GSMA

Language and Digital Humanitarian Action

The state of inclusion and exclusion for marginalised language speakers in digital humanitarian services

CLEAR Global

GSMA

The GSMA is a global organisation unifying the mobile ecosystem to discover, develop and deliver innovation foundational to positive business environments and societal change. Our vision is to unlock the full power of connectivity so that people, industry, and society thrive. Representing mobile operators and organisations across the mobile ecosystem and adjacent industries, the GSMA delivers for its members across three broad pillars: Connectivity for Good, Industry Services and Solutions, and Outreach. This activity includes advancing policy, tackling today's biggest societal challenges, underpinning the technology and interoperability that make mobile work, and providing the world's largest platform to convene the mobile ecosystem at the MWC and M360 series of events.

We invite you to find out more at www.gsma.com

Follow the GSMA on Twitter/X: @GSMA

CLEAR Global

CLEAR Global is a US nonprofit helping people get vital information and be heard, whatever language they speak. With its innovative language technology solutions, research, and community of over 120,000 language volunteers, CLEAR Global supports its partner organisations working in various contexts around the world. Envisioning a world without language barriers, CLEAR Global is set to improve global communication and information access.

Find out more at www.clearglobal.org

Follow CLEAR Global on Twitter/X: @CLEARGlobalOrg

UK International Development

Partnership | Progress | Prosperity

GSMA Mobile for Humanitarian Innovation

The GSMA Mobile for Humanitarian Innovation programme works to accelerate the delivery and impact of digital humanitarian assistance. This is achieved by building a learning and research agenda to inform the future of digital humanitarian response, catalysing partnerships and innovation for new digital humanitarian services, advocating for enabling policy environments, monitoring and evaluating performance, disseminating insights and profiling achievements.

Learn more at www.gsma.com/m4h or contact us at m4h@gsma.com

Follow GSMA Mobile for Development on Twitter/X: @GSMAm4d

Lead author

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

Contributors

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

The irony of only publishing this report in English, the world's most powerful language, is not lost on the research team. We aim to republish in additional languages in the future, but recognise they are likely to also be globally powerful languages, excluding marginalised language speakers.

Contents

"To start off any type of interaction, language is the first consideration. So, we need to be able to make services available in the language that people need."

– Humanitarian digital provider, Europe
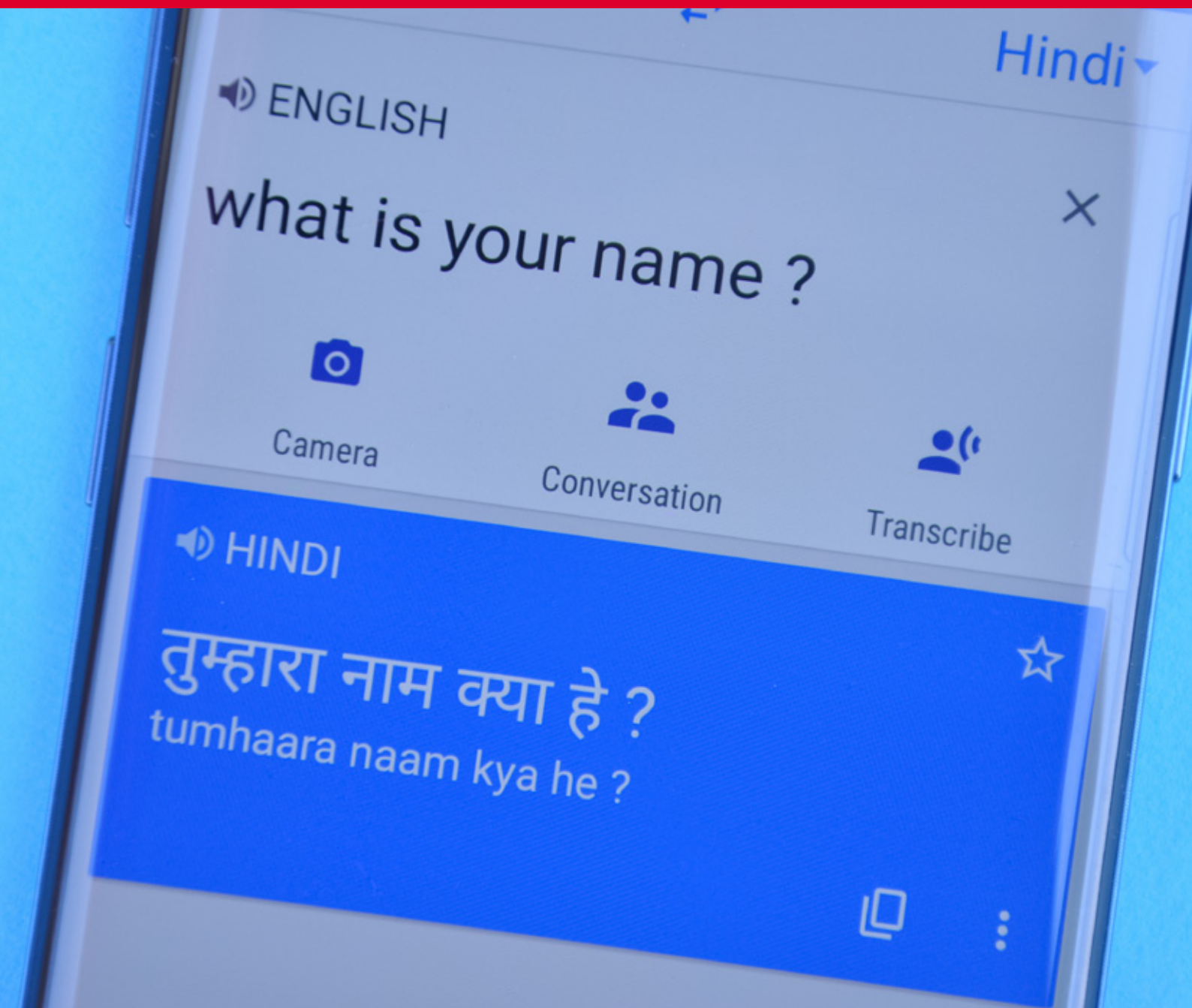
Why does language marginalisation matter in crisis settings?

Language marginalisation

Of the estimated 7,000 human languages, only a handful have a strong online presence and just a few hundred are available on digital platforms. These are broadly the dominant languages of the world's more economically and politically powerful nations. In this context, a language like Swahili, spoken by 100 million people, is marginalised globally.

