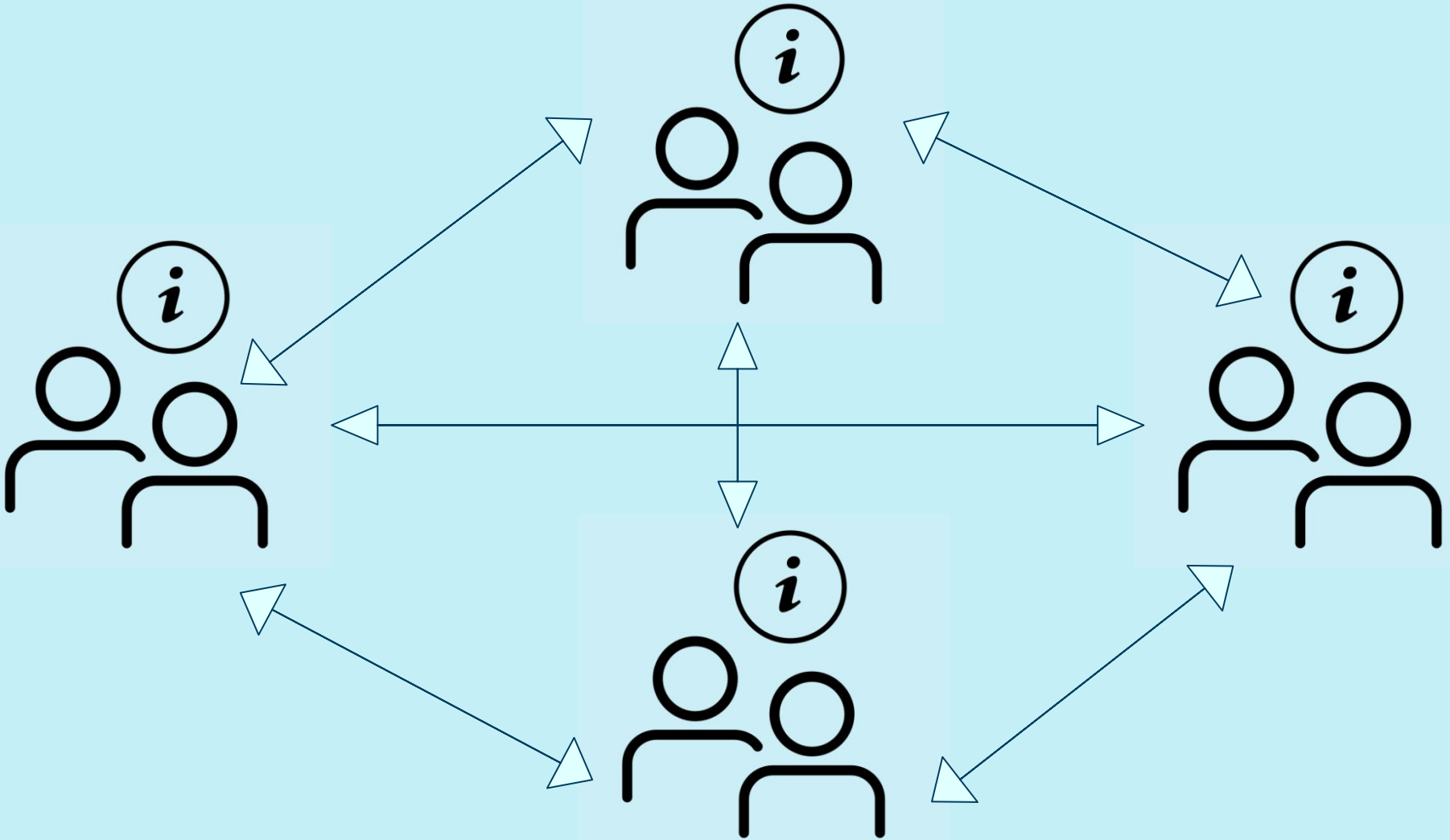
Better, Multilingual, Libraries and Information Services are Possible

