Mother Language Day 2021: Improving data on mother-tongue languages for better learning outcomes

https://edtechhub.org/2021/02/20/mother-language-day-2021-improving-data-on-mother-tongue-languages-for-better-learning-outcomes/

To celebrate International Mother Language Day 2021, EdTech Hub's Björn Haßler spoke with Alice Castillejo and Mia Marzotto of Translators without Borders. They talked about the importance of mother tongue-based education and the need to support speakers of marginalised languages. Do children who speak certain languages drop out early? What languages need to be supported in teaching and learning materials, as well as digital learning platforms? This blog post captures their conversation.

"How do we know that speakers of marginalised languages are left behind?"

Alice: Research around the world shows that educational outcomes are worse for students studying in a second language. The use of an unfamiliar language is linked to high dropout rates and low academic achievement.
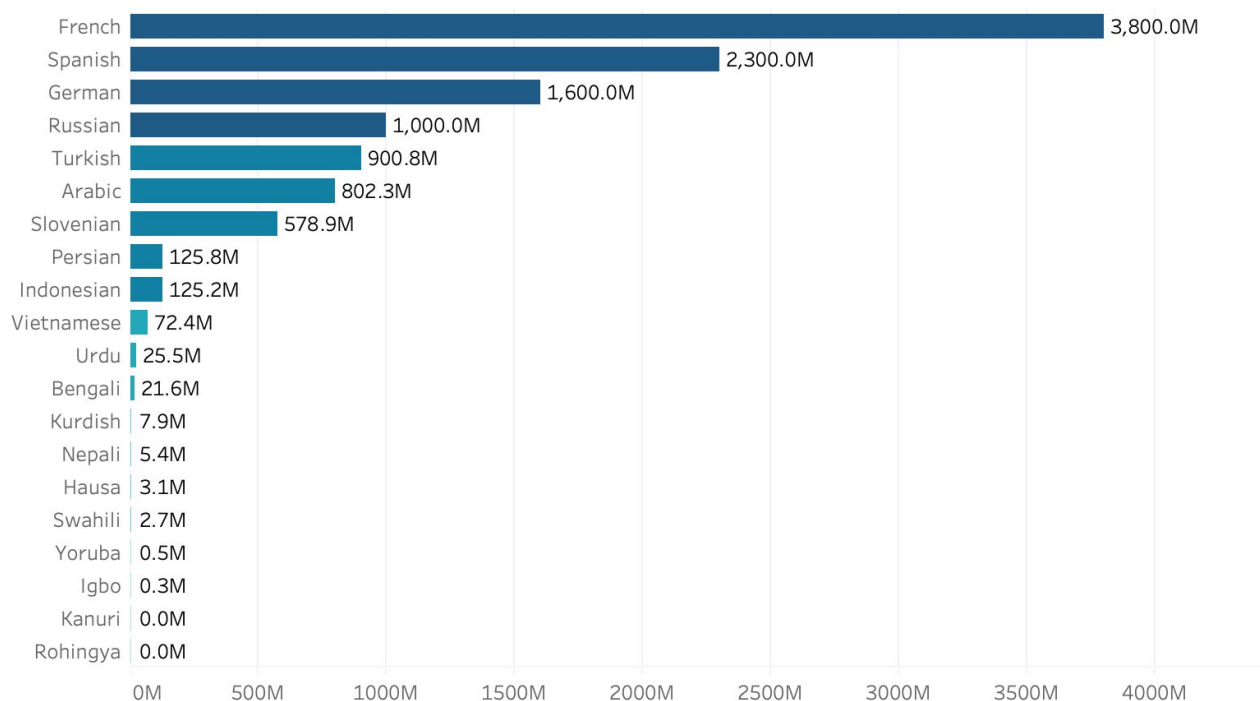
We also know that 40% of children worldwide are not educated in a language they speak at home. Crucially — if you want to know which learners are left behind, you need to measure it. If you collect language data, you can see which language groups are left behind and adapt your education programming accordingly.

Mia: It's not only the language of instruction that's a problem. Many people aren't reached at all due to other types of language barriers. Children or their parents might not know when to enrol, how to use the school web platform, or how to play the audiobooks. If you can't use the web, you will have a limited ability to find and access information. Can you imagine school projects without access to at least a book in your own language?

