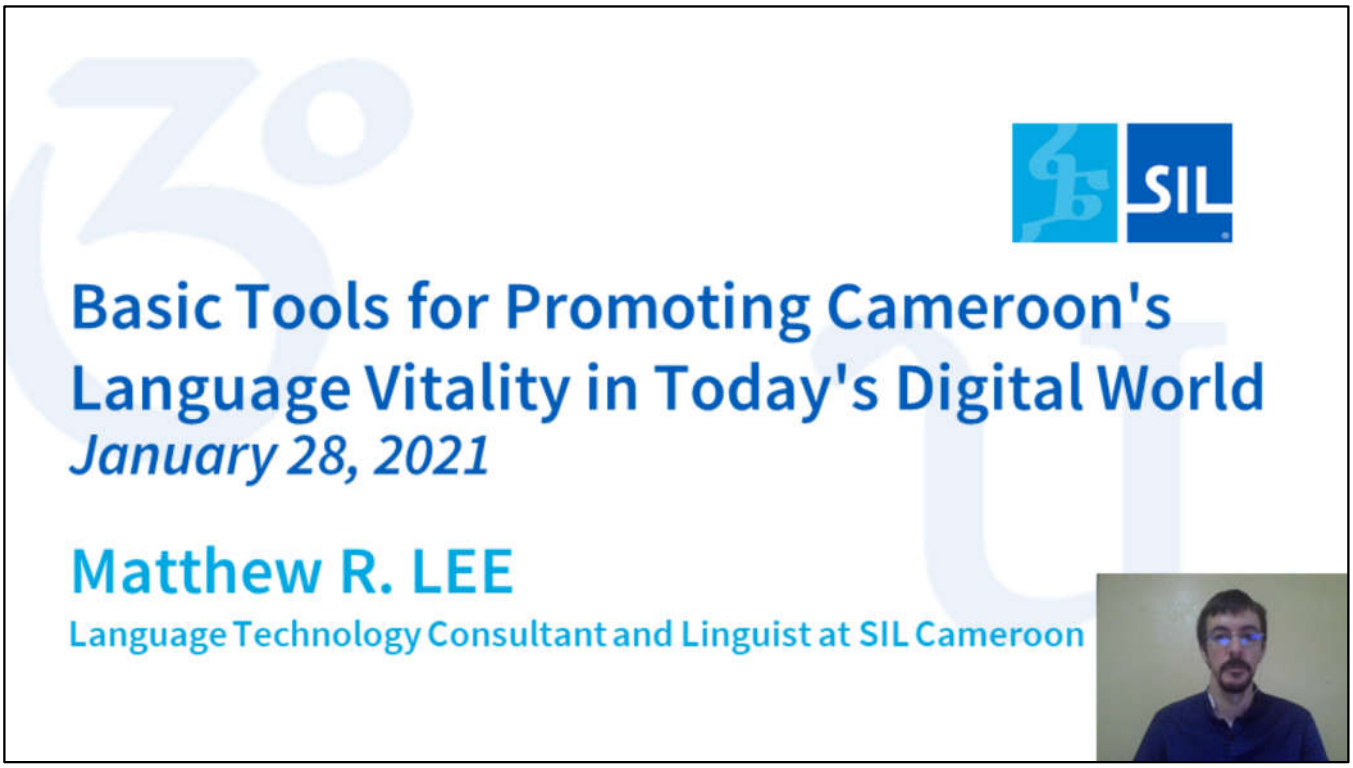


Basic Tools for Promoting Cameroon’s Language Vitality in Today’s Digital World, by Matthew R. Lee

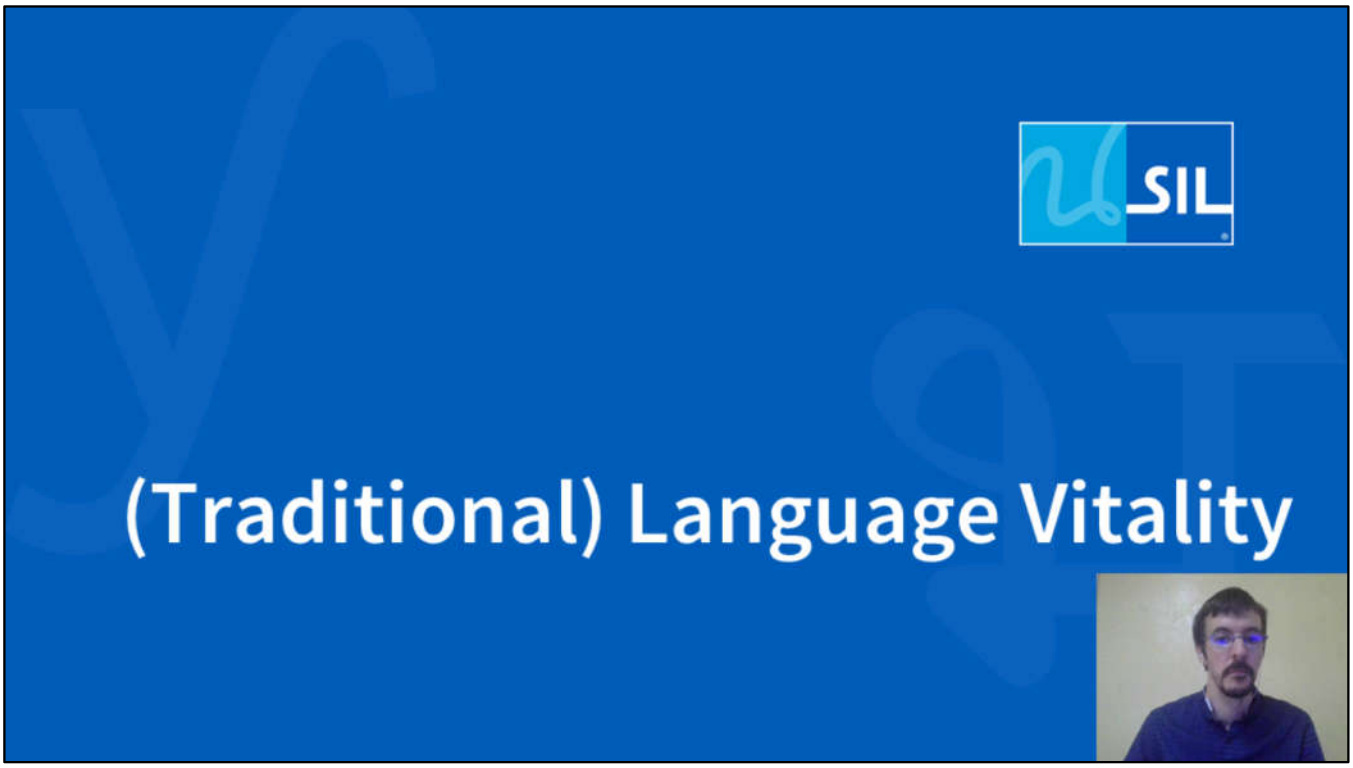
Who am I? I am a Language Technology Consultant and Linguist here at SIL Cameroon. I've been serving SIL Cameroon. I've been serving here since 2020, and I serve with my wife Teresa. We're Americans. I have a Bachelor in Integrated Science and Technology from James Madison University, and I am working on Finishing a Master's in Descriptive Linguistics from Dallas International University.

CMB Language Technology

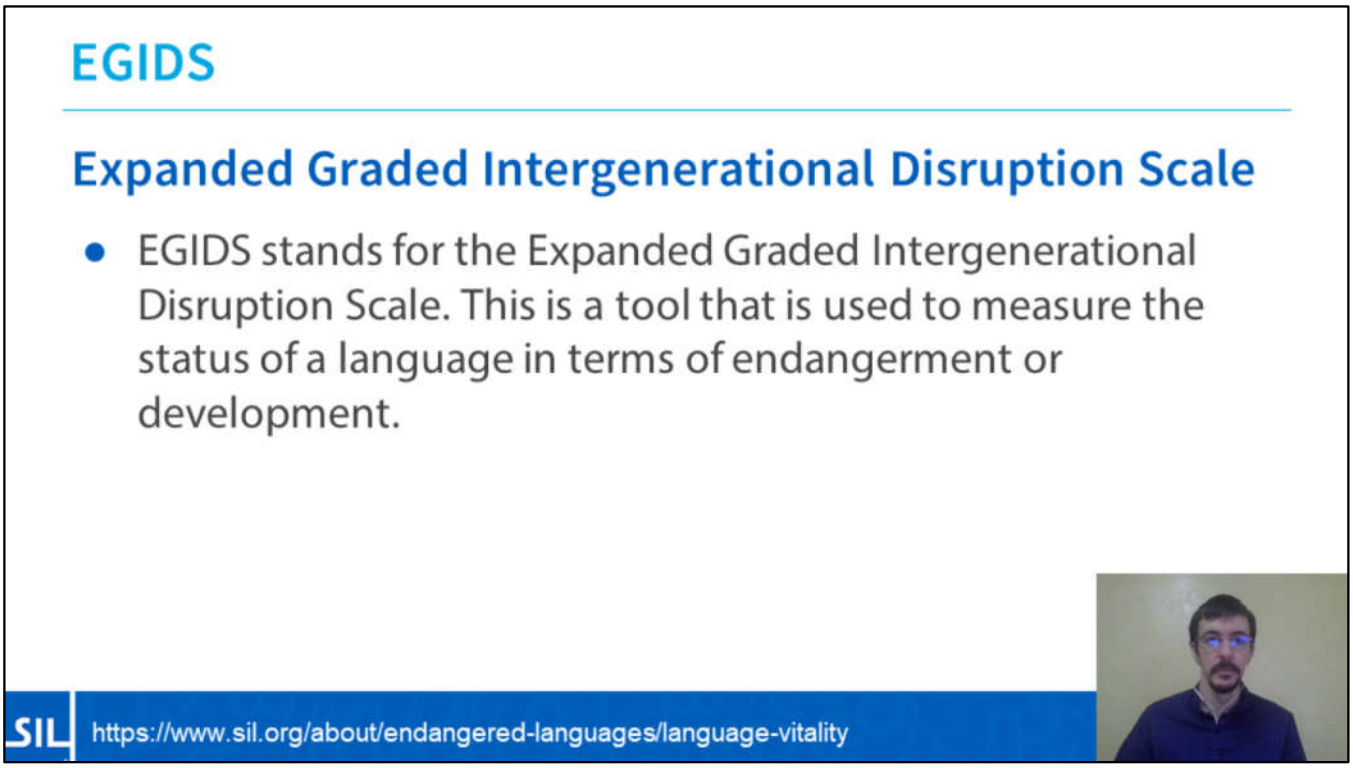
- Current Team
  - Matthew LEE
  - NGONO Louis Pascal

SIL

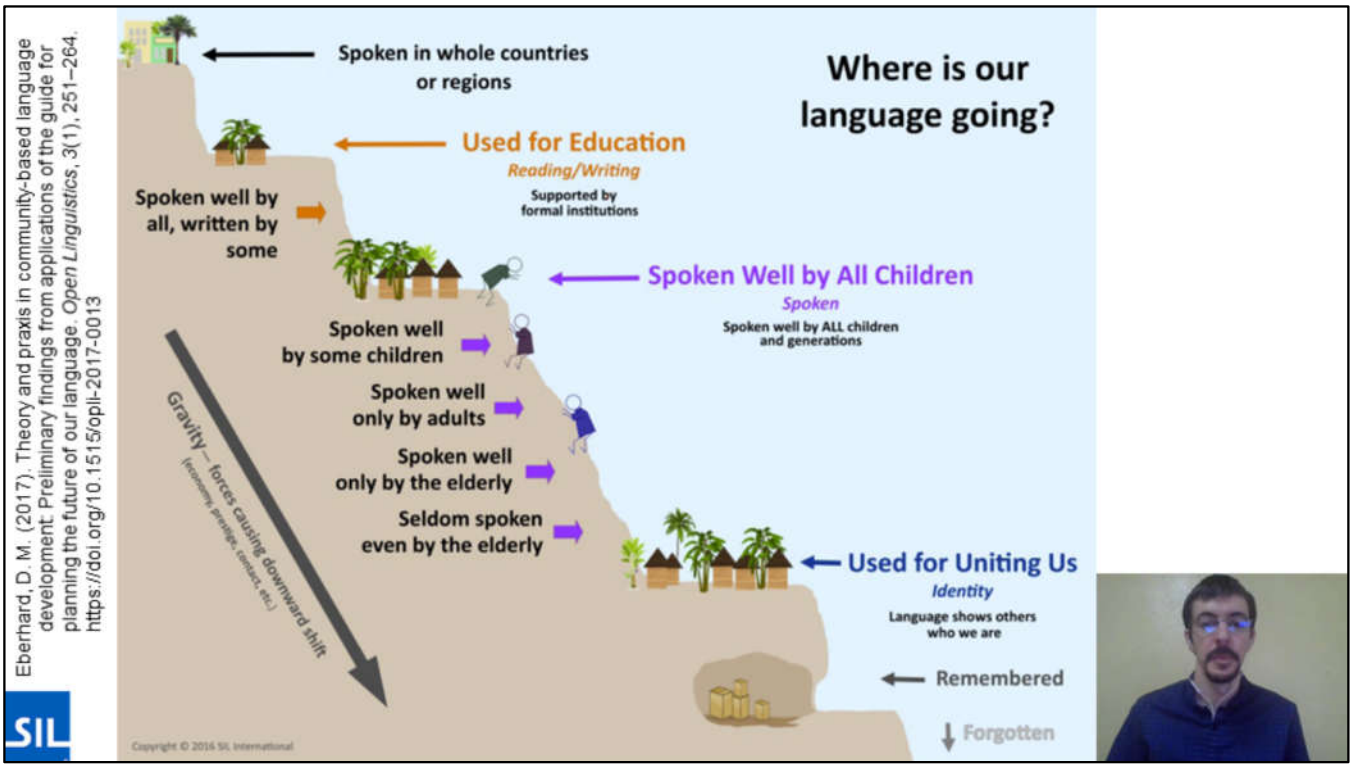
The current Language technology team consists of myself and Louis Pascal Ngono.



