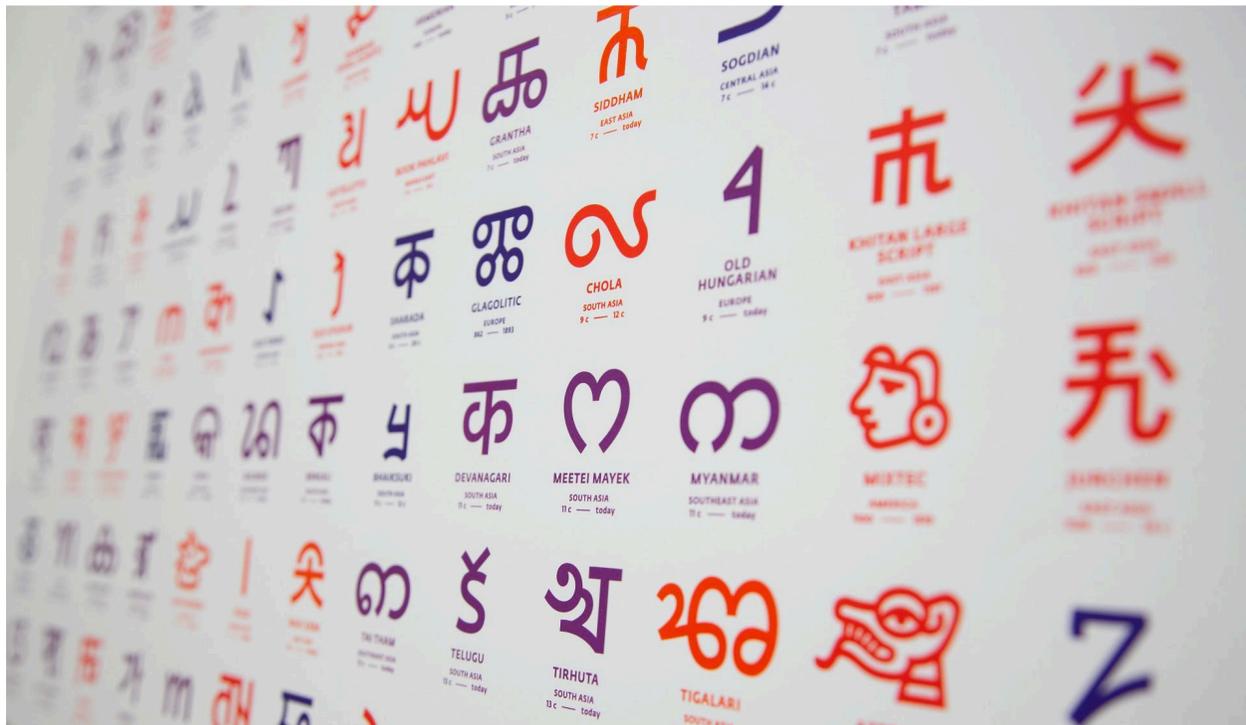


Les scripts manquants de l'Unicode



The Missing Script

À Nancy, en France, des designers typographiques travaillent avec des chercheurs et des communautés linguistiques minoritaires pour offrir une vie numérique à des écritures menacées. Portrait d'une initiative au confluent de l'informatique, de la linguistique, de l'anthropologie et de la justice sociale.

« Pour moi, c'est une manière de garder vivante la mémoire d'un peuple ». Adam Yeo a le ton assuré, le regard vif. « L'écriture, c'est l'incarnation visuelle d'une sensibilité commune. Quand on a pas la possibilité de l'utiliser dans le monde d'aujourd'hui, c'est toute une culture qui se tait ».

Adam Yeo est designer typographique. Dans les dernières années, il a travaillé à un projet d'envergure: digitaliser l'écriture de sa langue maternelle, le bété, parlée par près d'un million de locuteurs en Côte d'Ivoire. « C'est une écriture qui est née en 1956, au moment des grandes vagues d'indépendances,

pour transcrire une langue jusque-là uniquement orale. Sauf que depuis, elle n’a jamais été rendue disponible sur les ordinateurs, les téléphones, ce qui restreint sa diffusion. C’est précisément ce que je voulais changer ».