There's massive information asymmetry if you look at online content by language. On Wikipedia, for example, English is the largest edition in terms of users, followed by German and then French. On the other side of the spectrum,

there is an almost complete absence of content in many African and Asian languages.

Size of publicly available language datasets

Language	Value
French	3,800.0M
Spanish	2,300.0M
German	1,600.0M
Russian	1,000.0M
Turkish	900.8M
Arabic	802.3M
Slovenian	578.9M
Persian	125.8M
Indonesian	125.2M
Vietnamese	72.4M
Urdu	25.5M
Bengali	21.6M
Kurdish	7.9M
Nepali	5.4M
Hausa	3.1M
Swahili	2.7M
Yoruba	0.5M
Igbo	0.3M
Kanuri	0.0M
Rohingya	0.0M

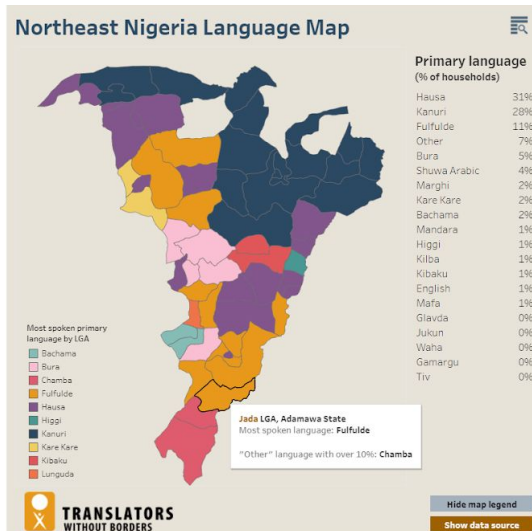Source: https://translatorswithoutborders.org/language-data-disparity/

And due to low levels of existing digital content, some widely spoken languages have limited language technology support, including machine translation and search functions.

This diagram shows that one can easily find publicly available parallel datasets to develop machine translation tools between English and French. But it is much harder to develop the same tools between English and Hausa, a language spoken by 63 million people globally. So speakers of Hausa and other marginalized languages find themselves on the wrong side of the digital divide - unable to access the information they need, when they need it.

"How do we address this? What are the steps?"

Mia: Let's first figure out what languages are spoken in the targeted communities. Then we can understand if education materials and services are taken up and used by speakers from all the language groups. As a first step, we need to capture accurate data about which languages are spoken. You can do this by asking, "What are the main languages you speak at home?" in education assessments and other surveys. It might surprise you to learn that there is no freely accessible dataset to find out which languages are spoken where.

Here are some examples of comparative maps.

Nigeria ([Wikipedia](#))	North-East Nigeria detail after TWB analysis ([link](#))
Mozambique ([Wikipedia](#))	Mozambique after TWB analysis ([link](#))
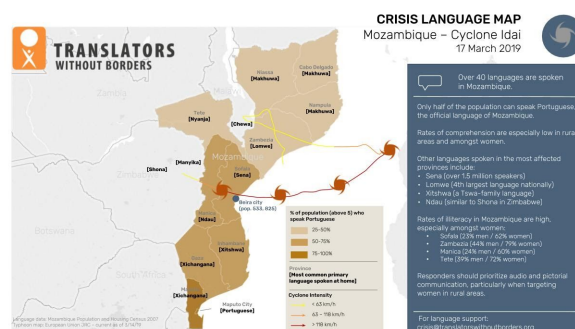
Alice: So, we need to put aside our assumptions about what languages people speak and understand. For example, Portuguese is only one of many languages spoken in Mozambique. To really understand the diversity of languages, we need to collect, share, and visualise the real picture. Translators Without Borders is trying to help do this through our language data initiative.

We can then apply this information: what languages need to be supported in a learning platform? Who drops out early and when? And we can then develop multilingual educational tools for learners and teachers, based on improved data about relevant languages. We can also track educational outcomes by primary language, to ensure no learner is left behind.

"This is very interesting. But which organisations can actually do this, and how much will it cost?"

