

Langue et action humanitaire numérique

L'état de l'inclusion et de l'exclusion des locuteurs de langues marginalisées dans les services humanitaires numériques



GSMA

La GSMA est une organisation mondiale unifiant l'écosystème du mobile pour découvrir, développer et proposer des innovations fondamentales pour des environnements commerciaux positifs et un changement sociétal. Notre vision est de libérer toute la puissance de la connectivité afin que les personnes, l'industrie et la société prospèrent. Représentant les opérateurs de téléphonie mobile et les organisations de l'écosystème du mobile et des industries adjacentes, la GSMA offre à ses membres trois grands piliers: la Connectivité pour le bien, des Services et Solutions pour l'Industrie, et la Sensibilisation. Cette activité consiste notamment à faire progresser les politiques publiques, à relever les plus grands défis sociétaux d'aujourd'hui, à étayer la technologie et l'interopérabilité qui font fonctionner la technologie mobile, et à fournir la plus grande plateforme au monde pour réunir l'écosystème du mobile lors des séries d'événements MWC et M360.

Nous vous invitons à en savoir plus sur www.gsma.com

Suivez la GSMA sur Twitter/X : [@GSMA](https://twitter.com/GSMA)

CLEAR Global

CLEAR Global est une organisation à but non lucratif américaine qui aide les personnes à obtenir des informations essentielles et à se faire entendre, quelle que soit la langue qu'elles parlent. Grâce à ses solutions technologiques linguistiques innovantes, à ses recherches et à sa communauté de plus de 120 000 bénévoles linguistiques, CLEAR Global soutient ses organisations partenaires qui travaillent dans divers contextes à travers le monde. Envisageant un monde sans barrières linguistiques, CLEAR Global a pour objectif d'améliorer la communication et l'accès à l'information à l'échelle mondiale.

Pour en savoir plus, consultez le site www.clearglobal.org

Suivez CLEAR Global sur Twitter/X : [@CLEARGlobalOrg](https://twitter.com/CLEARGlobalOrg)



Partnership | Progress | Prosperity

Ce document a été financé par UK International Development du gouvernement britannique et est soutenu par la GSMA et ses membres.

Les opinions exprimées ne reflètent pas nécessairement les politiques officielles du gouvernement britannique.

GSMA Mobile pour l'innovation humanitaire

Le programme GSMA Mobile for Humanitarian Innovation a pour objectif d'accélérer la mise en œuvre et l'impact de l'aide humanitaire numérique.

Cela se fait en élaborant un programme d'apprentissage et de recherche pour éclairer l'avenir de la réponse humanitaire numérique, en catalysant les partenariats et l'innovation pour de nouveaux services humanitaires numériques, en plaidant pour des environnements politiques favorables, en surveillant et en évaluant les performances, en diffusant des informations et en dressant le profil des réalisations.

Pour en savoir plus, rendez-vous sur www.gsma.com/m4h ou contactez-nous à l'adresse m4h@gsma.com

Suivez GSMA Mobile for Development sur Twitter/X : [@GSMAm4d](https://twitter.com/GSMAm4d)

Auteur principal

Ellie Kemp, Head of Research, Evidence and Advocacy, CLEAR Global

Contributeurs

Matthew Downer, Senior Insights Manager, Mobile for Humanitarian Innovation, GSMA

Milena Haykowska, Program Technology Lead, CLEAR Global

Alp Öktem, Computational Linguist, CLEAR Global

Christian Resch, Language Technology and Information Exchange Platform Specialist, CLEAR Global

Jason Symons, Data Solutions Manager, CLEAR Global

Emily Elderfield, Advocacy Officer, CLEAR Global

Andrea Restine Portugal, Research Intern, CLEAR Global

Note

L'ironie de la publication de ce rapport uniquement en anglais et en français, deux des langues les plus puissantes du monde, n'échappe pas à l'équipe de recherche. Nous avons l'intention de le republier dans d'autres langues à l'avenir, mais nous reconnaissons qu'il s'agira probablement aussi de langues puissantes à l'échelle mondiale, excluant les locuteurs de langues marginalisées.

Citation recommandée

Kemp, E. (2024). *Langue et action humanitaire numérique*. GSMA and CLEAR Global.

Table des matières

Pourquoi la marginalisation linguistique est-elle importante dans les situations de crise ?	3
Marginalisation linguistique	3
Marginalisation linguistique dans les réponses humanitaires numériques	4
Étude de cas : Mobile Vaani	6

Obstacles aux services humanitaires numériques dans les langues marginalisées	7
Prise de conscience du problème et manque de données	7
Donner la priorité au problème	8
Considérations relatives aux coûts	9

Un manque de solutions technologiques inclusives	9
Des services humanitaires numériques plus inclusifs	12
Définir le rôle des services numériques	12
Étude de cas : Talk to Loop	13
Identifier les caractéristiques et les bonnes pratiques des services humanitaires numériques inclusifs	14
Des possibilités inexploitées d'élargir l'accès grâce à des pratiques collaboratives et fondées sur des données probantes	16

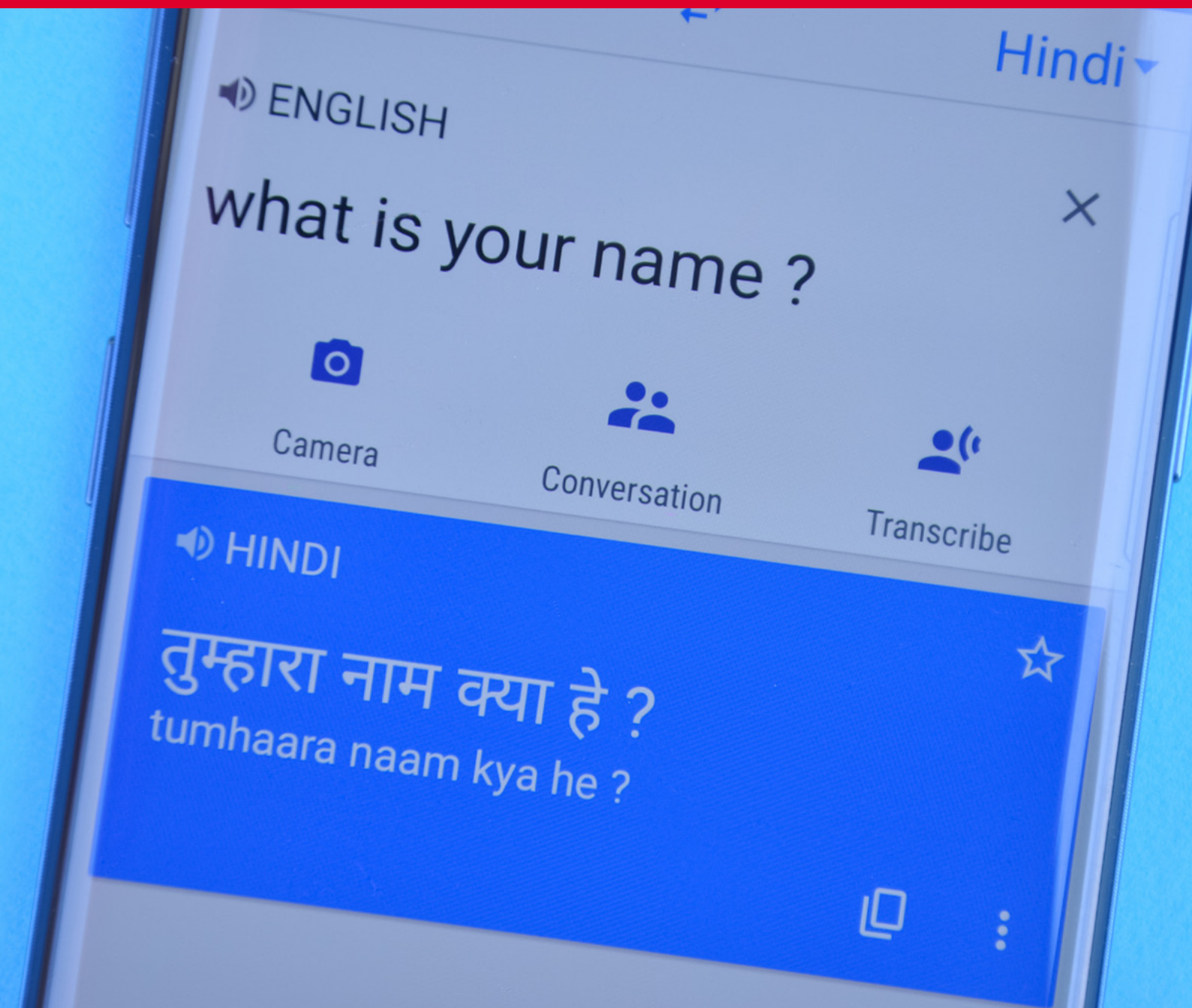
Conclusions	18
--------------------	-----------

Recommandations	19
------------------------	-----------

Annexe : Méthodologie	22
------------------------------	-----------

« Pour démarrer tout type d'interaction, la langue est la première considération. Nous devons donc être en mesure de proposer des services dans la langue dont les personnes ont besoin. »

- Fournisseur numérique humanitaire, Europe



Pourquoi la marginalisation linguistique est-elle importante dans les situations de crise ?

La marginalisation linguistique

Sur les 7 000 langues humaines estimées, seules une poignée ont une forte présence en ligne et quelques centaines seulement sont disponibles sur des plateformes numériques. Il s'agit généralement des langues dominantes des nations les plus économiquement et politiquement puissantes du monde. Dans ce contexte, une langue comme le swahili, parlée par 100 millions de personnes, est marginalisée à l'échelle mondiale.

La marginalisation numérique de la plupart des langues du monde et de leurs milliards de locuteurs reflète la marginalisation réelle des communautés linguistiques dans leur propre pays. Par exemple, une langue comme le nande, parlée par quelque 10 millions de personnes, est marginalisée dans l'est de la République démocratique du Congo (RDC), où la communication officielle se fait en grande partie en français et en swahili congolais.

