Recording voices and local languages with Lingualibre

Hugo Lopez
hugo.lopez@univ-toulouse.fr
Hugo Lopez

- Elearning and language professional
- Open education resources
- Wikimédian in University
- Occitan, not French.
Outline

- Language diversity
- LinguaLibre & objectives
- Demo of the tool (10mins)
- Current progresses, limits & biases
- Q&A
Languages diversity : dimensions

Document languages diversity and voices.

- Languages
- Accents
- Voices
- Genders
The diversity of our languages, their words, expressions, voices, are poorly documented and accessible. We want to record, share and make visible those expressions at large scale, in an easy and quick fashion (800 audio/hour).
LINGUALIBRE
What is Lingua Libre

The case of Alsacian!

- Wikimedia's open source recording tool
- ...to document Alsacian
- For language e-learning services
Oral languages' learning chain

Learners ← Speakers

e-services

Data

Lingualibre Studio

Learning  Audio & text  Sharing
DEMONSTRATION
(10MINS)
Lingua Libre: audio recording studio

Record a voice

[Image of a webpage showing a recording interface with a word list and recording options]

Page: Lingualibre.org Recording Studio
Lingua Libre: Apps

List: Lingua Libre Apps. Future: Moedict.tw?
PROGRESSES, LIMITS & BIASES

After 5 year and 900,000+ recordings, we would like to share past progresses, current analysis and future actions.
Languages typology & specifics

<table>
<thead>
<tr>
<th>Large</th>
<th>Medium</th>
<th>Minorities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resourced</td>
<td>Low resources</td>
<td>Written?</td>
</tr>
</tbody>
</table>

[Images of people working on computers]
## Languages typology & specifics

<table>
<thead>
<tr>
<th>Large</th>
<th>Medium</th>
<th>Minorities</th>
<th>Written?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resourced</td>
<td>Low resources</td>
<td>Written</td>
<td>Written?</td>
</tr>
</tbody>
</table>

![Large Resourced](image1.png)

![Medium Low resources](image2.png)

![Minorities Written?](image3.png)
In numbers

Production

Languages gallery

Contributors

Lingua Libre map
In numbers

Production

Languages gallery

Reuses

2022 review

Languages gallery

Languages with over 20k recordings (10)

For a detailed list of languages, see the list on the gallery page. Most languages have vocabulary lists ready to record; at least one is ready to record.

Languages (150)

Languages with over 20k recordings (10)
Qualitative

- Per language: large vs minorities
- Per gender
- Per age
- Per per area, income, etc.
## Languages typology coverage

<table>
<thead>
<tr>
<th>Demographic[1]</th>
<th>World languages</th>
<th>Lili languages</th>
<th>Supported language's profile</th>
<th>Examples</th>
<th>Community's presence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major (&gt;30M)</td>
<td>30</td>
<td>20</td>
<td>Mostly major Western or Indian languages.</td>
<td>FRA, SPA, BEN</td>
<td>Solid: Several productive speakers. Sustained or periodic.</td>
</tr>
<tr>
<td>Large (1~30M)</td>
<td>350</td>
<td>100</td>
<td>Mostly Western languages, other notable languages</td>
<td>NLD, AFR, CAT</td>
<td>Emerging: One productive speaker, few not-retained speakers. Fragile.</td>
</tr>
<tr>
<td>Marginalized (&lt;1M)</td>
<td>6500</td>
<td>40</td>
<td>Mostly larger minorities in Western countries.</td>
<td>ATJ, BRE, EUS</td>
<td>Contact point: No productive speaker, one not-retained speaker. Below fragile.</td>
</tr>
</tbody>
</table>
Key needs

- **Mobile e-dictionaries** for local communities
- For **revitalisation**! Not documentation.
- Outreach to 6,500 local communities?
Wikimedia Movement

Best global network to support languages diversity
Q&A
# Keep in touch

<table>
<thead>
<tr>
<th>Role</th>
<th>Contacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lead, dev</td>
<td>User:Yug</td>
</tr>
<tr>
<td>Wikimedia France</td>
<td>User:Adélaïde Calais WMFr</td>
</tr>
<tr>
<td>Home</td>
<td>Lingualibre.org</td>
</tr>
<tr>
<td>Code</td>
<td>github.com/lingua-libre</td>
</tr>
</tbody>
</table>
THANK!
## Credits

The following Wikimedia Commons images have been used:

<table>
<thead>
<tr>
<th>File</th>
<th>Licence</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>File:Human_Language_Families.png</td>
<td>CC-BY-SA</td>
<td>JFDP13</td>
</tr>
<tr>
<td>File:WikiLucas00_à_l’Institut_international_pour_la_Francophonie.jpg</td>
<td>CC-BY-SA</td>
<td>WikiLucas00</td>
</tr>
<tr>
<td>File:Daramlagon_and_Maerui-sama_session_on_Bikol_Wiktionary_and_Lingua_Libre_03.jpg</td>
<td>CC-BY-SA</td>
<td>Daramlagon</td>
</tr>
<tr>
<td>File:LinguaLibre_2022_Paris_Surui_training-03.jpg</td>
<td>CC-BY-SA</td>
<td>Yug</td>
</tr>
<tr>
<td>File:Forom_des_langues,_Toulouse,_2023-01.jpg</td>
<td>CC-BY-SA</td>
<td>Yug</td>
</tr>
<tr>
<td>File:éance_Lingua_Libre_à_Cotonou_en_Mars_2021_-_Photo_17.jpg</td>
<td>CC-BY-SA</td>
<td>Fawaz.tairou</td>
</tr>
<tr>
<td>File:Lingua_Libre_Atikamekw_at_Wikimania_2017_Montreal.jpg</td>
<td>CC-BY-SA</td>
<td>Benoit Rochon</td>
</tr>
</tbody>
</table>