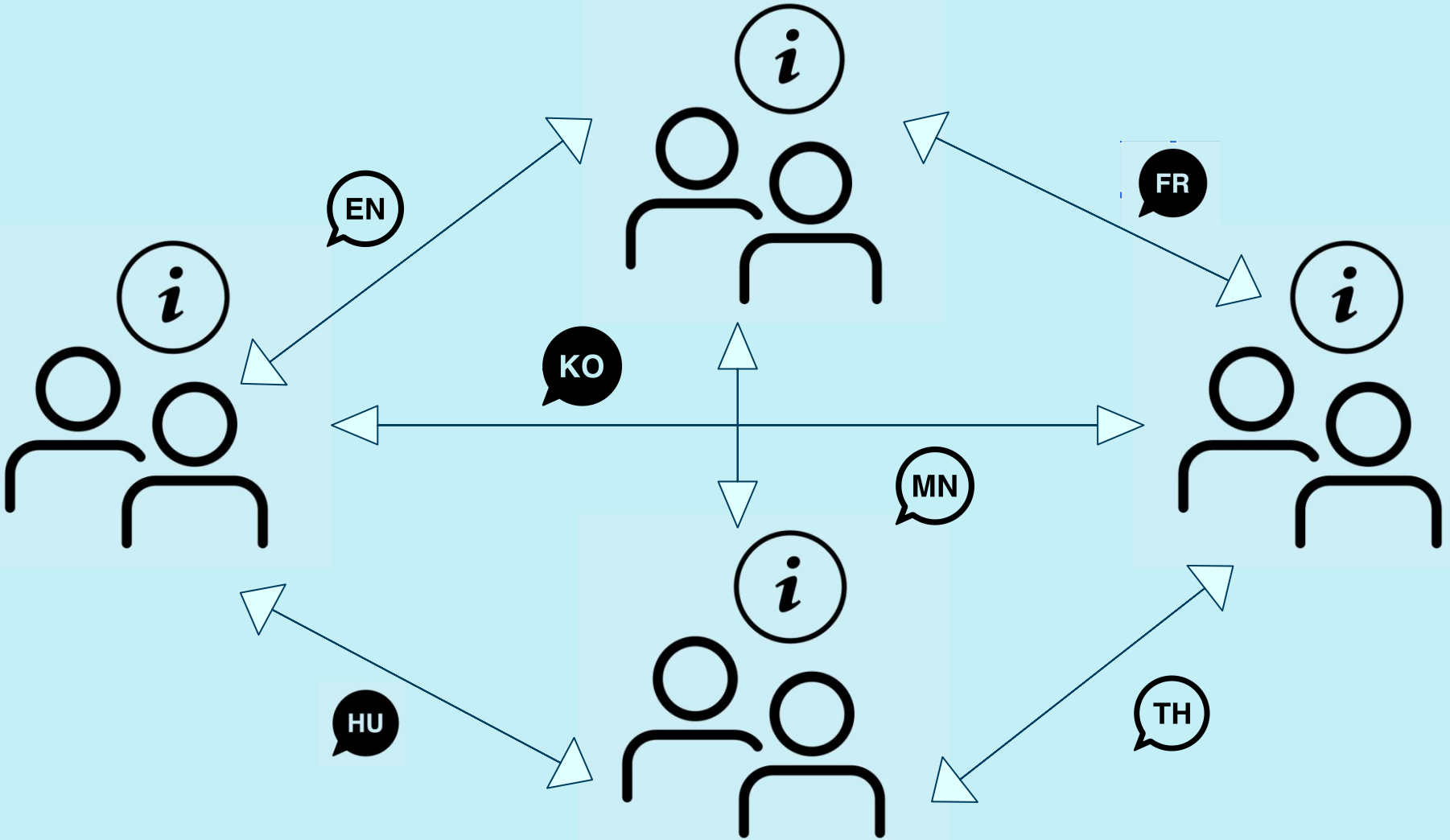
An exploration of cross language information exchange among researchers on the web

