

Cahiers **GUT** *enberg*

☞ FONTES LATINES EUROPÉENNES ET T_EX 3.0

☞ Michael J. FERGUSON

Cahiers GUTenberg, n° 7 (1990), p. 29-31.

<http://cahiers.gutenberg.eu.org/fitem?id=CG_1990__7_29_0>

© Association GUTenberg, 1990, tous droits réservés.

L'accès aux articles des *Cahiers GUTenberg*

(<http://cahiers.gutenberg.eu.org/>),

implique l'accord avec les conditions générales

d'utilisation (<http://cahiers.gutenberg.eu.org/legal.html>).

Toute utilisation commerciale ou impression systématique
est constitutive d'une infraction pénale. Toute copie ou impression
de ce fichier doit contenir la présente mention de copyright.

Fontes latines européennes et T_EX 3.0*

Michael J. FERGUSON

*TUG, Coordinator for Multilingual Activities
INRS-Télécommunications, Montréal, Canada*

Résumé Les normes ISO Latin 1 et Latin 2 d'encodage des caractères ont été modifiées pour T_EX 3.0 et permettent de traiter toutes les langues latines européennes.

Abstract Translation of the Report on the Extended T_EX Font Encoding which has been officially approved by TUG

Le groupe *Multilingual Activities* du *T_EX Users Group* a adopté mi-octobre 1990 l'*Extended T_EX Font Encoding* (encodage étendu des fontes en T_EX). Cette norme est un premier pas vers l'utilisation des nouvelles possibilités de T_EX 3.0 dans un environnement multilingue. L'emploi de cette norme permettra l'emploi direct en T_EX des lettres accentuées et la césure des mots accentués. Elle créera, pour la première fois, une interface consistante pour toutes les polices de caractères pour tout système T_EX. Ce qui signifie qu'un code donné produira le même résultat quelle que soit la fonte. Ainsi, une fonte machine à écrire tt produira une séquence de caractères à espacement proportionnel ffi alors que, pour le même code avec une fonte romain gras, on aura la ligature ffi. Cette norme devrait aussi servir d'encodage standard pour les noms de variables, dans le mode maths, qui ne sont pas en grec.

Cette nouvelle norme reprend les lettres de la norme ISO 8859 (dite Latin 1) et les lettres spécifiques de ISO Latin 2 ainsi que les ligatures et la ponctuation relevant des langues latines européennes et complète

certaines codes (notamment en position octal '000 à octal '037). Cette norme permet de traiter les langues suivantes sans avoir à expliciter les accents (c'est-à-dire qu'en français, par exemple, on tape directement é et non \ 'e à condition bien sûr d'avoir un clavier ad hoc) : albanais, allemand, anglais, danois, espagnol, féringien, finnois, français complet (ISO 8859 avait oublié « œ » et « Œ » — on les trouve ici en position '367 et '327), hollandais, hongrois, irlandais, islandais, italien, norvégien, polonais, portugais, roumain, serbocroate, slovaque, slovène, suédois, tchèque et turc. Cette norme permet aussi de manipuler une série de signes diacritiques ce qui, avec le mécanisme d'accentuation de T_EX, augmente encore le nombre de caractères « de base ». On pourra ainsi traiter d'autres langues, comme le lapon, mais T_EX ne sait toujours pas couper les mots en présence d'accents explicites.

Cette norme, par ailleurs, comprend certaines innovations permettant d'accroître la flexibilité d'emploi et d'augmenter la consistance lexicale sémantique.

- On dispose à la fois du trait d'union (*dash*, code octal '055) et d'un caractère explicite de césure (*hyphenation char*, code '177). Ceci permet aux dessinateurs de fontes de remplacer un « - » par un « = » sans perdre le trait d'union. Puisque « - » n'est plus le caractère de coupure de mots, il permet aux mots ayant un trait d'union explicite, par exemple INRS-Télécommunications, d'être coupés (actuellement en T_EX, ce mot ne peut être coupé qu'au niveau du trait d'union ce qui est fort gênant pour un mot aussi long). Ceci permet aussi de respecter l'usage serbocroate

* Cette note doit paraître sous le titre *Report on Multilingual Activities* dans le *TUGboat*, vol.11 n° 4, 1990. Il est reproduit ici avec l'aimable autorisation de l'auteur ; adaptation française de Jacques André.

qui veut que l'on mette en début de ligne un signe « - » si l'on a coupé la ligne précédente au niveau d'un trait d'union.

- La norme comprend des codes pour les divers guillemets : les guillemets droits (*ASCII double quote* : " , code '042), les guillemets français ouvrants et fermants¹ : « et » (respectivement de code '023 et '024), les guillemets allemands „ , (code '022) et les guillemets anglo-saxons “ et ” (de code '020 et '021). T_EX traite normalement ces derniers comme des ligatures et n'utilise les guillemets droits que dans certains cas, par exemple pour désigner les nombres hexadécimaux. Ce caractère devrait vraiment correspondre dans toute fonte à des guillemets droits, comme aujourd'hui dans la fonte machine à écrire tt de T_EX. Ceci rend disponible un nouveau caractère dans les fontes romanes et supprime le problème irritant de l'emploi de guillemets droits dans ces fontes. C'est au dessinateur de caractères de décider si les guillemets ouvrant, fermant et droit doivent, ou non, être différents. Cette norme, enfin, comprend les signes < et > à leur position ASCII normale.
- La norme définit un petit zéro qui, accolé au « % » permet de faire « pour mille », « pour dix mille », etc. Elle comprend aussi un « espace visible » (code '040) et des versions diacritiques ou non de certains caractères tels que « ~ » ou « ^ ».
- L'apport sans aucun doute le plus important, et le plus utile, est l'inclusion du caractère *Compound Word Mark* (marque de mot composé) – <cwm>. Cette marque, dont l'image est invisible, est un caractère de chasse nulle. Cette marque permettra, par exemple, de supprimer certaines ligatures dans les mots composés allemands : ainsi le mot « auflage » ne devrait pas avoir (comme ici) de ligature « fl » car il est composé des deux mots « auf » et « lage ». Par contre, un mot comme « flach » doit avoir cette ligature. Il y a plusieurs façons de faire intervenir la marque <cwm>. On peut par exemple définir \fl comme f<cwm>1, ce qui traduira la séquence au\fl age en « auflage ». On peut aussi utiliser la marque <cwm> pour appeler certaines ligatures. Si on veut, par exemple, que dans une fonte la ligature « ff » ne soit pratiquement jamais employée, la séquence d'accès à la ligature sera f<cwm>f. Par contre, en mettant cette <cwm> explicitement en saisie, on

