

ISO 639-3, la nouvelle liste de langues

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 14 février 2007. Dernière mise à jour le 11 janvier 2010

<https://www.bortzmeyer.org/iso-639-3.html>

Après une longue genèse, la nouvelle itération de la norme ISO 639 sur les langues humaines a été publiée le 5 février 2007.

ISO 639 normalise une liste de langues et un code (d'abord à deux lettres, désormais à trois) pour chaque langue. `fr` est le français, `cop` le copte, `cy` le gallois, etc. Cette norme sert de base à bien d'autres normes, par exemple les étiquettes de langue que l'IETF a normalisées dans le RFC 5646¹. La nouvelle itération de ISO 639 marque deux changements importants :

- Une nouvelle agence de maintenance, SIL, dont le travail est très apprécié mais qui a suscité des controverses <http://celia.cnrs.fr/FichExt/Am/A_10_03.pdf>, notamment pour son activité trop... missionnaire dans les pays du Sud,
- Un bond considérable du nombre de langues, qui passe de moins de cinq cents à plus de sept mille.

Cette augmentation n'a pas été sans mal puisque les premières itérations, ISO 639-1 et ISO 639-2, ne prenaient en compte que des langues établies, normalisées et bien connues. Désormais, des langues nettement plus marginales sont enregistrées, et il est souvent très difficile de savoir si deux de ces langues sont différentes ou pas.

D'autres changements font leur apparition comme la notion de macrolangue, permettant de traiter des cas complexes d'apparement de langues.

Parmi les nouvelles langues, notons au hasard :

- `chj`, "*Ojitl*[Caractère Unicode non montré]² *In Chinantec*"
- `jmx`, "*Western Juxtlahuaca Mixtec*",
- `kme`, *Bakole*,
- `hva`, "*San Lu*[Caractère Unicode non montré] *Potos*[Caractère Unicode non montré] *Huastec*",

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc5646.txt>

2. Car trop difficile à faire afficher par L^AT_EX

- hvc, "*Haitian Vodoun Culture Language*",
- nhj, "*Tlalitzlipa Nahuatl*",
- scf, "*San Miguel Creole French*".
- svc, "*Vincentian Creole English*",

Opinion personnelle : la distribution des macrolangues dans ISO 639-3 est assez arbitraire. Il devrait logiquement y avoir une macrolangue « allemand » qui couvre l'allemand standard et l'alémanique. Mais elle n'existe pas, alors qu'elle serait aussi raisonnable que la macrolangue Cri, qui couvre apparemment des langues très différentes.

Comme d'habitude, l'ISO, organisation privée et fermée, ne distribue pas la norme, mais la vend cher <<http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=39534&ICS1=1&ICS2=140&ICS3=20>>. Heureusement, elle est disponible via le site Web de l'agence de maintenance <<http://www.sil.org/iso639-3/>>.

Plusieurs systèmes d'exploitation intègrent un fichier contenant tous les codes définis par cette norme. Par exemple, pour Debian, c'est en CSV et c'est dans le paquetage `iso-codes` pour lequel une version intégrant 639-3 est en préparation <<http://www.perrier.eu.org/weblog/2008/01/19#how-many-languages>>.

Le RFC intégrant ISO 639-3, le RFC 5646, préparé par le groupe de travail LTRU <<http://www.ietf.org/html.charters/ltru-charter.html>> de l'IETF, a été publié en septembre 2009.

Pendant ce temps, les gens qui travaillent sur la norme ISO n'étaient pas restés inactifs et avaient publié ISO 639-5 <<https://www.bortzmeyer.org/iso-639-5.html>>, qui normalise les relations entre les langues. ISO 639-6, pour les dialectes, est encore en cours. On peut avoir une idée des développements actuels en lisant l'article "*Developments in Language Codes standards*" <<http://www.cs.surrey.ac.uk/BIMA/People/L.Gillam/downloads/publications/GLDV2007.final.LGDGCC.pdf>>, écrits par Gillam, Garside et Cox, trois des participants à l'ISO 639.