À mesure que les services numériques se développent, les locuteurs de langues minoritaires sont de plus en plus exclus.¹ Le contenu et le support

numériques sont concentrés dans une fraction des langues du monde, les lacunes les plus notables étant dans les langues africaines² et même dans les langues majeures telles que le bengali/bangla et l'hindi.³ Lorsque du contenu numérique est disponible, il est souvent dans un format inutilisable pour ceux qui ne savent pas lire et écrire ou dont la langue est en grande partie non écrite.⁴ Cela peut entraîner une plus grande exclusion numérique des locuteurs de langues minoritaires, y compris dans les communautés touchées par les crises.⁵

Le risque d'exclusion linguistique numérique est intersectionnel. Dans le monde entier, les femmes et les filles, les personnes âgées, les personnes handicapées et les groupes ethniques minoritaires ont moins accès à l'éducation⁶ et aux possibilités qu'elle offre d'acquérir des compétences en lecture, en calcul, en informatique et en langue seconde. Ces mêmes groupes ont systématiquement moins accès aux technologies numériques et les utilisent moins bien.⁷

1 Nations Unies. (2020). [Report of the Secretary-General Roadmap for Digital Cooperation](#); Jones, S. (28 novembre 2022). "[Understanding the Impact of Digital Language Support](#)". SIL Blog; CLEAR Global: [4 billion conversations website](#).

2 Jones, S. (28 novembre 2022). "[Understanding the Impact of Digital Language Support](#)". SIL Blog.

3 Dittus, M. and Graham, M. (2022). "[A platform survey: interface language support by widely-used websites and mobile apps](#)" in [State of the Internet's Languages Report](#). Oxford Internet Institute and Centre for Internet and Society; Simons, G.F., Thomas, A.L. and White, C.K. (2022). [Assessing Digital Language Support on a Global Scale](#). arXiv.

4 Caswell, J. (2019). [The Digital Lives of Refugees](#). GSMA; Translators without Borders. (2017). [Language profile of five IDP sites in Maiduguri, north-east Nigeria](#); Rocca, R. et al. (2023). [Natural language processing for humanitarian action: Opportunities, challenges, and the path toward humanitarian NLP](#). Frontiers.

5 Caswell, P. and Downer, M. (2022). [Digital Access and Barriers in Displacement-Affected Communities in Sudan](#). GSMA and NRC.

6 UNESCO. (2016). [40% don't access education in a language they understand](#).

7 Par exemple : Satari, A. (2021) [The Mobile Disability Gap Report 2021](#). GSMA; Jeffrie, N. (2023) [The Mobile Gender Gap Report 2023](#). GSMA.

Marginalisation linguistique dans les réponses humanitaires numériques

Les facteurs à l'origine de l'exclusion numérique sont souvent aggravés dans les situations de crise,⁸ ce qui signifie que l'exclusion numérique basée sur la langue laisse de côté de nombreuses personnes que les organisations humanitaires souhaitent privilégier en cas d'urgence. Les organisations de la société civile (OSC) répondantes à cette étude, dont la plupart travaillent directement avec des locuteurs de langues marginalisées, ont estimé que les personnes âgées étaient le principal segment de la population ayant du mal à utiliser la technologie numérique, suivies des femmes, des personnes handicapées et des groupes minoritaires.

Les locuteurs de langues marginalisées sont souvent confrontés à des défis supplémentaires en cas de crise, comme la difficulté à comprendre leurs droits. Cela peut être aggravé par les désavantages auxquels ils sont déjà confrontés du fait qu'ils ne parlent pas une langue dominante, comme des possibilités limitées de subsistance, d'éducation et de participation civique. Cela se recoupe souvent avec d'autres marqueurs identitaires, comme le sexe ou le statut de handicap, ce qui exacerbe les désavantages intersectionnels⁹ et les rend particulièrement vulnérables en cas d'urgence.

Lorsque les organisations humanitaires ne sont pas sensibles à la diversité linguistique, la langue peut créer une barrière, laissant les locuteurs de langues marginalisées dans l'incapacité d'accéder aux services et aux informations ou de faire entendre leur voix.¹⁰ Les suppositions sur qui parle une langue nationale, dominante ou de contact¹¹ peuvent aggraver ces problèmes,¹² et si les données ne sont pas systématiquement collectées sur la langue maternelle ou préférée des personnes, cette exclusion peut facilement passer inaperçue.¹³

Étant donné que la plupart des services numériques humanitaires ne sont disponibles que dans un nombre limité de langues, des millions de personnes sont exclues. Cependant, le secteur humanitaire manque souvent d'informations sur les personnes exclues de ses services. Cela peut aggraver l'exclusion, renforcer les inégalités et présenter un défi éthique aux principes humanitaires d'humanité et d'impartialité.

« Ce que nous avons constaté chez les communautés marginalisées avec lesquelles nous travaillons, c'est que la technologie les éloigne au lieu de les rapprocher. »

– Fournisseur numérique humanitaire, Afrique

Par exemple, les barrières linguistiques ont empêché les femmes réfugiées Twa de la RDC d'utiliser un service mobile de surveillance des droits de l'homme en Ouganda qui traite de l'exploitation et des abus sexuels. Dans une communauté de 200 personnes, seuls trois jeunes hommes parlaient une langue commune avec les OSC.

Dans l'est de la RDC, un groupe WhatsApp utilisé par une agence des Nations Unies pour coordonner les activités et les retours d'information des partenaires et des communautés communique uniquement en français et en swahili : « Le français pour les expatriés qui ne parlent pas les langues locales. [...] et les dirigeants communautaires qui ne savent pas écrire en français le font en swahili, puis d'autres traduisent pour eux. » Seuls les représentants communautaires qui savent lire et écrire en swahili ou en français et qui possèdent un smartphone peuvent y participer. Cela exclut en grande partie les femmes et les locuteurs des dizaines d'autres langues locales.

8 GSMA. (2022). [The Digital Worlds of Displacement Affected Communities](#). GSMA and UNHCR.

9 Par exemple : Translators Without Borders. (2017). [Language profile of five IDP sites in Maiduguri, north-east Nigeria](#).

10 CLEAR Global. (2022). [Hospital cleaner as interpreter: language and cultural awareness in Rohingya access to health care](#).

11 Une langue de contact, ou lingua franca, est un moyen de communication entre des locuteurs de différentes langues maternelles.

12 Par exemple : CLEAR Global. (2023). [A dangerous information gap](#); Lough, O. (2022). [Social media and inclusion in humanitarian response](#). HPG Working Paper. Overseas Development Institute (ODI).

13 Thomas, C. and Anonymous. (2024). [No Data is Data](#). Minority Rights Group International; IFRC. (2018). [World Disasters Report 2018](#); Uekusa, A. (2019). [Disaster Linguicism: Linguistic minorities in disasters](#). Cambridge University Press.

Environ un tiers des répondants des OSC ont indiqué que les membres des communautés qu'ils soutiennent n'ont pas accès aux services numériques dans leur propre langue, et un prestataire de services a indiqué que dans cinq des six pays où ils opèrent, les membres des communautés affirment que les barrières linguistiques les empêchent d'accéder aux mécanismes de retour d'information.

Cela illustre les défis auxquels sont confrontés les locuteurs marginalisés : ceux qui le peuvent doivent souvent naviguer dans les services numériques dans une deuxième ou même une troisième langue. Cela peut signifier que les personnes passent à côté d'informations vitales, qu'elles n'ont pas accès à l'assistance qui leur est destinée et qu'elles peuvent être confrontées à un risque accru de préjudice, comme la fraude. Il est essentiel que les organisations humanitaires soient conscientes de ces risques lors de la conception de services numériques tels que les systèmes d'alerte précoce (SAP), les programmes d'aide financière ou l'enregistrement à distance.

C'est particulièrement important lorsqu'il n'y a pas d'interface physique pour les services humanitaires.

Par exemple, lors des inondations au Pakistan en 2022, les locuteurs de langues telles que le saraiki, le baloutche et le sindhi ont manqué des alertes et des mises à jour vitales qui n'avaient été envoyées qu'en ourdou et en anglais par SMS.¹⁴ Certains propriétaires de smartphones parlant le saraiki ont utilisé WhatsApp pour envoyer des messages et des clips vidéo afin d'alerter la société civile, mais l'interface ne prenait pas en charge leur script et ceux qui le pouvaient ont été obligés d'improviser une translittération pour communiquer.

« Les dommages et les pertes économiques augmentent à cause du manque d'information. Nous sommes à l'ère du numérique, tout est numérisé. Les autorités disent donc : nous avons fait circuler ces informations. Mais les bonnes personnes ne les reçoivent pas. »

- Informateur de la société civile, Asie

Dans le nord-est du Nigéria, où les SMS sont utilisés pour notifier et enregistrer les personnes déplacées à l'intérieur du pays (PDI) pour obtenir de l'aide, les locuteurs du mafa dépendent souvent d'autres personnes pour traduire les messages en anglais et en haoussa. Cela les rend plus vulnérables aux abus et à l'exploitation.

« [Les voleurs] comprennent que les gens n'ont aucune compréhension de base du processus, alors ils essaient de voir comment ils peuvent les manipuler pour obtenir ce qui leur appartient. »

- Informateur de la société civile, Afrique

¹⁴ Voir : CLEAR Global. (2023). [A dangerous information gap](#).

Étude de cas :

Mobile Vaani



Mobile Vaani est un réseau de plateformes de médias communautaires basées sur la voix, indépendantes de la langue, qui permet une communication bidirectionnelle sur des questions cruciales. Développée par Gram Vaani, la plateforme partage du contenu audio avec et depuis les communautés et les organisations partenaires dans les langues des utilisateurs, sur des sujets tels que la santé, la nutrition et la violence sexiste. La langue de l'interface est déterminée par les organisations partenaires des communautés avec lesquelles elles souhaitent s'engager. Le réseau Mobile Vaani, qui fonctionne sous forme d'instances hyperlocales dans plus de 120 districts, a servi plus d'un million d'utilisateurs dans environ 20 États et 12 langues. Les partenaires ont également été autorisés à déployer la plateforme en Afghanistan, en Éthiopie, en Namibie, au Pakistan et en Afrique du Sud. On estime que 5 millions d'utilisateurs ont jusqu'à présent généré plus d'un million de rapports de voix communautaires, sous forme d'opinions, de requêtes, de griefs, d'histoires et d'informations.

