



**[Masterclass]  
[Comment inclure les  
citoyens allophones ?]**

— *[Abèdjè Sinatou Saka]*

**NUMÉRIQUE EN COMMUN[S]**

# Qui sommes-nous ?

Idemi Africa est un collectif né de l'envie de rendre les langues africaines plus visibles sur Internet.

## **Nos missions sont les suivantes:**

- Veiller à l'intégration des langues africaines par les politiques africains (Etude sur l'usage des langues africaines par les 54 Etats Africains)
- Produire des gros volumes de données open source pour faciliter la création d'outils numériques pour les populations locutrices des langues africaines. (Atelier de traduction)
- Sensibiliser l'opinion publique à travers des articles d'experts ou d'acteurs de la société civile.
- Visibiliser les initiatives qui lient le numérique et les langues africaines.

## La fracture numérique linguistique en France:

→ 6,8 millions d'**immigrés** vivent en **France**

→ **44,6% des immigrés vivant en France sont nés en Afrique et 35,4% sont nés en Europe.** Les pays de naissance les plus fréquents sont l'Algérie (12,8%), le Maroc (12%), le Portugal (10,1%) et l'Italie (4,6%)

→ **Comment intégrer les langues maternelles de ces différentes populations ?**

# Le français en France

- Le plurilinguisme est considéré d'une manière particulière et différente d'un pays à un autre.
- En France, le français, depuis ses origines a toujours été considéré comme une « belle langue », qu'il convient de préserver.
- L'un des principaux facteurs d'unité nationale est bien la langue française
- L'écrit est premier par rapport à l'oral, qui est considéré comme second

# Le français en France

- Le français vécu comme une « belle langue », participe ainsi à l'exclusion des personnes qui ne possèdent pas les compétences nécessaires pour utiliser la langue française de manière parfaite, ou du moins suffisamment maîtrisée du point de vue des exigences qu'elle comporte.
- Ces différents facteurs contribuent à rendre le territoire de la France monolingue, presque hermétique aux apports des autres langues. C'est ce monolinguisme dominant qui est à l'origine de la reconnaissance problématique des plurilinguismes des individus.

# Le français en France

- Il existe tout d'abord des langues socialement ou économiquement plus prestigieuses, comme l'explique Christine Hélot.
- C'est cette affirmation qui va créer ce que nous appelons aujourd'hui le bilinguisme de masse, en opposition au bilinguisme d'élite, mieux reconnu en France. Les langues souffrant de dévalorisation sont ainsi les langues issues de l'immigration. Il existe également une certaine « peur » du bilinguisme, en raison de la culture qu'elle véhicule, ou bien du système politique différent dont sont issus les immigrés.

## **Notre politique de plurilinguisme numérique**

→ **La loi du 4 août 1994 relative à l'emploi de la langue française et de ses textes d'application prévoient que les services de l'Etat ou les établissements publics placés notamment sous la tutelle du ministère de la communication doivent traduire leur site Internet, en au moins deux langues étrangères.**

→ **L'Internet constitue un outil privilégié d'accès des non-francophones à l'information administrative et pour faire connaître les politiques publiques.**

→ **le plurilinguisme sur les sites publics français demeure très inégal, en raison en premier lieu de la rareté des structures de traduction présentes dans les ministères.**

## **Notre politique de plurilinguisme numérique**

→ **Le site du Premier ministre, portail du gouvernement français, est disponible en cinq langues : français, anglais, allemand, espagnol et portugais. Les ministères des affaires étrangères et de l'économie, des finances et de l'industrie offrent sur leur site des informations en quatre langues : français, anglais, allemand et espagnol..**

→ **Aucun site public en arabe.**

→ **Dans le domaine de la recherche publique, les sites ne proposent, en règle générale, qu'une seule traduction en langue anglaise.**



**La fracture numérique linguistique et ses enjeux:**

→ **7000 langues reconnues dans le monde**

→ **La première langue utilisée sur internet est l'anglais. Au milieu des années 90, l'anglais représentait 80 % du contenu web.**

Aujourd'hui, seulement 10 langues dominent internet.