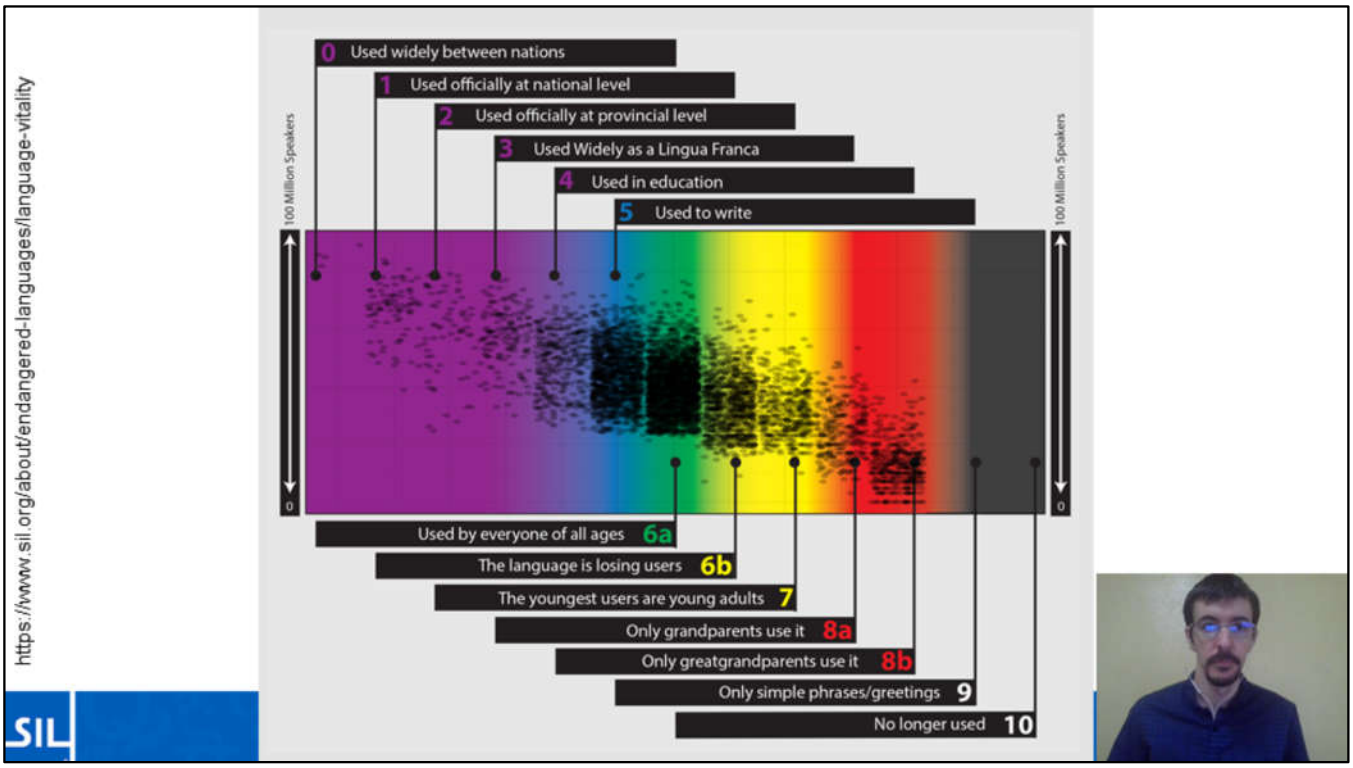
We will start with “Traditional” Language Vitality.



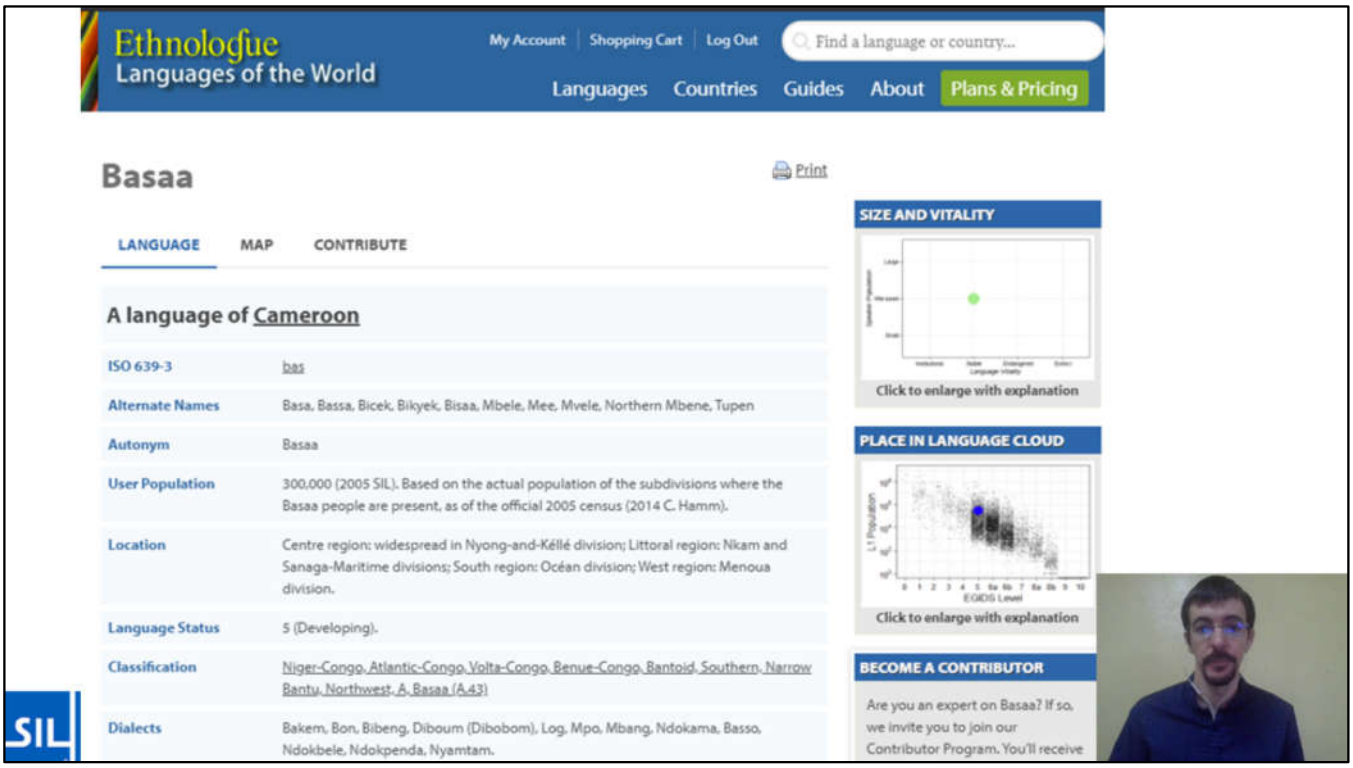
If you are familiar with linguistics or sociolinguistics, you may be familiar with EGIDS, the Expanded Graded Intergenerational Disruption Scale. EGIDS is a tool that is used to measure the status of a language in terms of endangerment or development.



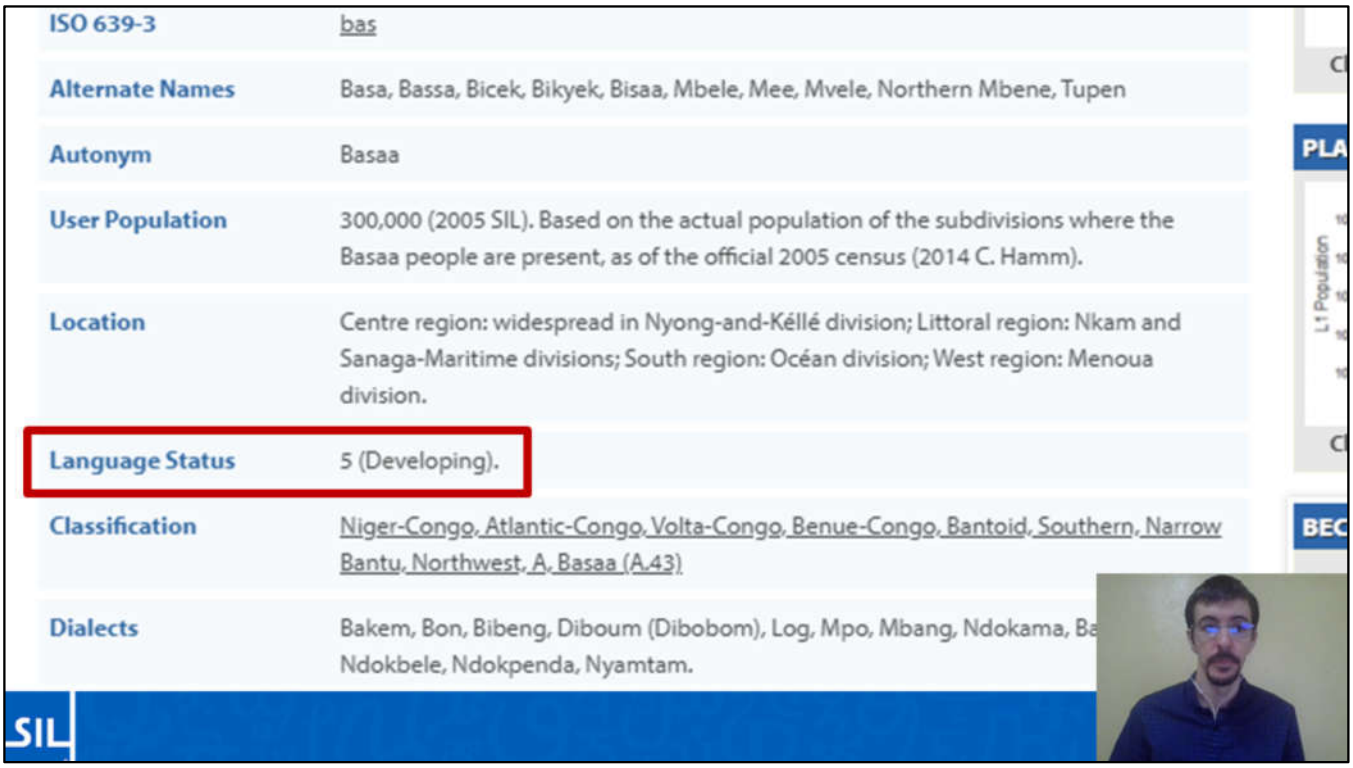
This is an illustration of the EGIDS scale. At the left, we would see national languages like Swahili, English, or French. At the right, you would find languages like Latin or Babylonian that no longer have a population. The most important indicator of future language vitality is not the number of speakers, but (as you see marked in violet) whether young people are learning and using a language.