« Si vous parlez de l'information comme pouvoir, ce pouvoir n'est pas accessible à une très large partie de la population parce que l'information n'est pas vraiment démocratisée au sens propre. L'information est accessible à une certaine élite, des personnes qui maîtrisent le numérique, des personnes qui ont accès à ces canaux et à des contenus disponibles dans des langues [puissantes]. [...] Il s'agit d'une combinaison non seulement de la disponibilité du contenu dans plusieurs langues, mais également des interfaces technologiques. »

- Vijay Sai Pratap, cofondateur, Gram Vaani

Mobile Vaani fonctionne sur les téléphones de base et les smartphones, en utilisant la voix pour surmonter les obstacles liés à l'alphabétisation et aux compétences numériques. Les utilisateurs accèdent aux informations audio et les partagent via un IVR basé sur un clavier en utilisant seulement trois touches numériques. Un système de rappel automatique des appels manqués rend le service gratuit pour les utilisateurs. Les organisations peuvent utiliser une application Android pour créer et partager du contenu dans les langues locales. Les contraintes liées à l'alphabétisation numérique sont traitées par une navigation simplifiée dans les menus et un renforcement des capacités en personne parallèle :

- Le contenu audio préenregistré est diffusé en boucle continue et les utilisateurs appuient sur 1, 3 ou 5 pour ignorer, aimer ou enregistrer une réponse.
- Les bénévoles de la communauté font la promotion de la plateforme, accompagnent les nouveaux utilisateurs et organisent des ateliers en personne pour garantir un retour d'information de diverses voix sur le contenu et la plateforme.

« [Avec l'audio] les gens peuvent s'engager, s'exprimer, poser des questions. [...] Personne n'est laissé de côté, qu'il soit alphabétisé ou illettré. »

- Deepak Kumar, Gestionnaire de programme principal, Gram Vaani

→ Lire l'affaire complète ici.

Obstacles aux services humanitaires numériques dans les langues marginalisées

Bien que de nombreux fournisseurs de services humanitaires numériques reconnaissent l'importance de l'inclusion des langues marginalisées, cette pratique est loin d'être la norme. D'énormes progrès ont été réalisés récemment dans le domaine des technologies linguistiques, mais peu de communautés linguistiques marginalisées en ont profité. Plusieurs facteurs semblent entraver les progrès.

Prise de conscience du problème et manque de données

Les organisations humanitaires surestiment souvent le nombre de personnes capables d'utiliser une deuxième langue en toute confiance, et manquent de données sur les langues les plus pertinentes dans lesquelles investir. Sans ces données, les hypothèses sur la portée des services, en particulier parmi les groupes marginalisés, ne sont pas fiables. L'exclusion numérique basée sur la langue peut rester invisible si elle n'est pas explicitement surveillée.

Les langues utilisées par une organisation humanitaire pour développer un service détermineront qui s'y engage et à qui elle répond. Cela commence par des recherches auprès des utilisateurs et des évaluations des besoins. La plupart des informateurs ont convenu que pour que ces consultations capturent les besoins des groupes marginalisés, elles doivent être menées dans les langues avec lesquelles ils sont le plus à l'aise. Les prestataires de services humanitaires consultés n'étaient pas convaincus que les recherches et les évaluations se déroulent systématiquement dans les langues marginalisées. Lorsque ce n'est pas le cas, ou lorsque les évaluations sont menées par des personnes extérieures, elles ne refléteront pas avec précision les besoins, les préoccupations et les priorités des communautés marginalisées.¹⁵

Les organisations humanitaires peuvent s'appuyer sur les données linguistiques existantes pour les populations locales comme point de départ pour les évaluations des besoins en matière d'information et de communication. Cependant, dans de nombreux contextes, les informations détaillées ne sont pas disponibles, notamment sur la portée des langues de

contact ou des lingua francas.¹⁶ Dans d'autres cas, les collègues des bureaux de pays ou des organisations partenaires sont censés savoir quelles langues sont pertinentes, même s'ils n'ont pas d'expérience directe de l'utilisation des langues dans les communautés marginalisées.

Étant donné que les non-utilisateurs ne sont pas pris en compte dans les données de la plateforme, l'exclusion linguistique peut facilement passer inaperçue. Un prestataire de services ne découvrira que des communautés linguistiques qui ne sont pas atteintes que s'il mène des recherches spécifiques et distinctes.

En cas d'urgence soudaine, les services peuvent être mis en ligne sans consultation préalable pour répondre rapidement aux besoins les plus urgents. Idéalement, les commentaires des utilisateurs permettraient d'apporter des corrections de cap si nécessaire.

« Nous avons mis tout le contenu en ukrainien pour la réponse à l'Ukraine dès le début, et les utilisateurs ont dit : "Pourquoi ce n'est pas en russe ?" Parce que nous essayions de répondre rapidement, il n'y avait pas d'évaluation des besoins. Les utilisateurs ont demandé du russe, nous avons donc réussi à le faire et avons trouvé ces ressources. »

– Fournisseur numérique humanitaire, Amérique du Nord

¹⁵ Translators without Borders. (5 décembre 2018). [The words between us: How well do enumerators understand the terminology used in humanitarian surveys?](#); Ground Truth Solutions. (27 mai 2021). [For Rohingya, trust begins with who is asking the questions, May 2021.](#)

¹⁶ CLEAR Global. (2022). [Global Language Data Review](#); Translators Without Borders. (2019). [Why we need to collect data on the languages of crisis-affected people.](#)

Donner la priorité au problème

Bien que les prestataires de services numériques humanitaires valorisent l'inclusion, ils doivent généralement donner la priorité à la portée et à l'échelle d'un service pour se conformer aux attentes des donateurs ou de l'organisation. Cela peut être aggravé par des objectifs de rapport qualité-prix (RQP) ou de retour sur investissement (ROI), une approche unique qui donne la priorité aux langues majoritaires pour répondre aux besoins du plus grand nombre de personnes, renforçant ainsi l'exclusion linguistique numérique.¹⁷

Par exemple, les informateurs ont décrit des services numériques disponibles uniquement en espagnol pour les pays d'Amérique centrale et d'Amérique du Sud, et en birman pour le Myanmar, malgré le fait que les minorités linguistiques soient victimes de discrimination, de désavantages et parfois de violences, ce qui en fait des groupes prioritaires pour les services humanitaires.



Ces dynamiques entrent en jeu lorsqu'un service est répliqué dans un nouveau contexte. Lorsqu'une application de signalement et d'orientation des cas de VBG, développée au Nigéria, devait être reproduite en RDC, son bailleur de fonds des Nations Unies et ses partenaires ont voulu donner la priorité au français. Cependant, moins de 2,5 % des personnes parlent le français comme langue principale à la maison en RDC.¹⁸

Ces décisions reposent souvent sur l'hypothèse selon laquelle les locuteurs non natifs de la langue dominante auront suffisamment de connaissances en deuxième langue pour accéder aux informations et aux services. Dans un contexte comme celui de la Somalie, cela suppose que les personnes affectées peuvent comprendre et communiquer dans le dialecte dominicain Mahaatiri dominant. Les tests de compréhension de la communication sur la vaccination contre la polio en langue mahaatiri par des locuteurs du dialecte maay ont révélé que cette hypothèse était erronée.¹⁹ On peut s'attendre à ce que les locuteurs de langues non apparentées, comme le mushunguli, trouvent la communication en somali mahaatiri encore plus difficile à comprendre.

« J'estime qu'environ 5 à 6 % des communautés minoritaires et marginalisées utilisent les services numériques, peut-être en écoutant la radio. La raison en est que tous les médias grand public et les réseaux sociaux parlent [un dialecte] ou une langue qui signifie que les communautés ne s'intéressent pas. Et certaines d'entre elles ne comprennent même pas bien. Et certaines d'entre elles n'ont même aucune information, aucun message de sensibilisation, aucune information clé partagée par les organisations internationales, les organisations des Nations Unies et les entités gouvernementales. Les communautés marginalisées n'ont pas la possibilité de s'y intéresser. »

- Informateur de la société civile, Afrique

¹⁷ Bryant, J. (2022). [Digital technologies and inclusion in humanitarian response](#). ODI HPG.

¹⁸ CLEAR Global [Language map of DRC](#).

¹⁹ Minority Rights Group. (2023). [Language barriers in polio vaccine campaigns in Somalia: Focus on Maay speakers in Banadir](#).

Considérations relatives aux coûts

« Certaines personnes sont laissées de côté simplement à cause du fonctionnement du financement. Pas parce que nous ne savons pas qu'il y a un problème. »

- Fournisseur numérique humanitaire, Amérique du Nord

Pour la plupart des prestataires de services consultés, le coût était le principal facteur limitant la fourniture de services humanitaires numériques dans les langues marginalisées, en particulier le coût de la traduction, de la transcription et du doublage.

Même les prestataires de services bien financés ont déclaré avoir du mal à faire face au coût de la communication dans les langues minoritaires et marginalisées, en particulier celles qui ne sont pas prises en charge par la technologie linguistique. L'un d'eux a estimé qu'il faudrait environ 10 000 USD pour traduire, adapter et exprimer tout leur contenu dans une seule langue. En conséquence, certains services sont déployés dans des langues dominantes dans plusieurs pays, les langues moins dominantes étant un objectif secondaire.

« Parce que les ressources sont si limitées [...] nous accordons la priorité au nombre de personnes atteintes [...] mais pour une organisation à but non lucratif, l'inclusion devrait être une évidence. »

- Fournisseur numérique humanitaire, Amérique du Nord

Maintenir un service dans plusieurs langues au fil du temps constitue également un obstacle financier, tant pour le contenu que pour l'interface utilisateur. Cela est particulièrement difficile pour les contenus qui sont sujets à des changements rapides, comme les conditions de sécurité le long d'une route de migration. Pour les langues moins disponibles dans le commerce, l'expertise linguistique et technique requise peut être à la fois plus difficile à trouver et potentiellement plus coûteuse.