## Les langues les plus utilisées sur internet:

### Top Ten Languages Used in the Web - March 31, 2020

( Number of Internet Users by Language )

TOP TEN LANGUAGES IN THE INTERNET	World Population for this Language (2021 Estimate)	Internet Users by Language	Internet Penetration (% Population)	Internet Users Growth (2000 - 2021)	Internet Users % of World (Participation)
<a href="#">English</a>	1,531,179,460	1,186,451,052	77.5 %	742.9 %	25.9 %
<a href="#">Chinese</a>	1,477,137,209	888,453,068	60.1 %	2,650.4 %	19.4 %
<a href="#">Spanish</a>	516,655,099	363,684,593	70.4 %	1,511.0 %	7.9 %
<a href="#">Arabic</a>	447,572,891	237,418,349	53.0 %	9,348.0 %	5.2 %
<a href="#">Portuguese</a>	290,939,425	171,750,818	59.0 %	2,167.0 %	3.7 %
<a href="#">Indonesian / Malaysian</a>	306,327,093	198,029,815	64.6 %	3,356.0 %	4.3 %
<a href="#">French</a>	431,503,032	151,733,611	35.2 %	1,164.6 %	3.3 %
<a href="#">Japanese</a>	126,476,461	118,626,672	93.8 %	152.0 %	2.6 %
<a href="#">Russian</a>	145,934,462	116,353,942	79.7 %	3,653.4 %	2.5 %
<a href="#">German</a>	98,654,451	92,525,427	93.8 %	236.2 %	2.0 %
<b>TOP 10 LANGUAGES</b>	5,273,725,132	3,525,027,347	66.8 %	1,188.2 %	76.9 %
Rest of the Languages	2,522,890,578	1,060,551,371	42.0 %	1,114.1 %	23.1 %
<b>WORLD TOTAL</b>	7,796,615,710	4,585,578,718	58.8 %	1,170.3 %	100.0 %

## La fracture numérique linguistique et ses enjeux:

- **La recherche de « emploi » sur Google dans une certaine langue peut vous rapporter 10 fois plus de résultats que dans une autre.**
- **Sur les 7 000 langues utilisés, vous ne pouvez rechercher sur Google que dans un peu plus de 130 langues différentes.**

## La fracture numérique linguistique et ses enjeux:

- Les recherches en anglais rapportent également entre quatre et cinq fois plus de résultats qu'en arabe par exemple.
- Sur les 288 éditions de Wikipedia en langue officielle, l'anglais est de loin la plus grande édition en termes d'utilisateurs.
- De l'autre côté du spectre, il y a une quasi-absence de tout contenu dans de nombreuses langues africaines et asiatiques.

# La fracture numérique linguistique et ses enjeux:

- Il existe d'énormes vides d'informations dans les langues non dominantes, où les gens, les lieux et les cultures sont engloutis dans l'obscurité. « **Les pays riches se définissent en grande partie eux-mêmes et les pays pauvres se définissent en grande partie par les autres.** »
- L'inégalité en matière d'information et de représentation dans différentes langues en ligne peut également affecter notre compréhension des lieux et même la manière dont nous y agissons

## La fracture numérique linguistique et ses enjeux:

→ Plus de **10 pour cent** de Wikipédia est écrit en anglais, par exemple, et près de la moitié des articles du site sont en langues européennes.

→ Avec la domination de l'anglais, Freuler a fait valoir qu'Internet a facilité «l'homogénéisation culturelle»

→ Facebook a été critiqué pour ne pas employer suffisamment de locuteurs natifs pour surveiller le contenu dans les pays où il compte des millions d'utilisateurs. Au Myanmar, par exemple, l'entreprise n'a eu pendant des années qu'une **poignée de locuteurs birman**s alors que les discours de haine proliféraient.