Adam Yeo et l'écriture bété. Photo: Adam Yeo

Le travail d'Adam s'inscrit dans le cadre du projet The Missing Scripts, une initiative hors-norme menée par l'Atelier national de recherche typographique français (ANRT), l'Université Berkeley et l'Université de sciences appliquées de Mainz, en Allemagne.

L'objectif ? Rescaprer les scripts manquants de l'Unicode.



(Compléter l'Unicode)

Que ce soit pour écrire un texto, envoyer un courriel ou gratifier les photos de vacances de son collègue d'un émoji de cœur, nous utilisons chaque jour le standard Unicode, sans vraiment le réaliser.

« L'Unicode, c'est un peu l'ADN du texte sur Internet », explique Thomas Huot-Marchand, chercheur à l'ANRT et co-instigateur du projet The Missing Scripts. « Pour afficher du texte informatique, il faut traduire les signes d'écriture en code binaire, c'est-à-dire en 0 et en 1. Dans les années 70 et 80, avec la globalisation d'Internet, on a réalisé qu'il fallait trouver un standard, pour être certains que ces codes puissent être lus de manière universelle. C'est là qu'est apparu le standard Unicode ». Créé en 1991, l'Unicode constitue donc en quelque sorte la partition qui supporte tous les écrits numériques, ainsi que l'autorité mondiale en la matière. Il est régulé par un comité central qui valide chaque année l'ajout de nouveaux signes, à condition de prouver leur pertinence et leur nécessité. C'est par exemple par là que les nouveaux émojis font leur apparition dans nos claviers.

Or, bien qu'il soit omniprésent dans nos vies, l'Unicode est loin d'être complet. C'est le constat que font les chercheurs Johannes Bergerhausen et Thomas Huot-Marchand en 2014. Avec des linguistes de l'Université Berkeley, ils découvrent que sur plus de 300 systèmes d'écriture répertoriés dans l'histoire de l'humanité, du cunéiforme aux émojis, seuls 146 y sont alors encodés. C'est plus de la moitié des écritures humaines, passées ou présentes, qui n'ont aucune possibilité de vie numérique ! Pour eux, c'est tout de suite clair: il s'agit d'un enjeu de préservation, mais aussi un grave problème d'inclusion.

Et il va falloir s'y atteler.



Les différents systèmes d'écritures manquants, représentés chacun par un signe. *The Missing Scripts*

Systèmes d'écritures, langues, scripts... De quoi on parle ?

Un système d'écriture renvoie à la manière dont une langue est transcrite pour être lue et écrite. Cette transcription est faite à l'aide de symboles, appelés des *signes* (scripts en anglais). On distingue trois grandes catégories :

- **Les alphabets**, où chaque signe représente un son. Par exemple, l'alphabet latin (A, B, C), utilisé pour écrire cet article, est employé par près de 3 000 langues.
- **Les syllabaires**, où chaque signe, appelé syllabique, représente une syllabe. Un exemple notable est le système d'écriture de l'inuktitut (ᐃᓄᐅᐆᐇᐈᐉ).
- **Les logogrammes**, constitués de signes représentant des idées. Parmi eux, on trouve les caractères chinois (字) ou ceux du cunéiforme mésopotamien (𐎶).

Détail crucial: les systèmes d'écriture sont liés aux langues qu'ils transcrivent, mais ils en sont indépendants. Par exemple, si l'alphabet latin est apparu avec la langue latine, celle-ci est aujourd'hui considérée comme une langue morte, tandis que son alphabet demeure l'un des plus utilisés au monde !



(Rapatrifier les écritures disparues)

Une des missions que s'est donné The Missing Scripts est de redonner vie à des systèmes d'écritures aujourd'hui disparus. Un défi au croisement du design et de l'archéologie. « Au début, j'essaie de recenser tous les écrits qui peuvent exister », explique Morgane Pierson, designer typographique à l'ANTR. Depuis son arrivée au projet, elle a pu travailler sur de nombreuses écritures, dont le maya, l'élymaïque et le nsibidi. C'est elle qui dessine les formes visuelles qui seront ensuite encodées.