« Lorsque vous apportez une petite correction, vous devez la faire dans toutes les langues. C'est le principal défi, pas la faisabilité technique. Vous devez d'abord tout traduire, puis gérer les mises à jour et les correctifs. »

- Fournisseur numérique humanitaire, Europe

Cela peut poser des problèmes éthiques si le budget est disponible pour la mise en place initiale, mais pas pour la maintenance : « Il n'est pas acceptable que nous commençons à créer du contenu dans une nouvelle langue que nous ne pouvons pas maintenir. » Ce contenu deviendrait inexact au fil du temps s'il n'y avait pas la capacité de le maintenir à jour, et « l'inexactitude est une désinformation. »

Ces problèmes sont aggravés lorsque la capacité de partager les coûts du soutien linguistique marginalisé est minime. Au regret de certains informateurs, l'investissement d'une organisation dans le multilinguisme n'a généralement pas contribué à la capacité de communication à l'échelle du secteur dans les langues concernées, par exemple, par le biais de glossaires partagés de terminologie testée par le public.

Un manque de solutions technologiques inclusives

Les technologies linguistiques inclusives, comme la reconnaissance vocale et la traduction automatique, pourraient contribuer à surmonter les obstacles financiers tout en élargissant l'accès aux services humanitaires numériques et aux services numériques de manière plus générale pour des millions de locuteurs de langues marginalisées. La plupart des fournisseurs de services numériques consultés pour cette étude utilisent déjà les technologies linguistiques, entre autres pour fournir des services en dehors des heures de travail, permettre aux utilisateurs de communiquer comme ils le souhaitent,

fournir des informations cohérentes dans plusieurs langues et réduire la dépendance à l'égard des fournisseurs de services pour les tâches et les contenus répétitifs. Cependant, un locuteur kanuri ou somali, par exemple, est encore loin de pouvoir accéder aux services humanitaires numériques grâce à la reconnaissance vocale automatique dans sa propre langue.

Le rythme de développement des technologies linguistiques s'accélère, avec Meta et Google, par exemple, publiant des modèles linguistiques pour

plusieurs centaines de langues.²⁰ Cependant, ce développement reste profondément inégal. Les résultats de haute qualité sont encore limités à une poignée de langues économiquement puissantes comme l'anglais, le français, l'espagnol et le chinois. Les langues des communautés touchées par la crise sont en grande partie négligées.

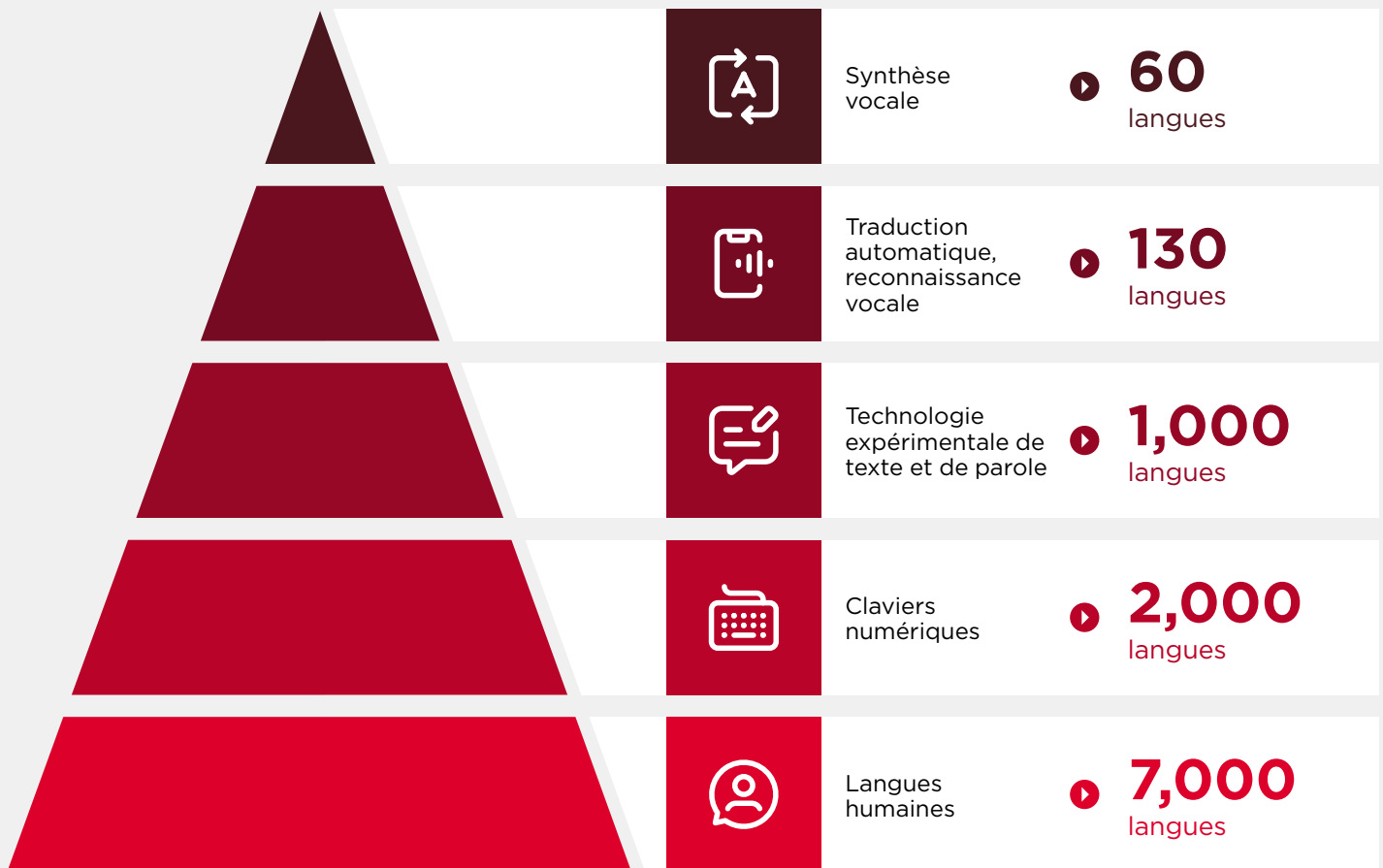
La plupart des 7 000 langues humaines du monde ne sont pas du tout prises en charge par les technologies linguistiques. Des technologies expérimentales sont en cours de développement pour environ 1 000 d'entre elles.

La norme Unicode, qui encode 161 systèmes d'écriture pour des milliers de langues, permet aux utilisateurs d'écrire, de lire et de stocker du texte sur un téléphone mobile. Elle constitue la base du traitement de la langue basée sur le texte, comme la traduction automatique, la transcription ou la classification des intentions. Mais ce traitement nécessite des claviers numériques, disponibles uniquement pour environ 2 000 langues.

La technologie de traitement elle-même est disponible pour un groupe de langues encore plus restreint. La traduction automatique pour traiter et répondre à une requête d'utilisateur, par exemple sur la manière d'accéder à l'aide humanitaire, n'est disponible que dans 130 langues et pour la plupart, la qualité est médiocre. La reconnaissance vocale, qui permettrait à une personne touchée par une crise de recevoir des réponses conversationnelles à une requête verbale d'un chatbot, est disponible dans un nombre similaire de langues.

La synthèse vocale, qui fournit à l'utilisateur une version orale générée par ordinateur d'un texte écrit, couvre moins de 1 % des langues humaines. C'est la technologie qui se cache derrière les lecteurs d'écran pour les utilisateurs malvoyants et peut également rendre le contenu plus accessible à un locuteur non alphabétisé ou non natif. Pour la plupart des langues, la technologie n'a pas été développée pour imiter les modèles de parole humaine, ce qui donne une impression robotique et peut être rebutant pour les nouveaux utilisateurs.

Figure 1 : La portée limitée des technologies linguistiques



²⁰ Meta. (n.d.). [No Language Left Behind website](#); Bapna, A., et al. (2022). [Building machine translation systems for the next thousand languages](#). Google Research.

L'absence ou la mauvaise qualité des technologies linguistiques pour toutes les langues, à l'exception de quelques langues dominantes, constitue une limitation fondamentale aux services humanitaires numériques inclusifs.²¹ Par exemple, un service d'assistance numérique pour les survivants de violences basées sur le genre dans le nord-est du Nigéria propose des fonctions de lecture à voix haute et de commande vocale, utiles dans une région où de nombreux utilisateurs ne savent pas lire. Cependant, en raison d'un manque de technologie linguistique, la lecture à voix haute se fait en anglais et seules les commandes vocales sont également disponibles en haoussa. L'anglais n'est pas largement compris et les femmes sont surreprésentées parmi les quelque 17 % de personnes du nord-est du Nigéria qui ne parlent pas bien ou pas du tout le haoussa.²²

Ce fossé en matière d'accès aux technologies reflète des inégalités plus larges entre les différents locuteurs de langues. Contrairement à la plupart des personnes touchées par la crise, les locuteurs de langues économiquement puissantes sont plus susceptibles d'être en ligne et d'offrir un marché prêt pour les services numériques commerciaux. Cela fournit à la fois l'incitation commerciale et l'opportunité de développer des technologies linguistiques dans leurs langues.