The digital marginalisation of most of the world's languages and their billions of speakers reflects the real-world marginalisation of language communities in their own countries. For example, a language like Nande, spoken by some 10 million people, is marginalised in eastern Democratic Republic of the Congo (DRC), where official communication is largely in French and Congolese Swahili.

As digital services expand, minority language speakers are becoming more excluded.[1] Digital content and support are concentrated in a fraction of the world's languages, with the most noticeable gaps in African languages[2] and even major languages such as Bengali/Bangla and Hindi.[3] Where digital content is available, it is often in an unuseable format for those who are not literate or whose language is largely unwritten.[4] This can result in minority language speakers being more digitally excluded, including in crisis-affected communities.[5]

The risk of digital language exclusion is intersectional. Worldwide, women and girls, older adults, people with disabilities and minority ethnic groups have less access to education[6] and the opportunities it presents to acquire literacy, numeracy, digital and second-language skills. These same groups are consistently found to have lower levels of access to, and use of, digital technologies.[7]

1   United Nations. (2020). Report of the Secretary-General Roadmap for Digital Cooperation; Jones, S. (28 November 2022). "Understanding the Impact of Digital Language Support". SIL Blog; CLEAR Global: 4 billion conversations website.
2   Jones, S. (28 November 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   For example: Satari, A. (2021) The Mobile Disability Gap Report 2021. GSMA; Jeffrie, N. (2023) The Mobile Gender Gap Report 2023. GSMA.

Language marginalisation in digital humanitarian responses

Factors driving digital exclusion are often compounded in crisis settings,[8] meaning language-based digital exclusion leaves out many of the people humanitarian organisations want to prioritise in an emergency. Civil society organisation (CSO) respondents to this study, most of whom work directly with marginalised language speakers, felt that older adults were the main population segment struggling to use digital technology, followed by women, people with disabilities and minority groups.

Marginalised language speakers often face additional challenges during a crisis, such as difficulty understanding their entitlements. This can be compounded by the disadvantages they already face from not speaking a dominant language, such as limited livelihood, educational and civic participation opportunities. This often intersects with other identity markers, such as gender or disability status, exacerbating intersectional disadvantages[9] and leaving them particularly vulnerable during an emergency.

When humanitarian organisations are not sensitive to linguistic diversity, language can create a barrier, leaving marginalised language speakers struggling to access services and information or to make their voices heard.[10] Assumptions about who speaks a national, dominant, or contact language[11] can compound these issues,[12] and if data is not routinely gathered on people's first or preferred language, this exclusion can easily go unnoticed.[13]

Since most humanitarian digital services are available in only a narrow range of languages, many millions of people are excluded. However, the humanitarian sector often lacks information on who is being excluded from their services. This can worsen exclusion, entrench inequality, and present an ethical challenge to the humanitarian principles of humanity and impartiality.

> "What we've seen from the marginalised communities we are working with is that technology is pushing them away instead of bringing them close."
>
> **– Humanitarian digital provider, Africa**

For example, language barriers prevented Twa refugee women from the DRC from using a mobile-based human rights monitoring service in Uganda addressing sexual exploitation and abuse. In a community of 200, only three young men spoke a common language with CSOs.

In eastern DRC, a WhatsApp group used by a United Nations agency to coordinate activities and feedback from partners and communities communicates only in French and Swahili: "French for expats who don't speak local languages. [...] and community leaders who can't write in French do so in Swahili, then others translate for them." Only community representatives who are literate in Swahili or French and own a smartphone can take part. This largely excludes women and speakers of the dozens of other local languages.

---

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

9   For example: 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 A contact language, or lingua franca, is a means of communicating between speakers of different first languages..

12 For example: 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.

Roughly a third of CSO respondents reported that members of the communities they support have no access to digital services in their own languages, and one service provider reported that in five of the six countries where they operate, community members say that language barriers prevent them accessing feedback mechanisms.

This demonstrates the challenges faced by marginalised language speakers: those who are able to often have to navigate digital services in a second or even third language. This may mean that people miss out on life-saving information, they may not access the assistance intended for them and they may face increased risk of harm, such as fraud. It is essential that humanitarian organisations are aware of these risks when designing digital services like early warning systems (EWS), cash programming or remote registration.

It is especially important when there is no physical interface for humanitarian services.

For example, during the 2022 Pakistan floods, speakers of languages such as Saraiki, Balochi and Sindhi missed vital warnings and updates that had only been sent in Urdu and English via SMS.[14] Some Saraiki-speaking smartphone owners used WhatsApp to send messages and video clips to alert civil society, but the interface did not support their script and those who could were forced to improvise transliteration to communicate.

> "Damage and economic losses increase because of the information gap. It's the digital era – everything is digitised. So the authorities say: we have circulated this information. But the right people aren't getting that information."
>
> – Civil society informant, Asia

In northeast Nigeria, where SMS is used to notify and register internally displaced people (IDPs) for assistance, Mafa speakers are often reliant on others to translate the English and Hausa messages. This leaves them more vulnerable to abuse and exploitation.

> "[Thieves] understand that people have no basic understanding of the process so they try to see how they can manipulate them to get what belongs to them."
>
> – Civil society informant, Africa

---

14   See: CLEAR Global. (2023). A dangerous information gap.

Case study:
Mobile Vaani

Mobile Vaani is a language-agnostic network of voice-based community media platforms that enables two-way communication on critical issues. Developed by Gram Vaani, the platform shares audio content with and from communities and partner organisations in user languages on topics like health and nutrition and gender-based violence (GBV). The interface language is determined by partner organisations in the communities they want to engage with.

The Mobile Vaani Network, operating as hyperlocal instances in over 120 districts, has serviced more than 1 million users in around 20 states and 12 languages. Partners have also been licensed to deploy the platform in Afghanistan, Ethiopia, Namibia, Pakistan and South Africa. An estimated 5 million users have so far generated over 1 million community voice reports, in the form of opinions, queries, grievances, stories and information.

Mobile Vaani works on both basic and smart phones, using voice to overcome literacy and digital skills barriers. Users access and share audio information through keyboard-based IVR using just three number keys. A "missed call"/ automated call-back system makes the service free of charge to users. Organisations can use an Android app to create and share content in local languages. Digital literacy constraints are addressed through simplified menu navigation and parallel in-person capacity building:

— Pre-recorded audio content plays on a continuous loop, and users press 1, 3 or 5 to skip, like or record a response

— Community volunteers promote the platform, support new users and run in-person workshops to ensure there is feedback from diverse voices on the content and platform.