¹Voir l'article de Fernand Baudin, "La typographie" entre « guillemets » dans ces *Cahiers GUTenberg* 7 {NdR}.

obtiendra la ligature. Cette marque peut aussi servir de caractère invisible de césure. Nous sommes persuadé que l'on n'est qu'au début des découvertes d'emplois de ce caractère.

Bien que nous ayons désormais une excellente extension des encodages de fonte, il y a encore beaucoup à faire. Le plus urgent est, probablement, de faire la même chose pour les fontes mathématiques. Une décision clé à prendre dans ce domaine est de décider s'il faut inclure les caractères grecs dans une fonte étendue de symboles mathématiques ou au contraire s'il faut accéder aux caractères grecs pour une formule de maths, soit depuis une fonte grecque, soit depuis une fonte de maths. En attendant, on peut utiliser les caractères grecs depuis les fontes actuelles cm.

Maintenant que nous disposons de cette norme de caractères, on peut créer un jeu complet de motifs de césure pour T_EX 3.0. Ces motifs devraient être exprimés en utilisant les seuls caractères imprimables du jeu ASCII de façon à pouvoir les transmettre par courrier électronique. Tous les caractères non ASCII tels que le « ç » ou le *thorn* islandais (code octal '376) devront alors être exprimés avec des barres inversées (par exemple \c c pour ç). Le motif de césure française permettant de couper avant tout ç sera alors 1\c c et \accenthyphcodes en amont des motifs devra inclure \c c{"e7} pour la définition du ç nécessaire à ces motifs. Enfin, les nouveaux motifs doivent tenir compte de l'emploi de <cwm>, soit simplement en incluant le motif équivalent à 1<cwm>, soit de façon plus élaborée. Je suis prêt à revoir ces motifs sur le plan syntaxique mais je préférerais que les personnes qui en sont à l'origine se chargent de la distribution. Je veux bien être le coordinateur de cette liste de « responsables ».

Je voudrais remercier Jean Michael Rynning pour son travail de collecte des informations détaillées qui ont été nécessaires pour la préparation de cette norme. Merci aussi à Jan Michael et à

Norbert Schwarz qui ont dirigé la réunion de Cork, travail difficile, et qui ont scruté tous les détails de cette norme qui, sans eux, n'aurait pas abouti.

	'0	'1	'2	'3	'4	'5	'6	'7	
'00x	·	·	·	·	·	·	·	·	"0x
'01x	˘	˘	˘	˘	˘	˘	<	>	
'02x	"	"	„	«	»	-	—	<cwmm>	"1x
'03x	o	l	j	ff	fi	fl	ffi	ffl	
'04x	□	!	"	#	\$	%	&	'	"2x
'05x	()	*	+	,	-	.	/	
'06x	0	1	2	3	4	5	6	7	"3x
'07x	8	9	:	;	<	=	>	?	
'10x	@	A	B	C	D	E	F	G	"4x
'11x	H	I	J	K	L	M	N	O	
'12x	P	Q	R	S	T	U	V	W	"5x
'13x	X	Y	Z	[\]	-	-	
'14x	'	a	b	c	d	e	f	g	"6x
'15x	h	i	j	k	l	m	n	o	
'16x	p	q	r	s	t	u	v	w	"7x
'17x	x	y	z	{		}	-	(hyph.) -	
'20x	Ǻ	Ą	Ć	Č	Ď	Ě	Ę	Ǧ	"8x
'21x	Ł	Ł	Ł	Ń	Ń	Ń	Ń	Ŕ	
'22x	Ŕ	Ś	Ŝ	Ş	Ţ	Ţ	Ů	Ů	"9x
'23x	Ÿ	Ž	Ž	Ž	IJ	İ	đ	§	
'24x	ǻ	ą	ć	č	d'	ě	ę	ǧ	"Ax
'25x	ł	ł	ł	ń	ń	ń	ń	ř	
'26x	ř	ś	ŝ	ş	ţ	ţ	ů	ů	"Bx
'27x	ÿ	ž	ž	ž	ij	i	đ	§	
'30x	À	Á	Â	Ã	Ä	Å	Æ	Ç	"Cx
'31x	È	É	Ê	Ë	Ì	Í	Î	Ï	
'32x	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	Œ	"Dx
'33x	Ø	Ù	Ú	Û	Ü	Ý	Þ	ŠS	
'34x	à	á	â	ã	ä	å	æ	ç	"Ex
'35x	è	é	ê	ë	ì	í	î	ï	
'36x	ð	ñ	ò	ó	ô	õ	ö	œ	"Fx
'37x	ø	ù	ú	û	ü	ý	þ	š	
	"8	"9	"A	"B	"C	"D	"E	"F	