Un autre défi est que les besoins technologiques des locuteurs de langues marginalisées touchées par les crises ne constituent pas un sujet de recherche populaire en dehors des situations d'urgence majeures à déclenchement soudain. Certaines crises, comme le tremblement de terre d'Haïti de 2010, ont galvanisé les efforts de collaboration pour créer des outils technologiques linguistiques.²³ Pourtant, peu d'efforts soutenus sont déployés pour la longue et difficile tâche de créer des technologies pour les langues des communautés touchées par des urgences récurrentes ou prolongées, ou pour développer des modèles qui peuvent être rapidement adaptés pour un déploiement urgent dans de nouvelles langues.²⁴

« Au cours de la dernière décennie, nous avons constaté que les gens souhaitent travailler sur une catastrophe lorsqu'elle se produisait [...]. Mais une fois la catastrophe terminée, ils voulaient reprendre le cours normal de leur vie. »

- Technologue linguistique, Amérique du Nord

Les données – texte simple, discours enregistré avec transcription et textes parallèles en paires de langues – sont la matière première pour la technologie linguistique. Un défi majeur pour étendre la portée de la technologie linguistique est que la plupart des langues, représentant des milliards de personnes dans le monde, ont à peine une représentation numérique, et donc peu de données disponibles.

21 Joshi, P., Santy, S., Budhiraja, A., Kalika, B. and Coudhury, M. (2021). [The State and Fate of Linguistic Diversity and Inclusion in the NLP World](#). arXiv.

22 Translators without Borders. (2019). [MSNA language data can help humanitarian workers communicate better with affected people](#).

23 Lewis, W.D., Munro, R. and Vogel, S. (2016). [Crisis MT: Developing A Cookbook for MT in Crisis Situations](#), Microsoft.

24 Cela fait écho au sous-investissement de longue date dans la préparation aux situations d'urgence et la réduction des risques de catastrophe dans le secteur humanitaire lui-même.

Des services humanitaires numériques plus inclusifs

Il est possible de créer des services humanitaires numériques plus inclusifs pour les locuteurs de langues marginalisées, mais cela nécessitera des changements dans l'ensemble du secteur. Toutes les parties prenantes devront s'attaquer aux questions fondamentales sur la manière dont elles utilisent les services numériques et les raisons pour lesquelles elles les utilisent, tirer les leçons des (rares) bons exemples de services numériques inclusifs et capitaliser sur les opportunités inexploitées pour élargir l'accès grâce à des pratiques collaboratives et fondées sur des données probantes.

Définir le rôle des services numériques

Les services numériques jouent de nombreux rôles dans le secteur humanitaire. Cependant, les participants à l'étude ont exprimé des points de vue très différents sur la manière dont la technologie numérique peut favoriser l'inclusion des locuteurs de langues marginalisées.

Surmonter l'exclusion

Les technologies numériques permettent notamment aux personnes qui ne peuvent pas se rencontrer en personne de communiquer, ce qui en fait un outil idéal pour atteindre des segments de la société qui sont autrement exclus. Cette exclusion peut être le résultat d'une inégalité sociale et peut toucher des pans plus larges de la société, notamment lorsque certains groupes – comme les femmes – ne bénéficient pas des mêmes droits que d'autres. Pour beaucoup, la valeur des services humanitaires numériques est donc de permettre aux groupes marginalisés et exclus d'obtenir les informations dont ils ont besoin et d'être entendus.

« [Le service] n'est pas utile pour quelqu'un qui sait lire et écrire, qui a une adresse électronique et qui parle la langue majoritaire. [...] Nous ne leur sommes pas utiles. Nous sommes utiles à ces personnes qui n'ont pas accès aux mécanismes de rétroaction à effet rapide. »

– Fournisseur numérique humanitaire, Europe

Certains informateurs considèrent les services numériques dans les langues marginalisées comme un moyen d'amplifier la voix de ceux qui reçoivent l'aide humanitaire et de modifier la dynamique du pouvoir en permettant aux citoyens de demander des comptes aux organisations humanitaires.

« Nous recueillons les points de vue des gens, car nous pensons que ces populations qui ont vécu une situation de crise savent, bien évidemment, ce dont elles ont besoin, mais elles peuvent aussi se rassembler et proposer des idées de solutionset de ce à quoi devrait ressembler leur avenir. »

– Fournisseur numérique humanitaire, Océanie

Atteindre ceux qui sont déjà en ligne pour libérer des ressources pour l'engagement en personne

Pour d'autres, l'objectif principal des technologies numériques est d'atteindre les locuteurs de langues couramment utilisées qui sont déjà en ligne. Les gains d'efficacité liés à l'automatisation, affirment-ils, libèrent des ressources pour fournir un soutien en face à face aux personnes moins connectées.

« Je pense toujours que [les organisations humanitaires] planifient en fonction des minorités – des populations les plus vulnérables, les moins représentatives du point de vue linguistique. Je leur réponds que ce service n'est pas pour eux. Nos outils s'adressent aux 60 à 70 % de la population que vous pouvez atteindre sans avoir à vous présenter en personne ; cela laisse à l'ONG toute liberté pour s'adresser aux autres en personne. »

– Fournisseur numérique humanitaire, Europe.



Étude de cas : **Talk to Loop**



Loop opère actuellement en 15 langues en Indonésie, aux Philippines, en Pologne, en Somalie, en Ukraine et en Zambie. L'arabe, l'anglais, le français et l'espagnol sont principalement disponibles pour le personnel des organisations humanitaires, tandis que les communautés touchées par la crise communiquent sur la plateforme dans les langues nationales et locales. Dans chaque pays d'opération, Loop travaille par l'intermédiaire d'un réseau d'organisations locales qui reconnaissent la valeur à long terme de la plateforme. En Zambie, par exemple, l'accent a été mis sur la responsabilisation des institutions publiques dans un contexte où l'un des partenaires de la société civile de Loop nous a dit que la confiance dans les autorités gouvernementales était faible.²⁵

Loop vise à contribuer à « une nouvelle ère de responsabilité dans les secteurs humanitaire et celui du développement international ». Les utilisateurs visés sont les membres de la communauté, les organisations humanitaires, les gouvernements et toute autre personne intéressée par la responsabilité. Les commentaires sur la plateforme deviennent une ressource collective : tout le monde peut les voir et les utiliser pour demander des comptes aux intérêts puissants.

Les barrières linguistiques empêchant de nombreuses personnes d'accéder aux mécanismes de feedback existants, pour Loop, la langue est « au cœur de ce que nous faisons pour essayer d'améliorer l'accessibilité ». Les canaux et

les langues utilisés sont initialement déterminés par des comités de pilotage locaux, puis de nouveaux ajouts sont effectués en fonction des commentaires de la communauté.

Loop adopte une approche pragmatique de la technologie linguistique. Malgré la qualité initiale médiocre des traductions en langues marginalisées, Loop utilise l'apprentissage automatique pour faire gagner du temps aux modérateurs. La transcription automatique est généralement médiocre pour les langues de Loop, à l'exception de l'anglais, de sorte que les modérateurs transcrivent manuellement ou corrigent les transcriptions automatiques et leurs traductions. La qualité s'est améliorée au fil du temps, car les traductions corrigées sont réinjectées dans la mémoire de traduction propriétaire du fournisseur de langues de Loop. En conséquence, Loop estime que, même avec la nécessité de correction, la technologie linguistique permet de gagner environ 25 % du temps d'un modérateur.

Loop collecte en permanence d'énormes quantités de données vocales sur les langues marginalisées, dont la plupart ne sont pas sensibles. L'entreprise estime qu'il est utile de partager ces données de manière appropriée pour soutenir le développement des technologies linguistiques, mais elle manque actuellement de ressources internes et d'expertise pour gérer les processus que cela impliquerait.

→ Lire l'étude de cas complète ici.

²⁵ Remarque : En février 2024, Talk to Loop a annoncé qu'elle mettait la plateforme en « hibernation » en raison de pénuries de financement. Voir : Talk to Loop (16 février 2024) [Loop enters hibernation amid pressing challenges](#).

Identifier les caractéristiques et les bonnes pratiques des services humanitaires numériques inclusifs

Même si tous les services humanitaires ne disposent pas des mêmes ressources, il devrait être courant de communiquer, dans la mesure du possible, dans les langues, les formats et les canaux que les groupes marginalisés sont à l'aise d'utiliser. Les services humanitaires numériques qui le font bien ont plusieurs points communs :

Créer de la valeur grâce à une conception centrée sur l'utilisateur

« Il était vraiment important que cette plateforme soit conçue par des femmes et des filles. Et pas seulement par des suppositions sur ce dont elles auraient besoin ou ce qui leur conviendrait. »

- Fournisseur numérique humanitaire, Europe

Les considérations linguistiques doivent faire partie d'une recherche approfondie auprès des utilisateurs sur la conception de nouveaux services humanitaires numériques, car cela garantira que les décisions sur les langues appropriées sont bien informées et prises tôt.

C'est important pour de nombreuses raisons. Dans certains cas, la langue de communication est moins une question de compréhension que de dialogue constructif basé sur la confiance. Par exemple, un service d'information pour les réfugiés s'est étendu au-delà des langues internationales comme le français au motif que « dans les situations de stress, ils préféreront de loin avoir l'information en lingala ou en kinyarwanda, par exemple ». Une plateforme numérique qui s'adresse aux jeunes au Kenya a constaté que, même si son public parle également le swahili et l'anglais, les questions sensibles sont mieux abordées en sheng :

« Les autres langues sont trop formelles et vous essayez d'aider un jeune à s'orienter dans la réalité de sa vie d'une manière qui lui importe. Faites cela en swahili ou en anglais et il aura l'impression de retourner à l'école. Mais le sheng est leur langue, celle dans laquelle il se sent à l'aise. Notre théorie du changement est basée sur les conversations – briser les mythes, faire émerger les sujets tabous. Nous devons donc parler dans leur langue. »

- Informateur de la société civile, Kenya

Les recherches sur les utilisateurs peuvent également aider à identifier les mots clés et les concepts auxquels les utilisateurs peuvent s'identifier. Un répondant a donné un exemple d'informations de service public destinées aux réfugiés et aux migrants pakistanais en Italie, qui devaient être localisées avec des références aux systèmes que les gens connaissent au Pakistan. Cela impliquait de fournir à la fois les termes en ourdou et en italien pour les concepts clés afin que les gens puissent les reconnaître en italien lorsqu'ils les voient.

