

Translation Commons: No Language and No Linguist Left Behind

Jeannette Stewart

Translation Commons

Cupertino, USA

jeannette@translationcommons.org

Abstract

This paper describes work conducted by Translation Commons to provide a reliable method of digitally rendering a language. It seeks to solve the problem that unrepresented languages have participating in a global communications network. We present a methodology that can easily be utilized by non-technical users to meet the particular needs of their language. Evidence of the success of this approach is cited with examples. The importance of this initiative is illustrated by recommendations for on-going and new projects.

Keywords: Translation Commons, Language Technology Project, Zero to Digital, indigenous communities, language status, language code, Unicode, Common Locale Data Repository (CLDR), University Outreach, Social Media, Global Language Digitization Initiative (GLDI), fonts, keyboards, digitization.

Περίληψη

Αυτό το άρθρο περιγράφει την εργασία που διεξήγαγε η μη κερδοσκοπική οργάνωση Translation Commons για να παρέχει μια αξιόπιστη μέθοδο ψηφιακής απόδοσης μιας γλώσσας. Επιδιώκει να λύσει το πρόβλημα ώστε οι μη αντιπροσωπευόμενες γλώσσες να μπορούν να συμμετέχουν σε ένα παγκόσμιο δίκτυο επικοινωνιών. Παρουσιάζουμε μια μεθοδολογία που μπορεί εύκολα να χρησιμοποιηθεί από μη τεχνικούς χρήστες για την κάλυψη των ιδιαίτερων αναγκών της γλώσσας τους. Η απόδειξη της επιτυχίας αυτής της προσέγγισης παρατίθεται με παραδείγματα. Η σημασία αυτής της πρωτοβουλίας αποδεικνύεται με συστάσεις για τρέχοντα και νέα έργα.

1. Introduction

Are all languages equal? Our language environment is teeming with life. An ecology of language involves devoted care and attention to conserve, guarantee sustainability, assure health and well-being with inarguable equality. Development of our digital domain has been unequal, endangering under-resourced languages. Strenuous effort is needed to future-proofing against exclusion and even extinction.

2. Translation Commons

[Translation Commons](https://translationcommons.org) is a nonprofit organization with a vision to help every language digitize to share equally in the benefits of a connected digital world, ensuring that “no language and no linguist is left behind”. Our online platform provides free access to language tools, language training and sharing of educational knowledge for all languages. During 2019, we are a Social-Civil society partner to International Year of Indigenous Languages and have been actively working on three projects: a language technology project, a university outreach project and a social media campaign.

2.1 Language Technology Project

The language technology team agreed on the following objectives:

- **Enable** indigenous communities to use internet technology
- **Engage speakers** and especially young learners to appreciate and honor their culture

- Identify **technology for educating**
- Encourage speakers to **create content** and community with digital tools and standards

All in their own language.

2.2 Zero to Digital Document

Their first project was a step-by-step document as technical guidelines on how to enable a language to be online. The document [Zero to Digital](https://translationcommons.org/zero-to-digital) is the brainchild of 4 of the most committed volunteers in the digitization of languages, all experts in their respective fields: Craig Cornelius, Craig Cummings, Deborah Anderson and Lee Collins. You can download it at translationcommons.org.

2.3 Objectives of the Document

The guidelines document helps communities:

- **Evaluate** their language’s situation with respect to digital support
- **Suggest approaches** to develop basic digital language support
- **Encourage practical usage** of language tools, even without formal language documentation, grammar, and educational standards
- **Point indigenous communities to available tools** and techniques to build digital capabilities.

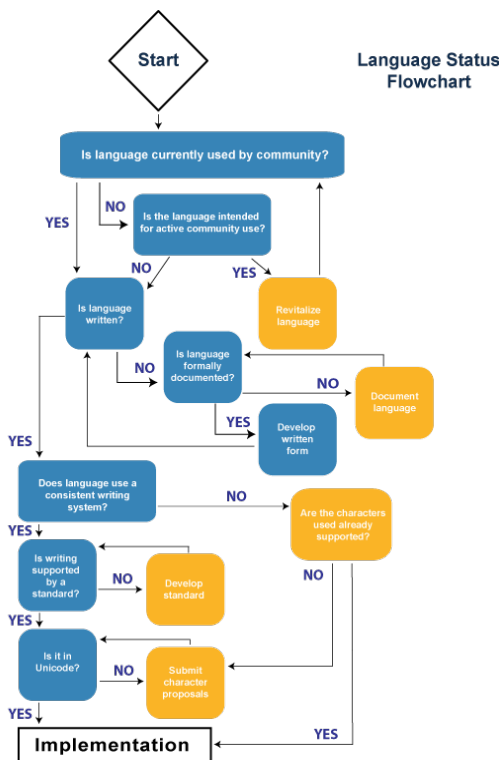
- **Engage community members** in decisions and process
- **Connect communities to standards** and technology professionals

2.4 Zero to Digital Flowcharts

With two main flowcharts to show the basic workflows in digitization, it is very easy to follow and has a more technical section at the end for the technical implementation. By using questions answered with Yes-No, the communities are guided to the next step they need to undertake with the available resources and people. It offers concrete actions for each response, a guide to perform specific tasks, pointers to resources and tools and suggestions for next steps.