"[With audio] people can engage, express themselves, ask questions. [...] Nobody is left out, literate or nonliterate."

– Deepak Kumar, Senior Programme Manager, Gram Vaani

⟶ **Read the full case study here.**

"If you're talking about information being power, that power is not available to a very large segment of the population because information is not really democratized in the true sense. Information is available to a certain elite, people who are digitally literate, people who have access to these channels and content available in [powerful] languages. [...] It is a combination of not just content availability in multiple languages, but also the technology interfaces."

– Vijay Sai Pratap, Co-founder, Gram Vaani

Barriers to digital humanitarian services in marginalised languages

Although many providers of digital humanitarian services recognise the importance of including marginalised languages, doing so is far from the norm. Huge advances in language technology have been made recently, but few marginalised language communities have reaped the benefits. Several factors seem to be impeding progress.

Awareness of the issue and lack of data

Humanitarian organisations often overestimate how many people can confidently use a second language, and lack data on the most relevant languages to invest in. Without that data, assumptions about the reach of services, especially among marginalised groups, are unreliable. Language-based digital exclusion can remain invisible if it is not explicitly monitored.

The languages a humanitarian organisation uses to develop a service will determine who engages with it and whose needs it serves. This begins with user research and needs assessments. Most informants agreed that for such consultations to capture the needs of marginalised groups, they need to be conducted in the languages they are most comfortable with. Humanitarian service providers consulted were not confident that research and assessment routinely take place in marginalised languages. When they do not, or when assessments are conducted by outsiders, they will not accurately reflect the needs, concerns and priorities of marginalised communities.[15]

Humanitarian organisations may rely on existing language data for local populations as a starting point for information and communication needs assessments. However, in many contexts, detailed information is unavailable, including on the reach

of contact languages or lingua francas.[16] In other instances, colleagues from country offices or partner organisations are assumed to know which languages are relevant, even if they do not have direct experience with language use in marginalised communities.

Because non-users are not captured in platform data, language exclusion can easily go unnoticed. A service provider will only learn there are language communities not being reached if they conduct specific, separate research.

In sudden-onset emergencies, services may go online without prior consultation to meet the most urgent needs quickly. Ideally, user feedback would provide course corrections where needed.

> "We put all content in Ukrainian for the Ukraine response from the start, and users said: 'Why is this not in Russian?' Because we were trying to respond fast, there was no needs assessment. Users asked for Russian, so we managed that, and found those resources."
>
> **– Humanitarian digital provider, North America**

---

15   Translators without Borders. (5 December 2018). The words between us: How well do enumerators understand the terminology used in humanitarian surveys?; Ground Truth Solutions. (27 May 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.

Prioritising the issue

Although humanitarian digital service providers value inclusion, they usually must prioritise the reach and scale of a service to conform to donor or organisational expectations. This can be compounded by value-for-money (VfM) or return-on-investment (RoI) targets, a one-size-fits-all approach that prioritises majority languages to meet the needs of the greatest number of people, entrenching linguistic digital exclusion.[17]

For example, informants described digital services being made available only in Spanish for Central and South American countries, and in Burmese for Myanmar, despite linguistic minorities there experiencing discrimination, disadvantage and sometimes violence, all of which make them priority groups for humanitarian services.

These dynamics come into play when a service is being replicated in a new context. When a GBV reporting and referral app, developed in Nigeria, was scheduled for replication in the DRC, its UN backer and their partners wanted to prioritise French. However, less than 2.5% of people speak French as their main language at home in the DRC.[18]

Often implicit in these decisions is an assumption that non-native speakers of the dominant language will have sufficient second-language knowledge to access information and services. In a context like Somalia, this assumes that affected people can understand and communicate in the dominant Mahaatiri dialect of Somali. Comprehension testing of Mahaatiri-language polio vaccination communication among speakers of the Maay dialect found this assumption to be flawed.[19] Speakers of unrelated languages, such as Mushunguli, can be expected to find communication in Mahaatiri Somali even harder to understand.

> "I might estimate that like 5% to 6% of the minority and marginalised communities use digital services, maybe listening to the radio. The reason being that all the mainstream media and social media speak [in a dialect] or a language that means communities are not interested. And some of them even are not understanding well. And some of them even do not have any information, any awareness message, any key information that is shared by international organisations, UN organisations and government entities. Marginalised communities don't have the possibility to get interested."
>
> **– Civil society informant, Africa**

---

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.

Cost considerations

> "Some people do get left out just because of how the funding works. Not because we don't know there's a problem."
>
> **– Humanitarian digital provider, North America**

For most service providers consulted, cost was the main limiting factor in providing digital humanitarian services in marginalised languages, particularly the cost of translation, transcription and voice-overs. Even well-funded service providers described struggling with the cost of communicating in minority and marginalised languages, especially those not supported by language technology. One estimated that it would take around USD 10,000 to translate, adapt and voice all their content into a single language. As a result, some services are deployed in dominant languages across several countries, with less dominant languages a secondary goal.

> "Because resources are so tight [...] we prioritise the number of people reached [...] but for a nonprofit, inclusivity should be a given."
>
> **– Humanitarian digital provider, North America**

Maintaining a service over time in multiple languages is also a cost barrier, for both content and the user interface. This is especially challenging for content that is subject to rapid change, like security conditions along a migration route. For less commercially available languages, the linguistic and subject matter expertise required can be both harder to find and potentially more expensive.

> "When you make a small fix, you have to do it in all languages – that's the main challenge, not the technical feasibility. You first get it all translated, and then you have to manage updates and fixes."
>
> **– Humanitarian digital provider, Europe**

This can present ethical issues if the budget is available for initial set-up but not maintenance: "It's not okay for us to start content in a new language that we can't sustain." That content would become inaccurate over time if there's not capacity to keep it up to date, and "inaccuracy is misinformation."

These issues are compounded when there is minimal capacity to share the costs of marginalised language support. To some informants' regret, one organisation's investment in multilingualism has generally not contributed to sector-wide communication capacity in the languages concerned, for instance, through shared glossaries of audience-tested terminology.

A lack of inclusive technological solutions

Inclusive language technology – like speech recognition and machine translation – could help overcome cost barriers while also expanding access to digital humanitarian services, and digital services more broadly, for millions of marginalised language speakers. Most digital service providers consulted for this study are already using language technology to, among other things, provide services outside working hours, allow users to communicate however they want, provide consistent information in several languages, and reduce reliance on service providers for repetitive tasks and content. However, a Kanuri or Somali speaker, for example, is still a long way from being able to access digital humanitarian services through automatic speech recognition in their own language.