« Ça demande un très long travail d'analyse. Je compare les sources, je fais un espèce de catalogue de formes, de proportions pour tenter de voir les constances, les subtilités ». Face à des langues disparues, le travail des designers typographiques consiste d'abord à tenter de retrouver les intentions originelles

des scribes et à recomposer le geste d'écriture. « Je vais essayer différents outils, différentes prises en main. Si on fait un travail trop rapide, il va y avoir des erreurs, ça va être naïf ».

Dans le cas des langues anciennes, Morgane travaille avec des historiens, qui s'assurent que son rendu est précis et conforme à la source. Pas de place à l'erreur: quand une écriture est intégrée à l'Unicode, le retour en arrière est impossible. « Il faut donc être extrêmement prudents », souligne-t-elle.

L'encodage des signes peut s'avérer très utile pour les historiens et les archéologues. « Les caractères numériques nous permettent de constituer des dictionnaires et des corpus de texte facilement accessibles en ligne. On est plus seulement dépendants des photos ou des dessins », explique Jean Revez, professeur au département d'histoire de l'Université du Québec à Montréal et spécialiste de l'Égypte ancienne. Il note que la digitalisation des hiéroglyphes, intégrés à l'Unicode en 2009, a permis une mise en commun et une analyse inédite des écrits, en plus d'ouvrir des voies de recherche, comme la possibilité de traduire les hiéroglyphes en langues modernes grâce à l'IA.

Rares sont les écritures anciennes qui ont néanmoins pu être numérisées de la sorte. « Beaucoup de chemin reste à faire », souligne M. Revez.

DS1WJ S1WJ JSIC JSV&0ET

(Langues vivantes cherchent voies numériques)



Alexey Deminov

Au Canada, les syllabaires de plusieurs langues autochtones (l'inuktitut, le dakelh, le cri, l'ojibwé, le pied-noir et les langues athapascanes) ont pu être inscrits dans l'Unicode dès 1999, facilitant le développement de leur vie numérique.

« C'est fondamental parce que ça permet de rejoindre les jeunes générations », explique Richard Compton, professeur en linguistique à l'UQAM et spécialiste de l'inuktitut. « Le danger, c'est que les jeunes se mettent à texter, à surfer sur internet ou à regarder des vidéos uniquement en anglais ou en français. Pour qu'une langue se transmette, il faut que ça soit la langue de tous les jours. Donc il faut qu'elle s'articule avec la réalité des jeunes, les réseaux, les textos — les choses *le fun* ! ». Selon le chercheur, la vitalité linguistique est également liée à la perception que les jeunes ont de leur langue. Si celle-ci n'est pas adaptée à leurs usages modernes, elle peut vite être perçue comme folklorique, voire dépassée. « Si la langue n'est associée qu'aux aînés et à la vie traditionnelle, il y a un écart qui se crée. On peut commencer à se dire "peut-être que c'est pas cool d'utiliser la langue..." ».

L'utilisation des langues n'est pas anodine, puisqu'elle impacte directement l'identité culturelle. « L'emploi continu de l'inuktitut est indéniablement lié à un fort sentiment d'identité inuite », relève en ce sens l'Institut Avataq, dédié à la culture inuite du Nunavik, dans ᐃᑦᑕᑎᐅᑦ (La langue que nous chérissons), un rapport publié en 2012. Celui-ci note le rôle des technologies modernes dans l'érosion de l'inuktitut et recommande qu'elles soient « mises à contribution de manière positive afin de favoriser la santé de la langue ».