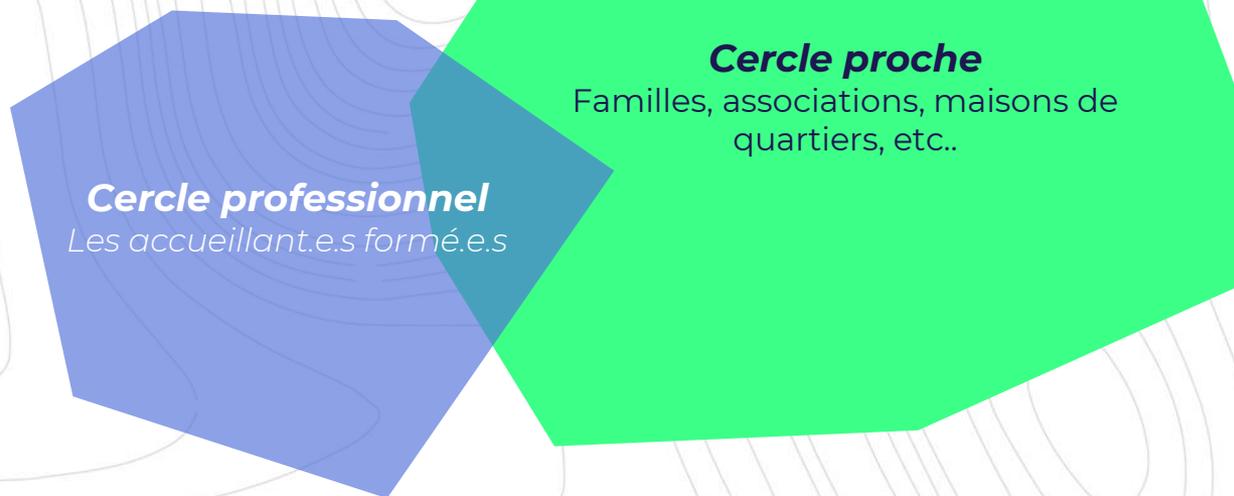
# La fracture numérique linguistique et ses enjeux:

- Une autre partie du problème vient du fait que relativement peu de jeux de données ont été créés dans ces langues adaptés à la formation d'outils d'intelligence artificielle.
- Par exemple, les algorithmes de Facebook, formés principalement en anglais ne s'adaptent pas très bien au cinghalais, qui est parlé par environ 17 millions de personnes au Sri Lanka et peut être écrit de quatre manières différentes.

Cela rend difficile pour le réseau social d'identifier automatiquement des choses comme le discours de haine dans le pays.

András Kornai, chercheur sur la mort numérique des langues, prédit que 95% de toutes les langues utilisées aujourd'hui ne gagneront jamais en popularité en ligne.

# Les usages numériques des populations allophones en France (échantillon non représentatif) et en situation d'illectronisme