Aisha Khan
INFO 601-2 / Spring 2023

Introduction

and statistics




EN
FR
KO
MN
HU
TH

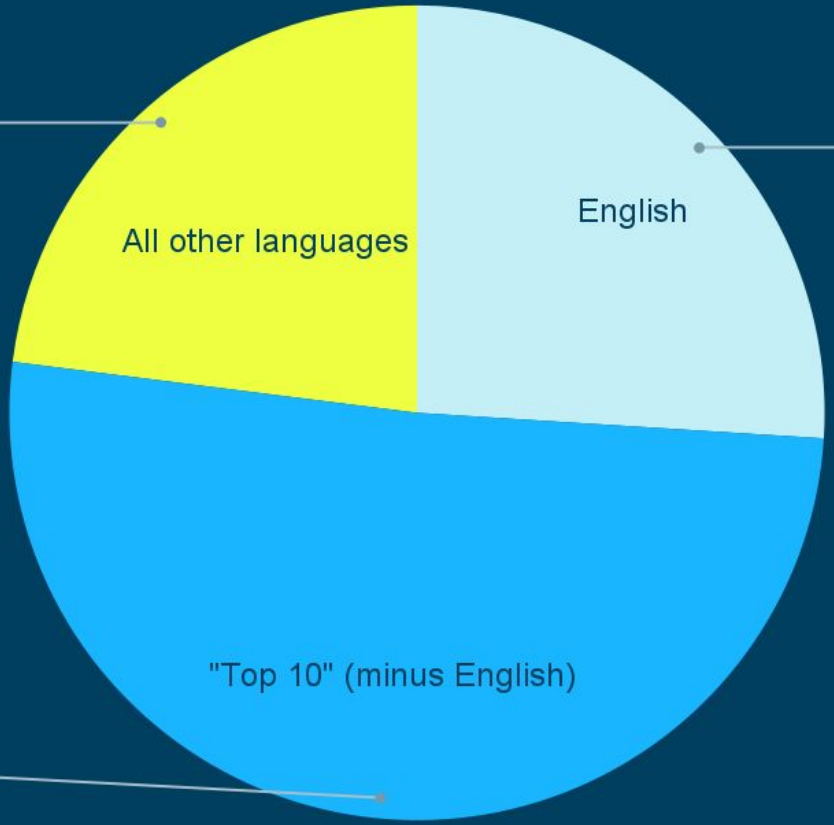
“[...] libraries help ensure that people can access the information they need – regardless of age, education, ethnicity, language, income, physical limitations or geographic barriers.”