The pace of language technology development is accelerating, with Meta and Google, for example, releasing language models for many hundreds of languages.[20] However, this development remains profoundly unequal. High-quality output is still limited to a handful of economically powerful languages like English, French, Spanish and Chinese. The languages of crisis-affected communities are largely unserved.

Most of the world's 7,000 human languages are not served by language technology at all. Experimental technology is being developed for roughly 1,000 of them.

---

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.
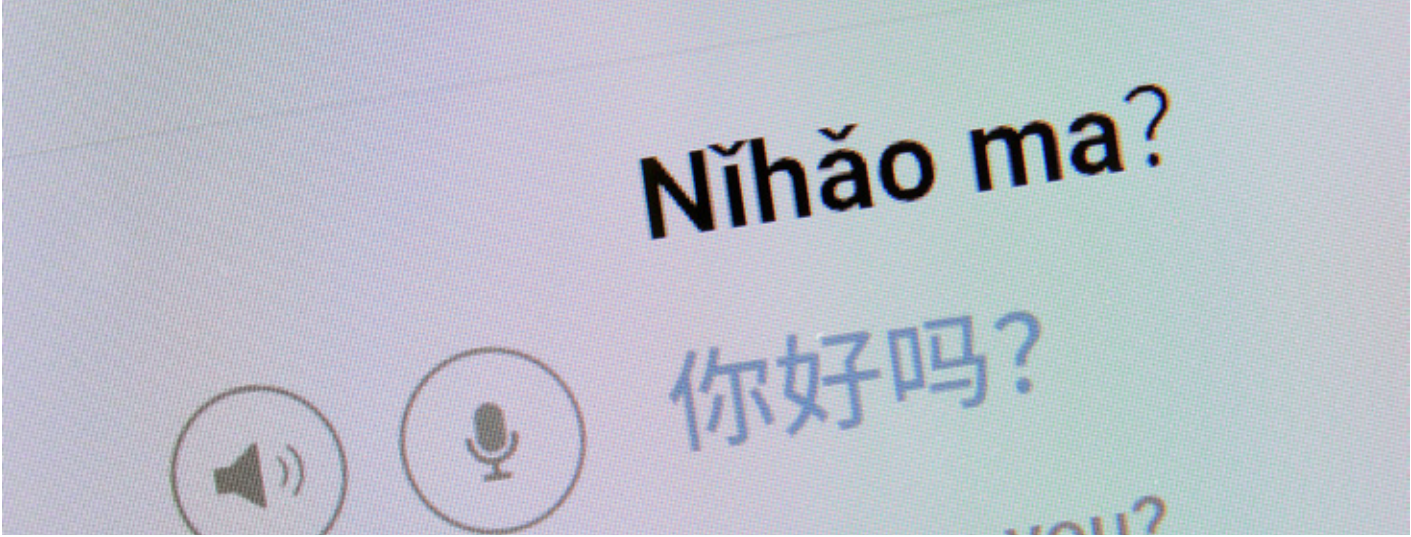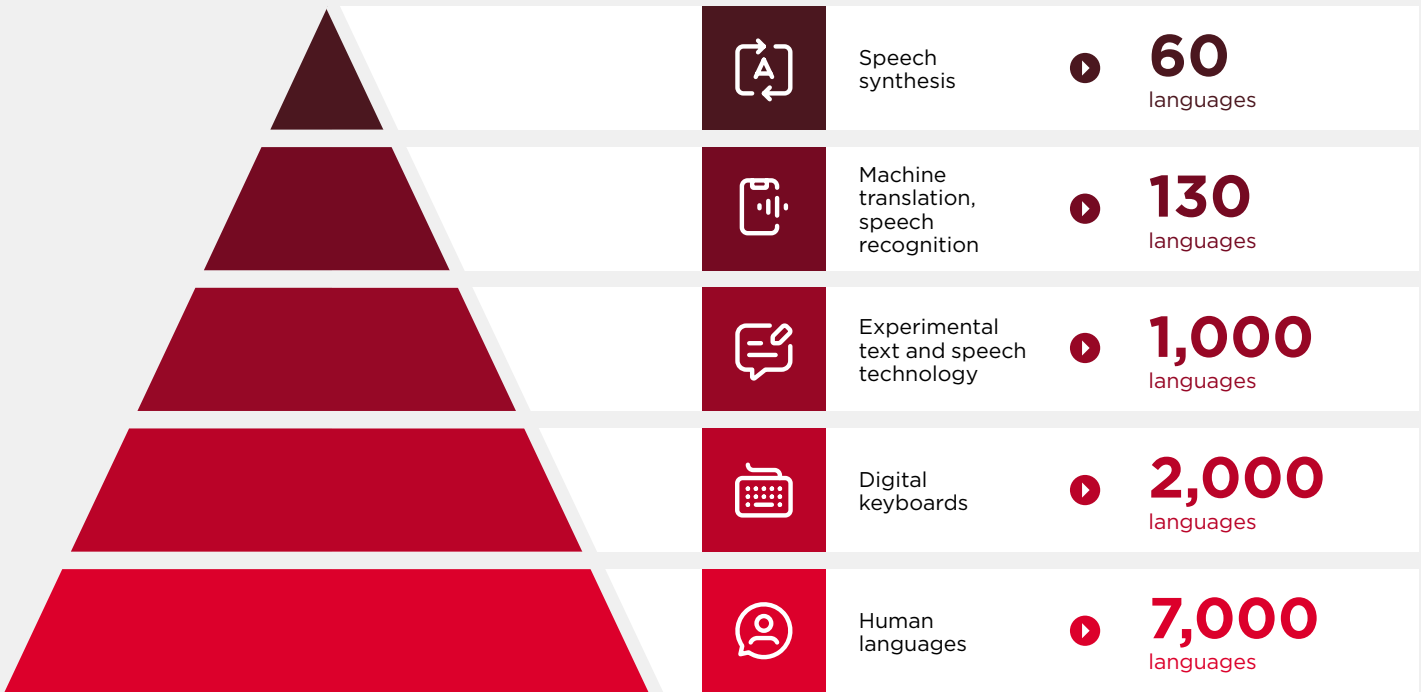
The Unicode Standard, which encodes 161 writing systems for thousands of languages, enables users to write, read and store text on a mobile phone. It is the foundation of text-based language processing like machine translation, transcription, or intent classification. But this processing requires digital keyboards, only available for around 2,000 languages.