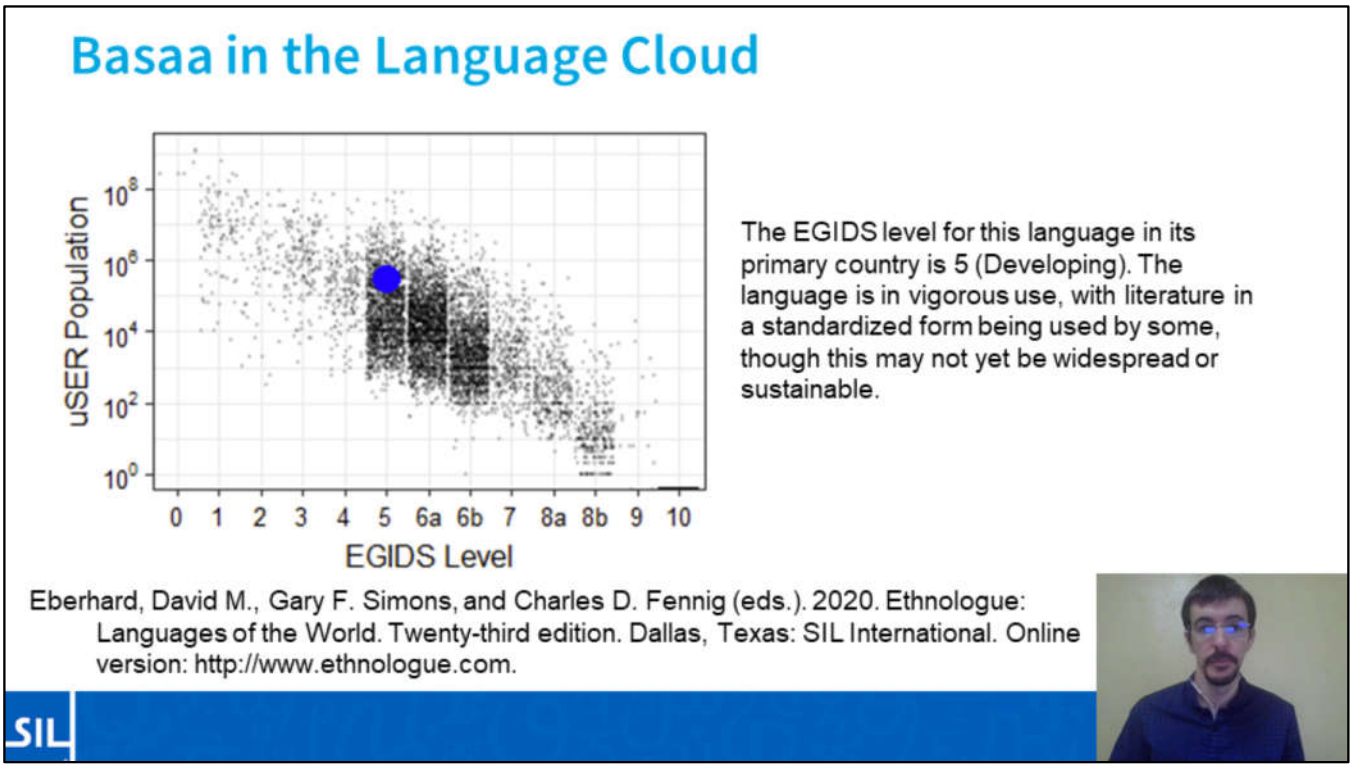
This chart shows the vitality of the known languages of the world. Each dark dot represents a language. The height of each dot represents the relative number of speakers. Languages to the left are likely more secure. Languages in the middle are in transition, and languages toward the right are in extreme danger. You can see that most of the languages of the world are between 5 and 7.



If we look at Basaa, for example, on the Ethonologue, a resource managed by SIL International…

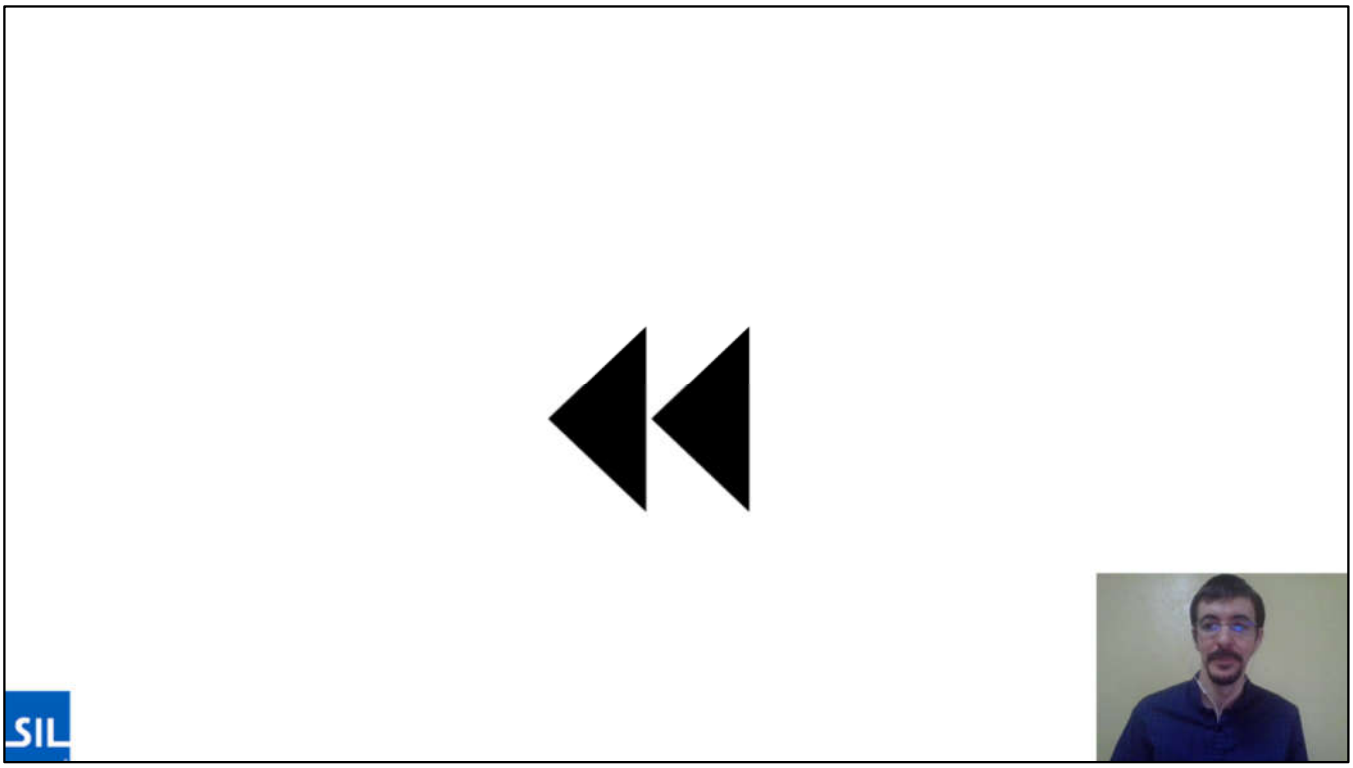


Here we see that the status is listed as 5, Developing.



According to the Ethnologue, The EGIDS level for this language in it's primary country is 5 (Developing). The language is in vigorous use with literature in a standardized form being used by some, though this may not yet be widespread or sustainable.
You can see the blue dot that represents Basaa among the world's languages.
[Read The Description]

I have mentioned SIL's Ethonologue, a database of the languages of the world. You may have heard recently that it is no longer available online for free. Actually, the Ethnologue is available for free in many countries of the world, such as Cameroon. All you need to do is register (for free), click the link in your email, and you will have instant access.



Wait a minute…I did say I work in language Technology, where's the technology part? Don't worry, it's coming!

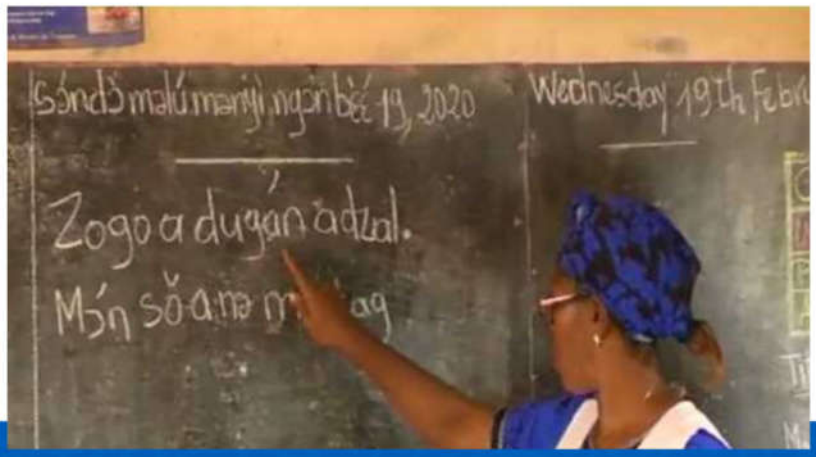
Limits of EGIDS

EGIDS measures the use of languages in social situations and includes written and spoken usage as appropriate.

Photo: https://unsplash.com/photos/oTrwlvPvpVo

Writing Systems Technology

But once an orthography is developed, accepted, documented, and shared, how does it get from here…



Kindzeka, M. E. (2020, February 21). How Cameroon Plans to Save Disappearing Languages | Voice of America - English. *VOA News*. Retrieved from https://www.voanews.com/africa/how-cameroon-plans-save-disappearing-languages

SIL

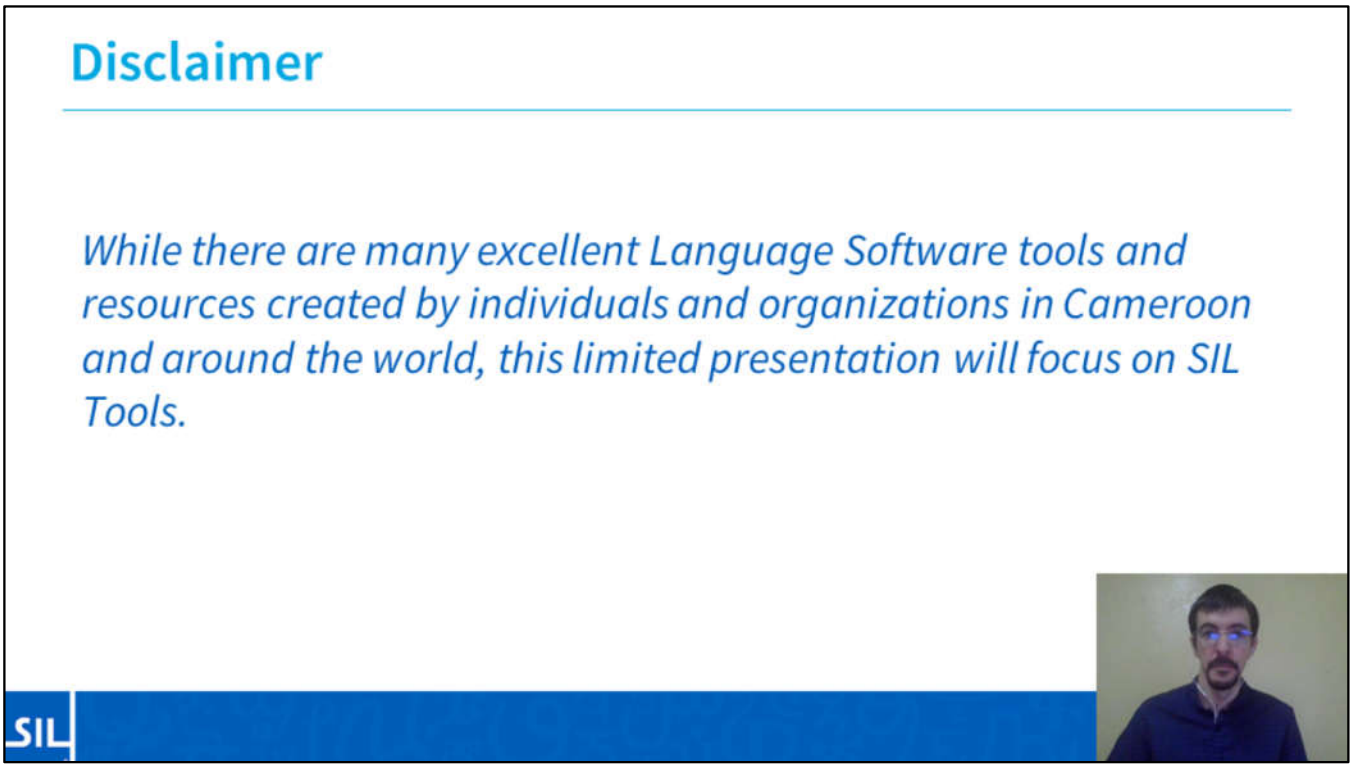
Writing Systems Technology

…to here?

https://unsplash.com/photos/z9mq3SP9uy4

SIL

Digital Language Support
Soutien numerique des langues



Before I get too far, I need to put in a little disclaimer.

