Challenges for language technologies in Ayapaneco

Jhonnatan Rangel, PhD

inalco — Institut national des langues et civilisations orientales

SeDyL UMR8202

Objectives

This poster aims to:

- Present challenges to implementing language technologies in Ayapaneco
- Illustrate how almost 10% of the world's languages could face these same challenges

Introduction

Ayapaneco is a critically endangered and low-resourced language. Like many other languages in this situation, it poses various fundamental challenges for the implementation of language technologies. These challenges limit the scope of its documentation, preservation, reclamation and revitalization.

Critically endangered

There are currently 577 critically endangered languages in the world, making up almost 10% of the world's 6,000+ languages [2].

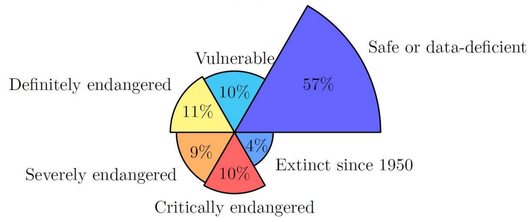

Figure 2: Vitality of the world's languages

Critically endangered languages present the following characteristics:

- Small proportion of speakers to larger community
- Remaining speakers are elders (grandparent generation)
- Inter-generational transmission interrupted for decades
- Limited use in daily language practices
- Lack of written tradition and literacy

Low-resourced

Low-resourced languages (LRL) lack the necessary annotated corpora and tools for their utilization in language technologies such as the fast growing field of Natural Language Processing. It is estimated that out of the world's 6,000+ languages, only about 20 of them could be considered high-resourced (HRL). An additional 60 languages have some language technology tools available or are medium-resourced (MRL)[1].

HRL	MRL	LRL
0.4%	1%	98.6%

Table 1: Resource distribution of world languages

Some examples include:

- HRL: English, German, French, Spanish
- MRL: Hebrew, Hindi, Czech
- LRL: Ayapaneco, Ixcateco, Kaurna

Ayapaneco

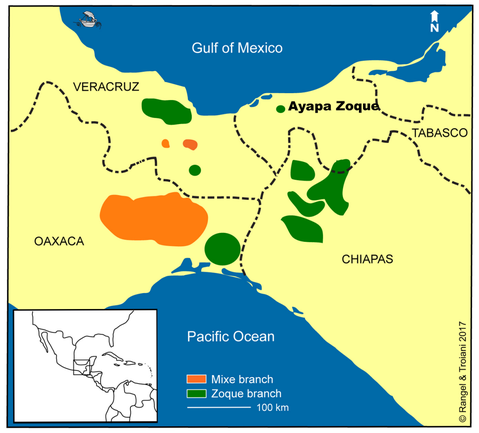

Figure 1: Geographic location

Ayapa Zoque, Ayapaneco or *numde 'oode* (autonym) belongs to the Mixe-Zoque language family and is spoken in Ayapa, a village of 5,500 inhabitants in the municipality of Jalpa de Méndez, Tabasco, Mexico [3].

- Approximately 11 elder speakers representing only 0.3% of Ayapa's total population
- Not transmitted for over 60 years
- Limited use of the language
- No written tradition, practical orthography in progress
- Under-documented & under-studied
- Recent language revitalization efforts
- Critically endangered & low-resourced

Challenges for Ayapaneco

Annotation bottleneck | Human resources | Financial resources

Capacity | Infrastructure | Linguistic

As Ayapaneco is a critically endangered, low-resourced, under-documented and under-studied language, we are confronted with multiple challenges to implementing language technologies.These can be categorized in six major groups.

1. Annotation bottleneck: 1-2 hours spent annotating for every minute of recording
2. Human resources: only 1-2 manual annotators who are not speakers; all speakers are elders with no digital literacy
3. Financial resources: no state funding for language technologies in critically endangered languages
4. Capacity: scarce computer access; written system still in progress
5. Infrastructure: limited, unreliable and expensive access to internet
6. Linguistic: limited language documentation and description; high proportion of unstructured variation [3]

References

[1] Duong, L. *Natural Language Processing forResource-Poor Languages*. PhD dissertation, University of Melbourne, 2017.

[2] Moseley, C. *Atlas of the World's Languages in Danger*, 3rd ed. UNESCO, 2010.

[3] Rangel, J. *Variations linguistiques et langue en danger. Le cas du numte oote ou zoque ayapaneco dans l'état de Tabasco, Mexique*. PhD dissertation, INALCO, 2019.

Future Perspectives

The implementation of language technologies could help us maximize the limited time we have left to engage with critically endangered languages. Given the multiple and complex challenges we face, we must explore from a theoretical and methodological point of view new multidisciplinary approaches. We must also create synergies among varied actors (academia, governments, NGOs, communities and civil society) to better address these challenges and support capacity building in the long term. Simultaneously, and more fundamentally, we must address the root causes of these challenges such as inequality, poverty and discrimination as critically endangered and low-resourced languages are spoken by minority and marginalized communities.

Contact Information

Email jhonnatan.rangelmurueta@inalco.com
Twitter @JhonnRan
Blog https://wils.hypotheses.org/
Academia /jhonnatanrangel
HAL jhonnatan-rangel
Research Gate jhonnatan_rangel
ORCID https://orcid.org/0000-0003-2493-7634