- **The American Library Association’s (ALA) Library Bill of Rights**

~7,168

Number of languages spoken today

World Internet Participation by Language

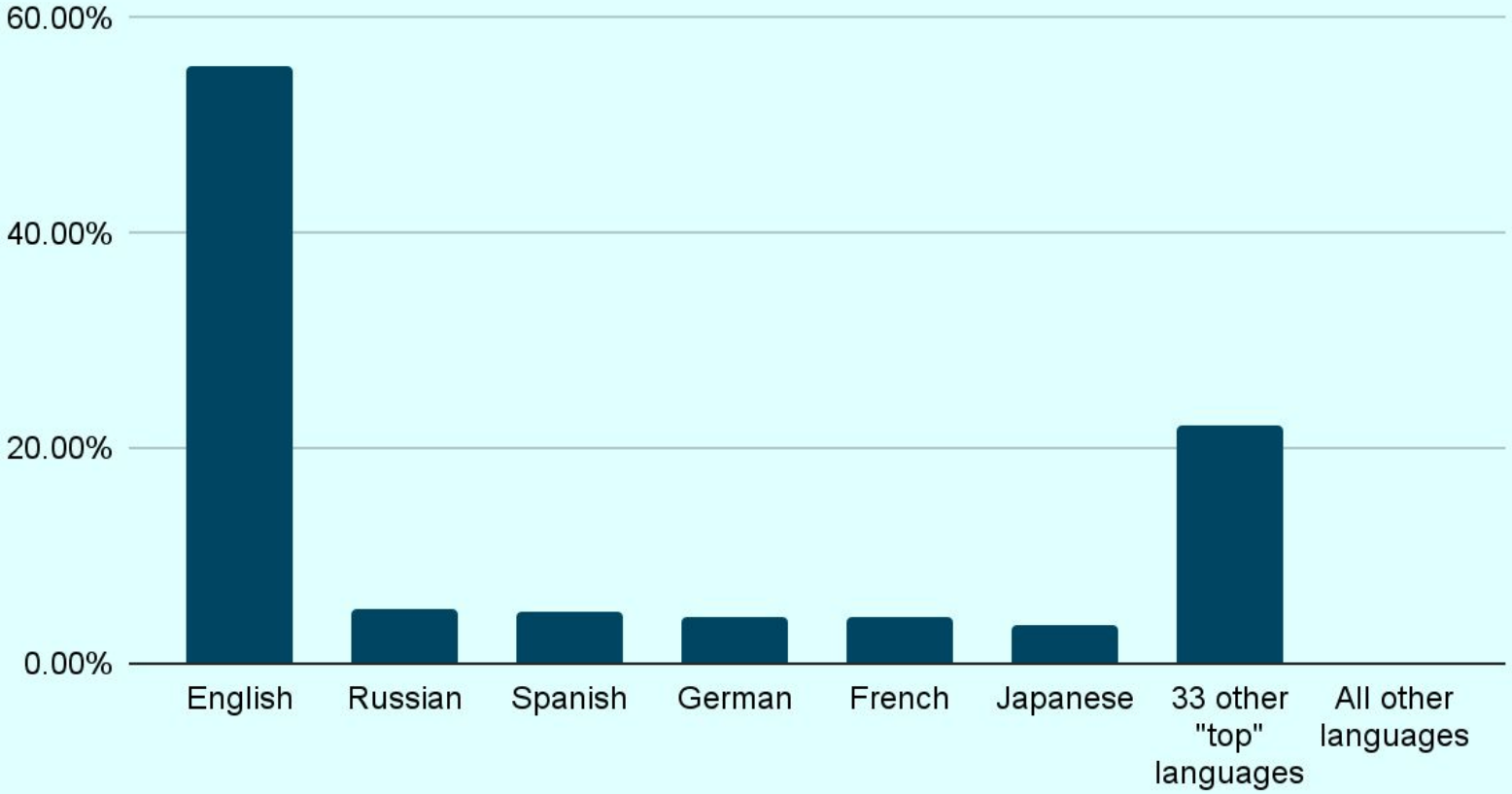


All other languages
23.0%
2.5 Billion people

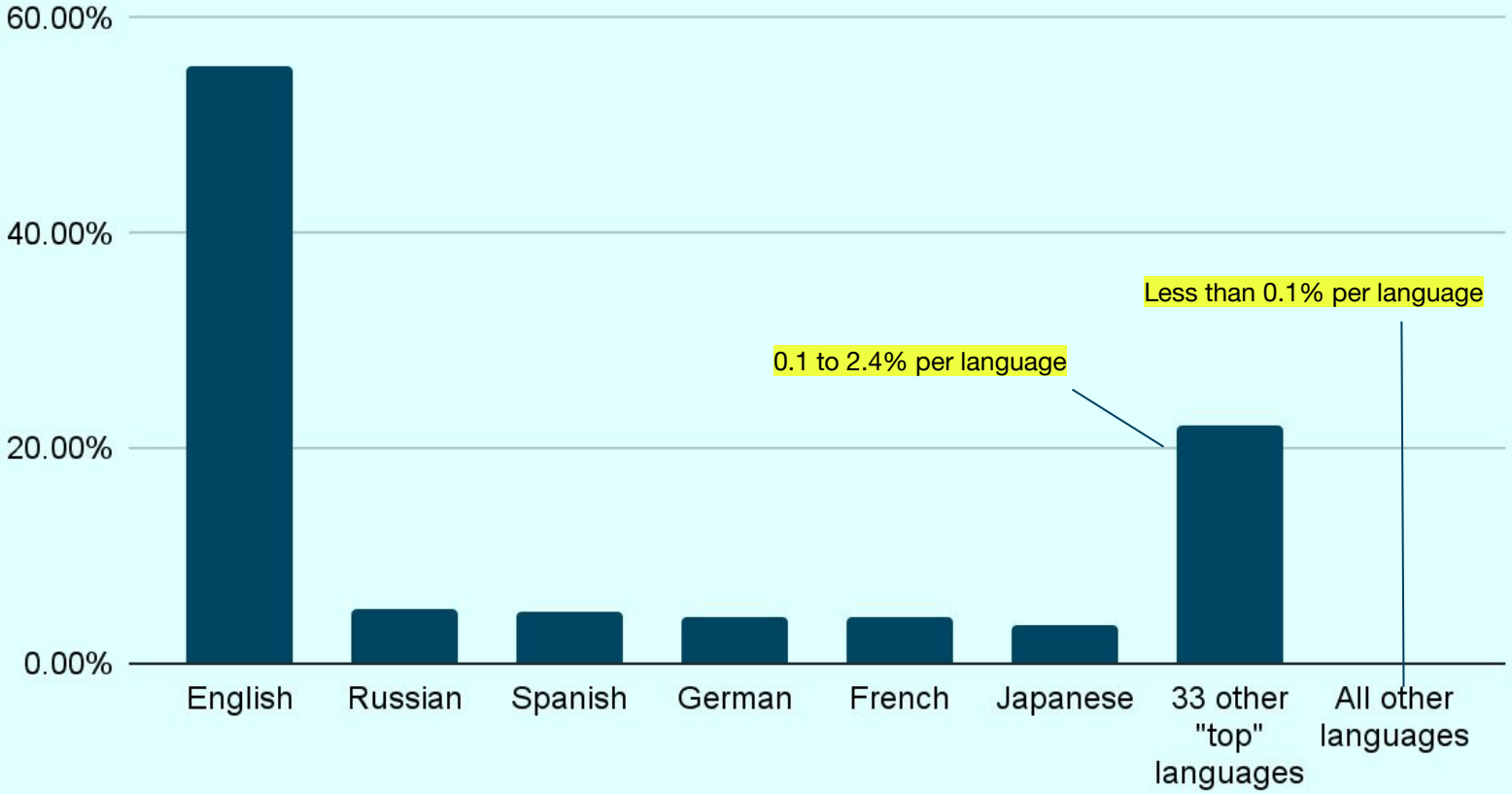
English
26.0%
1.5 Billion people

"Top 10" (minus
51.0%
3.8 Billion people

Known Internet by Language



Known Internet by Language



Only about **5%** of the world population has English as their first language. And estimates vary, but about 7 - 20% have English as their 2nd language.

Only about **5%** of the world population has English as their first language. And estimates vary, but about 7 - 20% have English as their 2nd language.

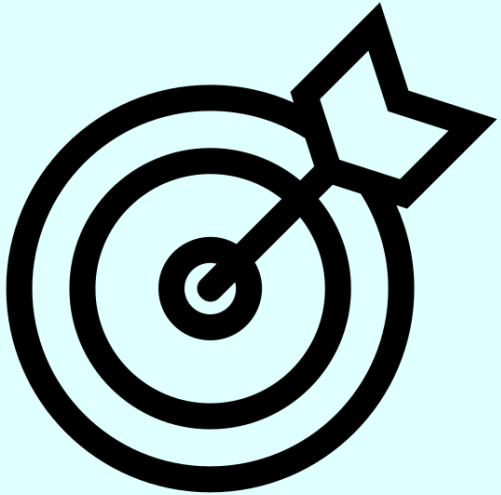
75% of information seekers on the web can’t access a majority of content in a native / preferred language.

Capitalism, colonialism, racism

- “Multilingualism, Neoliberalism, and Language Ideologies in Libraries” by Ean Henninger from In the Library with the Lead Pipe

- “On Dark Continents and Digital Divides: Information Inequality and the Reproduction of Racial Otherness in Library and Information Studies” by D.J. Hudson

Cross language information exchange among researchers in the context of an English dominant web and scientific community



Research Exchange in the Scientific Community

“Information comprehension is a prerequisite for developing sound health policies, doing research, and establishing fruitful linkages among health professionals at a global scale.”

From “Improving the availability of health research in languages other than English” by Pakenham-Walsh

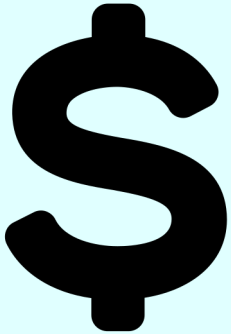
“[...] areas experiencing loss of biodiversity and thus in the greatest need of information, education, and conservation practices are often places where English is not spoken widely.”