La création de signes digitaux constitue donc un pas dans la bonne direction. Mais leur impact est à nuancer. « Les communautés n'attendent pas après Unicode », rappelle Annunziato Pio Mazzaferro, chercheur à l'ANTR. « La plupart du temps, elles vont faire sans et créer tout de même des adaptations typographiques, parce que c'est une question de nécessité ». Il note l'exemple de formes d'innovation scripturales, où des communautés détournent des outils numériques pour transcrire leur langue. On peut penser à usage de chiffres à la place de certains sons, qu'on retrouve dans le wendat ou dans l'arabizi, une adaptation de l'arabe utilisée en ligne.

El salamou 3alaykom

Kayfa l7aal ?

Sawfa nalta9i ghadan

Ma3a el salama !

السلام عليكم

كيف الحال ؟

سوف نلتقي غدا

مع السلامة !

Exemple d'arabizi. Image: Yasmine Khibri

Dans le cas de l'encodage de signes existants, la professeure d'anthropologie à l'Université Laval Michelle Daveluy rappelle la nécessité de s'assurer que ces outils soient réellement mis à disposition des communautés. « Il faut qu'il y ait des processus d'appropriation par les locuteurs. Ce travail, c'est louable, il faut le faire. Mais après ça, on se retrouve souvent avec un trou, il n'y a pas de transmission entre les acteurs concernés ».

հաշձ ՀՓԻՆՀԼ ԿՈՂԼՏԻ

(Co-construire de nouveaux codes, un art de la prudence)

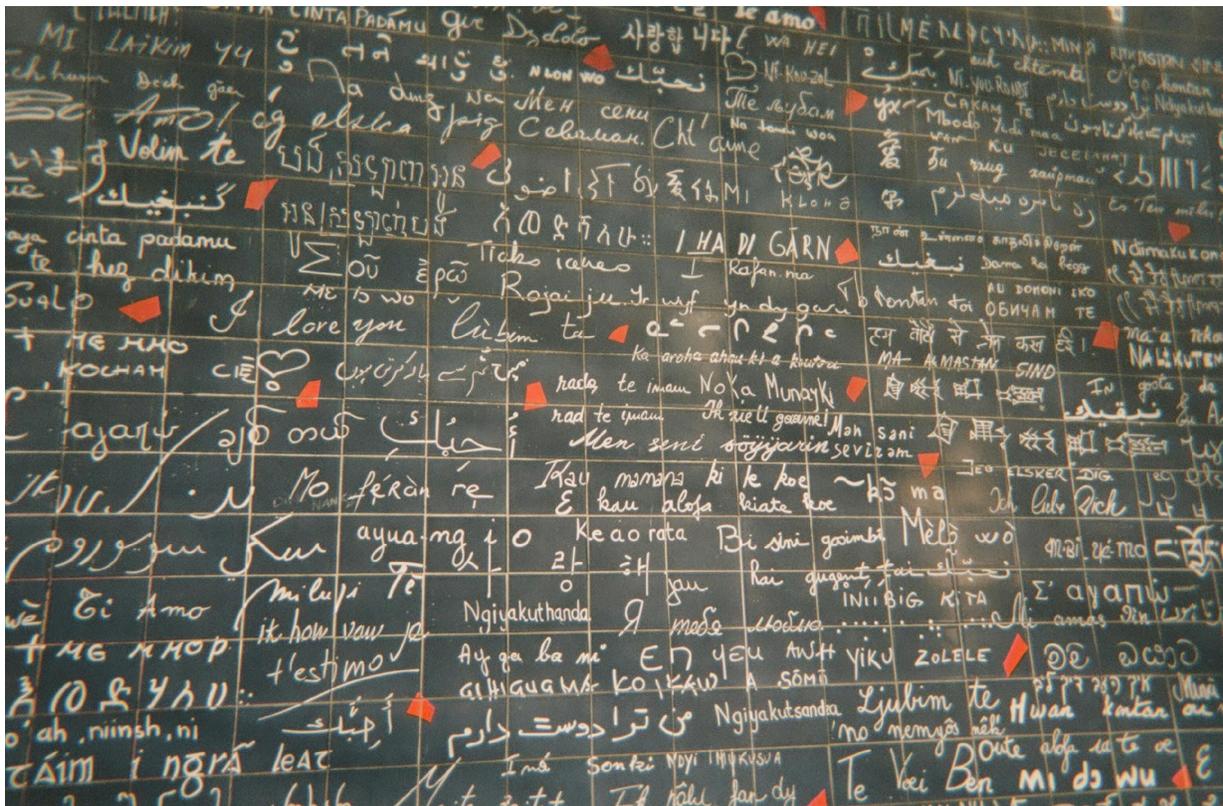
Pour les chercheurs de The Missing Scripts, c'est une évidence: le travail doit se faire dans une posture d'humilité et, surtout, de concert avec les communautés concernées. Car le travail avec des groupes linguistiques minoritaires, souvent autochtones, demande son lot de précautions.