Assessing Digital Language Support

(Simons and Thomas 2019)

- They began by identifying six categories of digital language support:
  - Encoding support (e.g. keyboards, fonts)
  - Localized user interface (e.g. OS, browsers, messaging)
  - Surface-level processing (e.g. spell-checking, completion)
  - Meaning-level processing (e.g. machine translation)
  - Speech processing (e.g. speech-to-text, text-to-speech)
  - Virtual assistance (e.g. Siri, Alexa)

SIL https://scholarspace.manoa.hawaii.edu/handle/10125/44838

What are some indicators of digital language support? Simons and Thomas sought to answer this question to add an index of Digital Language support.

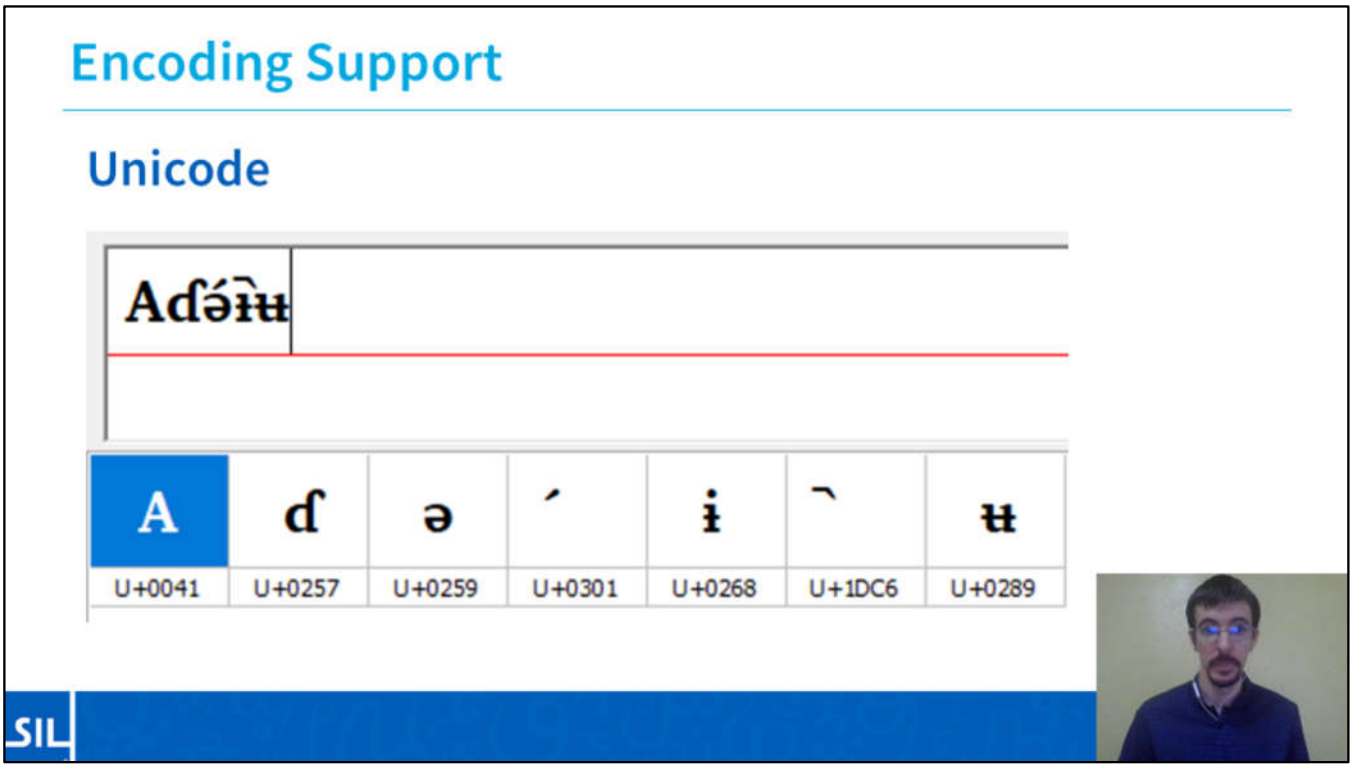
Assessing Digital Language Support

(Simons and Thomas 2019)

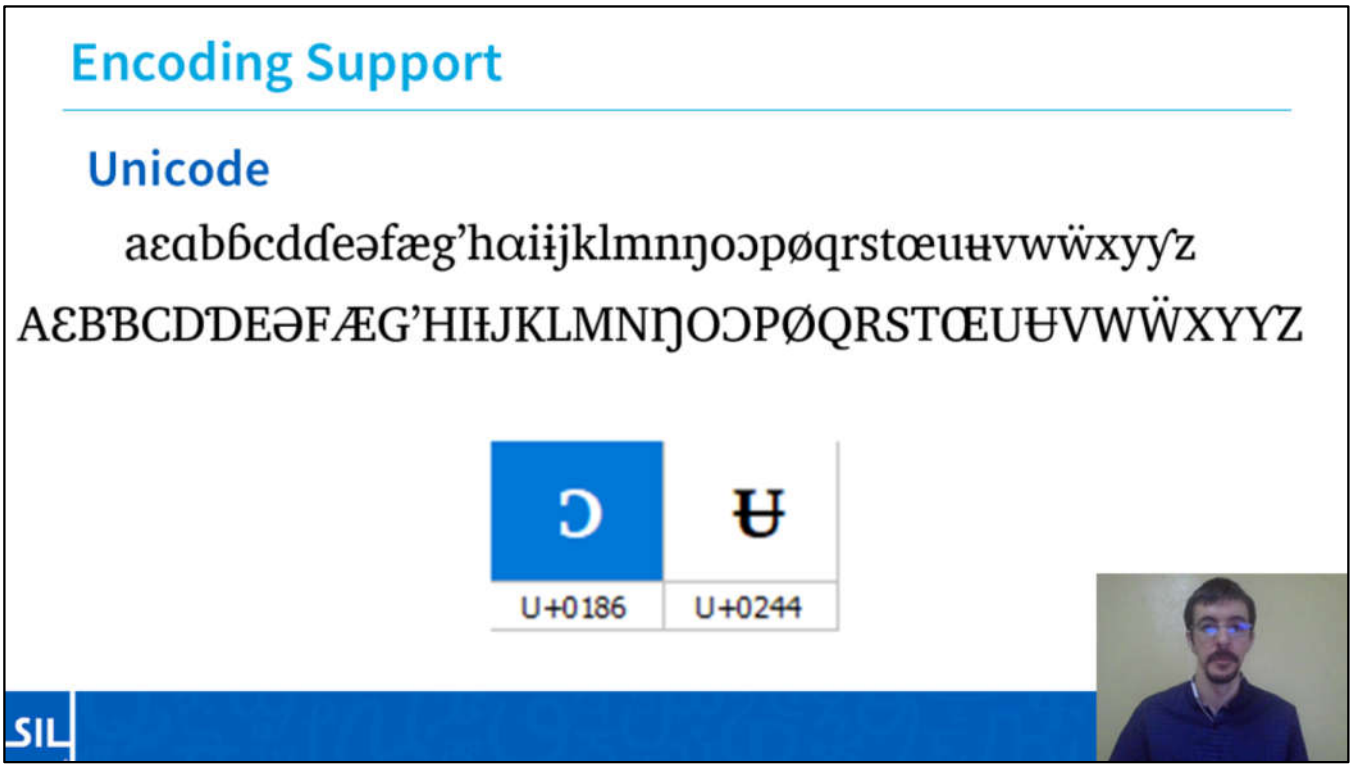
- They began by identifying six categories of digital language support:
  - **Encoding support (e.g. keyboards, fonts)**
  - Localized user interface (e.g. OS, browsers, messaging)
  - **Surface-level processing (e.g. spell-checking, completion)**
  - Meaning-level processing (e.g. machine translation)
  - Speech processing (e.g. speech-to-text, text-to-speech)
  - Virtual assistance (e.g. Siri, Alexa)

SIL https://scholarspace.manoa.hawaii.edu/handle/10125/44838

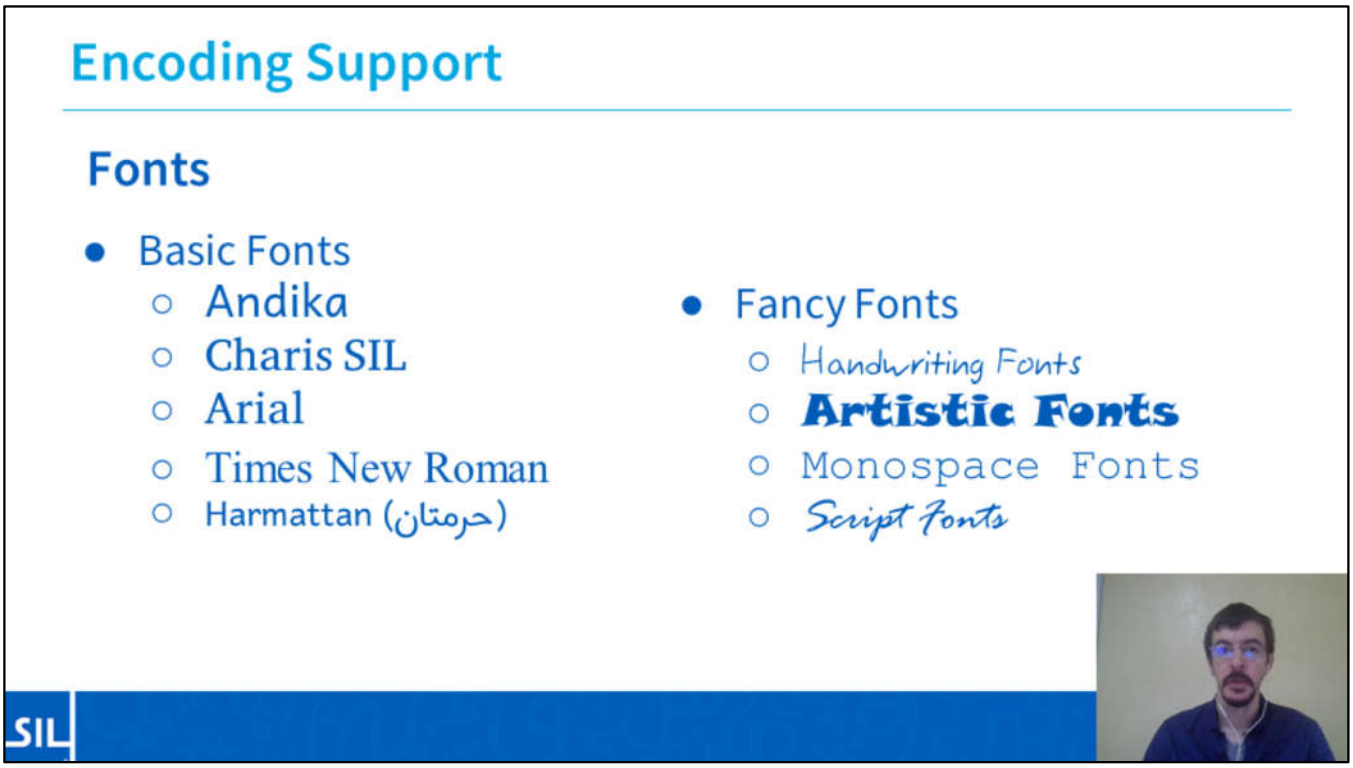
And we will discuss two of these today. Encoding Support (for example keyboards and fonts), and Surface level processing (for example, spell-checking or autocompletion)