Alice: Well, that depends on who you are. If you are within the government, you can ask for language to be included in censuses and build language into systematic data collection. You can ask children who register for education, "What is the primary language you speak at home?" and disaggregate data using language going forward.

What does this cost? Well, the time it takes to ask one more question and pull out the additional graphs. It's hard to quantify, but built within existing systems; it's a cheap, low-hanging fruit.

Mia: If you work for a humanitarian organisation, you can add 4 simple questions to your routine questionnaires. It costs no more than the assessment would have done anyway and provides a wealth of data for all sectors. And you should take language into account during your data collection. Across the board, we're learning that language-aware humanitarian action of this kind is an effective and cost-effective way to improve accountability, reach, and impact.

> **To fill the gap in humanitarian data on language, add 4 simple questions to needs assessments:**
>
> - **What is the main language you speak at home?**
> - **Which language do you prefer to receive written information in?**
> - **Which language do you prefer to receive verbal information in?**
> - **How do you prefer to receive information (in-person, radio, TV, poster, leaflet, phone call, SMS, etc.)?**

The added effort around the extra analysis is a few hours of someone's time. Yet, this can go a long way in filling a critical data gap for more effective programming.

"Okay, once we've collected the data, how can we turn it into action?"

Alice: We can use language data to take action on three levels: children, teachers, and systems. Let's start with children. Using language data, you will be able to tell which language groups didn't register for school, dropped out early, or consistently underperform. You can then adjust your communication and programming to meet their needs, leaving no one behind.

Mia: We can also take action to support teachers. Recent research in Bangladesh and Nigeria shows that teachers often don't speak the language of instruction well. When teachers aren't confident in the language of instruction, they can't get access to guidance to help them teach. And of course, such materials include critical guidance on child-friendly classrooms or Covid-19 school reopening plans.

Teacher training workshops can be challenging because they use technical language that's hard to understand. This means teachers can face challenges with teaching and have no available resources to help them overcome these challenges.

Translators Without Borders's ongoing research in collaboration with the Global Education Cluster and Save the Children shows teachers ask for materials:

- **By email and print**
- **In searchable formats**
- **Shorter and in plain language**
- **In pictorial and audio-visual formats**

Alice: With language data, we can also take action on a system level. If over 40% of children are educated in a language they don't speak at home, we need to know how this affects their education. Much as we did for gender, unless we collect data on language inequality, we cannot begin to address this. Collective

gathering and sharing of language data can allow us to understand how to increase learning outcomes for speakers of marginalised languages.

"Beyond education, fairer access, and closing digital divides, what else can we do? Give us an idea."

Alice: Many of us who speak major languages (like English) find it easier and easier to navigate the world. We have readily available machine translation apps conveniently located on our phones or laptops when we need them. This enables us to find and triangulate information, and carry out research. No-one invests in marginalised languages, and here is a demonstration of the difference. Take this source sentence and see what happens when you use machine translation, and what that does for life-saving information for millions of people.

Source sentence	
Your life jacket is under your seat. Place it over your head and tie the straps around you. Inflate your life jacket after leaving the aircraft.	
Language and population	**Machine translation**
Hausa, 44 million	Air pollution is under your control. Put it on top of your head and tie it to your body. You can only ventilate the air when you leave the boat.
Burmese, 32 million	Your life jackets are under your seat. This is putting your head to form on your own. Cotton. After leaving the plane on your life jackets in the air breathed.
Turkish, 75.7 million	Your life jacket is under your seat. Pass it over your head and tie it around the body. Only swallow the life vest after leaving the plane.

Mia: We urgently need to invest more in language technology solutions for marginalised languages. Natural-language-processing applications such as chatbots can support multilingual, interactive education for those learning

remotely. Machine translation and automatic speech recognition in local languages is a service that can be deployed rapidly in crisis situations. It enables people to access the information they need — both online and offline.

Thank you, Alice and Mia, for these insights. You can get in touch with <u>Alice Castillejo</u> if you have a use case for innovative language technology or other language support needs and you would like to collaborate with Translators Without Borders. You can also get in touch with <u>EdTech Hub</u> if you have any further questions or ideas.