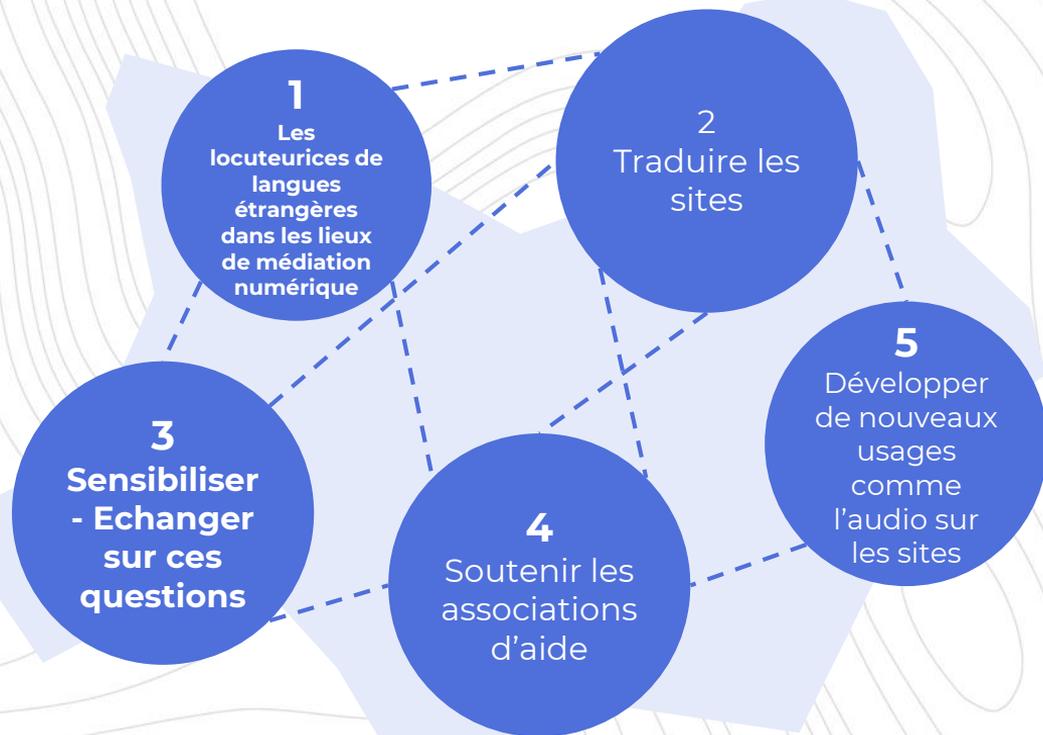
# Cercle proche

- Les enfants, les amis, la famille représentent les premières personnes aidantes pour celles et ceux qui ne parle pas la langue officielle de leurs différentes démarches.
- Ensuite, ils et elles vont généralement dans les bibliothèques ou maisons de quartiers, des lieux sociaux également car ils permettent de tisser des liens , de se familiariser à l'outil numérique mais surtout de sortir de l'isolement.

# Cercle professionnel

- Lorsqu'il s'agit d'une personne isolée, elle se rend directement à l'administration concernée pour poser ses questions et se renseigner sur ses droits.
- Lorsque l'administration est sensibilisée et dispose d'un personnel multilingue, l'utilisateur est pris en charge.
- Des fois, elle est souvent accompagnée d'une autre personne bilingue ou peut s'exprimer de manière sommaire dans la langue officielle.

## 5 POINTS CLÉS



## La traduction pour l'inclusion et la représentation

- La traduction permet de prendre en compte les normes linguistiques et culturelles des populations. **Disponibilité des langues étrangères sur le serveur vocal interactif (3400). Elle permet aussi de produire du savoir pour des populations dont les langues sont digitalement mortes.**
- **Ne pas se limiter aux traductions automatiques qui favorisent des contre-sens ou des biais de genre**
- Ces traductions devront être mises en valeur afin d'éviter une balkanisation de l'information en ligne. Il s'agit aussi de traduire tous les messages d'utilité publique (vidéos par exemple) sur internet.

# Echanger avec les personnes concernées

- Il s'agit ici de multiplier les espaces de discussions et de créer des temps de rencontres, **ateliers linguistiques** entre les personnes qui oeuvrent contre la fracture numérique linguistique, les personnes concernées et d'autres acteurs. Le but est de réaliser des échanges riches pour comprendre les problématiques, de questionner les représentations des uns et des autres, **créer du contenu** et tenter de trouver des solutions.
- **Ces différentes rencontres pourraient également stimuler la recherche universitaire en France sur la question de la fracture linguistique numérique.**

# Développer de nouveaux usages sur les sites internet

- Valoriser de nouvelles fonctionnalités comme des commandes vocales dans différentes langues étrangères permettrait aux populations ne sachant pas lire d'accéder à l'information de manière simple et efficace.
- **Il s'agit donc d'intégrer une politique linguistique dans les pratiques d'accessibilité numérique.**

# Soutenir les associations d'aide

- De nombreuses associations ou maisons de quartiers soutiennent des centaines de personnes allophones dans leurs démarches administratives.
- Inéluctablement, il s'agit de les identifier, les accompagner et les soutenir afin qu'elles puissent accomplir leurs missions **et les aider à mettre en ligne le savoir linguistique.**