Les spécialistes notent avant tout l'impératif de se mettre au service des besoins réels des communautés. Directement sollicité en 2020 par les locuteurs d'un dialecte de l'inuktitut, le nattilik, le studio néerlandais Typothèque a par exemple travaillé à la création de nouveaux signes Unicode adaptés, et ce de façon pro bono.

L'encodage pose également un risque d'influencer les langues, en réduisant leur complexité. « Il peut y avoir beaucoup de variations dans la façon dont les gens parlent et écrivent leur langue, notamment entre des communautés éloignées géographiquement », explique Michelle Daveluy. Puisque le standard écrit ne fait pas toujours consensus, l'Unicode présente le danger d'imposer une norme unique et d'aplanir la diversité. « C'est comme un couteau à double tranchant. Les langues sont variables parce qu'elles sont utilisées, mais pour les préserver, il faudrait les homogénéiser pour pouvoir les transcrire de façon uniforme », note la chercheuse.

Du côté de l'UNESCO, on prévient de la nécessité de s'assurer que les données linguistiques co-crées avec des communautés autochtones demeurent la propriété totale de celles-ci. Plusieurs scandales ont en effet émergé dans les dernières années, où des groupes de recherche se sont appropriés les données de nations autochtones, notamment en déposant une marque de propriété intellectuelle sur la langue Lakota. De manière générale, l'organisation prône la gratuité et l'accessibilité à toutes les ressources produites.

Finalement, gare aux solutions simplistes ! « Il ne faut pas oublier que le principal vecteur de revitalisation d'une langue, c'est l'éducation », rappelle Richard Compton. « C'est souvent les premières années de scolarité dans une langue qui déterminent sa survie à l'âge adulte ». Pour le chercheur, il faut se méfier des tentations de proposer des solutions technos à des enjeux sociaux complexes. Il rappelle qu'au Canada, l'accès à l'éducation dans la langue maternelle est encore très difficile pour de nombreuses communautés autochtones. « Le plus crucial pour la préservation des langues, c'est donc d'assurer un réel partage des ressources financières de ce côté-là ».



Suin Seong



(Encore du chemin à faire pour le droit à l'écriture digitale)

Aujourd'hui, le standard Unicode compte un total de 168 systèmes d'écritures. En presque dix ans d'activité, le projet The Missing Scripts n'a permis que d'en introduire deux, bien que de nombreuses autres propositions soient toujours à l'étude. C'est le cas du bété qui doit encore faire ses preuves. « Il faut faire montre de l'utilité de ces écritures, du nombre d'utilisateurs, de la volonté politique », explique Thomas Huot-Marchand. « Mais c'est difficile de démontrer un engouement populaire pour une écriture qui ne peut justement pas exister dans les canaux de communication modernes. C'est l'œuf ou la poule ».

Il nomme la responsabilité des entreprises technologiques à favoriser une véritable inclusion, notamment en assurant la prise en charge sur les ordinateurs, les sites web et les outils physiques, comme les claviers. « On est à la merci des grandes entreprises. Si elles ne veulent pas utiliser ces outils, on ne peut rien faire », relève Morgane Pierson.

Si le chemin reste long, de plus en plus de voix réclament aujourd'hui son caractère nécessaire. « C'est la survie d'imaginaires entiers qui en dépend » conclut Adam Yeo.