The processing technology itself is available for an even smaller group of languages. Machine translation to process and respond to a user query, for example, on how to access humanitarian assistance, is only available in 130 languages and for most the quality is poor. Speech recognition, which would enable a crisis-affected individual to receive conversational responses to a verbal query from a chatbot, is available in a similar number of languages.

Speech synthesis, which gives the user a computer-generated oral version of written text, covers less than 1% of human languages. This is the technology behind screen readers for visually impaired users and can also make content more accessible for a non-literate or non-native speaker. For most languages, technology has not been developed to mimic human speech patterns, which makes it sound robotic and can be off-putting for new users.

Figure 1: The limited range of language technology

Speech synthesis	**60** languages
Machine translation, speech recognition	**130** languages
Experimental text and speech technology	**1,000** languages
Digital keyboards	**2,000** languages
Human languages	**7,000** languages

The lack or poor quality of language technology for all but a few dominant languages is a fundamental limitation to inclusive digital humanitarian services.[21] For example, a digital support service for GBV survivors in northeast Nigeria provides read-out-loud and voice command features, valuable in a region where many intended users cannot read. However, due to a lack of language technology, read-out-loud is in English and only voice commands are also available in Hausa. English is not widely understood, and women are overrepresented in the approximately 17% of people in northeast Nigeria who do not speak Hausa well or at all.[22]

This gulf in technology provision reflects wider inequalities between different language speakers. Unlike most crisis-affected people, speakers of economically powerful languages are more likely to be online and offer a ready market for commercial digital services. This provides both the commercial incentive and opportunity to develop language technology in their languages.

Another challenge is that the technology needs of crisis-affected marginalised language speakers are not a popular research subject outside of major sudden-onset emergencies. Some crises, such as the 2010 Haiti earthquake, have galvanised collaborative efforts to build language technology tools.[23] Yet, there is little sustained effort for the long and difficult task of building technology for the languages of communities affected by recurrent or protracted emergencies, or for developing models that can be rapidly adapted for urgent deployment in new languages.[24]

“Over the past decade we’ve found that people are interested in working on a disaster when it’s happening [...]. But then when it’s over they want to get back to their lives.”

– **Language technologist, North America**

Data – simple text, recorded speech with transcription and parallel texts in language pairs – is the raw material for language technology. A key challenge in expanding the reach of language technology is that most languages, representing billions of people worldwide, have barely any digital representation, and hence little available data.

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 humanitarians 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   This echoes the long-standing underinvestment in emergency preparedness and disaster risk reduction in the humanitarian sector itself.

More inclusive digital humanitarian services

It is possible to create more inclusive digital humanitarian services for marginalised language speakers, but it will require shifts from across the sector. All stakeholders will need to grapple with fundamental questions about how and why they use digital services, draw lessons from the (few) good examples of inclusive digital services and capitalise on untapped opportunities to expand access through collaborative and evidence-based practices.

Defining the role of digital services

Digital services play a variety of roles in the humanitarian sector. However, research participants expressed a clear range of views on how digital technology can support the inclusion of marginalised language speakers.

Overcoming exclusion

Among other things, digital technology enables communication between people who cannot meet in person, making it ideal for reaching segments of society who are otherwise excluded. This exclusion might be the result of societal inequality, and can affect larger swathes of society, including when certain groups – such as women – are not afforded the same rights as others. For many, therefore, the value of digital humanitarian services is enabling marginalised and excluded groups to get the information they need and to be heard.

> "[The service] isn't useful for somebody who's literate, has an email address, speaks in the majority language. […] We're not useful to them. We are useful to these people who can't get access to the quick-win feedback mechanisms."
>
> – Humanitarian digital provider, Europe

Some informants view digital services in marginalised languages as a way to amplify the voices of those receiving humanitarian assistance and to shift power dynamics by enabling people to hold humanitarian organisations to account.

> "[We gather people's] insights because we think these populations who have been in a crisis situation know, obviously, what they need but they can also come together and provide ideas on solutions and what they think their future should look like."
>
> – Humanitarian digital provider, Oceania

Reaching those already online to free up resources for in-person engagement

For others, the main purpose of digital technology is to reach speakers of commonly used languages who are already online. The efficiency gains from automation, they argue, free up resources to provide face-to-face support to less connected individuals.

> "I always think [aid organisations] plan for the minorities – for the most highly vulnerable populations, the least representative from a language perspective. I push back and say: this service is not for them. Our tools engage with the 60–70% you can reach without having to be face-to-face; that leaves the NGO free to engage the others in person."
>
> – Humanitarian digital provider, Europe.

Case study:
Talk to Loop

Talk to Loop, or "Loop", is a digital platform that enables people to safely provide feedback on the humanitarian and public services they receive and report abuse of various kinds, through a range of channels including a low-data website, WhatsApp, Facebook, and voice calls.[25]

Loop currently operates in 15 languages across Indonesia, the Philippines, Poland, Somalia, Ukraine and Zambia. Arabic, English, French and Spanish are available mainly for the staff of humanitarian organisations, while crisis-affected communities communicate on the platform in national and local languages. In each country of operation, Loop works through a network of local organisations that recognise the long-term value of the platform. In Zambia, for instance, the focus has been on holding public institutions to account in a context where one of Loop's civil society partners told us trust in government authorities is low.

Loop aims to contribute to "a new era of accountability in the humanitarian and international development sectors". The intended users are community members, humanitarian organisations, governments and anyone else with an interest in accountability. Feedback on the platform becomes a collective resource – anyone can see it and use it to hold powerful interests to account.

Because language barriers prevent many from accessing existing feedback mechanisms, for Loop, language is "at the heart of what we're doing to try to increase accessibility". The channels and languages used are initially determined by local steering committees, with new additions based on community feedback.

Loop takes a pragmatic approach to language technology. Despite the initially poor quality of marginalised language outputs, Loop uses machine learning to save moderator time. Automated transcription is generally poor for Loop's languages, except English, so moderators either manually transcribe or correct automated transcriptions and their translations. Quality has improved over time as corrected translations are fed back into Loop's language provider's proprietary translation memory. As a result, Loop estimates that, even with the need for correction, language technology saves about 25% of a moderator's time.

Loop is continuously collecting huge amounts of marginalised language voice data, much of which is not sensitive. They see value in sharing this data appropriately to support language technology development, but currently lack the internal resources and expertise to manage the processes this would entail.