# Former le personnel

- La solution la plus évidente serait d'avoir des interprètes dans les administrations afin d'accueillir les populations locutrices de langues étrangères.
- Auquel cas, il faudrait permettre au personnel d'identifier par le biais de carte sonore comme c'est le cas dans la santé, la langue maternelle de l'utilisateur.
- Il est également possible de comparer les pratiques administratives entre les pays d'origine et la France.
- **Intégrer une dimension linguistique dans l'outil Aidant Connect**

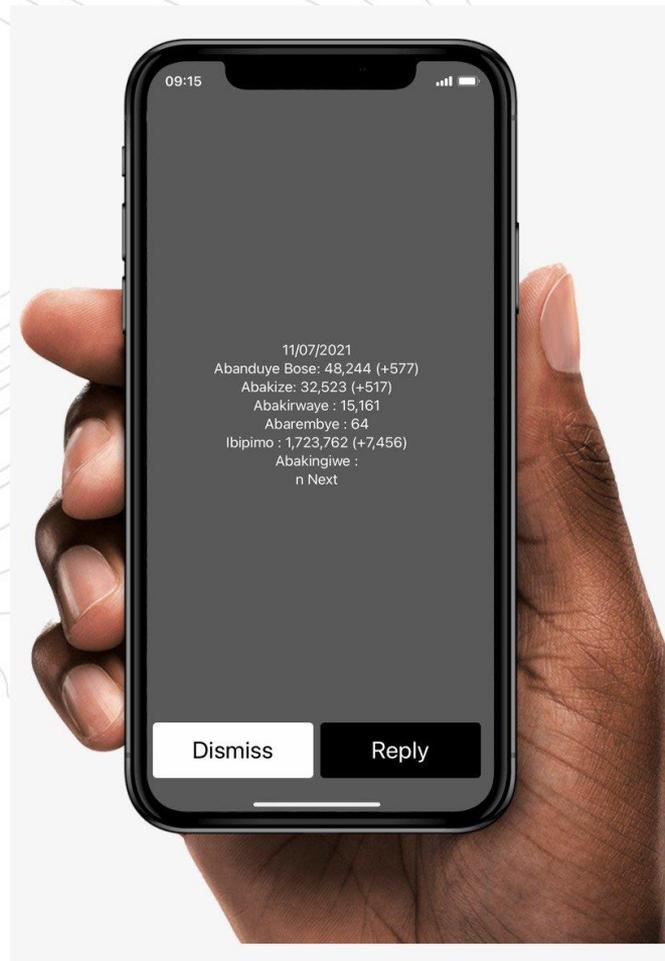
# Exemple d'inclusion linguistique numérique

- Les Rwandais disposent d'un nouvel outil dans la lutte contre le COVID : Mbaza, **un chatbot** qui fournit des informations opportunes et précises sur la pandémie en kinyarwanda, français et anglais.
- Mbaza est basé au Rwanda Biomedical Center et est une collaboration entre [Mozilla Common Voice](#), la start-up rwandaise [Digital Umuganda](#), la Rwanda Information Society (RISA) et la Coopération allemande au développement, ou « Deutsche Gesellschaft für Internationale Zusammenarbeit » ( GIZ).

## Service publique et langue

Mbaza :

- Mbaza est accessible à tous les Rwandais en utilisant le code court USSD \*114#
- **Toutes les infos sur la COVID 19**
- Mbaza est un excellent exemple de ce que peut réaliser une technologie vocale plus inclusive.
- En créant un ensemble de données vocales en kinyarwanda, des millions de Rwandais ont désormais accès à des informations essentielles sur la pandémie.



## Les inspirations

→ Wikitungues , une organisation qui promeut la diversité linguistique.