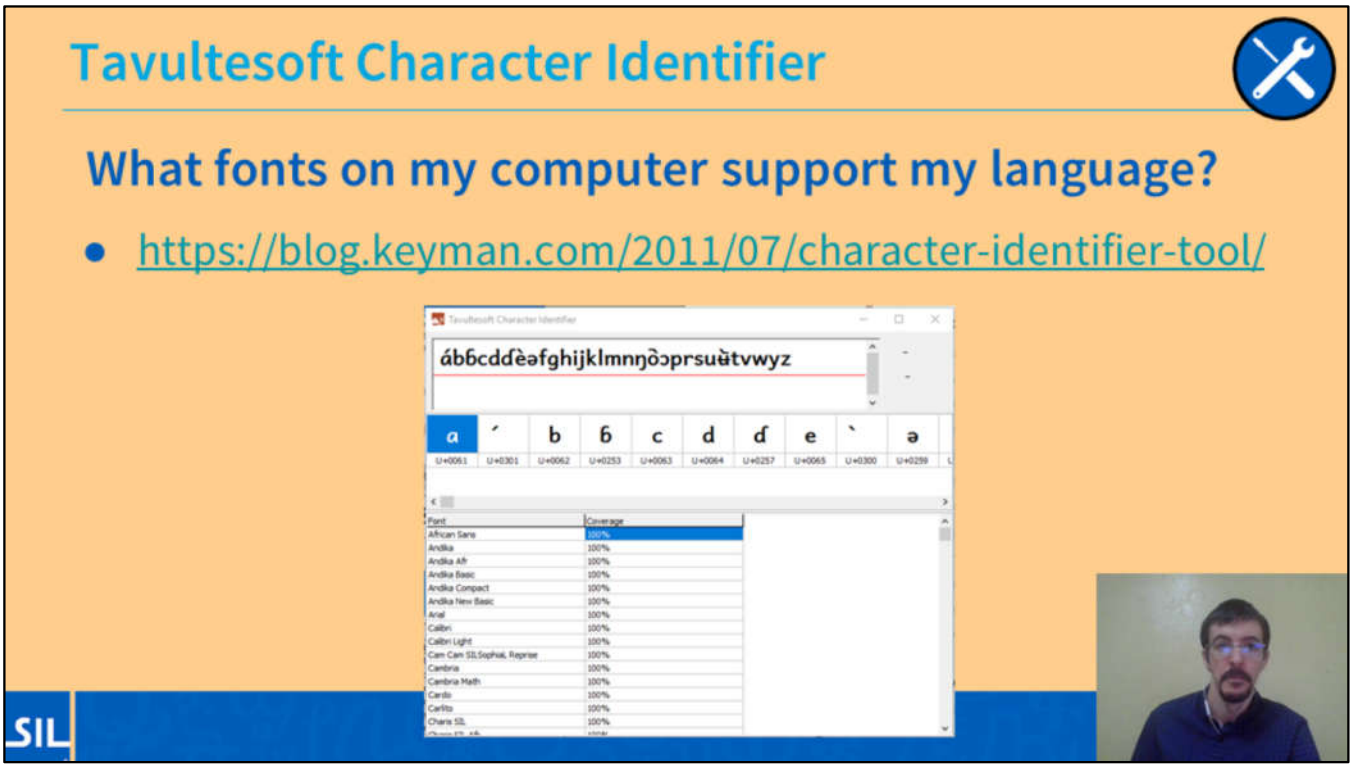
When we talk about Encoding support, we need to talk about Unicode. The Unicode Consortium takes each unique character from the world's languages and assigns it a static number to avoid confusion.
[Show a couple of numbers]
Most of the time, computers handle these hidden numbers properly without you having to worry about them.
If a new character or variation is being actively used by a community, partners can present this as a request to Unicode.



This is the General Alphabet of Cameroon. All of these characters have been added to Unicode, and in my understanding, some of the last ones added, in 2007, were capital Ɔ (open o) and capital Ʉ (barred u).



Typically, within a few years, SIL and some of the base Microsoft and open-source fonts will support most of the characters added in Unicode. Unfortunately, you will find various “fancy” fonts that may not support all of the characters in your language. Some of these “fancy” fonts may be handwriting fonts, Artistic fonts, monospace fonts, and script fonts.



What if you want to know which fonts on your computer can display all of the characters of a particular language? You can use the Tavultesoft Character Identifier, and it will show not-only the precise identity of each character you typed, but also a sorted list of fonts that support all of the needed characters.

Encoding Support

Keyboards

- Add-On Keyboards
  - Keyman (Mobile and Desktop, SIL)
  - Gboard (Mobile, Google)
  - Swype (Mobile, Microsoft)
  - MSKLC (Desktop, Microsoft)
- OS-Native Support
  - Windows
  - macOS
  - Android
  - iOS
  - Linux (XKB, SCIM, iBUS)

SIL

Where are we with Encoding Support?

Encoding

- All Cameroon Characters have been in Unicode since 2007
  - Font-makers are catching up.

SIL

Where are we with Encoding Support?

Fonts

- SIL Fonts (Charis, Doulos, Andika) have supported those characters and displayed them properly for many years.
- Microsoft fonts such as Calibri, Arial, and Times New Roman, contain all of the needed characters since about 2009, but don't always place diacritics properly (especially in italics).
- Google Fonts, such as Noto or Roboto, do a great job with Cameroonian characters since 2014 (Android 5.0).

SIL

Where are we with Encoding Support?

Unified Keyboards

- SIL's Unified Cameroon Keyboards are available for Windows, Linux, Android, Mac and iOS.
  - *You should be able to type any language that follows the GACL.*
- Other tools like the African or PanAfrican Keyboard support languages beyond Cameroon, but they can be a bit overwhelming.

Unified Keyboards are ideal for users of many languages.

SIL

Where are we with Encoding Support?

Language-Specific Keyboards

- Google's Gboard and Microsoft's Swype for mobiles are starting to release individual keyboards for some well-documented Cameroonian languages.
- We'd love to work with communities who wish to develop a simplified custom layout for their language.

SIL

Surface-Level Processing

Simple linguistic tools that work at the word level.

- Spell-checking
  - Hunspell

  Works in LibreOffice and other tools
- Autocompletion and Autocorrection
  - Keyman Mobile Keyboards

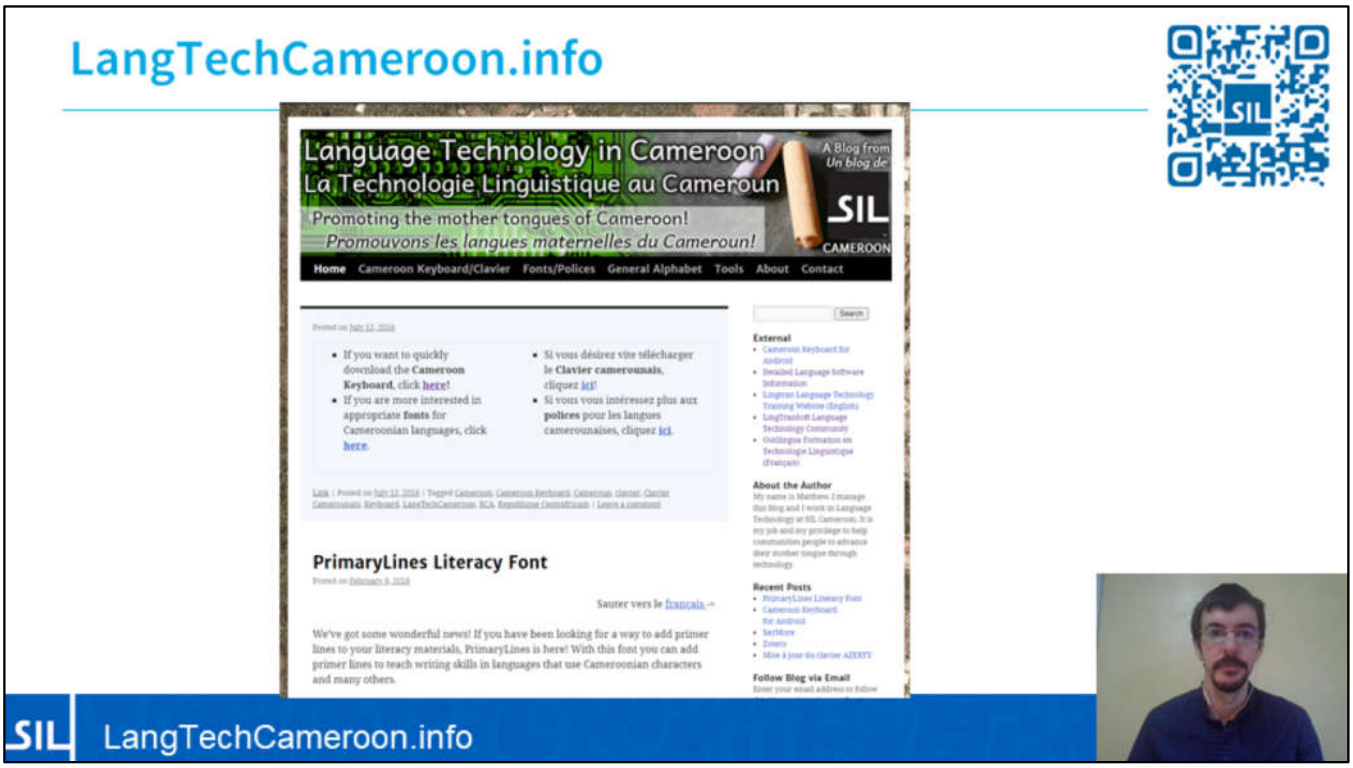
SIL

Digital Language Support is Measurable

All that is required is to think of a tool, and then read the documentation, or just try it out.

SIL

In this section, I will introduce you to several Foundational Technologies.

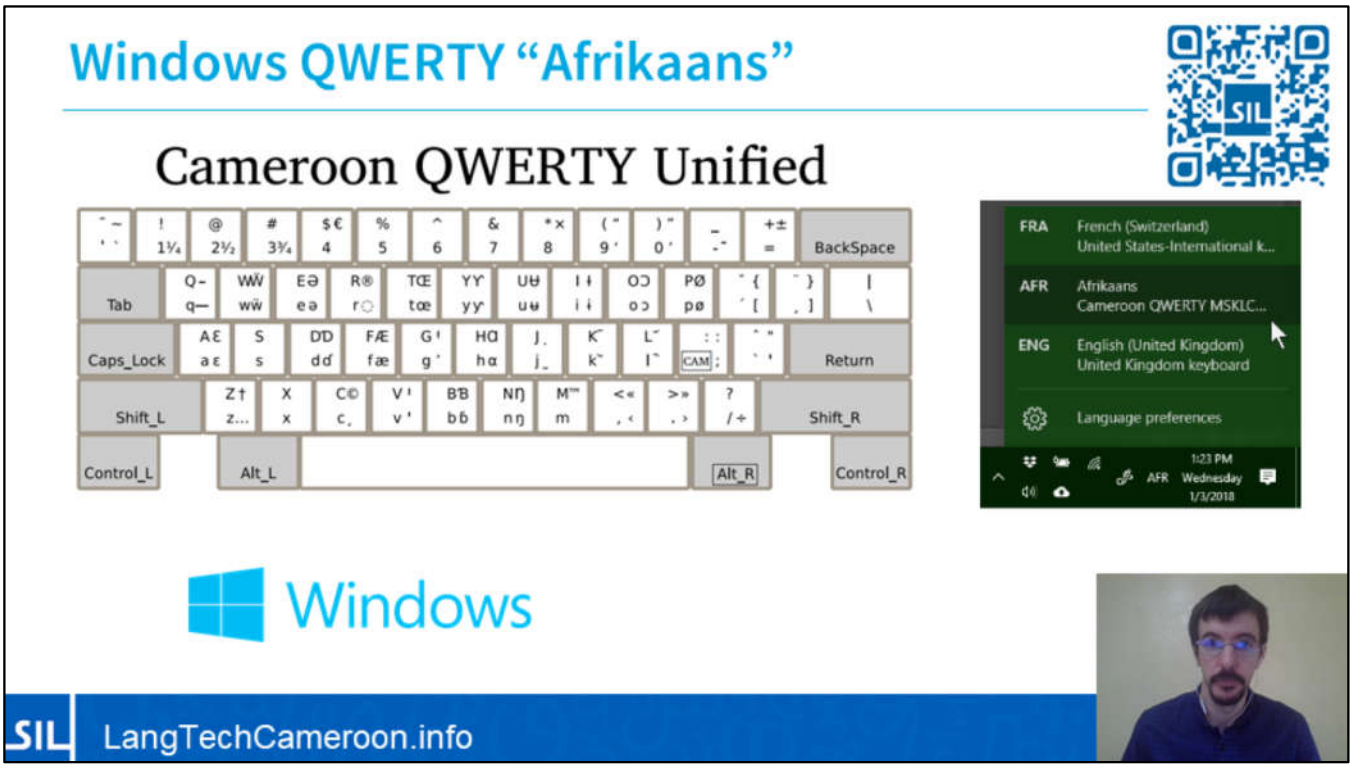


We start out with this site, http://LangTechCameroon.info This site is the home of Language Technology at SIL Cameroon, and you can find resources about keyboards, fonts, and Language Technology tools which might be interesting, and also presentations like this one.