Summary

To celebrate International Mother Language Day 2021, Björn spoke with Alice Castillejo and Mia Marzotto at Translators without Borders. We talked about the importance of mother tongue-based education and the need to support speakers of marginalised languages.

Research around the world shows that educational outcomes are worse for students studying in a second language. The use of an unfamiliar language is linked to high dropout rates and low academic achievement. We also know that 40% of children worldwide are not educated in a language they speak at home.

We need to put aside our assumptions about what languages people speak and understand. For example, Portuguese is only one of many languages spoken in Mozambique. To really understand the richness of different languages, we need to collect, share, and visualise the real picture. Having done that, we can then apply this information. Do children who speak a particular language drop out early? We can then develop multilingual educational tools for learners and teachers based on improved data about relevant languages. We can also track educational outcomes by primary language, to ensure no learner is left behind.

We can use language data to take action on three levels: children, teachers, and systems. With children, we can identify which language groups didn't register for school, dropped out early, or consistently underperform. We can then adjust our communication and programming to meet their needs, leaving no one behind. We can also take action to support teachers.

With language data, we can also take action on a system level. If over 40% of children are educated in a language they don't speak at home, we need to know how this affects their education. Much as we did for gender, if we don't gather the data on inequality, we cannot begin to address it. Collective gathering and sharing of language data can allow us to understand how to increase learning outcomes for speakers of marginalised languages and provide the support they need and deserve.

Resumo deste artigo do blog en portugués

De forma a se comemorar o Dia Internacional da Língua Materna neste ano de 2021, Björn conduziu uma conversou com Alice Castillejo e Mia Marzotto na *Translators without Borders*. Falamos sobre a importância da educação na língua materna e a necessidade de apoiar os falantes nativos de línguas marginalizadas.

Segundo pesquisas feitas ao redor do mundo indicam que os resultados acadêmicos são nefastos para alunos que estudam em um segundo idioma. O uso de uma língua desconhecida está associado a altas taxas de evasão e baixo desempenho acadêmico. É sabido igualmente que 40% das crianças no mundo não são educadas na língua que falam em casa ou seja maternal.

Precisamos deixar de lado nossas hipóteses sobre as línguas que as pessoas interagem e compreendem. Por exemplo, a língua portuguesa é apenas uma das muitas línguas faladas em Moçambique. Para realmente compreender a riqueza das diferentes linguagens, precisamos coletar, compartilhar e visualizar a imagem real. Feito isso, podemos aplicar essas informações. Será que é tendencioso que as crianças falem uma determinada língua e de seguida esquecê-la? Desta modo, seria possível desenvolver ferramentas educacionais poliglota para alunos e professores com base em dados aprimorados em idiomas relevantes. A outra hipótese seria de acompahar os resultados curriculares dos alunos na sua língua maternal de forma a garantir a inclusão de todos e o avanço de todos os alunos.

Podemos usar dados de linguagem para actuar em três níveis: crianças, professores e sistemas. Com as crianças, podemos identificar grupos linguísticos que não se matricularam na escola, que desistiram prematuramente ou que tiveram um desempenho consistentemente baixo. Podemos então ajustar nossa comunicação e programação para atender às necessidades deles, sem deixar ninguém para trás. Também podemos tomar medidas para apoiar os professores.

Com os dados linguísticos, também podemos actuar no nível do sistema. Se mais de 40% das crianças estão sendo educadas em um idioma que não falam em casa, precisamos saber como isso afecta sua educação. Assim como fizemos com o gênero, se não coletarmos os dados sobre as desigualdades, não podemos começar a consertar. Coletar e compartilhar dados sobre línguas coletivamente pode nos ajudar a entender como melhorar os resultados de aprendizagem de falantes e nativos de línguas marginalizadas. Precisamos de lhes dar o apoio de que necessitam e merecem.

ملخص هذه المدونة

للاحتفال باليوم العالمي للغة الأم 2021، تحدث بيورن مع أليس كاستيليجو وميا مارزوتو في مترجمون بلا حدود. تحدثنا عن أهمية التعليم القائم على اللغة الأم والحاجة إلى دعم المتحدثين باللغات المهمشة.

