Can a robot help save an endangered language?

Maximiliano DURAN
ELLIAD , Université de Franche Conté
Besançon, France
duran_maximiliano@yahoo.com

Abstract

A robot using artificial intelligence, a comprehensive set of linguistic resources and pedagogical functionalities may help to preserve Quechua. It can help in M.T of school texts and general culture documentation into Quechua. Written documentation, is essential to keep this language alive. I have been working on such a robot, for several years. I named it Yachaj/expert. The first stage of this project has the following functions: Automatic conjugation, lexical queries of Quechua-FR-SP; elementary spelling checking; and transliteration (alpha version) of texts written in the official spelling of Cuzco, Ecuador or Bolivia to that of Ayacucho and vice-versa.

Keywords: robot, quechua, Yachaj, automatic conjugation, transliteration

Runasimipi


1. A lack of compulsory quechua schooling

The rapid erosion of the Quechua language in just two generations (Fig. 1), in Peruvian territory, confirms its endangered language status. This relative decrease in population objectively shows the danger of Quechua becoming a dead language in the next two generations. History shows that Quechua has been the victim of many injustices of all kinds that have brought it to this tragic state: historical, administrative, social, linguistic and even psychological.

In Peru, there has been no State policy of compulsory schooling in Quechua in these regions. One consequence of this is the very large and traumatic negative social impact on this population. Since it is not obligatory to speak, and eventually to write correctly in QU, parents are inhibited from transmitting and deepening the knowledge of this language in their children, which inexorably implies the loss of their character of mother-paternal tongue. The language becomes marginalized. The lexical, morpho-syntactic level of the discourse is gradually impoverished. Socially, the child, and subsequently the adult Quechua-speaker, receives from his peers a discriminatory treatment and negative assessment from the monolingual Spanish-speaking speakers, instead of his bilingual knowledge (Quechua-Spanish) being positively weighted. Another consequence is that the incentive for the natural development of the language in the different fields of culture in the society is lost. Cultural production in QU becomes almost a clandestine activity or song in this language nowadays; and to aggravate the tragedy, the few who still do, use a lexicon with many “loans” of the language of contact favoring the “quechuallano” or the "medio-lengua" (half quechua-half Spanish).

1.1 Quechua speaking population in Peru

![Figure 1: Fall of Quechua speaking population.]

1.2 Other converging factors

The other factors that makes worse this erosion are:

- a. Incoherent spelling
- b. Scarcity of general cultural texts written in QU
- c. Social discrimination and racism against quechua speakers

2. Diversification of quechua spelling

In Peru, several spellings have been established by governmental decrees for the QU: one for the Cuzco region, others for Ayacucho and Ancash, among others. In Ecuador, Bolivia and Argentina (Santiago del Estero) have also decreed other orthographies. Thus, for the same QU word like language /simi/ there are the following spellings: shimi, simi, ŝimi, čimi. It is as if the various transcripts of...
Castilian were made official orthography in each region. That would give rise to hundreds of official spellings of the same Spanish language. In Puerto Rico, for example, one would have to write /Puerto Rico/, following the local pronunciation of /puert/o; in Argentina, one would have to write officially /chó canto/, to spell the conjugated form of “to sing” at the first singular person /yo canto/; or in Chile you would have to write /si pûe/ for /si pués/. Historically, of the dozens of transcriptions proposed before the Spanish golden century of literature, the transcription of the speech of the region of Castile was adopted, and since then it is called Castilian. This diversification of official orthographies for each Quechua dialect region, rather than unifying them, tends to reinforce the socio-linguistic separations. Can we imagine Castilian written in hundreds of official spellings in the world? If this is not possible for that language, why should it be acceptable or good for Quechua? Can a Quechua expert robot help save the language? A robot, expert in Quechua, using artificial intelligence, and a comprehensive set of linguistic and pedagogical resources, can be an effective aid in the task of safeguarding the language. Such a robot can perform elementary pedagogical tasks, serve for lexicographic mono and multi-language consultation. And, to respond to the lack of written documentation in quechua, the robot can significantly help in the automatic translation of texts written in French or Spanish into Quechua. With this help, in a relatively short time, we could obtain scientific, technical, historical or literary documentation written in this language, which I think is an essential pillar to keep the Quechua language alive (of course there will then be the hard stage of reviewing these translations).

In view of the real shortage of teachers trained for teaching with Quechua as the main language of communication, and the almost non-existent Quechua teachers specializing in technical subjects, scientific, literary or other areas of universal culture, the robot that I am preparing to be an expert on Quechua, whom I have called Yachaj expert, and who is being programmed to give lessons in QU could be a valuable aid for the educational system in the Andes. The current technological progress in computing and robotics is a stimulus for the creation of Yachaj and allows us to hope that in a relatively short time it can be equipped with many functionalities, linguistic and more pedagogical resources. Artificial intelligence (with its various components such as expert systems, the technique of machine learning through artificial neural networks, voice and graphs recognition techniques) applied to QU give us the hope of being able to gestate such an expert robot. After that, the cloning of the robot Yachaj, for its implantation in schools and other centers and teaching, can be done without great difficulty.