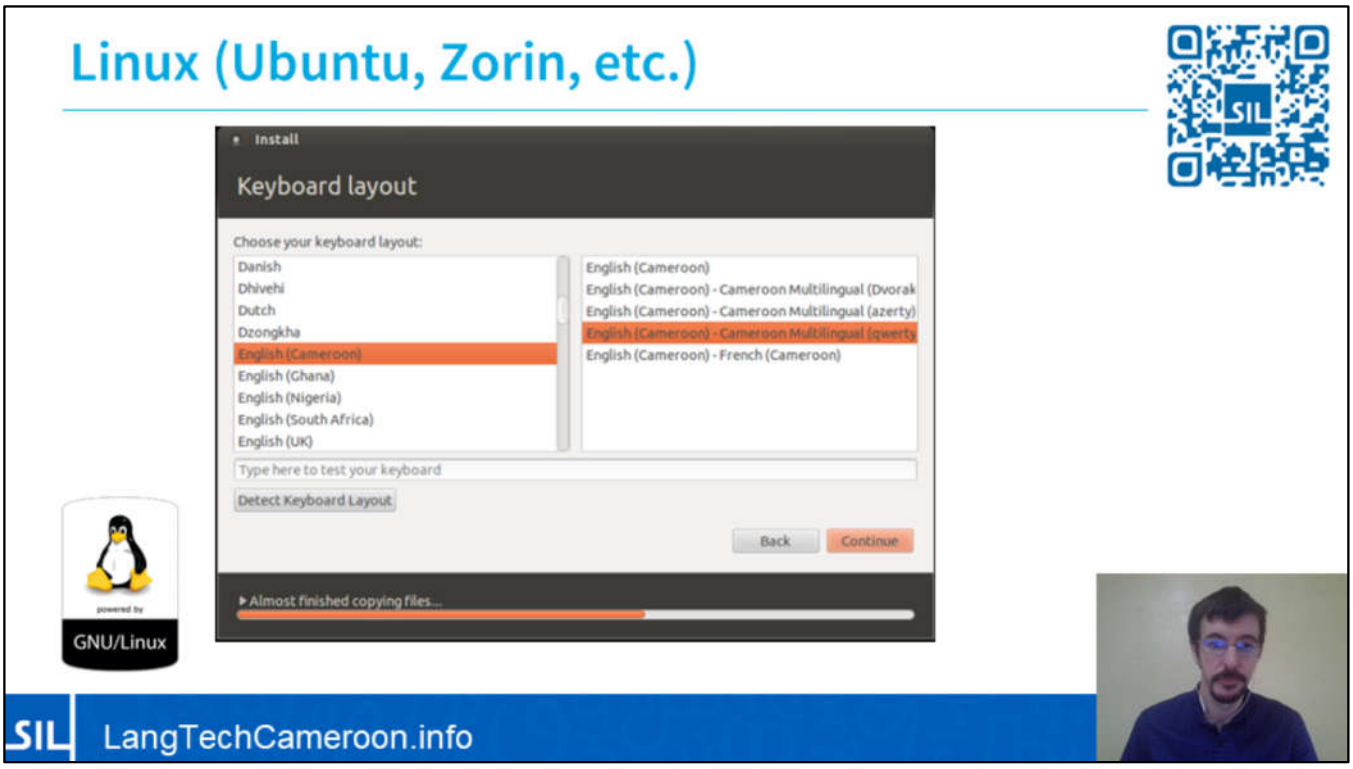
The site, http://LangTechCameroon.info, is also the home of the Cameroon Keyboard.

Here is the site, where you can see that this site helps you tc
and install the Cameroon Keyboard on your machine or devi

It may be helpful to explain that there are two versions of the Cameroon keyboard for desktop. This is the AZERTY version, that follows the French Keyboard, AZERTY. This one is known as “Galicien”, because we did not have the opportunity to add languages. The Azerty keyboard uses the exclamation point key to access most of the special characters.

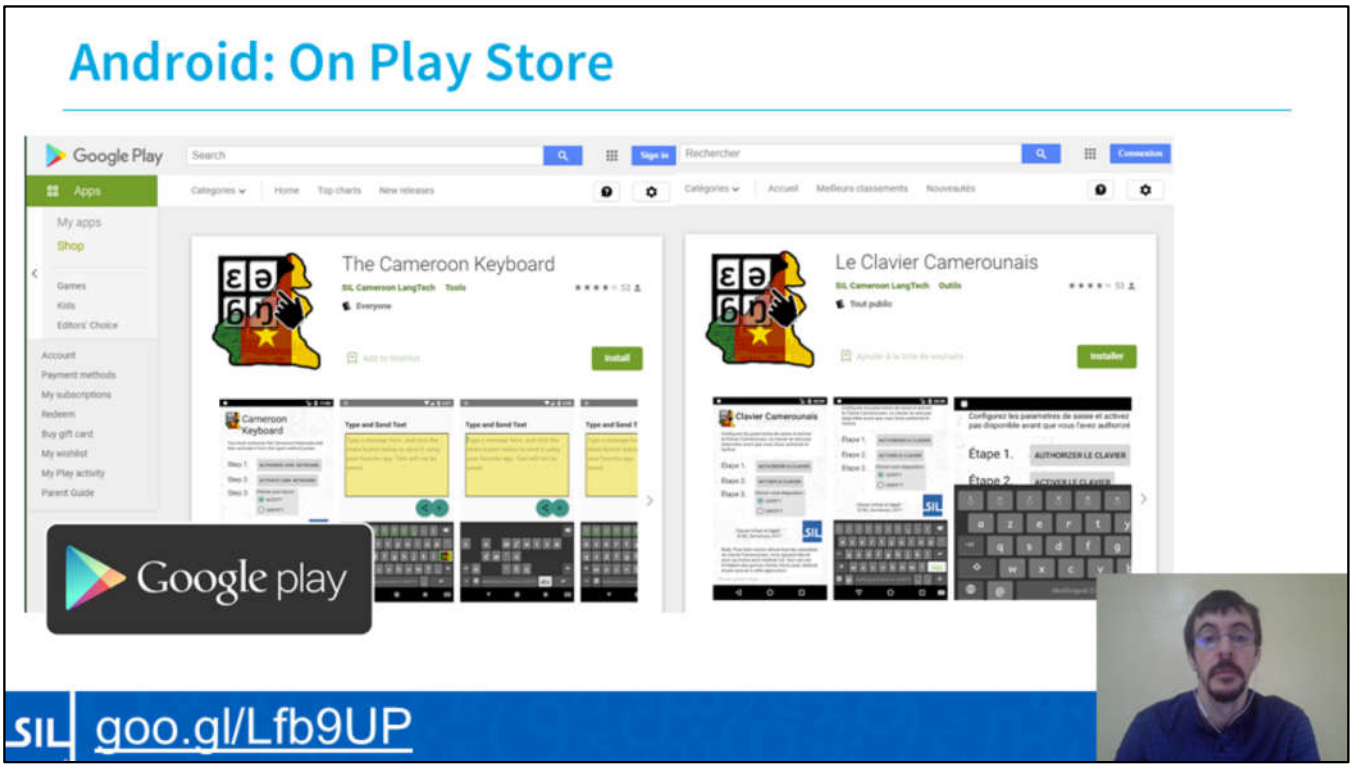


Also, we have the QWERTY Version that follows the English keyboard of the US and UK. This one is known as Afrikaans, for the same reason as before. The Qwerty Keyboard uses the semicolon key to access most special characters. These versions are less than 2 megabytes each, and can be found on http://LangTechCameroon.info

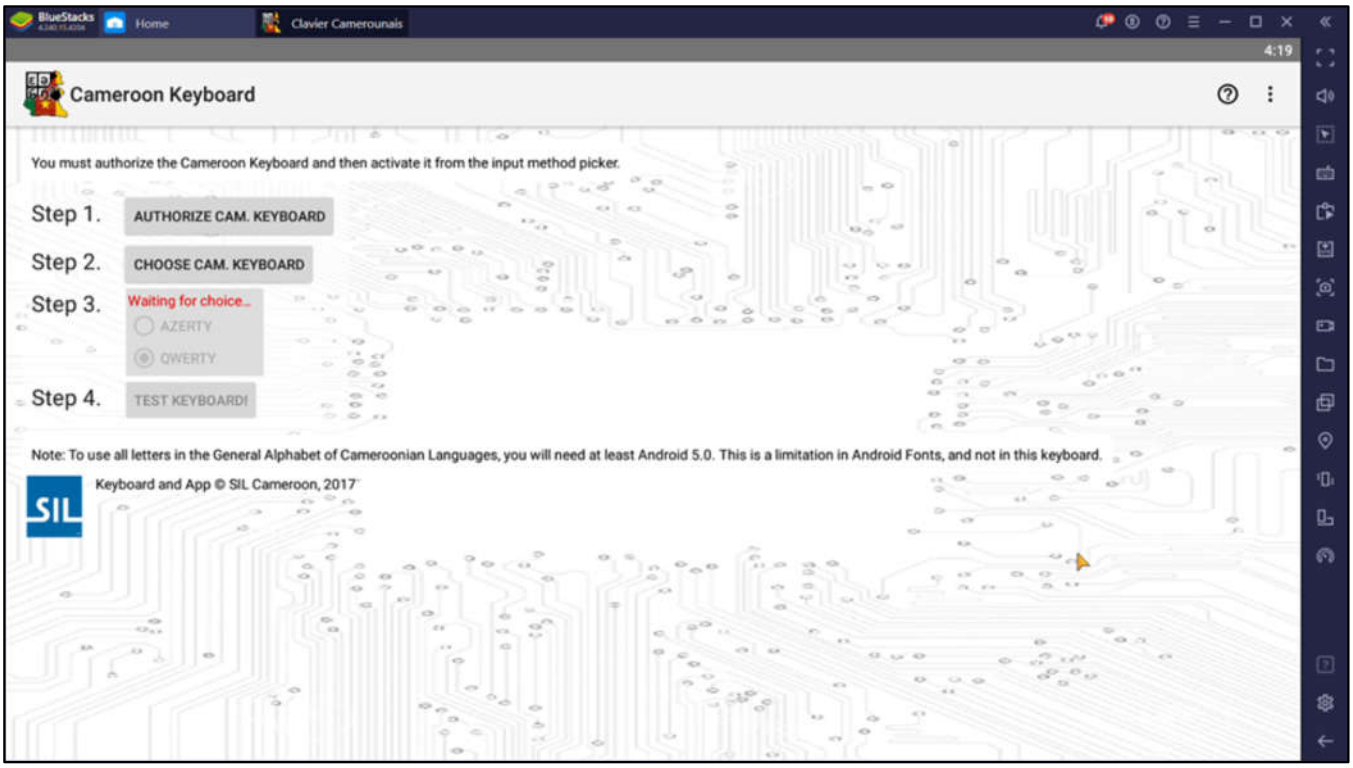


You can install the Cameroon keyboard on other operating systems. For Linux, if one installs Ubuntu or Zorin, and most “flavors” of Linux, you can choose the Cameroon Keyboard from the settings without having to install other applications on top.

Once installed, it will be available for typing in all of your programs.



For Android, you can install the Cameroon keyboard from the Play Store on tablets and telephones that have at least Android 5, called Marshmallow. Once installed, you can use the keyboard for text messages (SMS), to write emails, to post on Facebook, to chat on WhatsApp, or in any other application. You can write and work in your language on the telephone!
Just open up the Play Store, and search for Cameroon Keyboard, or Le Clavier Camerounais. When you see the Icon with the map and letters, you have found it!


