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 ccslc9, which has *beak_darkness* = 5/30. Also for ccslc9, *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.

T$_E$X and latin alphabet languages

Yannis Haralambous

> ('Εκείνη) μιλάει τή γλώσσα τῶν
> 'Ετρούσκων, μιάν ἄγνωστη, παράωενη
> γλώσσα, πού κανένας δέν τήν ξέρει καί
> κανένας ποτέ δέν θά τήν μιλήσει.
>
> —Κωστούλα Μητροπούλου[1]

T$_E$X provides enough accents and special letters to write in (almost) every latin alphabet language. But there are problems; following D. E. Knuth,[2] "... format designers should provide rules for how to obtain accents and special characters in their particular systems. plain T$_E$X works well enough when accents are infrequent, but (its) conventions are by no means recommended for large-scale applications of T$_E$X 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 T$_E$X, 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 T$_E$X français* and ein Deutsches T$_E$X 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). 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**[3]: č Č, ë Ë
2. **Basque**[4]: ñ Ñ
3. **Breton**[5]: é É, è È, ê Ê, ñ Ñ, ô Ô, ü Ü
4. **Catalan**[6]: à À, ç Ç, é É, è È, í Í, ó Ó, ò Ò, ú Ú, ü Ü
5. **Corsican**[7]: à À, é É, è È, î Î, ô Ô, ù Ù, û Û
6. **Creole**[8]: ä Ä, â Â, ã Ã, é É, ë Ë, ê Ê, ē Ē, ĕ Ĕ, ï Ï, ñ Ñ, ò Ò, ö Ö, õ Õ, ŏ Ŏ, ü Ü, ũ Ũ, ù Ù, ÿ Ÿ
7. **Croatian**[28]: ć Ć, č Č, đ Đ, š Š, ž Ž
8. **Czech**: á Á, č Č, ď Ď, é É, ě Ě, í Í, ň Ň, ó Ó, ř Ř, š Š, ť Ť, ú Ú, ů Ů, ý Ý, ž Ž
9. **Danish**[9]: å Å, æ Æ, ø Ø
10. **Dutch**[10]: é É, ë Ë
11. **English**: *none*
12. **Esperanto**[11]: ĉ Ĉ, ĝ Ĝ, ĥ Ĥ, ĵ Ĵ, ŝ Ŝ, ŭ Ŭ
13. **Estonian**[12]: ä Ä,, õ Õ, ö Ö, š Š, ü Ü, ž Ž
14. **Faroese**[9]: á Á, æ Æ, ð þ, í Í, ó Ó, ø Ø, ú Ú, ý Ý
15. **Finnish**[13]: ä Ä, ö Ö
16. **French**: à À, â Â, ç Ç, é É, è È, ê Ê, ë Ë, î Î, ï Ï, ô Ô, œ Œ, ù Ù, û Û, ü Ü (*and it seems that in pre-Louis XIV texts*[14] *there is also the letter* ÿ Ÿ)

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

17. **Gaelic**[15]: 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**[16]: á Á, ä Ä, é É, ë Ë, í Í, ï Ï, ñ Ñ, ó Ó, ö Ö, ú Ú, ü Ü, ý Ý, ÿ Ÿ
21. **Hungarian**[17]: á Á, é É, í Í, ó Ó, ö Ö, ő Ő, ú Ú, ü Ü, ű Ű
22. **Icelandic**[9]: á Á, æ Æ, đ þ, é É, í Í, ó Ó, ö Ö, ú Ú, ý Ý
23. **Indonesian**[18]: *none*
24. **Irish**[19]: *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**[20]: ç Ç, ê Ê, h̄ H̄, î Î, ş Ş, û Û, ẍ Ẍ
27. **Latin**: *none (sometimes æ Æ is used)*
28. **Latvian**[12,21]: ā Ā, č Č, ē Ē, ģ Ģ, ī Ī, ķ Ķ, ļ Ļ, ņ Ņ, š Š, ū Ū, ž Ž
29. **Lithuanian**[12,22]: ą Ą, č Č, ę Ę, ė Ė, į Į, š Š, ų Ų, ū Ū, ž Ž
30. **Maltese**[23]: ċ Ċ, ġ Ġ, ħ Ħ, ż Ż
31. **Norwegian**[9]: å Å, æ Æ, ø Ø
32. **Occitan**[24]: ç Ç, é É, è È, í Í, ì Ì, ó Ó, ò Ò, ú Ú
33. **Polish**[25]: ą Ą, ć Ć, ę Ę, ł L, ń Ń, ó Ó, ś Ś, ź Ź, 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**[26]: á Á, é É, ë Ë, í Í, ñ Ñ, ó Ó, ö Ö, ú Ú, ü Ü
36. **Romansh**[?]: á Á, ó Ó, ö Ö, ù Ù, ü Ü...?
37. **Romanian**[27]: â Â, ă Ă, î Î, ş Ş, ţ Ţ
38. **Slovakian**: à À, č Č, đ Ď, é É, ě Ě, í Í, ľ Ľ, ñ Ñ, ó Ó, ô Ô, ř Ř, š Š, ť Ť, ú Ú, ů Ů, ý Ý, ž Ž
39. **Slovenian**[29]: č Č, š Š, ž Ž
40. **Spanish**: á Á, é É, í Í, ñ Ñ, ó Ó, ú Ú, ü Ü
41. **Swedish**[9]: ä Ä, å Å, ö Ö
42. **Turkish**[30]: â Â, ç Ç, ğ Ğ, ı I, i İ, î Î, ö Ö, ş Ş, û Û, ü Ü
43. **Vietnamese**[31]: â Â, ă Ă, đ Đ, ê Ê, ô Ô, ơ Ơ, ư Ư and *the following accents:* ´,˜,`,`,.
44. **Welsh**[32]: â Â, è È, ê Ê, î Î, ï Ï, ô Ô, ö Ö, û Û, ŵ Ŵ, ŷ Ŷ

and finally,

MR transliteration of USSR cyrillic: è È, ë Ë, ĭ Ĭ *for* э Э, ë Ë, й Й.