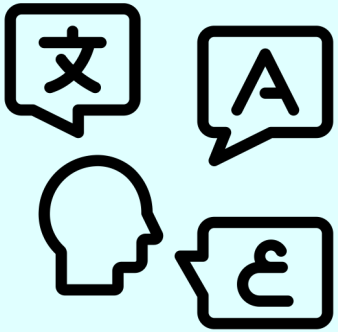
From “Multilingual access: Language hegemony and the need for discoverability in multiple languages” by McElroy and Bridges

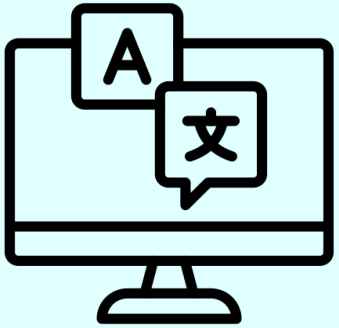
Challenges

Historically, it has been **a challenge to quantify and understand the multilingual / non-English speaking information seekers' needs** as research carried out from a monolingual / English speaking lead to incorrect conclusions.

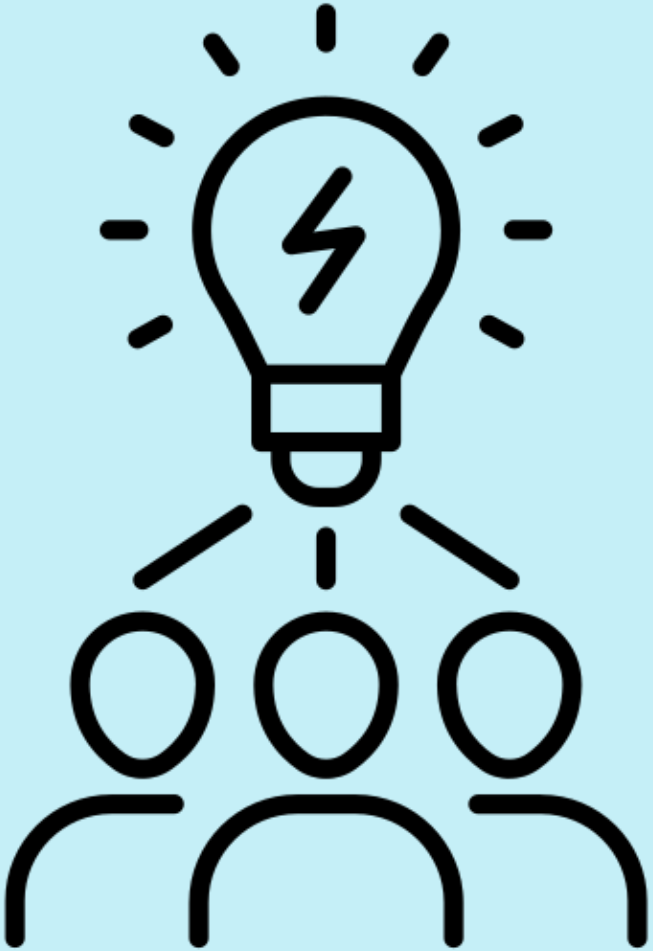
Creating & Maintaining Multilingual Repositories / Systems







Improve discoverability by translating titles & abstracts



New and rapidly improving technologies like neural machine translation can be implemented at relatively low costs to manage translation tasks like this at scale - something that would not be possible even about 10 years ago.

…but it’s not enough

Other issues include…

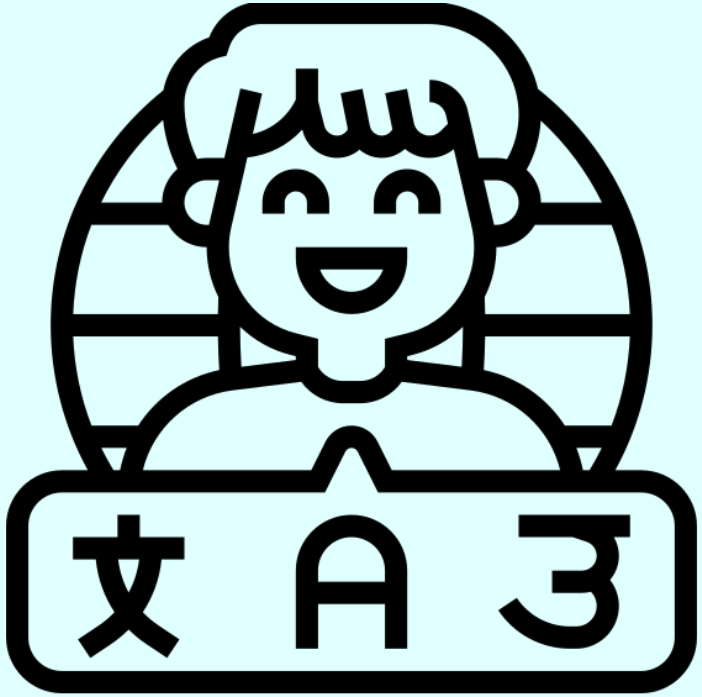
- Algorithmic bias
- Citation metrics
- Indexing
- Commodification of language
- Individual and community language ideologies