تظهر الأبحاث حول العالم أن النتائج التعليمية أسوأ للطلاب الذين يدرسون بلغة ثانية. يرتبط استخدام لغة غير مألوفة بارتفاع معدلات التسرب وانخفاض التحصيل الدراسي. نعلم أيضًا أن 40٪ من الأطفال في جميع أنحاء العالم لا يتلقون تعليمًا بلغة يتحدثونها في المنزل.

نحن بحاجة إلى وضع الافتراضات الخاصة بنا جانبًا حول ما هي اللغات التي يتحدث بها الناس ويفهمونها. على سبيل المثال، البرتغالية هي لغة واحدة فقط من بين العديد من اللغات المستخدمة في موزمبيق. لفهم ثراء اللغات المختلفة حقًا، نحتاج إلى جمع الصورة الحقيقية ومشاركتها وتصورها. بعد القيام بذلك ، يمكننا بعد ذلك تطبيق هذه المعلومات. هل الأطفال الذين يتحدثون لغة معينة يتركون الدراسة في وقت مبكر؟ يمكننا بعد ذلك تطوير أدوات تعليمية متعددة اللغات للمتعلمين والمعلمين بناءً على بيانات محسنة حول اللغات ذات الصلة. يمكننا أيضًا تتبع النتائج التعليمية من خلال اللغة الأساسية ، لضمان عدم ترك أي متعلم دون تعليم.

يمكننا استخدام بيانات اللغة لاتخاذ إجراءات على ثلاثة مستويات: الأطفال والمعلمين والأنظمة. بالنسبة للأطفال ، يمكننا تحديد المجموعات اللغوية التي لم تسجل في المدرسة ، أو التي لم تستمر في دراستها ، أو التي كان أداؤها ضعيفًا باستمرار. يمكننا بعد ذلك تعديل برامجنا لتلبية احتياجاتهم ، دون ترك أي شخص دون تعليم. يمكننا أيضًا اتخاذ إجراءات لدعم المعلمين. أحيانا لا يتحدث المعلمون لغة التدريس جيدًا.

باستخدام بيانات اللغة ، يمكننا أيضًا اتخاذ إجراءات على مستوى الأنظمة. إذا تم تعليم أكثر من 40٪ من الأطفال بلغة لا يتحدثونها في المنزل ، فنحن بحاجة إلى معرفة كيف يؤثر ذلك على تعليمهم. مثلما فعلنا مع الجنس ، إذا لم نجمع البيانات حول عدم المساواة ، فلا يمكننا البدء في معالجتها. يمكن أن يتيح لنا العمل الجماعي وتبادل بيانات اللغة فهم كيفية زيادة نتائج تعلم المتحدثين باللغات المهمشة.

Résumé de cet article de blog en français

Pour célébrer la Journée internationale de la langue maternelle 2021, Björn s'est entretenu avec Alice Castillejo et Mia Marzotto à Translators without Borders. Nous avons parlé de l'importance de l'éducation fondée sur la langue maternelle et de la nécessité de soutenir ceux qui défendent les langues minoritaires.

Des recherches menées dans le monde entier montrent que les résultats scolaires sont, tristement,moins favorables pour les étudiants qui étudient dans une deuxième langue. L'utilisation d'une langue étrangère comme langue d'enseignement principale est fortement liée aux taux d'abandon élevés et à de faibles résultats scolaires. Nous savons également que 40% des enfants dans le monde ne sont pas éduqués dans une langue qu'ils parlent à la maison.

Nous devons mettre de côté nos hypothèses sur les langues que les gens parlent et comprennent. Par exemple, le portugais n'est qu'une des nombreuses langues parlées au Mozambique. Pour vraiment comprendre la richesse des différentes langues, nous devons collecter, partager et visualiser l'image réelle. Cela fait, nous pouvons ensuite appliquer ces informations pour influencer la recherche, la politie et la pratique. Les enfants qui parlent une langue maternelle abandonnent-ils plus tôt? Nous pouvons ensuite développer des outils pédagogiques multilingues pour les apprenants et les enseignants. Cela sera basé sur des données améliorées sur les langues pertinentes. Nous pouvons également suivre les résultats scolaires par langue principale, pour nous assurer qu'aucun apprenant n'est laissé pour compte.

