ř Ř, š Š, ť Ě, ú Ú, ů Ů, ý Ý, ž Ž
9. Danish<sup>9</sup>: å Å, æ Æ, ø Ø
10. Dutch<sup>10</sup>: é É, ë Ě
11. English: none
12. Esperanto<sup>11</sup>: ĉ Ĉ, ĝ Ĝ, ĥ Ĥ, ĵ Ĵ, ŝ Ŝ, ŭ Ŭ
13. Estonian<sup>12</sup>: ä Ä, õ Ö, ö Ö, š Š, ü Ü, ž Ž
14. Faroese<sup>9</sup>: á Á, æ Æ, đ Þ, í Í, ó Ó, ø Ø, ú Ú, ý Ý
15. Finnish<sup>13</sup>: ä Ä, ö Ö
16. French: à À, â Â, ç Ç, é É, è È, ê Ê, ë Ě, î Î, ï Ĭ, ô Ô, œ Œ, ù Ù, ú Ú, ü Ü (and it seems that in pre-Louis XIV texts<sup>14</sup> there is also the letter ŷ Ÿ)

---

"composite" characters. Arrangements are being made to publish the details in a future TUGboat.

17. Gaelic<sup>15</sup>: à À, é É, è È, ì Ì, ó Ó, ò Ò, ù Ù  
 18. Galician: á Á, é É, í Í, ñ Ñ, ó Ó, ú Ú  
 19. German: ä Ä, ö Ö, ß, ü Ü  
 20. Guarani<sup>16</sup>: á Á, ä Ä, é É, ë È, í Í, ì Ì, ñ Ñ, ó Ó, ö Ö, ú Ú, ü Ü, ý Ý, ÿ Ÿ  
 21. Hungarian<sup>17</sup>: á Á, é É, í Í, ó Ó, ö Ö, ő Ő, ú Ú, ü Ü, ű Ű  
 22. Icelandic<sup>9</sup>: á Á, æ Æ, đ Þ, é É, í Í, ó Ó, ö Ö, ú Ú, ý Ý  
 23. Indonesian<sup>18</sup>: none  
 24. Irish<sup>19</sup>: *irish language has its own, most beautiful alphabet*  
 25. Italian: à À, é É, è È, í Í, ì Ì, ò Ò, ú Ú, ù Ù  
 26. Kurdish<sup>20</sup>: ç Ç, ê Ê, ê Ê, î Î, ş Ş, û Û, x X  
 27. Latin: none (sometimes æ Æ is used)  
 28. Latvian<sup>12,21</sup>: ā Ā, č Č, ē Ē, ģ Ģ, ī Ī, ķ Ķ, ļ Ļ, ņ Ņ, š Š, ū Ū, ž Ž  
 29. Lithuanian<sup>12,22</sup>: ą Ą, č Č, e E, é É, į Į, š Š, ū Ū, ž Ž  
 30. Maltese<sup>23</sup>: ċ Ċ, ġ Ġ, ħ Ħ, ż Ż  
 31. Norwegian<sup>9</sup>: å Å, æ Æ, ø Ø  
 32. Occitan<sup>24</sup>: ç Ç, é É, è È, í Í, ì Ì, ó Ó, ò Ò, ú Ú  
 33. Polish<sup>25</sup>: ą Ą, ć Ć, e E, ł Ł, ń Ń, ó Ó, ś Ś, ź Ż, ż Ż  
 34. Portuguese: á Á, à À, â Â, ã Ã, ç Ç, é É, ê Ê, í Í, ó Ó, ô Ô, õ Õ, ú Ú, ü Ü  
 35. Qhëshwa<sup>26</sup>: á Á, é É, ë È, í Í, ñ Ñ, ó Ó, ö Ö, ú Ú, ü Ü  
 36. Romansh<sup>7</sup>: á Á, ó Ó, ö Ö, ù Ù, ü Ü...?  
 37. Romanian<sup>27</sup>: â Â, ă Ă, î Î, ș Ș, ț Ț  
 38. Slovakian: à À, č Č, đ Ď, é É, ě Ě, í Í, ĺ Ľ, ň Ň, ó Ó, ô Ô, ř Ř, š Š, t T, ú Ú, ů Ů, ý Ý, ž Ž  
 39. Slovenian<sup>29</sup>: č Č, š Š, ž Ž  
 40. Spanish: á Á, é É, í Í, ñ Ñ, ó Ó, ú Ú, ü Ü  
 41. Swedish<sup>9</sup>: ä Ä, å Å, ö Ö  
 42. Turkish<sup>30</sup>: â Â, ç Ç, ğ Ğ, ı I, î İ, î Î, ö Ö, ş Ş, û Ü, ü Ü  
 43. Vietnamese<sup>31</sup>: â Â, ă Ă, đ Đ, ê Ê, ô Ô, o O, u U and the following accents: ‘, ’, ‘, ’, ..  
 44. Welsh<sup>32</sup>: â Â, è È, ê Ê, í Í, ì Ì, ô Ô, ö Ö, û Ü, w W, y Y