Développer des approches holistiques pour l'inclusion linguistique

Les barrières linguistiques aggravent souvent d'autres désavantages, et les locuteurs de langues marginalisées peuvent bénéficier de pratiques conçues pour surmonter ces barrières. Les nouveaux services devraient développer des stratégies d'inclusion en réponse aux différents types d'exclusion numérique que subissent les locuteurs de langues marginalisées. Les informateurs fournissant des services numériques destinés à la communauté ont souligné les exemples suivants :

- SMS, numéros de téléphone gratuits, appels manqués et systèmes de rappel compatibles avec les téléphones de base sans frais
- Communication vocale et picturale et IVR basé sur un clavier pour aider les utilisateurs moins alphabétisés, de langue seconde et malvoyants ou malentendants
- Chatbots et rapports via des plateformes de messagerie sur Internet que les utilisateurs connaissent déjà
- Partage de contenu hors ligne pour les utilisateurs ayant une mauvaise connectivité et un accès Internet

L'interaction humaine est souvent essentielle, tant pour l'accessibilité que pour instaurer la confiance dans des technologies inconnues et de nouveaux services. Plusieurs organisations humanitaires déploient du personnel ou des bénévoles qui parlent les langues locales pour informer les communautés sur leurs services et leur expliquer comment les utiliser.

D'autres utilisent la technologie pour faciliter l'interaction humaine, notamment dans des contextes linguistiquement divers où les membres du personnel ne parlent pas nécessairement toutes les langues des communautés desservies. Lorsque des questions sont posées à un service d'information à distance dans une langue que la plateforme du pays ne couvre pas, la communication en ligne permet aux opérateurs possédant les compétences linguistiques adéquates de répondre facilement depuis un autre pays..

Considérer le consentement et l'éthique

L'utilisation des seules langues dominantes dans les services humanitaires numériques pose des problèmes éthiques liés au consentement éclairé et au déséquilibre des pouvoirs au sein du secteur. Le consentement humanitaire est déjà problématique lorsque le refus peut entraîner le refus d'assistance, et le consentement à la collecte de données numériques est problématique lorsque l'intelligence artificielle (IA) permet de réutiliser les données pour développer de nouveaux systèmes.²⁶ Ce problème est amplifié lorsque les personnes doivent utiliser un service dans une deuxième langue et lorsque les concepts impliqués dans les discussions sur la protection des données n'ont pas d'équivalent local ou ne sont pas d'usage courant.²⁷ Cela a des répercussions à tous les niveaux, de l'évaluation des besoins à la conception des services, en passant par leur utilisation, la confiance et l'engagement.



« Le consentement éclairé est un véritable défi éthique s'il n'est pas exprimé dans la langue des personnes concernées. [...] Si les personnes ne comprennent pas à quoi elles consentent, [...] nous ne pouvons pas vraiment considérer ce consentement. »

– Fournisseur numérique humanitaire, Europe

Plusieurs informateurs fournissant des services numériques sont abordés ces problèmes dans les interfaces utilisateur, par exemple, en :

- Demander aux utilisateurs de sélectionner leur langue préférée avant de demander un consentement éclairé dans cette langue
- Permettre aux utilisateurs de refuser de partager des informations personnelles
- Pour une plateforme conçue pour les utilisateurs à risque de VBG, ne collecter aucune donnée personnelle
- Protéger l'identité des utilisateurs des plateformes de messagerie ou de médias sociaux tierces utilisées pour communiquer avec eux

Le développement des technologies linguistiques soulève également des questions de protection des données et d'éthique, qui n'ont pas été largement abordées à mesure que l'utilisation humanitaire des technologies linguistiques s'est développée. La technologie vocale est particulièrement difficile car les données vocales permettent d'identifier un locuteur individuel. Étant donné que les données vocales sont nécessaires pour créer la reconnaissance vocale, la transcription automatique et d'autres outils de communication non textuelle, le défi de la protection des données a un impact disproportionné sur le développement technologique pour les personnes moins alphabétisées et les locuteurs de langues orales.

Certaines communautés linguistiques marginalisées rejettent tout développement de technologie linguistique qui diminue leur contrôle et leur propriété sur leur langue et leur culture.²⁸ D'autres commentateurs ont critiqué la « colonisation » des langues autochtones par une approche de la technologie linguistique qui les traite comme une marchandise.²⁹ Le développement de la technologie linguistique pour l'action humanitaire doit répondre à ces préoccupations.

26 Beduschi, A. (2022). [Harnessing the potential of artificial intelligence for humanitarian action: Opportunities and risks](#). ICRC.

27 Iacucci, A.A. (2021). [Using Social Media in Community-Based Protection](#). UNHCR.

28 Te Mana Raraunga. (2018). [Principles of Māori Data Sovereignty](#).

29 Bird, S. (2020). [Decolonising Speech and Language Technology](#). ACL Anthology.

Des opportunités inexploitées pour élargir l'accès grâce à des pratiques collaboratives et fondées sur des données probantes

Cette recherche a identifié plusieurs opportunités pour rendre les services humanitaires numériques plus inclusifs pour les locuteurs de langues marginalisées qui, apparemment, restent inexploitées.

Travailler avec les locuteurs de langues marginalisées et la société civile

Les informateurs de la société civile étaient en grande partie enthousiastes, mais pas particulièrement bien informés, sur les possibilités offertes par les technologies linguistiques. Les communautés qu'ils soutiennent peuvent raisonnablement craindre leur impact sur leur identité culturelle. Pourtant, leur participation peut combler les lacunes en matière de connaissances et leur consentement sera essentiel pour exploiter les données linguistiques humanitaires en vue d'obtenir des résultats humanitaires.

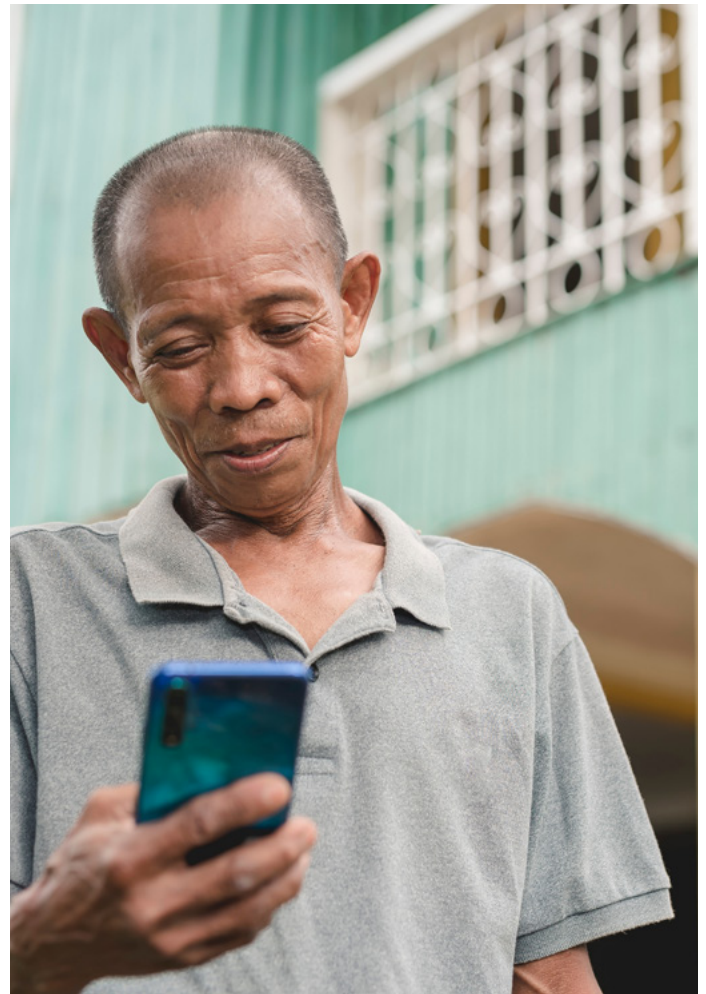
Sans travailler directement avec les locuteurs de langues marginalisées pour développer des services et des technologies numériques, la qualité, la pertinence et la durabilité seront compromises.

« Si vous n'impliquez pas quelqu'un qui parle la langue lorsque vous créez quelque chose, les choses se gâtent. Prenez la séparation des mots [pour la mise en forme imprimée] : un Européen dirait : il suffit de séparer les espaces. Mais en thaï ou en chinois, cela ne fonctionne pas, car leurs langues n'utilisent pas les espaces de la même manière. »

- Technologue linguistique, Amérique du Nord

Sans cette collaboration, le résultat pourrait être totalement ou partiellement sans intérêt pour les membres des communautés linguistiques concernées, qui peuvent ne pas avoir intérêt à se connecter à Internet, ne pas vouloir que leur langue soit marchandisée pour un marché mondial ou préférer une deuxième ou troisième langue pour la communication numérique.³⁰

Des espaces de collaboration se sont ouverts entre les communautés linguistiques marginalisées et les technologues linguistiques. Des plateformes comme [Lanfrica](#), [Common Voice](#) et [Keyman](#) rendent le processus de développement d'un modèle vocal ou d'un clavier pour une langue non prise en charge accessible à de petites équipes de linguistes et d'informaticiens. Des espaces de collaboration comme le [Hugging Face Hub](#) permettent aux développeurs de s'appuyer sur le travail des autres, tandis que des entreprises centrées sur la communauté comme [Karya](#) visent à générer et à organiser des données linguistiques pour les langues marginalisées sans exploiter les communautés qui les parlent. Bien qu'elles aient leurs limites, les technologues ont décrit ces innovations comme des « changeurs de jeu », élargissant rapidement la disponibilité des données linguistiques et accélérant le développement d'applications.³¹



³⁰ Bird, S. (2022). [Local Languages, Third Spaces, and other High-Resource Scenarios](#). Association for Computational Linguistics.

³¹ Cependant, les défis restants incluent la dépendance à l'égard des bénévoles pour les données linguistiques et des smartphones et ordinateurs, le processus de configuration à forte intensité de main-d'œuvre et la nécessité de collecter d'abord suffisamment de données textuelles pour être lues pour les données vocales, ce qui peut exclure certaines des langues les moins prises en charge.