BlueStacks
Home
Clavier Camerounais
4:19
Cameroon Keyboard
You must authorize the Cameroon Keyboard and then activate it from the input method picker.
Step 1.
AUTHORIZE CAM. KEYBOARD
Step 2.
CHOOSE CAM. KEYBOARD
Step 3.
Waiting for choice...
AZERTY
QWERTY
Step 4.
TEST KEYBOARD!
Note: To use all letters in the General Alphabet of Cameroonian Languages, you will need at least Android 5.0. This is a limitation in Android Fonts, and not in this keyboard.
SIL
Keyboard and App © SIL Cameroon, 2017

Note

The fact that the last Cameroon Characters were added in 2014 is quite important, as phones released before Android 5 (Lollipop) probably use old fonts that don't support all Cameroon characters.

Even if you could type all the characters, in Facebook, for example, sometimes they might not display at all.

This is why SIL Cameroon's Cameroon Keyboard requires Android 5.0.

SIL

You can use the Cameroon Keyboard on iPhones, iPads, and Macs, but until someone offers me a Mac, I cannot release a dedicated Cameroon Keyboard App. You can install Keyman (www.keyman.com), an application made by SIL International, and the Cameroon Keyboards, both AZERTY and QWERTY, are available inside.

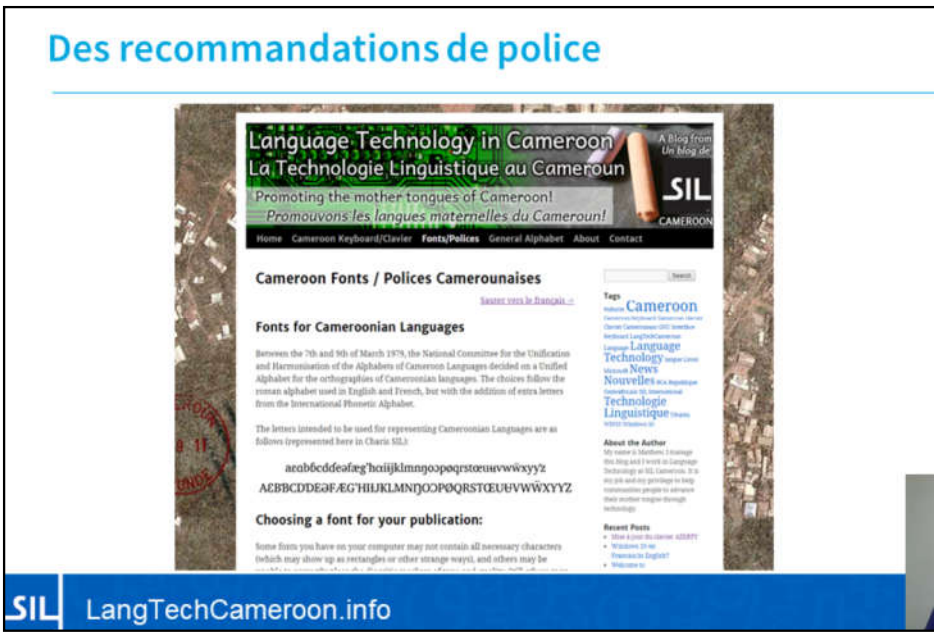
IPA Keyboards / Claviers API

- Windows/Mac/Linux:
  - https://scripts.sil.org/uniipakeyboard
- Android?
  - https://play.google.com/store/apps/details?id=com.edwardgreve.ipakeyboard

It's good to know that there are other keyboards available. If you need a keyboard for the IPS, here are some links.

Best Fonts for Cameroon Languages

Des recommandations de police



On the site, LangTechCameroon.info, there is a guide fo

Once characters are added to Unicode, it takes a few years for those who develop fonts (or typefaces) to catch up. Have you ever tried to type something like this

…and gotten something like this, or even worse, like this.

Disclaimer (Again)

Just because a font contains a set of letters and symbols doesn't mean it will always show them correctly.

SIL

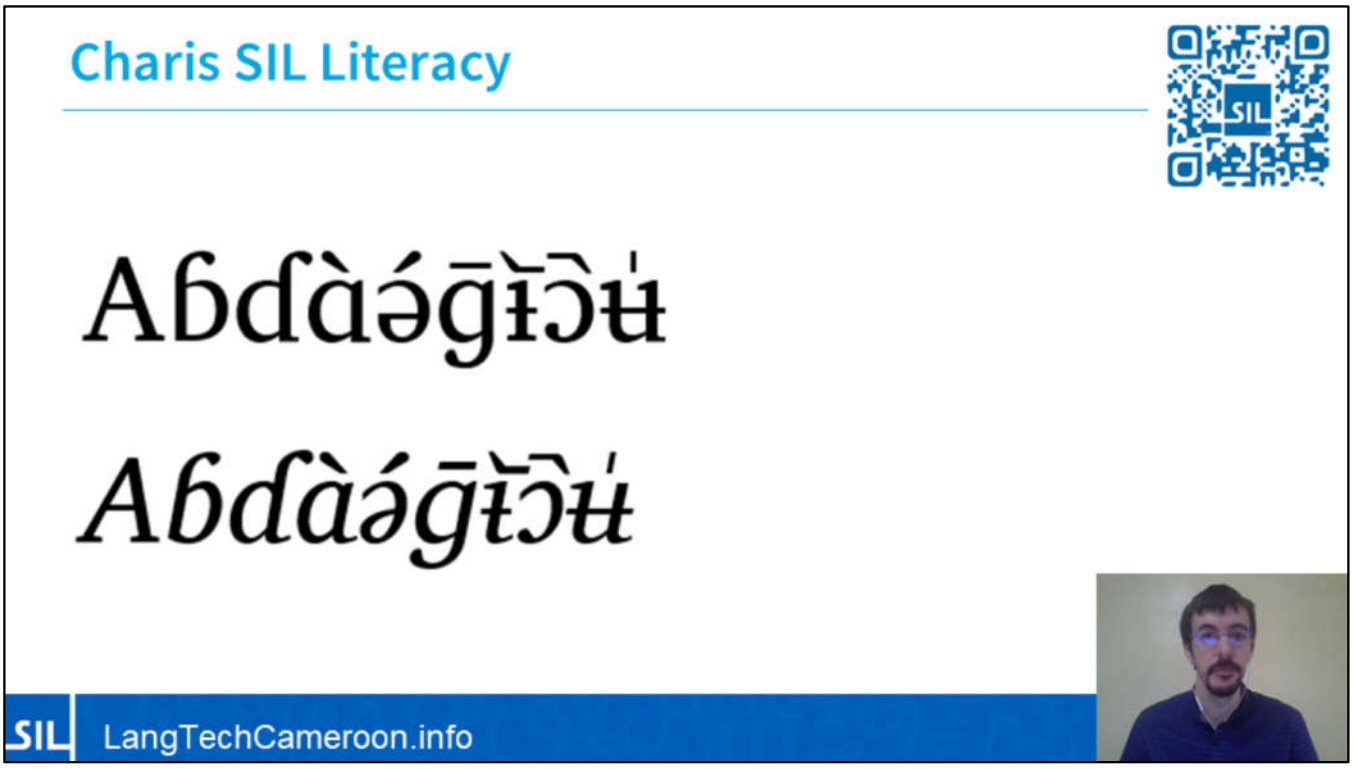
Disclaimer (Again)

Just because a font contains a set of letters and symbols doesn't mean it will always show them correctly.

The following fonts are some that should contain all of the charecters you need, and display them properly…avoiding some of the issues shown here.

First, there is Charis SIL. This is a serif font that is similar to Times New Roman, but it has a full set of characters for many languages that use Roman, or Latin, Script.

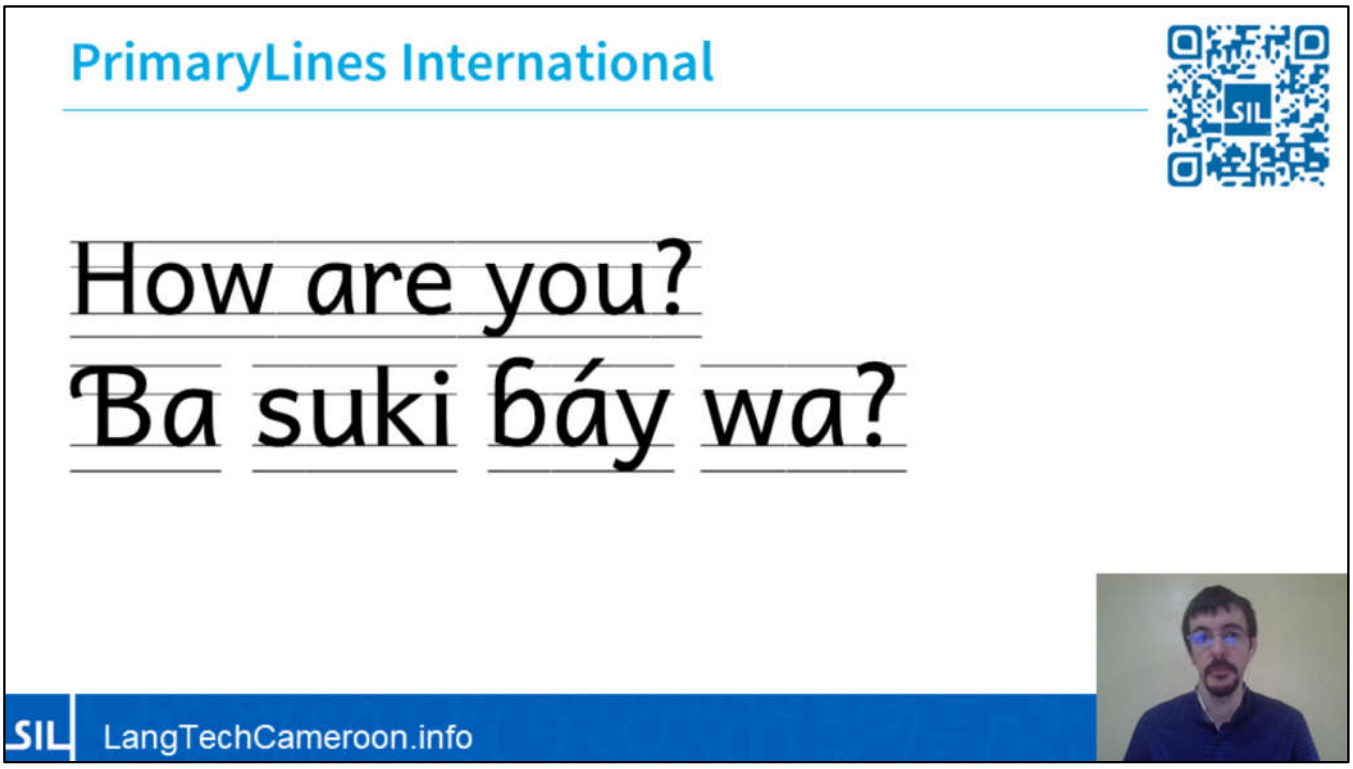
Doulos is often used for the phonetic alphabet because there are signs, like the diacritics that are easier to see than with Charis SIL.



For those who work in Literacy, it can be useful to have the simpler literacy forms of several characters. There are Literacy versions of several fonts. In Charis SIL Literacy, we see that we have the simplified “a” instead of the double-story a. like we write by hand.

Doulos SIL also has the simple a and the simplified hooked “g”, as we write by hand.

For those that are learning to write, we also have Andika. Andika is a simplified font without the hooks of serif fonts, and we see simple forms that could almost be written with a pen. We see the same simple characters.



For those that are working on Primers, the font PrimaryLines is Developed here in Cameroon that gives you the option to display a font that already has lines showing the extents of each letter.