⟶ **Read the full case study here.**

---

25   Note: In February 2024 Talk to Loop announced that they were putting the platform into 'hibernation' due to funding shortages.
      See: Talk to Loop (February 16th 2024) Loop enters hibernation amid pressing challenges.

Identifying features and good practices of inclusive digital humanitarian services

While not every humanitarian service is equally resourced, it should be standard practice to communicate, as far as possible, in the languages, formats and channels that marginalised groups are comfortable using. Digital humanitarian services that do it well have several things in common:

Building value through user-centred design

> "It was really important that this platform was designed by women and girls. And not just making assumptions [about] what they would need or what would be suitable for them."
>
> **– Humanitarian digital provider, Europe**

Language considerations need to be part of extensive user research on the design of new digital humanitarian services, as this will ensure decisions on appropriate languages are well informed and taken early.

This is important for many reasons. In some cases, the language of communication is less about comprehension than enabling meaningful dialogue based on trust. For example, an information service for refugees was expanding beyond international languages like French on the grounds that "actually in stress situations they will much prefer to have the information in, for example, Lingala or Kinyarwanda." A digital platform engaging young people in Kenya found that, although their audience also speak Swahili and English, sensitive issues are best discussed in Sheng:

> "Other languages are too formal and you're trying to help a young person navigate the reality of their lives in ways that matter to them. Do that in Swahili or English and they feel they're back to school. But Sheng is their language, the one they are comfortable in. Our theory of change is based on conversations – busting myths, surfacing taboo issues. So we have to speak in their language."
>
> **– Civil society informant, Kenya**

User research can also help identify key words and concepts that users can relate to. One respondent gave an example of public service information for Pakistani refugees and migrants in Italy that needed to be localised with references to the systems people were familiar with in Pakistan. This involved providing both the Urdu and the Italian terms for key concepts so that people could recognise them in Italian when they see them.

Developing holistic approaches to language inclusion

Language barriers often compound other disadvantages, and marginalised language speakers can benefit from practices designed to overcome these barriers. New services should develop inclusion strategies in response to the different types of digital exclusion that marginalised language speakers experience. Informants providing community-facing digital services highlighted the following examples:

— SMS, toll-free phone numbers, missed call and call-back systems compatible with basic phones at no cost

— Voice and pictorial communication and keyboard-based IVR to assist less literate, second-language and visually or hearing-impaired users

— Chatbots and reporting through internet-based messaging platforms that users are already familiar with

— Offline content sharing for users with poor connectivity and internet access

Human interaction is often key, both to accessibility and to building trust in unfamiliar technology and new services. Several humanitarian organisations deploy staff or volunteers who speak local languages to tell communities about their services and explain how to use them.

Others use the technology to facilitate human engagement, particularly in linguistically diverse contexts where staff members may not speak all the languages of the communities served. When questions come to a remote information service in a language the country platform does not cover, online communication makes it easy for operators with the right language skills to respond from another country.

Considering consent and ethics

The use of only dominant languages in digital humanitarian services presents ethical challenges related to meaningful informed consent and power imbalance within the sector. Humanitarian consent is already problematic where refusal may result in the denial of assistance, and consent to digital data collection is problematic where artificial intelligence (AI) enables data to be reused to develop new systems.[26] This is amplified when individuals must use a service in a second language, and where the concepts involved in discussions of data protection have no local equivalents or are not in mainstream use.[27] This has impacts all the way from needs assessment to service design, service use, trust and engagement.

> "Informed consent is such an ethical challenge if it's not in people's own language. […] If they don't understand what they are consenting to, […] then we can't really consider that consent."
>
> – Humanitarian digital provider, Europe

Several informants providing digital services were addressing these issues in user interfaces, for instance, by:

— Asking users to select their preferred language before seeking informed consent in that language

— Enabling users to opt out of sharing personal information

— For one platform designed for users at risk of GBV, collecting no personal data at all

— Shielding users' identities from third-party messaging or social media platforms used to communicate with them

The development of language technology also presents data protection and ethical concerns, which have not been widely discussed as humanitarian use of language technology has expanded. Voice technology is especially challenging because voice data can identify an individual speaker. Since voice data is needed to build speech recognition, automated transcription and other tools for non-text communication, the data protection challenge has a disproportionate impact on technology development for less literate individuals and speakers of oral languages.

Some marginalised language communities reject any development of language technology that diminishes their control and ownership of their language and culture.[28] Other commentators have criticised the "colonisation" of Indigenous languages by an approach to language technology that treats them as a commodity.[29] Language technology development for humanitarian action must address these concerns.

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.

Untapped opportunities to expand access through collaborative and evidence-based practice

This research identified several opportunities to make digital humanitarian services more inclusive for marginalised language speakers that, seemingly, remain untapped.

Working with marginalised language speakers and civil society

Civil society informants were largely enthusiastic, but not particularly well informed, about the possibilities of language technology. The communities they support may reasonably fear its impact on their cultural identity. Yet, their participation can bridge knowledge gaps and their consent will be essential to harness humanitarian language data for humanitarian outcomes.

Without working directly with marginalised language speakers to develop digital services and technologies, the quality, relevance and sustainability will be impaired.

> "If you don't involve someone who speaks the language when you're building something, things just break. Take splitting words [for print formatting]: a European would say: just split on white space. But in Thai or Chinese that doesn't work because their languages don't use white space in the same way."
>
> – Language technologist, North America

Without that collaboration, the outcome may be wholly or partially irrelevant to members of the language communities concerned, who may have no interest in going online, may not want their language commodified for a global market or may prefer a second or third language for digital communication.[30]