Other issues include…

- Algorithmic bias
- Citation metrics
- Indexing
- Commodification of language
- Individual and community language ideologies

Invest in

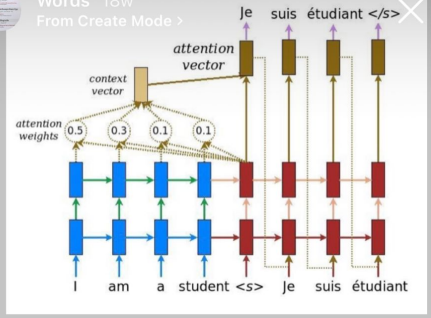
- Multilingual information infrastructure
- Multilingual metadata standards
- Encourage open access publishing
- Provide translation services
- And more…!



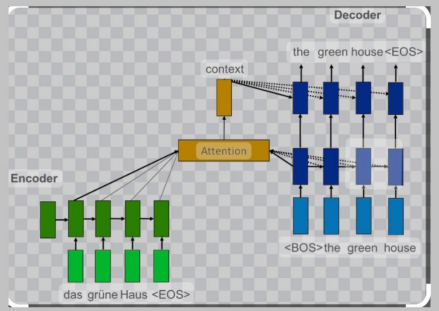
Thank you!

7:09

Words 18w
From Create Mode >



ok. so there are TWO ways to talk about neural machine translation. This is one.



Create More

7:10

Words 18w
From Create Mode >

this is two. And there is absolutely no in between (I'm being dramatic, but also not)

Activity Create More

7:10

Words 18w

when used responsibly, like many other AI systems, we should be making workers lives easier. There should be less work and that should be a good thing. At the very least, less drivel to translate making space for human translators to work on actually interesting translations - poetry t9n is one example.

Machine translation is by no means solved. GNMT can still make significant errors that a human translator would never make, like dropping words and mistranslating proper names or rare terms, and translating sentences in isolation rather than considering the context of the paragraph or page. There is still a lot of work we can do to serve our users better. However, GNMT represents a significant milestone. We would like to

there's a lot of reasons to genuinely be concerned about AI, sure, but one of the most touted ones is that human jobs and livelihoods are being taken away. We are made to fear machines and automation in the very specific context of capitalism because in this system you have to be "working" - even if it's absolute nonsense work - in order to "earn" a paycheck that may or may not even provide folks with enough to have basic needs met.

Activity Create More

7:10

's my technical PSA for the day. also if you have NMT experience (generally / specifically, engine training) or just any interest in it, please DM me

this is an old article, but it's important to point out that problems like this are still handled on an ad-hoc basis in practical applications of NMT

we will also continue to rely on the expertise of human translators for a very very long time. If anyone tells you otherwise, they are wrong. If you genuinely think I'm wrong, please DM me.

Neural Machine Translation (NMT) is a new approach to machine translation that has shown promising results that are comparable to traditional approaches. A significant weakness in conventional NMT systems is their inability to correctly translate very rare words: end-to-end NMTs tend to have relatively small vocabularies with a single *unk* symbol that represents every possible out-of-vocabulary (OOV) word. In this paper, we propose and implement an effective technique to address this problem. We train an NMT system on data that is augmented by the output of a word alignment algorithm, allowing the NMT system to emit, for each OOV word in the target sentence, the position of its corresponding word in the source sentence. This information is later utilized in a post-processing step that translates every OOV word using a dictionary. Our experiments on the WMT'14 English to French

Activity Create More