and finally,

MR transliteration of USSR cyrillic: è È, ë È, ì Ì for э Э, ë È, н Њ.

## Conclusions

There are 190 different symbols on this list, and 8 (æ Æ ı ø Ø œ Æ ß) are already in the CM fonts. So the remaining 182 are far too many to be included in the 128 empty boxes. But, looking closer, we see that for the most important “european” languages (languages of the European Community, central europe socialist countries and Turkey) we only need 112 symbols from this list, and automatically most of the other languages are covered too. So I think the best thing to do is to take these 112 symbols and leave the 16 empty boxes for arbitrary use (also I wait for your comments and suggestions). These 112 symbols are the following:

á Á, à À, â Â, ä Ä, ã Ã, ą Ą, å Å, ă Ă  
 é É, ç Ç, ċ Ć  
 đ Ď, đ Þ Đ  
 é É, è È, ê Ê, ë È, e E, e Ě  
 ğ Ğ  
 í Í, ì Ì, î Î, ï Ï, ï Ï, ï  
 ĺ Ľ, ĺ Ľ  
 ń Ń, ñ Ñ, ñ Ñ  
 ó Ó, ò Ò, ô Ô, ö Ö, õ Õ, ő Ő, o O  
 ř Ř  
 ś Ś, ș Ș, ș Š  
 ț Ț, ț T  
 ú Ú, ù Ù, û Ü, ü Ü, ű Ű, ů Ů, u U  
 ý Ý, ŷ Ÿ  
 ź Ż, ż Ż, ż Ż

## Which ligatures to choose

Once we have the new font, the problem is to assign the right ligatures, according to the three basic rules:

- compatibility with T<sub>E</sub>X instructions (for example, ~n is not admissible for ñ, since ~ is active).
- compatibility with 128-character fonts (the example

```
e'crire de cette manie're
nai"ve en franc/ais
```

disobeys this rule; to use it we would have to change the use of ', ', and /, and texts written in “classical” T<sub>E</sub>X couldn't be typeset anymore).

Taking (a) and (b) into account, the only ASCII characters we can use are <, >, and, by slightly disobeying rule (b), we could (and will have to) take also ", +, and = (this could be done because

' ' also gives " , and \$\$\$ and \$=\$ are even better choices for + and =).

(c) readability.

The solution I propose is the following:

- 1) take < and > for the acute and grave accents respectively (they are easy to find on the keyboard, symmetric, and give the direction of the accent).
- 2) take = for the ^ accent; also take =n for ñ.
- 3) take " for " on vowels. Also take "s for ß, following Part1<sup>33</sup>.
- 4) take the plus sign + for both the háček and tilde accents (the portuguese language doesn't have háčeks and czech has no tildes, so there is no essential risk of confusion). So for example +a means ā, and +z, ž (this is why we took =n for ñ: it's the only letter that takes both háček and tilde accents).
- 5) take > also for the "french-style" cedilla on c, s, t and == for the "polish-style" hook on a and e (a is a very wicked letter; it takes almost every existing diacritical mark).
- 6) since <o, <u give ó, ú, it's reasonable to put <<o, <<u for ő, ű.
- 7) take " before d, l, t for đ, ľ, Ț; take <>o and <>u for ȝ, ȕ (o is at least as wicked as a).
- 8) the rest of the distribution is purely random.