Evaluating Digital Language Support

Basaa Digital Language Support

Encoding Support

- Entry in the Common Locale Data Repository
- Entry in the SIL Locale Data Repository
- Keyman and Gboard Keyboards

SIL

Basaa Digital Language Support

Localized Content

- Indigenous Blogs
- Online Dictionaries
- Health Resources
- Scripture Publications
- More

SIL

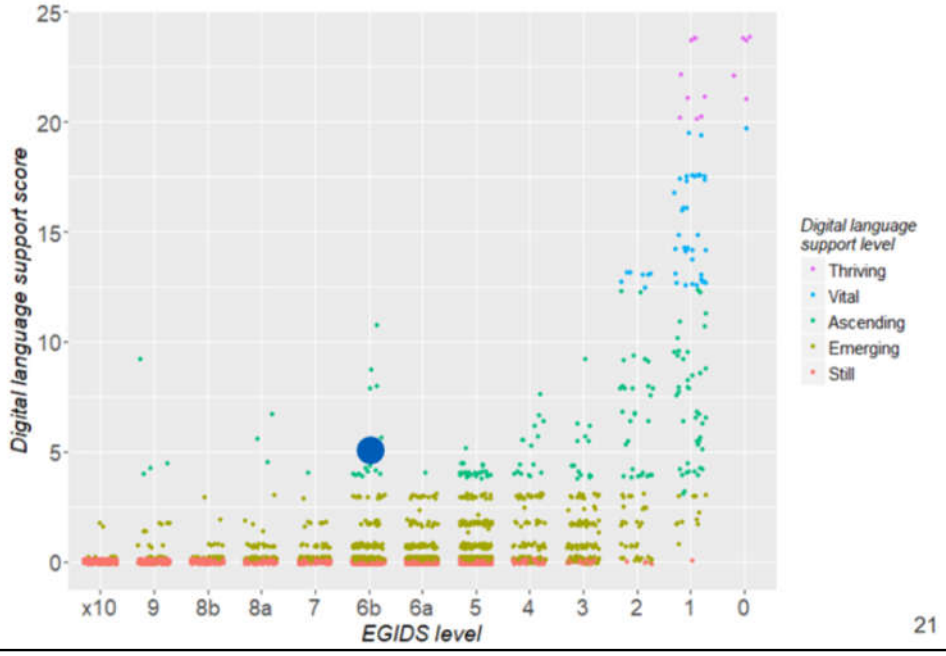
Basaa Digital Language Support

Surface-Level Language Processing

- Gboard Spelling Dictionary

SIL

DLS score by EGIDS level



Review: Digital Language Support

So now we've established that Cameroonian Languages (following the GACL) have basic digital support.

- All Cameroon Characters are in Unicode
- Most base fonts, SIL and non-SIL, contain these characters, and many of them align the diacritics correctly.
- Some "fancy" fonts will contain most Cam. Characters.
- Virtual Keyboards are available on Windows, Linux, MacOS, Android and iOS.
- Autocorrection is now possible on Mobile Keyboards with a basic wordlist.

SIL

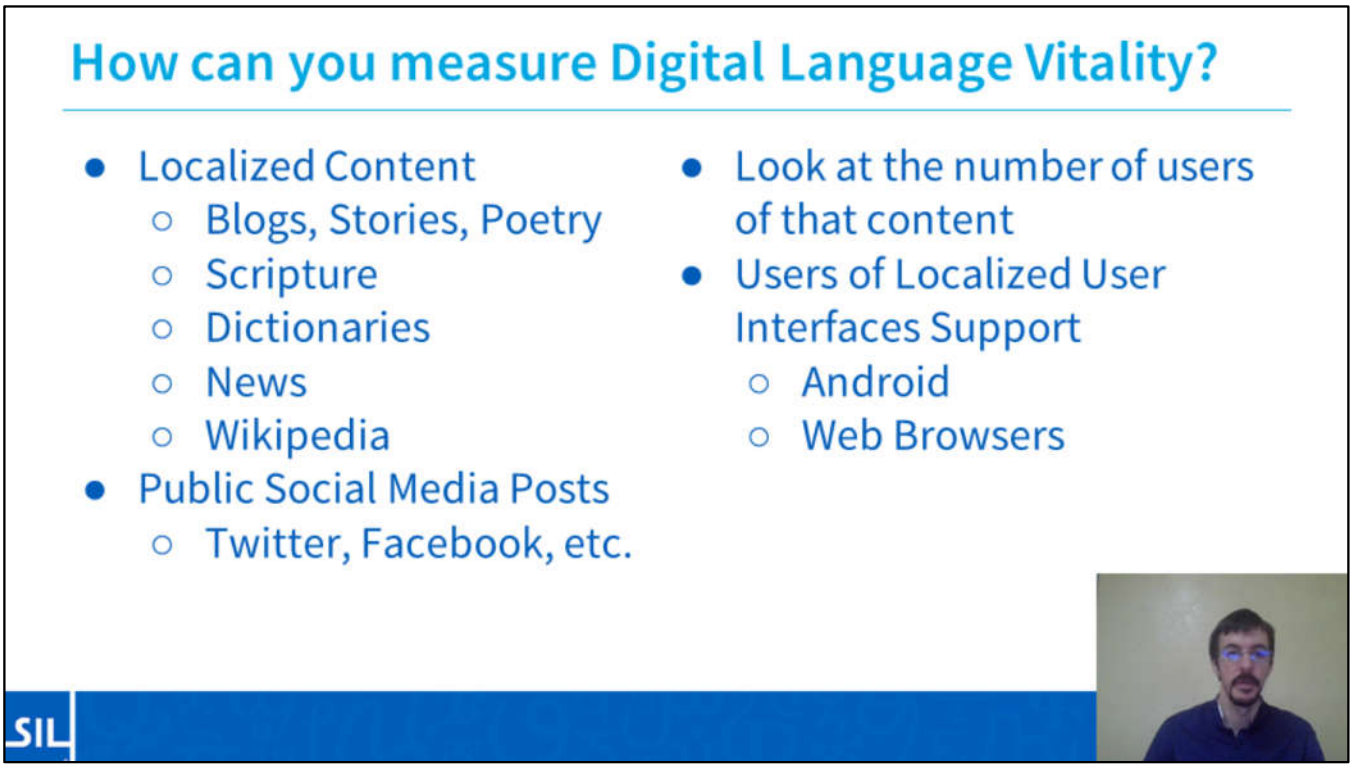
As you know, larger cities tend to benefit first from internet technologies. While there is always room for improvement in speed, cost, and reliability, the telephone networks are reaching further and further into Cameroon's distant villages. More and more people are finding ways to access the internet, through mobile technologies, office and home connections, and at cyber-cafes.

Review: Digital Language Support

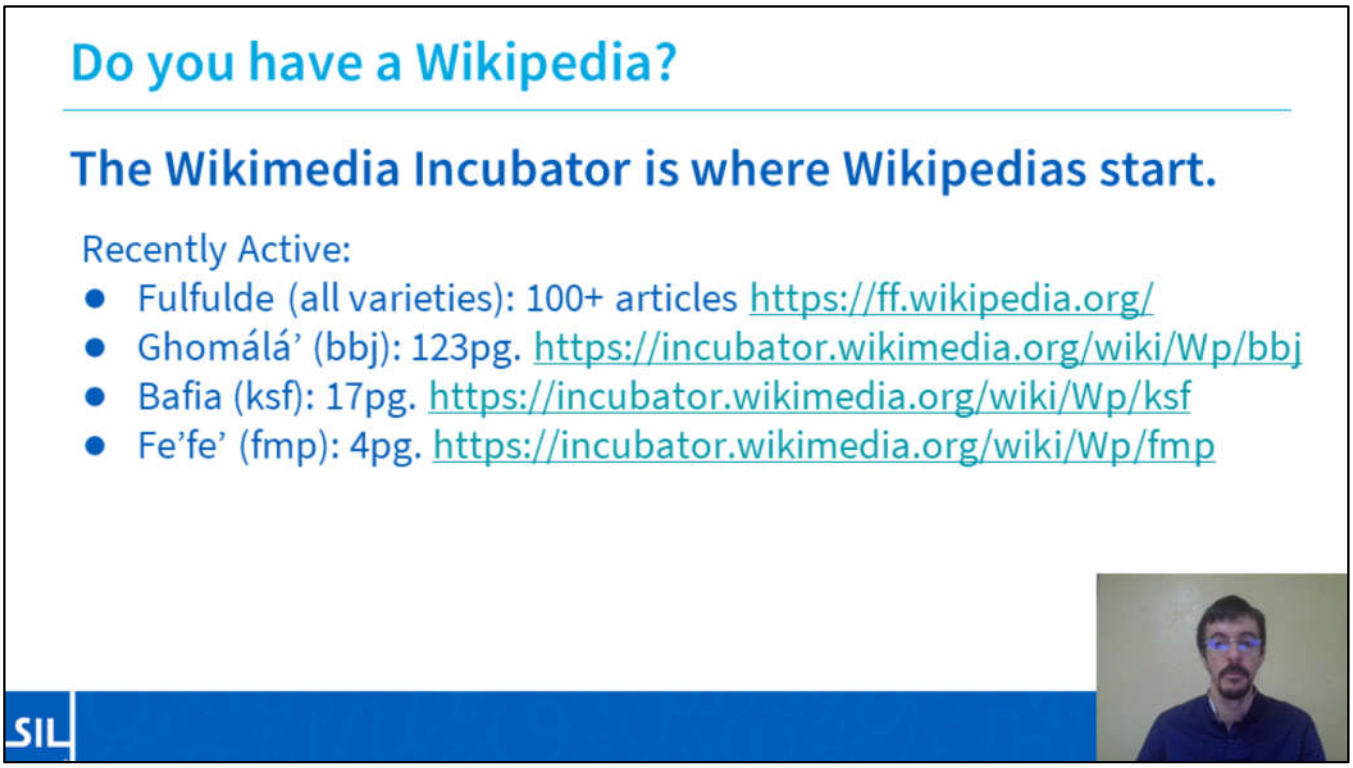
So…what is next?

Digital Language Vitality

How are languages used in the digital world?



First, Simons and Thompson looked for minority languages where other languages are found.



The Wikimedia Incubator is a place to start a Wikipedia. Once a community reaches over 100 pages of quality content, they can petition to have their project moved to Wikipedia.

Do you have a Wikipedia

Wikimedia Incubator (as of Jan 2021)

Inactive:

- Bafia (ksf): 17pg.
- Ewondo (ewo): 7pg.
- Basaa (bas): 4pg.
- Bulu (bum): 4pg.
- Medumba (byv): 4pg.
- Akoose (bss): 2pg.
- Ngomba (jgo): 3pg.
- Aghem (agq) 2pg.
- Duala (dua): 2pg.
- Kwasio (nmg): 2pg.
- Mundang (mua): 1pg.
- Yangben (yav): 1pg.

SIL

Wp/agq (2 P)
Wp/bas (4 P)
Wp/bbj (1 C, 123 P)
Wp/bss (2 P)
Wp/bum (4 P)
Wp/byv (4 P)
Wt/dua (2 P)
Wp/ewo (7 P)
Wb/fmp (4 P)
Wp/fmp (4 P)
Wp/jgo (3 P)
Wp/ksf (17 P)
Wp/nmg (2 P)
Wp/yav (1 P)

How can you measure Digital Language Vitality?

- By searching the language name, they could find some local content, but this content is spread across many places.
- Smaller social networks, blogs, and sites, are everywhere.

- It is likely that most communications in each language are interpersonal, direct messages between individuals sharing a language or language family.

How is Your Community Communicating online?

- When communicating digitally with a friend that speaks the same language, are you more likely to communicate using French or English?
- Are people in your community encouraged to write, publish, and share printed and online works in the language?
- If you wanted to read a book in your language, where would you go?
- Is your language community represented on sites where other languages are spoken?

SIL

These are questions that NGOs like SIL can't answer for you.

How is Your Community Communicating online?

- Are there digital literacy classes and discussion groups?
- Do you have community websites? How are they promoted?
- Are there community centers where people can connect to the Internet and share in communication?
- Where else would you like to see your languages?

SIL

Putting this into Perspective

- SIL Cameroon continues to improve support for using Cameroon Languages in basic technologies on major operating systems.
- SIL maintains tools to make publishing simpler, such as Bloom.
- But while SIL has the capacity to assist you when you need an advocate or partner to make things possible, the authors, listeners, readers, and users are in your communities.

SIL

Review

- The reality is that the future Digital Vitality of your language, or languages, starts with you.
  - Local members and diaspora members
  - Speakers and listeners
  - Authors and programmers
  - Elders and youths

The New Model

- Traditional Language Vitality is driven by the adults of a community, and how much value they put on passing on their language and culture to the next generations.
- In contrast, Digital Language Vitality may be in the hands of the youths, who have the opportunity and the obligation, to preserve the cultural history of their parents and to develop new forms of digital expression.

SIL

Traditional Language Vitality is driven by the adults of a community, and how much value they put on passing on their language to the next generations. In contrast, Digital Language Vitality may be in the hands of the youths (the Android Generation), who have the opportunity, and the obligation, to preserve the cultural history of their parents and to develop new forms of digital expression.

What part will you play in the Digital Language Vitality of your community?