Nous pouvons utiliser les données linguistiques pour agir à trois niveaux: les enfants, les enseignants et les systèmes. Avec les enfants, nous pouvons identifier les groupes linguistiques qui ne sont pas inscrits à l'école, qui ont abandonné prématurément ou qui ont constamment sous-performé. Nous pouvons alors ajuster notre communication et notre programmation pour répondre à leurs besoins, sans laisser personne de côté. Nous pouvons également agir pour soutenir les enseignants. Les enseignants ne maîtrisent pas souvent la langue d'enseignement, qui n'est pas, parfois, leur langue maternelle.

Avec les données linguistiques, nous pouvons également agir au niveau du système. Si plus de 40% des enfants sont éduqués dans une langue qu'ils ne parlent pas à la maison, nous devons savoir comment cela affecte leur éducation. Tout comme nous l'avons fait pour le genre, si nous ne collectons pas les données sur les inégalités, nous ne pouvons pas commencer à y remédier. La collecte et le partage collectifs de données linguistiques peuvent nous permettre de comprendre comment améliorer les résultats d'apprentissage des locuteurs de langues marginalisées.

Resumen de esta entrada de blog en castellano

Para celebrar el "Día Internacional de la Lengua Materna", Björn habló con Alice Castillejo y Mia Marzotto de 'Traductores sin Fronteras'. Hablamos sobre la importancia de la educación basada en la lengua materna y la necesidad de apoyar a los hablantes de lenguas marginadas. We talked about the importance of mother tongue-based education and the need to support speakers of marginalised languages.

Investigaciones realizadas alrededor del mundo muestran que los resultados educativos son peores para los estudiantes que estudian en una segunda lengua. El uso de una lengua poco conocida está relacionado con altos índices de deserción escolar y bajo rendimiento académico. También sabemos que el 40% de los niños de todo el mundo no reciben educación en la lengua que hablan en casa.

Necesitamos dejar de lado nuestras suposiciones sobre qué idiomas habla y entiende la gente. Por ejemplo, el portugués es solo uno de los muchos idiomas que se hablan en Mozambique. Para comprender realmente la riqueza de las diferentes lenguas, debemos recopilar, compartir y visualizar la situación real. Una vez hecho esto, podemos aplicar esta información. ¿Los niños que hablan una determinada lengua abandonan pronto la escuela? Con ello podremos desarrollar herramientas educativas multilingües para alumnos y profesores basadas en datos mejorados sobre las lenguas relevantes. También podemos hacer un seguimiento de los resultados educativos por lengua principal, para garantizar que ningún alumno se quede atrás. Podemos utilizar datos de las diferentes lenguas para actuar en tres niveles: niños, profesores y sistemas. Con los niños, podemos identificar cuáles fueron los grupos de lenguas que no se matricularon en la escuela, la abandonaron temprano o tuvieron un bajo rendimiento académico constante. Es así que podemos adaptar nuestra comunicación y diseño de programas educativos para satisfacer sus necesidades, sin dejar a nadie atrás. También podemos tomar medidas para apoyar a los maestros. Con datos de las diferentes lenguas, también podemos actuar a nivel del sistema. Si más del 40% de los niños son educados en una lengua que no hablan en casa, tenemos que saber cómo afecta esto a su educación. Al igual que en el caso del género, si no recopilamos los datos sobre la desigualdad, no podemos empezar a abordarla. La recopilación y el intercambio colectivo de los datos de lenguas habladas, pueden permitirnos comprender cómo mejorar los resultados de aprendizaje de los hablantes de lenguas marginadas.

Reference

Haßler, Björn, Castillejo, A., Marzotto, M., El-Serafy, Y., Khlayaleh, A., Langa, A., Koomar, S., Saadeddin, Z., Tegha, G., & Villavicencio Peralta, X. A. (2021). Mother Language Day 2021. Blog post. DOI: 10.5281/zenodo.4544985 Available at https://docs.edtechhub.org/lib/ZQCX7TX3. Available under Creative Commons Attribution 4.0 International, https://creativecommons.org/licenses/by/4.0/.