Conclusions

There are 190 different symbols on this list, and 8 (æ Æ ı ø Ø œ Œ 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
ć Ć, ç Ç, č Č
đ Ď, đ þ Đ
é É, è È, ê Ê, ë Ë, ę Ę, ě Ě
ğ Ğ
í Í, ì Ì, î Î, ï Ï, ĭ Ĭ, İ
ľ Ľ, ł L
ń Ń, ñ Ñ, ň Ň
ó Ó, ò Ò, ô Ô, ö Ö, õ Õ, ő Ő, ơ Ơ
ř Ř
ś Ś, ş Ş, š Š
ţ Ţ, ť Ť
ú Ú, ù Ù, û Û, ü Ü, ů Ů, ű Ű, ư Ư
ý Ý, ÿ Ÿ
ź Ź, ž Ž, 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:

(a) compatibility with TEX instructions (for example, ~n is not admissible for ñ, since ~ is active).

(b) 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" TEX 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 ˆ accent; also take =n for ñ.

3) take " for ¨ on vowels. Also take "s for ß, following Part1[33].

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 ã, 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 ď, ľ, ť; take <>o and <>u for o̕, u̕ (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
$\left\{\begin{matrix} a & A \\ c & C \\ e & E \\ i & I \\ n & N \\ o & O \\ s & S \\ u & U \\ y & Y \\ z & Z \end{matrix}\right\}$
gives
$\left\{\begin{matrix} á & Á \\ ć & Ć \\ é & É \\ í & Í \\ ń & Ń \\ ó & Ó \\ ś & Ś \\ ú & Ú \\ ý & Ý \\ ź & Ź \end{matrix}\right\}$

> followed by
$\left\{\begin{matrix} a & A \\ c & C \\ e & E \\ i & I \\ o & O \\ s & S \\ t & T \\ u & U \end{matrix}\right\}$
gives
$\left\{\begin{matrix} à & À \\ ç & Ç \\ è & È \\ ì & Ì \\ ò & Ò \\ ş & Ş \\ ţ & Ţ \\ ù & Ù \end{matrix}\right\}$

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

" followed by
$\left\{\begin{matrix} a & A \\ d & D \\ e & E \\ i & I \\ l & L \\ o & O \\ s & \\ t & T \\ u & U \\ y & Y \\ z & Z \end{matrix}\right\}$
gives
$\left\{\begin{matrix} ä & Ä \\ ď & Ď \\ ë & Ë \\ ï & Ï \\ ľ & Ľ \\ ö & Ö \\ ß & \\ ť & Ť \\ ü & Ü \\ ÿ & Ÿ \\ ż & Ż \end{matrix}\right\}$

+ followed by
$\left\{\begin{matrix} a & A \\ c & C \\ d & D \\ e & E \\ g & G \\ i & I \\ l & L \\ n & N \\ o & O \\ r & R \\ s & S \\ z & Z \end{matrix}\right\}$
gives
$\left\{\begin{matrix} ã & Ã \\ č & Č \\ đ & Đ \\ ě & Ě \\ ğ & Ğ \\ ĭ & Ĭ \\ ł & Ł \\ ñ & Ñ \\ õ & Õ \\ ř & Ř \\ š & Š \\ ž & Ž \end{matrix}\right\}$

== followed by $\left\{\begin{matrix} a & A \\ e & E \end{matrix}\right\}$ gives $\left\{\begin{matrix} ą & Ą \\ ę & Ę \end{matrix}\right\}$

++ followed by $\left\{\begin{matrix} a & A \\ & D \\ o & O \end{matrix}\right\}$ gives $\left\{\begin{matrix} æ & Æ \\ & þ \\ ø & Ø \end{matrix}\right\}$

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

<> followed by $\left\{\begin{matrix} o & O \\ u & U \end{matrix}\right\}$ gives $\left\{\begin{matrix} o̕ & O̕ \\ u̕ & U̕ \end{matrix}\right\}$

"" followed by $\left\{\begin{matrix} a & A \\ i & I \\ o & O \\ u & U \end{matrix}\right\}$ gives $\left\{\begin{matrix} å & Å \\ ı & İ \\ œ & Œ \\ ů & Ů \end{matrix}\right\}$

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\"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.
```

(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 fût d'ores et déjà une idée dégénérée
et ambigüe.

References

1. K. Mhtropo'ulou, Za'ar 19, K'edroc, ¿Aj'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
    Guarani 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. А. 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, Dicţionar Francez–Român, Editura
    Ştiinţifică, Bucureşti 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, Langen-
    scheidt, 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@frcitl71