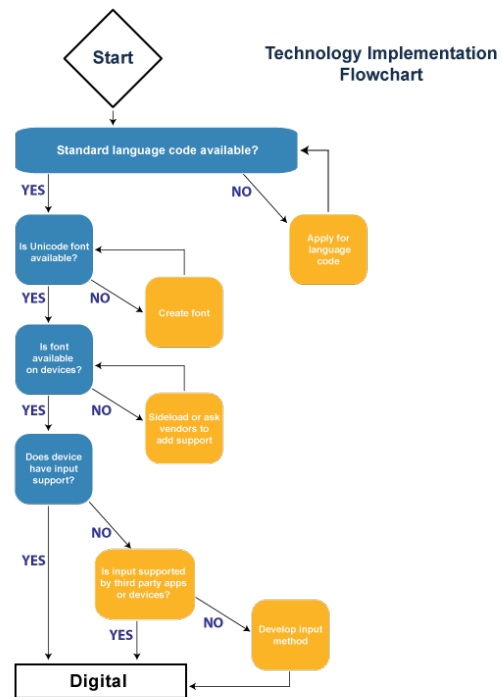
1.4.1 Language Status Flowchart

This workflow describes the steps to determine the current status of a language in preparation for using it online. The aim is to determine the current level of support for your language on computers and mobile devices. It includes suggested next steps to help you take your language online. You do not need a technical background to answer these questions.



1.4.2 Technology Implementation Flowchart

When you have determined the status of your language, use this workflow to get started with the technology available to take your language online.



Any written language used in digital devices must therefore include:

- a standard encoding system for characters, for example Unicode,
- rendering systems plus fonts for characters used in the language, and
- methods or applications to render the code points to desired media.

1.4.3 Unicode

Unicode standardization covers many of the text processing aspects of a writing system and language. With font and input support, many things will just work. This includes most aspects of basic word processing, spreadsheets, and email.

1.4.4 Common Locale Data Repository (CLDR)

The Common Locale Data Repository (CLDR) provides “key building blocks for software to support the world's languages” by collecting useful information for different locales (language and country). This data can provide the names of languages, countries, months, weekdays, and other information. It also enables locale-aware formatting of date, time, numbers, and other commonly formatted information. Although CLDR data is not required for basic text communication in indigenous languages, this information enhances language functionality. Almost all tools such as email, texting, social media, and so on will work well when fonts and keyboard are present.

This information is used by programmers to create output for online applications in specific context such as localized calendars, spreadsheets, numeric output, menu selections, and other user interface contexts.

3. University Outreach Project

For our second project, we reached out to the Language Departments of universities around the world and asked them to create an awareness event in their institution. To host classroom presentations, lectures, workshops and conferences during Fall 2019 to discuss cultural diversity, focusing on indigenous languages. We created a multicultural intern group and reached out to over 650 universities, of which many created events. They are still creating events until the end of the year and we just met with a group of them to brainstorm on next steps for the years to come.

4. Social Media Campaign

Our third project of a social media campaign reached 200,000 people with over 15% engagement. This was done by volunteers with no budget. The significance of this success has made us now aim for 1M viewers.

4.1 Campaign objectives

The campaign was named “Sharing Inspiration”:

- Helping indigenous groups to inspire each other
 - Language education
 - Leveraging modern technology
 - Cultural pride and resilience
- Inspiring the audience to see the potential of indigenous speakers
 - Entrepreneurship and innovation
 - Indigenous knowledge (sciences, historical knowledge)
- Successful language revitalization projects

5. Success Stories

5.1 Cherokee Nation

The Cherokee Nation in North America. From creating their syllabary in 1836 to contracting out the creation of their fonts they were always committed to the latest technology, from typewriter balls to early mobile and today they are fully digitized, from operating systems and desktop applications to texting, mobile apps like maps and mango, social media and everything else we do in the main 20 languages.

5.2 ADLaM

Another example is from the early 1990s when two teenagers created a writing system for the African language of the Fulani people. Their work became “Bindi Pulaar” and eventually “Adlam”. An article entitled “The Alphabet that Will Save a People from Disappearing” describes how Adlam became widespread across many African countries. This had a huge impact on a couple of technologists who helped the brothers digitize it. The ADLaM journey started first with handwriting, then implemented it as a font encoding with Arabic code points. Technical challenges and support by technology vendors are enabling rapid growth of this alphabet in a previously under-served language community implementation. The challenges are many as the newly standardized script meets the technical infrastructure of

the Internet. ADLaM holds great potential to improve literacy for millions as it is spreading rapidly in more than 20 countries.