Forger des partenariats axés sur les cas d'utilisation pour les technologies linguistiques

Les communautés marginalisées sont à la fois la source et les bénéficiaires potentiels des données linguistiques, qui sont essentielles au développement des technologies linguistiques. Les organisations humanitaires qui utilisent les technologies numériques pour recevoir et répondre aux informations, aux commentaires et aux questions des communautés marginalisées génèrent et collectent toutes des données linguistiques au cours du processus.

Un écosystème de communication humanitaire dans lequel chaque message vocal non sensible, transcription ou traduction était utilisé en toute sécurité pour améliorer la précision et la rapidité de la communication multilingue pour tous les utilisateurs aurait des avantages considérables pour les locuteurs de langues marginalisées.

Les efforts de collecte de données linguistiques universitaires manquent généralement de se concentrer sur un cas d'utilisation réel. Lorsque très peu de données sont disponibles, l'expérience montre qu'il est plus efficace de collecter uniquement les données nécessaires à une application ou un service spécifique. Pour créer les modèles de langage et de parole pour des applications réellement utiles, la collecte de données axée sur les cas d'utilisation doit se concentrer sur le domaine (santé publique, VBG, aide financière, etc.) et ses utilisateurs visés (âge, sexe, dialecte, alphabétisation, etc.). Les modèles de langage et de parole doivent également être évalués par rapport à des cas d'utilisation réels, car les références universitaires et les données linguistiques générées en laboratoire peuvent ne pas refléter l'utilisation et le contexte réels de la langue.

Cette approche ne serait pas sans défis. Il y aurait un besoin inhérent de donner la priorité au consentement, à la protection des données et à la propriété des communautés locutrices marginalisées pour lesquelles ces services sont conçus. De même, des cadres liés à l'éthique et à la durabilité de telles collaborations devraient être développés entre divers acteurs (communautés, organisations humanitaires, donateurs et fournisseurs de technologies linguistiques commerciales et universitaires).

Pour que ces modèles soient efficaces, la société civile et les communautés marginalisées devraient mieux comprendre comment les technologies

linguistiques sont construites et ce qu'elles peuvent faire. Enfin, les organisations humanitaires devraient veiller à ce que les services hors ligne et en personne dont de nombreux groupes marginalisés ont besoin soient fournis en parallèle et ne soient pas déclassés.

Utilisation pragmatique des langues de contact

De nombreux locuteurs de langues marginalisées parlent également dans une certaine mesure une langue de contact ou une lingua franca. La langue de contact qu'ils parlent peut dépendre de la géographie et de l'ethnicité. Ceux qui ont peu ou pas d'éducation formelle dans cette langue la parleront et la comprendront avec plus de confiance et de précision dans les domaines pour lesquels ils l'utilisent régulièrement, comme les transactions commerciales. Les informations sur la santé et la bureaucratie gouvernementale restent un défi pour la plupart.

Certains n'auront pas de compréhension fonctionnelle ou de capacité à parler dans la langue de contact. Les tests de compréhension au Nigéria et en RDC ont révélé que c'était le cas de manière disproportionnée pour les personnes âgées et les femmes de tous âges.³² Ces groupes peuvent utiliser des appels et des notes vocales ou des SMS dans leur propre langue lorsqu'ils ont accès à des téléphones portables.

« Le portugais brésilien et européen est bien couvert, mais les Angolais et les Cap-Verdiens ne tireront pas les mêmes avantages de ces outils. L'arabe des Saoudiens et des Égyptiens est mieux modélisé que celui des Marocains et des Irakiens, car la plupart des données proviennent d'Arabie saoudite et d'Égypte. »

- Technologue linguistique, Amérique du Nord

Des efforts systématiques pour collecter et partager des données sur l'utilisation des langues pour des communautés linguistiques marginalisées spécifiques pourraient maximiser la portée des services numériques pour les utilisateurs ayant des capacités en seconde langue et identifier les langues de contact les plus pertinentes dans lesquelles investir. Les organisations humanitaires et les développeurs de technologies linguistiques auraient tous deux intérêt à travailler avec la société civile pour recueillir ces informations.

³² Translators without Borders. (2017). [Language profile of five IDP sites in Maiduguri](#). Translators without Borders (2019) [Missing the mark? People in eastern DRC need information on Ebola in a language they understand](#).

Conclusions

Les services numériques ne sont pas la réponse à tous les besoins ou à tous les individus. Le soutien en personne sera toujours essentiel pour répondre aux crises humanitaires. Cependant, lorsqu'ils sont bien exécutés, les services numériques ont le potentiel d'accroître la participation, la responsabilisation et l'accès aux services pour les communautés marginalisées du monde entier.

La plupart des services numériques humanitaires ne sont disponibles que dans un nombre limité de langues. Des millions de personnes touchées par des crises sont ainsi exclues. Les prestataires de services ont souvent peu d'informations sur l'identité de ces personnes et sur l'ampleur du problème. Les barrières linguistiques aggravant d'autres désavantages, les groupes potentiellement vulnérables sont touchés de manière disproportionnée par l'exclusion fondée sur la langue.

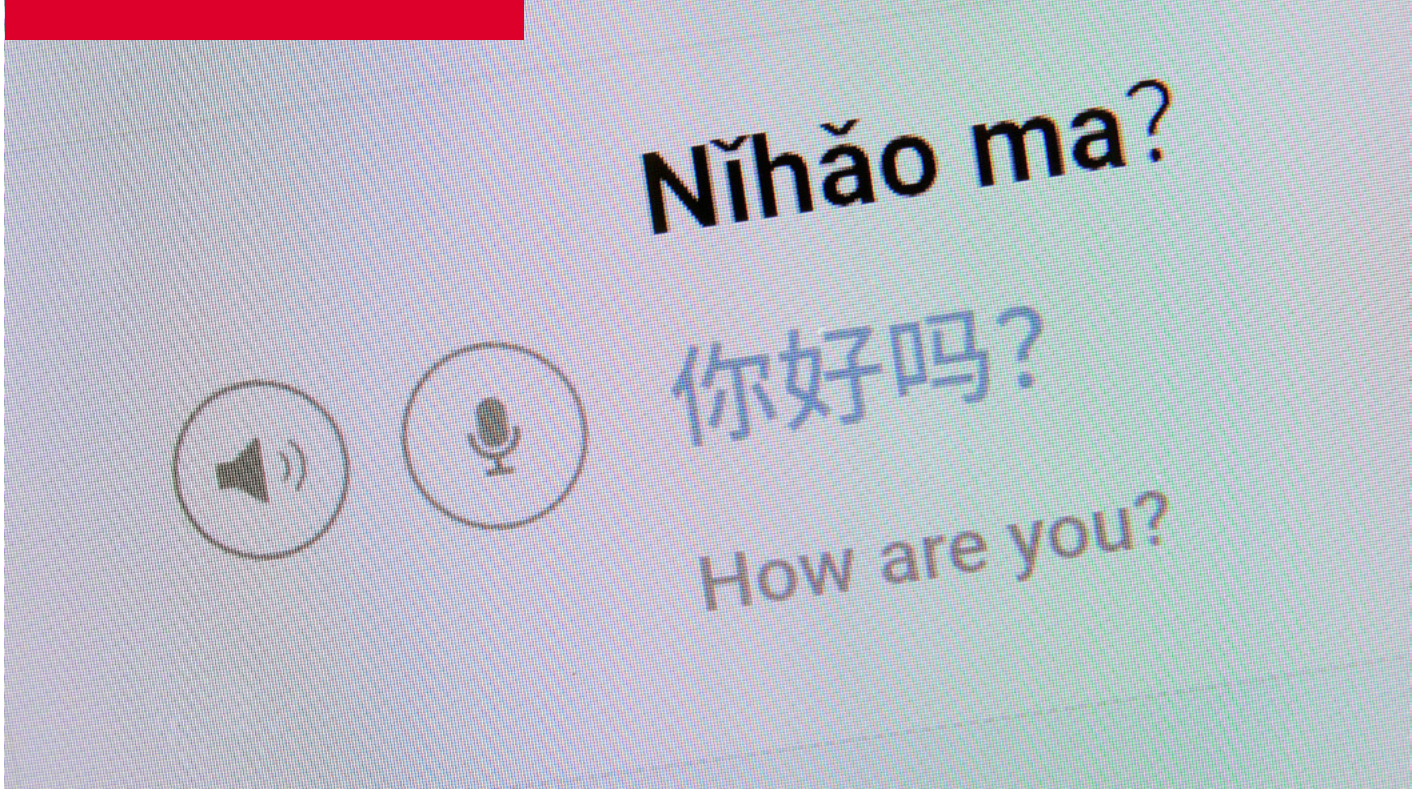
Le succès, le rapport qualité-prix ou le retour sur investissement des services numériques se mesurent principalement en fonction du nombre d'utilisateurs atteints plutôt que de la compréhension des besoins les plus importants.

Cela peut réduire leur impact potentiel et, pour les locuteurs de langues marginalisées, cela représente une perte bien plus grande.

Des informations et une efficacité peuvent être obtenues en s'associant à des organisations ayant des liens avec des communautés marginalisées et en soutenant le développement de services numériques par ces dernières. Les bonnes pratiques existantes peuvent également maximiser l'accessibilité, comme l'utilisation d'un langage simple, de graphiques testés par les utilisateurs et d'une combinaison de formats et de canaux dans la mesure du possible, y compris des services parallèles hors ligne et en personne.

Développer des services humanitaires numériques plus accessibles et plus efficaces pour les locuteurs de langues marginalisées nécessitera une approche multidimensionnelle et intersectorielle qui place les besoins et les expériences des utilisateurs visés au centre. Cela implique d'utiliser de meilleures données comme base pour un ciblage plus éclairé et transparent, et de rechercher des gains d'efficacité grâce à un développement de technologies linguistiques plus collectif et axé sur les cas d'utilisation.





Recommandations

Les évaluations des besoins en matière de services numériques (et d'autres services humanitaires) doivent être menées dans les langues marginalisées.