3. The Quechua expert robot Yachaj
For more than 7 years I have been working on the development of such a robot. Several of the basic linguistic resources, such as electronic dictionaries, have been prepared since 1990. In this work, I count with the scientific council of researchers of the Laboratory of Linguistics and Informatics of the University of Zurich, of Grenoble and Besançon and some colleagues of INALCO of Paris. At the current stage of its development, Baby robot QU, is able to perform, in laboratory, the following linguistic and pedagogical functions.

3.1 Lexicography
The Yachaj robot have around twenty bilingual electronic dictionaries such as:

- Dictionary DG-SP-QU containing more than 43,000 QU-SP entries and 34,000 QU-SP ones.
- Dictionary DG FR-QU containing around 36,500 entries and DG QU-FR containing around 21,200 entries.
- Dictionary of Quechua conjugated forms (more than 2 million forms) with their respective translations FR-QU and QU-FR.
- MWU Dictionary, containing multi-word linguistic units in QU-FR and QU-SP.
- An original Quechua scientific-technical lexicon, etc.

![Figure 2: QU-SP general dictionary (38 000)](image)

![Figure 3: LVF FR-QU verbs (8 600)](image)
3.2 Morphology

Yachaj “knows” how to apply thousands of rules of QU morphology, and he knows how to apply them. It knows exhaustively all the nominal suffixes SUF_N, adjective suffixes Suf_A, and verb suffixes inter and post positional IPS and PPS, and knows how to make the grammatically valid combinations between two or more suffixes. It knows all the semantic values of about 240 language suffixes. In the following figure we show the extracts of the inflections (655) of a noun like wasi/ house and the verbal forms (7743) that Yachaj obtains by applying conjugation rules, of derivation with 1-3 suffixes.

Figure 4: Inflectional forms of wasi (1-3 suf.) (655)

Figure 5: Inflected forms of verb rimay (1-3 suf.) (7743)

4. Spelling checker

The robot contains a user interface to get help in spelling correction of texts written in official QU of Ayacucho. Work is under way on spell-checkers for the other dialectal versions.

5. Automatic inter-dialectal transliterator

Yachaj is able to carry out the inter-dialectal transliteration between the Cuzco>Ayacucho versions and vice versa, in the laboratory. Following the same technology, the programming of automatic transliteration has been initiated to help pass from an official spelling, Cuzco, Ancash, Ecuador, Santiago del Estero or Boliviano to Ayacuchano and vice versa. This system can be used to solve this writing controversy of “Into which written quechua? and allow us to overcome that divisive barrier that each idiosyncrasy claims to make prevail the quechua and its writing of ‘its’ region is correct and the rest is not. If a Cuzco way of writing is not readable to the Ayacucho-Quechua-speaking, by pressing the "transliterate" button to the Ayacuchano, he will be able to get his text in official Ayacucho spelling. And vice versa.

Here is an excerpt from the automatic transliteration of a poem by Jayme Araoz Chacón (2008) written in official Cuzco spelling into Ayacucho spelling.

<table>
<thead>
<tr>
<th>Cuzco spell</th>
<th>Ayacucho sp.</th>
<th>Castilian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rumi</td>
<td>Rumi</td>
<td>Piedra</td>
</tr>
<tr>
<td>Waklaw chimpaman</td>
<td>Waklaw chimpaman</td>
<td>La piedra que al frente lancé</td>
</tr>
<tr>
<td>Rumi chamqasqay</td>
<td>Rumi chamqasqay</td>
<td>La piedra que al frente arrojó</td>
</tr>
<tr>
<td>Waklaw chimpaman</td>
<td>Rumi wischusqay</td>
<td>Porqué ahora no aparece?</td>
</tr>
<tr>
<td>Rumi wischusqay</td>
<td>Rumi wischusqay</td>
<td>Porqué hoy ya no regresa?</td>
</tr>
<tr>
<td>Maytaq kunan</td>
<td>Maytaq kunan</td>
<td>De modo parecido</td>
</tr>
<tr>
<td>Rikurimunchu</td>
<td>Rikurimunchu</td>
<td>La mujer que quiero,</td>
</tr>
<tr>
<td>Maytaq kunan</td>
<td>Maytaq kunan</td>
<td>Del mismo modo mi amada ves cómo ya no vuelve, ves cómo</td>
</tr>
<tr>
<td>Kutiimunchu</td>
<td>Kutiimunchu</td>
<td>hoy ya no se aparece</td>
</tr>
<tr>
<td>Maytaq kunan</td>
<td>Maytaq kunan</td>
<td></td>
</tr>
<tr>
<td>rikurimunchu</td>
<td>rikurimunchu</td>
<td></td>
</tr>
</tbody>
</table>

Piedra
La piedra que al frente lancé
La piedra que al frente arrojó
Porqué ahora no aparece?
Porqué hoy ya no regresa?
De modo parecido
La mujer que quiero,
Del mismo modo mi amada ves cómo ya no vuelve, ves cómo hoy ya no se aparece
6. Conclusion

We have presented the first stage of a quechua expert robot. To build it, we are using an important number of linguistic resources obtained mainly with the aid of the linguistic platform NooJ (Silberztein 2003, 2015). We are using also some machine learning techniques. For the time being it is capable to perform a certain number of pedagogical functionalities which may be useful for elementary school learning in quechua. The first stage of this project have the following functions: Automatic conjugation, lexical queries of Quechua-FR-SP; elementary spelling checking; and an alpha version of a transliterator of texts written in official spelling of Cuzco, into Ayacucho spelling and vice-versa.

7. Bibliographical References