**Elle souhaite documenter toutes les langues du monde.**

## Les inspirations

**Mozilla** est une organisation qui travaille au crowdsourcing d'ensembles de données linguistiques pouvant être utilisés gratuitement par n'importe quel développeur, comme [Common Voice](#), qui est l'ensemble de données vocales le plus diversifié au monde.

Il comprend des enregistrements de plus de 42 000 personnes dans des langues dominantes comme l'anglais et l'allemand, mais aussi le kabyle.

Le projet est conçu pour donner aux ingénieurs les outils dont ils ont besoin pour créer des choses comme des programmes de synthèse vocale dans différentes langues.

## Les inspirations

- [Le projet Langues en danger](#) est une plateforme en ligne qui permet de documenter et de sauvegarder les langues en danger. L'idée est de leur donner seconde vie.
- Le projet Muysccubun, par exemple, a travaillé [pour documenter et partager la langue muisca éteinte](#), parlée historiquement dans le centre de la Colombie, en créant des dictionnaires en ligne et en créant une communauté autour de leur page Facebook.

# Conclusion

- En 2011, l'ONU a déclaré l'accès à Internet comme un droit humain fondamental.
- Il est clair cependant que l'accès à lui seul ne suffit pas à mettre tout le monde sur un pied d'égalité numérique.
- Alors qu'Internet et les médias sociaux sont de plus en plus intégrés dans la façon dont nous nous connectons et comprenons le monde qui nous entoure, il en va de même pour la langue que nous utilisons pour accéder à cette expérience.

## Conclusion

- L'accès à Internet offre également la possibilité d'autonomisation linguistique : pour documenter et préserver les langues, partager du matériel pédagogique pour encourager de nouveaux locuteurs, traduire des informations importantes pour les groupes marginalisés et même créer des communautés virtuelles de locuteurs où ils peuvent avoir du mal à exister hors ligne.

## Conclusion

Google peut mettre en œuvre des outils pour traduire instantanément des pages Web, mais ce n'est pas la même chose que d'avoir une multiplicité de perspectives humaines.

# Conclusion

Mettre plus de langues en ligne peut en fin de compte être un exercice de préservation culturelle.

Malgré tous les efforts des défenseurs, il est peu probable qu'il y ait jamais autant de sites Web en yoruba, disons, qu'il y en a en français ou en arabe.

Mais les nouveaux internautes peuvent simplement choisir de naviguer dans leur deuxième ou troisième langue au lieu de leur langue maternelle.

# Bibliographie

- <https://www.sante-solidarite.org/des-outils-pour-mieux-accueillir-personnes-allophones.html>
- <https://hal-univ-fcomte.archives-ouvertes.fr/hal-02385167/document>
- [https://foundation.mozilla.org/fr/blog/building-speech-recognition-models-for-global-languages-with-the-mozilla-common-voice-dataset-and-nvidia-nemo/?utm\\_source=twitter&utm\\_campaign=commonvoice&utm\\_medium=social&utm\\_content=1630940104](https://foundation.mozilla.org/fr/blog/building-speech-recognition-models-for-global-languages-with-the-mozilla-common-voice-dataset-and-nvidia-nemo/?utm_source=twitter&utm_campaign=commonvoice&utm_medium=social&utm_content=1630940104)

# Bibliographie

- <https://rhinocc.fr/comprendre-la-strategie-nationale-pour-linclusion-numerique-en-france/>
- <https://www.senat.fr/questions/base/2003/qSEQ031210447.html>
- <https://foundation.mozilla.org/en/blog/in-rwanda-voice-technology-innovation-helps-fight-covid/>
- <http://labs.theguardian.com/digital-language-divide/>

**MERCI**

The background of the image is a topographic map. It features contour lines representing elevation. Two large regions are highlighted with semi-transparent colors: a blue region on the left and a green region on the right. The word "MERCI" is printed in a bold, dark blue, sans-serif font in the upper left quadrant of the image.



**NEC**

**NUMÉRIQUE  
EN COMMUN[S]**

**NUMÉRIQUE EN COMMUN[S]**