Spaces have opened for collaboration between marginalised language communities and language technologists. Platforms like Lanfrica, Common Voice and Keyman make the process of developing a speech model or keyboard for an unsupported language accessible to small teams of linguists and computer scientists. Collaboration spaces like the Hugging Face Hub enable developers to build on one another's work, while community-centred ventures like Karya aim to generate and curate language data for marginalised languages without exploiting the communities that speak them. While they have their limitations, technologists described such innovations as "game changers", rapidly expanding the availability of language data and speeding the development of applications.[31]
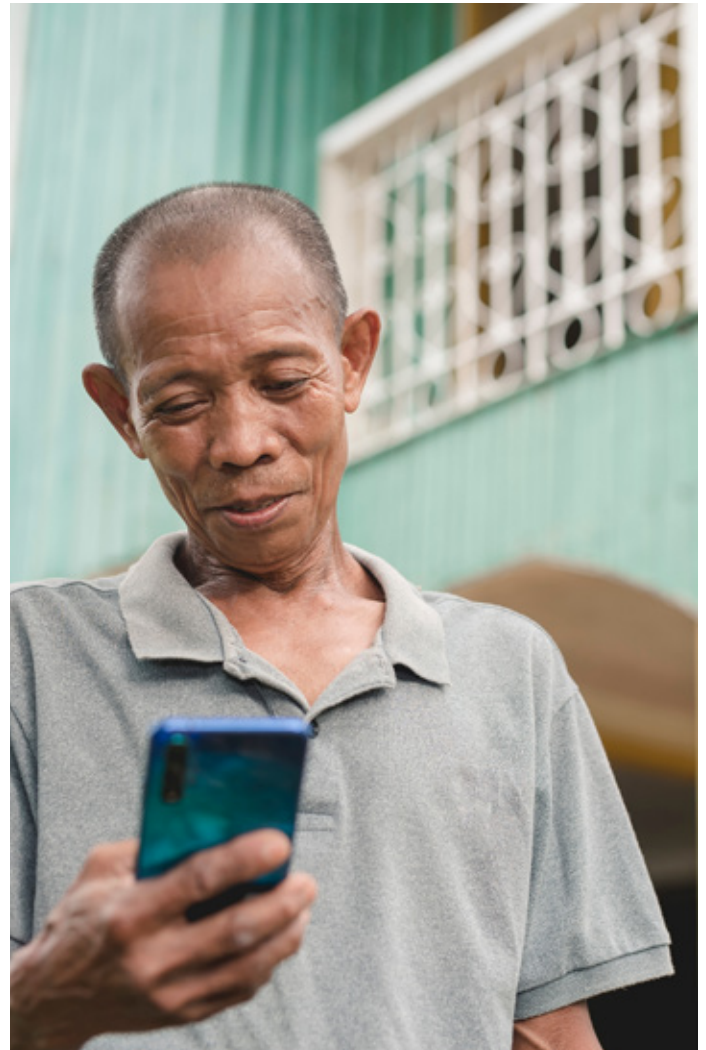
30 Bird, S. (2022). Local Languages, Third Spaces, and other High-Resource Scenarios. Association for Computational Linguistics.
31 However, remaining challenges include the reliance on volunteers for language data and on smartphones and computers, the labour-intensive set-up process and the need to first gather sufficient text data to be read out for voice data, which may exclude some of the least-supported languages.

Forging use case focussed partnerships for language technology

Marginalised communities are both the source and potential beneficiaries of language data, which is key to the development of language technology. Humanitarian organisations that use digital technology to receive and respond to information, feedback and questions from marginalised communities all generate and gather language data in the process.

A humanitarian communication ecosystem in which every non-sensitive voice message, transcription or translation was safely used to improve the accuracy and speed of multilingual communication for all users would have far-reaching benefits for marginalised language speakers.

Academic language data collection efforts typically lack the focus of an actual use case. Where only very little data is available, experience shows that it is most efficient to collect only the data needed for a specific application or service. To build the language and speech models for genuinely useful applications, use case-oriented data collection should focus on the domain (public health, GBV, cash assistance, etc.) and its intended users (age, gender, dialect, literacy, etc.). Language and speech models should also be evaluated against actual use cases, as academic benchmarks and laboratory-generated language data may not reflect actual language use and context.

This approach would not be without challenges. There would be an inherent need to prioritise the consent, data protection and ownership of marginalised language-speaking communities for whom such services are designed. Similarly, frameworks related to the ethics and sustainability of such collaborations would need to be developed across a range of actors (communities, humanitarian organisations, donors and commercial and academic language technology providers).

For these models to be successful, civil society and marginalised communities would need a much better understanding of how language technology is built and what it can do.

Finally, humanitarian organisations would need to ensure that the offline and in-person services that many marginalised groups require are provided in parallel, and not deprioritised.

Pragmatic use of contact languages

Many marginalised language speakers also speak a contact language or lingua franca to some degree. Which contact language they speak may depend on geography and ethnicity. Those with little or no formal education in that language will speak and understand it with greater confidence and accuracy in the areas they routinely use it for, such as market transactions. Health information and government bureaucracy remain a challenge for most.

Some will not have functional comprehension or speaking ability in the contact language. Comprehension testing in Nigeria and the DRC found that this was disproportionately the case for older adults and women of all ages.[32] These groups may use calls and voice notes or text in their own language when they have access to mobile phones.

Non-native speakers of the contact language might use simple and sometimes ungrammatical sentence structure, with a limited vocabulary that mixes in words from other languages. Their spelling might be unorthodox and inconsistent, and their accent when speaking may not be that of a first-language speaker. This means that existing language technology in a contact language is unlikely to work for them – speech recognition will not understand their voice commands, search engines will not understand their written queries and messaging apps will mangle their communication with autocorrection. Functionality also tends to be stronger only for more economically powerful variants of contact languages.

> "You have Brazilian and European Portuguese covered well, but people from Angola and Cape Verde will not get the same benefit out of the tools. Saudis' and Egyptians' Arabic is modelled better than Moroccans' and Iraqis' because most of the data comes from Saudi Arabia and Egypt."
>
> **– Language technologist, North America**

Systematic efforts to collect and share data on language use for specific marginalised language communities could maximise the reach of digital services for users with some second-language capabilities, and identify the most relevant contact languages to invest in. Humanitarian organisations and language technology developers would both have an interest in working with civil society to gather that information.

---

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

Digital services are not the answer for every need or individual. In-person support will always be essential when responding to humanitarian crises. However, when done well, digital services have the potential to expand participation, accountability and service access for marginalised communities around the world.

Most humanitarian digital services are only available in a narrow range of languages. Many millions of crisis-affected people are excluded as a result. Service providers often have little insight into who those people are and the scale of the problem. Because language barriers compound other disadvantages, potentially vulnerable groups are disproportionately affected by language-based exclusion.

The success, value for money or return on investment of digital services is primarily measured by the number of users reached rather than an understanding of who has the greatest needs. This can diminish their potential impact and, for marginalised language speakers, represents a far greater loss.

Insights and efficiency can be gained from partnering with, and supporting digital service development by, organisations with connections to marginalised communities. Existing good practices can also maximise accessibility, such as the use of plain language, user-tested graphics and a mix of formats and channels to the extent possible, including parallel offline and in-person services.