Des outils comme Kobo ToolBox, qui prend en charge les questions graphiques et audio préenregistrées ainsi que la transcription et la traduction automatiques, peuvent favoriser des évaluations des besoins plus inclusives. Les prestataires de services doivent baser leur sélection de langues sur les données d'utilisation des langues de la population cible et fournir aux collecteurs de données une formation et un soutien linguistique adéquats. Les donateurs, quant à eux, peuvent reconnaître l'importance de données d'évaluation des besoins de qualité en définissant leurs exigences de qualité et en fournissant le financement nécessaire pour y répondre.

La communication et la conception des services doivent être basées sur la recherche et le suivi au-delà des utilisateurs existants.

Les informations recueillies auprès des personnes qui n'utilisent pas le service sont essentielles pour identifier et combler les lacunes. Elles doivent être systématiquement recueillies dans des langues autres que celles actuellement utilisées par le service. Les recherches sur l'utilisation des langues peuvent conduire à une plus grande efficacité et à une plus grande transparence en clarifiant les besoins de communication des utilisateurs potentiels. Les prestataires humanitaires et leurs donateurs doivent réfléchir à la meilleure façon d'intégrer ces deux composantes dans la conception et la prestation des services.

Les donateurs pourraient soutenir une plus grande efficacité, un plus grand impact et une plus grande durabilité en encourageant et en soutenant le partage de données et la collaboration entre leurs partenaires.

Les organisations devraient partager leurs données sur l'utilisation des langues et les préférences de communication de la communauté, ainsi que sur les personnes qui ne peuvent pas accéder aux services, afin d'éclairer les efforts de réponse à l'échelle de la communauté et de sensibiliser. Un financement pourrait être mis à disposition pour des collaborations qui s'appuient sur et contribuent à soutenir les plateformes et les services fonctionnant dans les langues marginalisées, conformément aux Principes pour le développement numérique.³³

Les organisations humanitaires devraient appliquer dans toute la mesure du possible les bonnes pratiques existantes en matière de communication inclusive.

Il faut que les communications – numériques ou autres – utilisent des langues, des formats et des canaux qui atteindront les groupes marginalisés. Cela commence par des données sur les besoins en communication. De telles pratiques ne sont pas toujours coûteuses : l'utilisation d'un langage clair par défaut, l'embauche de personnel parlant les langues locales et l'élaboration de glossaires de concepts clés dans les langues locales sont des pratiques très utiles mais peu coûteuses.

Une approche plus cohérente et centrée sur l'utilisateur est nécessaire en matière de consentement et de protection des données.

Ces préoccupations concernent en particulier les locuteurs de langues marginalisées qui communiquent dans une deuxième langue. Certains participants à la recherche investissent dans des mesures de protection technologiques et des processus de consentement en langage clair, car la compréhension est essentielle à un consentement éclairé et significatif. Les bonnes pratiques doivent être diffusées et soutenues pour devenir la norme dans l'ensemble du secteur.

Affecter des ressources directes au développement de technologies linguistiques qui répondent aux besoins des communautés linguistiques marginalisées dans les situations de crise.

Les donateurs et les entreprises technologiques devraient envisager de financer des recherches sur des cas d'utilisation réels dans le domaine humanitaire. Ces recherches devraient s'appuyer sur des données relatives à l'utilisation des langues maternelles et de contact pour des communautés linguistiques marginalisées spécifiques. La technologie linguistique qui en résulterait devrait être largement accessible aux intervenants humanitaires.

Identifier les obstacles qui empêchent les organisations humanitaires de participer à l'utilisation et à l'amélioration collectives des technologies linguistiques pour les langues marginalisées.

Le développement de formats et de flux de travail communs pour que les humanitaires puissent partager des données vocales et textuelles, ainsi que la coordination de la mise à jour et de la communication des processus de consentement éclairé, pourraient améliorer l'adoption, l'efficacité et la transparence.³⁴ L'expertise requise pour gérer ces processus devrait devenir une composante standard des budgets de projet. Des plateformes collectives telles que Common Voice et Hugging Face pourraient commencer à être utilisées en collaboration avec les chercheurs en technologies linguistiques sans attendre une coordination à l'échelle du secteur.

³³ Principles for Digital Development. (n.d.). [Build for Sustainability website](#).

³⁴ Reconnaître les sensibilités concernant le risque d'exploitation, qui est grave et doit toujours être pris en compte avant d'utiliser les données de quiconque.

Les donateurs humanitaires, les gouvernements et les principaux développeurs de technologies linguistiques devraient envisager d'établir un cadre sectoriel coordonné pour le développement des technologies linguistiques.

Les avantages pourraient être considérables, tant au sein qu'au-delà de l'action humanitaire. En plaçant l'utilisateur de services au centre d'un écosystème technologique, les prestataires de services de toutes sortes pourraient adapter progressivement leurs services aux préférences d'un éventail plus diversifié d'utilisateurs. Les citoyens pourraient utiliser des technologies linguistiques accessibles pour participer aux débats nationaux et internationaux qui leur sont actuellement fermés, pour s'organiser au sein des communautés et pour comprendre et revendiquer leurs droits. Les gouvernements pourraient les utiliser pour consulter et communiquer plus efficacement avec les citoyens sur les politiques et les risques pour la santé publique. Les technologies linguistiques pourraient compléter, et non remplacer, les services en personne afin d'étendre leur portée et leur efficacité.

Impliquer les communautés et la société civile dans les décisions relatives au développement des technologies linguistiques pour leurs langues.

Le consentement libre et éclairé des communautés sera nécessaire pour la collecte de données et le développement de technologies.³⁵ Pour éviter de participer à l'exploitation des communautés marginalisées, les humanitaires doivent s'assurer que toute technologie linguistique développée pour leurs langues est adaptée à leurs besoins et leur sera bénéfique. Cela nécessitera des discussions transparentes et réalistes sur les risques et les avantages potentiels du développement de technologies linguistiques, ainsi que sur la manière dont les communautés souhaitent que cela se produise.

Envisager un soutien ciblé aux communautés linguistiques les plus marginalisées pour développer les données linguistiques comme point de départ de la technologie linguistique.

Même avec une meilleure coopération, il y aura toujours des langues qui ne seront pas couvertes dans un délai raisonnable. Certaines langues peuvent tout simplement être excessivement chères pour le secteur humanitaire et non rentables pour le secteur privé. Toutefois, les donateurs dont le mandat dépasse le cadre de l'action humanitaire peuvent investir dans des interventions ciblées pour aider ces communautés linguistiques à créer des données linguistiques, éventuellement en collaboration avec des chercheurs universitaires.

Investir dans les services numériques locaux.

Bon nombre de problèmes évoqués dans ce rapport pourraient être mieux traités si les organisations proches des communautés marginalisées jouaient un plus grand rôle dans la conception et la fourniture de services numériques à leur intention.

³⁵ OHCHR. (2013). [Free, Prior and Informed Consent of Indigenous Peoples](#).

Annexe : Méthodologie

Ce rapport est basé sur une étude documentaire détaillée, ainsi que sur des enquêtes en ligne et des entretiens avec des informateurs clés (KII) avec trois groupes de participants. Ces groupes ne s'excluaient pas mutuellement, par exemple, plusieurs organisations étaient à la fois des organisations de la société civile et des fournisseurs de services numériques.

- **Organisation de la société civile (OSC)** : OSC des pays à revenu faible et intermédiaire ayant une connaissance directe de l'accès et de l'expérience des locuteurs de langues marginalisées en matière de technologie numérique.
- **Commissaire aux technologies/fournisseur de services numériques** : organisations internationales qui commandent et fournissent des services numériques, avec une connaissance globale.
- **Technologue** : experts en technologies linguistiques ayant une connaissance de la faisabilité et des aspects pratiques d'applications spécifiques.

Enquêtes

Des enquêtes ont été élaborées pour recueillir des données quantitatives et qualitatives auprès des OSC et des commissaires aux technologies (humanitaires)/fournisseurs de services numériques.

L'enquête auprès des OSC était disponible en anglais, français, espagnol et arabe. L'enquête auprès des commissaires aux technologies était disponible en anglais. Les enquêtes ont été diffusées en ligne, via des partenaires, des newsletters et des invitations directes par courrier électronique de CLEAR Global et de la GSMA. Le travail sur le terrain a duré du 12 septembre au 15 novembre 2023.

Entretiens avec des informateurs clés

Les entretiens avec des informateurs clés (EIC) ont été menés sur Zoom entre le 2 octobre et le 10 novembre 2023. Des guides d'entretien semi-structurés ont été utilisés pour assurer la cohérence tout en permettant l'exploration. Les participants ont été identifiés par le biais de contacts connus, de répondants à l'enquête et d'un échantillonnage en boule de neige, et sélectionnés pour leur diversité géographique et leur expertise.

Figure 2 : Participants à la recherche par méthode et catégorie

	Enquêtes	EIC	Etude de cas EIC	Total
Organisation de la société civile	75	8	4	80
Fournisseur de services numériques	24	9	4	30
Technologue linguistique	0	5	0	5
Total	99	22	8	115

Limitations

Les répondants à l'enquête et les informateurs clés identifiés par le biais de l'enquête se sont auto-sélectionnés et, par conséquent, les résultats ont une valeur indicative. Les informateurs clés ont été sélectionnés parce que CLEAR Global ou la GSMA savaient qu'ils avaient une expertise et une expérience pertinentes. Bien que tous les efforts aient été faits pour atteindre les OSC du monde entier, les résultats sont déséquilibrés au niveau régional, avec une absence de représentation de l'Asie de l'Est et une surreprésentation de l'Ouganda et du Guatemala.

Bien que l'objectif de l'étude était de se concentrer sur l'inclusion linguistique dans les services numériques humanitaires, il s'agissait également d'une étude préliminaire. Nous étions donc ouverts à entendre un éventail plus large d'expériences, et certains participants ont apporté des points de vue extérieurs à la sphère humanitaire.

Siège social de la GSMA

1 Angel Lane

Londres

EC4R 3AB

Royaume-Uni

Tél. : +44 (0)20 7356 0600