You can see the whole of it in the following table:

<	followed by	$\left\{ \begin{array}{l} a A \\ c C \\ e E \\ i I \\ n N \\ o O \\ s S \\ u U \\ y Y \\ z Z \end{array} \right\}$	gives	$\left\{ \begin{array}{l} á Á \\ é É \\ í Í \\ ñ Ñ \\ ó Ó \\ ś Ś \\ ú Ú \\ ý Ý \\ ź Ź \end{array} \right\}$
>	followed by	$\left\{ \begin{array}{l} a A \\ c C \\ e E \\ i I \\ o O \\ s S \\ t T \\ u U \end{array} \right\}$	gives	$\left\{ \begin{array}{l} à À \\ ç Ç \\ è È \\ ì Ì \\ ò Ò \\ ş Ş \\ Ț Ț \\ ù Ù \end{array} \right\}$

=	followed by	$\left\{ \begin{array}{l} a A \\ e E \\ i I \\ n N \\ o O \\ u U \end{array} \right\}$	gives	$\left\{ \begin{array}{l} â Â \\ ê Ê \\ î Î \\ ñ Ñ \\ ô Ô \\ û Û \end{array} \right\}$
---	-------------	--	-------	--

"	followed by	$\left\{ \begin{array}{l} a A \\ d D \\ e E \\ i I \\ l L \\ o O \\ s \\ t T \\ u U \\ y Y \\ z Z \end{array} \right\}$	gives	$\left\{ \begin{array}{l} ä Ä \\ đ Đ \\ ë È \\ ï Ì \\ ľ Ľ \\ ö Ö \\ ß \\ Ț Ț \\ ü Û \\ ý Ý \\ ź Ź \end{array} \right\}$
---	-------------	---	-------	---

+	followed by	$\left\{ \begin{array}{l} a A \\ c C \\ d D \\ e E \\ g G \\ i I \\ l L \\ n N \\ o O \\ r R \\ s S \\ z Z \end{array} \right\}$	gives	$\left\{ \begin{array}{l} ā Ā \\ ċ Ć \\ đ Đ \\ ē Ę \\ ğ Ğ \\ ĩ Ĭ \\ ł Ł \\ ñ Ñ \\ ō Ŏ \\ ř Ř \\ ś Ś \\ ź Ź \end{array} \right\}$
---	-------------	--	-------	--

==	followed by	$\left\{ \begin{array}{l} a A \\ e E \end{array} \right\}$	gives	$\left\{ \begin{array}{l} a A \\ e E \end{array} \right\}$
----	-------------	--	-------	--

++	followed by	$\left\{ \begin{array}{l} a A \\ D \\ o O \end{array} \right\}$	gives	$\left\{ \begin{array}{l} æ Æ \\ b \\ ø Ø \end{array} \right\}$
----	-------------	---	-------	---

<<	followed by	$\left\{ \begin{array}{l} a A \\ o O \\ u U \end{array} \right\}$	gives	$\left\{ \begin{array}{l} ă Ă \\ ő Ő \\ ű Ű \end{array} \right\}$
----	-------------	---	-------	---

<>	followed by	$\left\{ \begin{array}{l} o O \\ u U \end{array} \right\}$	gives	$\left\{ \begin{array}{l} ȝ Ȥ \\ ȕ Ȥ \end{array} \right\}$
----	-------------	--	-------	--

""	followed by	$\left\{ \begin{array}{l} a A \\ i I \\ o O \\ u U \end{array} \right\}$	gives	$\left\{ \begin{array}{l} â Â \\ ı İ \\ œ Œ \\ û Û \end{array} \right\}$
----	-------------	--	-------	--

## An example

By typing

Zde se v+semo+zn+e sna+z<i m+e p+remluvit,  
 abych z""ustal je+st+e n+ekolik m+es<ic""u  
 a napsal je+st+e jednu operu. Hay""ir!  
 ""I>s "oyle de+gil. B"uy"u+g"u  
 k"u>c"u+g"une takilmay""i pek severdi.  
 Ce f=ut d'ores et d<ej>a une id<ee  
 d<eg<en<er<ee et ambig"ue.

instead of

Zde se v\v semo\v zn\v e sna\v z'\i\ m\v  
 e p\v reluvit, abych z{\accent'027u}stal  
 je\v st\v e n\v ekolik m\v es'\i  
 c{\accent'027u} a napsal je\v st\v e jednu  
 operu. Hay|i r! \D I\c s \"oyle de\u  
 gil. B\"uy\"u\"u g\"u k\"u\"c c\"u\"u g\"u  
 takilmay|i\ pek severdi. Ce f\ut d'ores  
 et d\ej\"a une id\ee d\eg\en\er\ee  
 et ambig\"ue.

(which looks better?) you get

Zde se všemožně snaží mě přeluvit, abych zůstal  
 ještě několik měsíců a napsal ještě jednu operu.  
 Hayır! İş öyle değil. Büyüğü küçüğüne takilmayı  
 pek severdi. Ce fut d'ores et déjà une idée dégénérée  
 et ambiguë.