Developing more accessible and impactful digital humanitarian services for marginalised language speakers will require a multifaceted, cross-sectoral approach that puts the needs and experiences of intended users at the centre. This includes using better data as a basis for more informed and transparent targeting, and looking for efficiencies through more collective, use case-focused language technology development.

Recommendations

Needs assessments for digital services (and other humanitarian services) should be conducted in marginalised languages.

Tools like Kobo ToolBox, which supports graphic and pre-recorded audio questions and automated transcription and translation, can support more inclusive needs assessments. Service providers should base their language selection on language use data for the target population and provide data collectors with adequate training and language support. Donors, meanwhile, can recognise the importance of quality needs assessment data by defining their quality requirements and providing the funding to meet them.

Communication and service design should be based on research and monitoring beyond existing users.

Insights from those not using the service are key to identifying and addressing gaps and should be routinely gathered in languages other than those the service currently caters for. Language use research can lead to greater efficiency and transparency by clarifying the communication needs of potential users. Humanitarian providers and their donors should consider how best to incorporate both components in service design and delivery.

Donors could support greater efficiency, impact and sustainability by encouraging and supporting data sharing and collaboration between their partners.

Organisations should share their data on community language use and communication preferences, and on who cannot access services, to inform response-wide efforts and raise awareness. Funding could be made available for collaborations that build on and help sustain platforms and services operating in marginalised languages, in line with the Principles for Digital Development.[33]

Humanitarian organisations should apply existing good practices on inclusive communication to the greatest extent possible.

It needs to be standard practice for communications – digital or otherwise – to use languages, formats and channels that will reach marginalised groups. This begins with data on communication needs. Such practices are not always expensive: using plain language by default, hiring staff who speak local languages and developing glossaries of key concepts in local languages is high value but not high cost.

A more consistent, user-centred approach is needed for consent and data protection.

These are particular concerns for marginalised language speakers communicating in a second language. Some research participants are investing in technological safeguards and plain language consent processes, as comprehension is vital to meaningful informed consent. Good practices need dissemination and support to become standard across the sector.

Direct resources to develop language technology that meets the needs of marginalised language communities in crisis settings.

Donors and technology companies should consider funding research for real-life use cases in the humanitarian field. This should be informed by data on the use of first and contact languages for specific marginalised language communities. The resulting language technology should be widely available to humanitarian responders.

---

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

Identify bottlenecks preventing humanitarian organisations from participating in collective use and improvement of language technology for marginalised languages.	The development of common formats and workflows for humanitarians to share voice and text data, and coordination on updating and communicating informed consent processes, could improve uptake, efficiency and transparency.[34] The expertise required to manage these processes should become a standard component of project budgets. Collective platforms like Common Voice and Hugging Face could begin to be used in collaboration with language technology researchers without waiting for sector-wide coordination.
Humanitarian donors, governments and the major language technology developers should consider establishing a sector-wide, coordinated framework for language technology development.	The benefits could be profound, both within and beyond humanitarian action. By placing the service user at the centre of a technology-enabled ecosystem, providers of all kinds could progressively tailor their services to the preferences of a more diverse range of users. Citizens could use accessible language technology to participate in national and international debates that are currently closed to them, to organise across communities and to understand and claim their rights. Governments could use it to consult and communicate with citizens more effectively about policies and public health risks. Language technology could complement, not replace, in-person services to expand their reach and effectiveness.
Involve communities and civil society in decisions on language technology development for their languages.	Free and informed consent of communities will be needed for data collection and technology development.[35] To avoid colluding in the exploitation of marginalised communities, humanitarians must ensure any language technology developed for their languages is geared to their needs and will benefit them. This will require transparent and realistic conversations about the potential risks and benefits of developing language technology, and if and how communities want that to happen.
Consider targeted support for the most marginalised language communities to develop language data as the starting point for language technology.	Even with better cooperation, there will still be languages that will not be covered in a reasonable time frame. Some may simply be prohibitively expensive for the humanitarian sector and unprofitable for the private sector. However, there may be scope for donors with a broader remit than humanitarian action to invest in targeted interventions to support those language communities in creating language data, potentially in collaboration with academic researchers.
Invest in local digital services.	Many of the issues outlined in this report could be better addressed if the organisations close to marginalised communities played a greater role in designing and delivering digital services for them.

---

34 Acknowledging sensitivities about the potential for exploitation, which is grave and must always be addressed before using anyone's data.
35 OHCHR. (2013). Free, Prior and Informed Consent of Indigenous Peoples.

Annex: Methodology

This report is based on a detailed desk review, as well as online surveys and key informant interviews (KIIs) with three groups of participants. These groups were not mutually exclusive, for example, several organisations were both civil society and digital service providers.

— **Civil society organisation (CSO):** CSOs from low- and middle-income countries with direct insights into marginalised language speakers' access to and experience of digital technology

— **Technology commissioner/digital service provider:** international organisations commissioning and providing digital services, with global insights

— **Technologist:** language technology experts with insights into the feasibility and practicalities of specific applications

Surveys

Surveys were developed to gather quantitative and qualitative data from CSOs and (humanitarian) technology commissioners/digital service providers. The CSO survey was available in English, French, Spanish and Arabic. The technology commissioners survey was available in English. The surveys were disseminated online, via partners, newsletters and direct email invitations from CLEAR Global and GSMA. Fieldwork lasted from 12 September until 15 November 2023.

Key informant interviews

KIIs were conducted on Zoom between 2 October and 10 November 2023. Semi-structured interview guides were used for consistency while allowing for exploration. Participants were identified through known contacts, survey respondents and snowball sampling, and selected for diversity of geography and expertise.

Figure 2: **Research participants by method and category**

	Survey	KII	Case study KII	Total
Civil society organisation	75	8	4	**80**
Digital service provider	24	9	4	**30**
Language technologist	0	5	0	**5**
Total	99	22	8	**115**

Limitations

Survey respondents and key informants identified through the survey were self-selecting, and as such, findings have indicative value. Key informants were selected because they were known by CLEAR Global or the GSMA to have relevant expertise and experience. Although every effort was made to reach CSOs worldwide, the results are regionally imbalanced, with no representation of East Asia and an overrepresentation of Uganda and Guatemala.

Although the intention of the study was to focus on language inclusion in humanitarian digital services, it was also a preliminary scoping study. We were therefore open to hearing from a wider range of experience, and some participants provided insights from outside the humanitarian sphere.

GSMA Head Office
1 Angel Lane
London
EC4R 3AB
United Kingdom
Tel: +44 (0)20 7356 0600