6. What’s next ?

Digitization is practically possible when the community is fully committed.

6.1 Internet Connectivity

There are over 4,000 written languages (many more oral languages) in the world today and only 1,000 of them are supported online. This is because most of these languages are not yet digitized.

What would the world be like if we were all online? With so many international agencies and companies working towards 100% connectivity, we ask: In what language will the next 7 billion people use the internet? This is an answer only we can give, all of us in this room and many more from the communities themselves, the academic world, the language professionals and most importantly our language technologists.

Bringing connectivity to emerging countries will solve many problems and create great opportunities. BUT connectivity alone will not allow people to be online unless their language is digitally supported. To fully partake in the benefits of online connectivity, people will need to be able to type, speak and read the online material in their own language. What if we could bring together everything we all learned and work at a global scale?

7. Global Language Digitization Initiative

Translation Commons’ mission is to take all necessary steps to bring every language online so emerging countries can benefit from connectivity.

It is a vision we all share and we have all worked towards. But we are each an expert in a different field and it is only when we combine our knowledge, passion and fortitude, we will accomplish our vision.

Let’s have our technologists create the technical support and our academics the linguistic support. Let’s have our educators and researchers work on teaching the next generations and let’s get the language professionals to embrace and include new linguists from thousands of languages.

We have created the **Global Language Digitization Initiative (GLDI)**. Our network of volunteers comes from all language specializations, our educational network of universities and schools is rapidly expanding and our platform offers free training and practice for new linguists, putting Translation Commons in a unique position to be the Language Technology partner among the team of organizations striving for 100% global connectivity.

7.1 ‘It takes a village to raise a child’

Inspired by this famous African proverb, we acknowledge the critical importance of nurturing newly digitized languages, parenting them, as it were, allowing them to grow in safety, educating them and providing for their needs until they reach sufficient maturity to take care of themselves.

Our curiosity about these children’s future leads us to ask: “What if we could make a bigger difference than anything we’ve done so far?”

“What if we could honor languages of communities past and present?”

“What if we could work on something so important that it could lead us to solve some of the world’s toughest problems?”

We can do this by partnering, creating coalitions and working together: “a broad, multi-stakeholder alliance for digital cooperation” as suggested by the UN report on June 2019. This is the best way to achieve the UN Sustainable Development Goals.

Today Translation Commons is two thousand volunteers strong, a million tomorrow!

Everyone in Translation Commons is committed to this vision and we pledge to be the coordinating force between all of us, to connect and channel our expertise where, when and how it is needed.

7.2 GLDI Phases

For a language to be supported on the internet, a series of steps must occur:

Language Technology

- Create the stack of documents on how to digitize. Language Digitization means encoding their script so it can be recognized digitally; creating software for their keyboard and other means of input; creating templates for them to start compiling digital archives of voice and text; creating digital dictionaries and digital courses for children to learn their language through modern tools.
 - [Zero to Digital](#), a step-by-step guide to implement language digitization
- Reach out to communities and assess their digital needs
- Guide the communities through implementation of language digitization.
-

Linguistic and Cultural Support

- Create teams of Language Ambassadors to connect communities with local educational institutions and schools
- Create groups of Project Managers to liaise between communities and various expert organizations that will help them digitize
- Provide education on internet safety in their own language
- Create collaboration between universities to share knowledge and research
 - Lecture exchange
 - Curriculum enhancement
 - Incorporate new digitized languages in standard settings

Create Economic Equal Opportunities

Facilitate the creation of new jobs by training translators to help communities share knowledge and information through translating from and into their language

- Offer access to free tools for linguists already on TC Platform
 - Computer-assisted Translation
 - Translation Memory technology
 - Machine Translation
 - Glossaries/Terminology Management
- Offer free tutorials, courses, videos for linguists through the TC Learning Center
- Create a marketplace for linguists of new languages for
 - Internships and jobs with industry
 - Internships and jobs with language companies

8. Conclusion

The Zero to Digital document presents the global language community with a broad vision to bring together many diverse organizations, communities and institutions. Progress can only be achieved when all of the aforementioned are bridged by a coordinating body. Translation Commons is pledged to build and maintain this bridge, providing the means for all language users to share equally in the opportunities offered by digital life.

9. Copyright

ELRA’s policy is to acquire copyright for all LT4All contributions. In assigning your copyright, you are not forfeiting your right to use your contribution elsewhere. This you may do without seeking permission and is subject only to normal acknowledgement to the LT4All proceedings. The LT4All Proceedings are licensed under CC-BY-NC, the Creative Commons Attribution-Non Commercial 4.0 International License.