## References

1. K. Mhtropo'ulou, *Za'ar 19*, K'edroc, ĵAj'hna 1980.
2. D. E. Knuth, *The TeXbook*, Addison-Wesley, 1983.
3. H. Boissin, *Grammaire de l'Albanais moderne*, chez l'Auteur, Paris 1975.
4. X. Kintara, *Diccionario vasco-español*, Hiztegia bi mila, Bilbo 1984.
5. E. Breiz, *Dictionnaire Breton*, Garnier, Paris 1986.
6. B. Margarit, *Gramatica Catalana*, Gredos, Madrid, 1962.
7. P. Ciavatti (Ed.), *Dizziunariu Corsu-Francese*, Albiana, Levie 1985.
8. *this alphabet comes from a project of R. Germain*, cf. R. Germain, *Grammaire créole*, l'Harmattan, Paris 1980.
9. M. O'C. Walshe, *Introduction to the Scandinavian Languages*, Andre Deutsch, London 1965.
10. F. van Passel, *Le Néerlandais pour tous*, Presses Pocket, Paris 1983.
11. L. L. Zamenhof, *Fundamento de Esperanto*, Esperantoj Francaj Eldonoj, Marmande 1963.
12. *communicated to me by V. Rosenhaus*.
13. H. Fromm & M. Sadeniemi, *Finnisches Elementarbuch*, Carl Winter Universitätsverlag, Heidelberg 1956.
14. *as for example in the lovely text:*  
*Le dit du Saige trop le feiz*  
*Favorable, bien n'en puis mais*  
*Qui dit: "Esjoÿs toy, mon filz,*  
*En ton adolescence"; mais*

*Ailleurs sert bien d'ung autre mes,*  
*Car "Jeunesse et adolescence",*  
*C'est son parler, ne moins ne mais,*  
*"Ne sont qu'abus et ignorance".*

taken from F. Villon (1431-1463?), *Le Grand Testament*, dtv 2078, München 1982.

15. G. Calder, *A Gaelic Grammar*, Gairn Publications, Glasgow 1972.
16. B. M. Lliteras, A. P. Peñasco & L. F. Maluquer, *El Guaraní a su Alcance*, Loyola, Asuncion 1976.
17. G. Darai, *Universal-Wörterbuch Ungarisch*, Langenscheidt, Berlin 1988.
18. A. L. N. Kramer, *Kamus Inggeris Ketijl*, Ch. E. Tuttle, Tokyo 1975.
19. Rev. P. S. Dinneen, *Irish-English Dictionary*, The educational company of Ireland, Dublin 1927.
20. E. D. B. Khan & R. Lesot, *Grammaire Kurde*, Librairie d'Amérique et d'Orient, Paris 1970.
21. A. Гутмание, *Русско-Латышский Разговорник*, Лиесма, Рига 1975.
22. F. Bielenstein, *Lettische Grammatik*, Zentralantiquariat der DDR, Leipzig 1972.
23. S. Paterson, *Guide de voyage: Malte*, Berlitz, Lausanne 1987.
24. R. Barthe, *Lexique Français-Occitan*, Collection des Amis de la Langue d'Oc, Paris 1973.
25. L. Bielas, *Słownik minimum Polsko-Francuski*, Wiedza Powszechna, Warszawa 1987.
26. J. Lara, *Diccionario Qhëshwa-Castellano*, Los Amigos del Libro, La Paz 1978.
27. N. P. Condeescu, *Dicționar Francez-Român*, Editura Științifică, București 1967.
28. V. Javarek & M. Sudijć, *Serbo Croat*, Teach Yourself Books, London 1972.
29. F. Tomšič, *Slovensko-Nemški Slovar*, Državna Založba Slovenije, Ljubljana 1973.
30. H. Wendt, *Praktisches Lehrbuch Türkisch*, Langenscheidt, Berlin 1979.
31. N. D. Hoa, *Vietnamese-English Dictionary*, Ch. E. Tuttle, Tokyo 1966.
32. S. J. Williams, *A Welsh Grammar*, University of Wales Press, Cardiff 1980.
33. H. Partl, German TeX, *TUGboat* 9 (1988) 70-72.

◇ Yannis Haralambous  
 U.F.R. de Mathématiques  
 Université de Lille-Flandres-Artois  
 59655 Villeneuve d'Ascq Cedex  
 France  
 Bitnet: yannis@frcit171