FORMALIZING QUECHUA NOUN INFLEXIONS

MAXIMILIANO DURAN

Abstract

This paper presents a sample of a dictionary of Quechua nouns (IQN) that contains inflectional information and will constitute the basis for a new electronic Quechua-French dictionary. We have taken a set of 1500 nouns out of a Quechua dictionary[1], to which we have applied morphological NOOJ grammars and obtained 173000 inflected forms. Evaluation of this IQN dictionary on corpus shows good results.

Generalities about Quechua

The Quechua language was the official language of the Inca civilization. It was originated in the central Andes of Peru around the first half of the first millennium of the present era.

According to the first European historians (XVI century), it was spoken across the whole territory of the Inca Empire which covered the area between the province of Pasto in the south of Colombia, all the territory of Ecuador and Peru, part of Bolivia and the north of Argentina and Chile.

Some years after the installation of the Spanish colonial system, the Quechua speaking-population started to decline. From nearly 100% of people who understood the Inca language in the region (estimated at around 12 million), nowadays only 6,280,000 do (census 2010), that is to say 13% of the population in the same region.

[1] M. Duran (2009) Diccionario Quechua-Castellano.

If the tendency shown in the following statistics is maintained in the coming decades, we should fear the worst concerning the survival of Quechua language. This is one of the reasons why UNESCO has declared it in danger of extinction in its World Languages Atlas published in 2009 in Paris.

Quechua speakers in the Andes region according to census of the last 8 years	
country	Quechua speakers
Perú	3 360 000
Bolivia	2 100 000
Ecuador	500 000
Argentina	300 000
Colombia	20 000
Chile	unknown
Total	6 280 000

How to save this language which is in danger of extinction?

In this context, I think that any effort to save this "linguistic species" which is in danger of **extinction** should be welcomed. I recognize that the most significant contribution has to be governmental legislation making education in the Quechua language from an early age mandatory in the Quechua speaking regions.

The publication of different kind of dictionaries (of words, etymological dictionaries, of toponyms, proper nouns, metaphors, sayings, guesses, synonyms, etc.) would certainly also contribute to this.

Radio and TV broadcasting in these regions should also be in Quechua.

Computer science may also help: taking advantage of current advanced computing technology, it might be possible to develop programs which help generate an automatic translation from a language, French for example, into Quechua and vice versa, based on short paragraphs or articles at the beginning.

A dictionary of Quechua noun inflexions

We try to include our present work within this rescue movement. In the first time, starting with a general Quechua-Spanish dictionary I have extracted some hundreds of Quechua nouns to build a computer program module capable of generating a dictionary of the inflections of these nouns.
In a second place I worked on the French translation of these nouns.
My first corpus contained 1500 Quechua nouns (which will be enhanced in the near future). The use of NOOJ has allowed us to obtain our first complete dictionary of inflected nouns for the 1500 nouns. We thus obtain 173 000 inflected nouns. It will be used it in our current project of automatic translation from Quechua to French.

The Quechua grammar paradigms

To build the first module of our quechua lexical NOOJ dictionary we associate to each one of its entries a paradigm that formalizes the inflection. Before, we have written up a certain number of paradigms and programmed them in NOOJ.

For instance, to form the plural of a noun in quechua we simply add to it the suffix "kuna" at the end of the entry. To formalize this rule we write the PLU paradirm as follows
PLU= <E>/N+s | kuna/N+p

Here <E> is the empty string of the concatenation, and "|" is the disjunction operator which allows us to select alternate sequences.

This program states that if we add an empty string to the lexical entry (e.g. wasi, *house*) we get the singular form of the noun (wasi) and if we add "kuna" to it we obtain the plural form (wasikuna, *the houses*).

To obtain de possessive forms of a noun corresponding to the three persons in singular, we apply the rule: we add at the end of the noun an **i**, for the first person, **iki** for the second, and **n** for the third person. Which we put in the NOOJ format and call it POSSIFS

POSSIFS = i/POS+s+1 | iki/POS+s+2 | n/POS+s+3
Which will gives us for wasi
wasii, POS+s+1, *my house*
wasiiki, POS+s+2, *your house*

wasin, POS+s+3, *his(her) house*

The cases, the negation, the possessives and their combinations

The programming of the combinations of 21 inflexion cases of Quechua (accusative, genitive, ablative, directional, etc.), the possessives, the plural, the question, the negation, allowed by the language, permitted us to generate new inflected nouns. For example the following combinations are regular in Quechua:
N+ACC+ question,
N+ orientation + ACC + question,
N+ diminutive + possessive + orientation,
N+all the preceding ones combined with the negation (Neg) , etc.

In fact, the quechua language contains complex inflection forms like
wasi-cha-iki-paq *for your little house*
which is one of th 22 forms generated by the paradigm

CHAPOSSIFSPAJ = | :CHA | :POSSIFS | : PAJ;
This includes, besides POSSIFS, the following paradigms
CHA = cha/DIM; #Dim diminutive
PAJ = paq/BEN; #BEN gets benefit Ex. wasipag *for the house*

For inflections like
wasichallaikichikpaqhinaraq : *it is something that may fit well to your house*
we have more complex paradigms like

CHAPLURCAS_1 = <E>/N+s | kuna/N+p
| :TA | :CHA | :SHA | :ablatifMANTA | :PAS | :MAN | :POSSIF | :contrasTAQ | :PASSHA | :RAJSHA | :prioriteRAJ
| :allatifMAN |:locatifPI | :gnitifPA | :instruWAN | :benefPAQ | :cmpartifQINA | :termintifKAMA
| :kausatifRAIKU | :POSPI | :POSTA | :CHU | :assertMI | :thematissQA | :CHA :POSSIF | :PURAPI

The New <Q> operator

In Quechua enunciation it is always possible to duplicate one noun or an adjective. If one noun is duplicated, the new syntagm obtained means the *abundance of what is indicated by the noun*. Ex. sacha: arbre, (tree); sachasacha: forêt (forest). The duplication of an adjective give us *the superlative of this adjective*. Ex. taksa: petit, (small); taksataksa: minuscule (very small). In order to be able to program this particular characteristic of the Quechua language, Max Silberztein has created and introduced a new operator for NOOJ, which from now on will of course helpful for any other language. This operator, now symbolized by <Q> allows us to obtain, from a noun or an adjective a new Quechua lexeme without category modification.

In our NOOJ paradigms it appears as in the following examples

DUPV = <Q>/N; # where the operator <Q> duplicates and juxtapose the nouns ending in a vowel

DUPC = <L><Q>/N; #for the nouns ending in a consonant

The dictionary of inflected Quechua nouns

We present below a sample of our dictionary of inflected nouns:

maqma,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+N+s
maqmakuna,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+N+p
maqmata,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+ACC
maqmacha,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+DIM
maqmachá,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+PRO

maqmamanta,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+ABL
maqmapas,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+INC
maqmaiku,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+POS+PEX+1
maqmaikichik,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+POS+p+2
maqmainku,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+POS+p+3
maqmataq,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+CON
maqmapaschá,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+PSL
maqmaraqchá,maqma,N+FR="jarre
grande"+FLX=CHAPLURCASVOY+PRI+PRO

This first dictionary of inflected Quechua nouns, obtained by our partial module, is not exhaustive. We continue to work on it in order to complete it.

Having obtained encouraging results for the nouns with NOOJ, we have also obtained a sample of a similar module for the inflection of the verbs. For example , the following paradigm

conjugVERBES = <E>/INF
|<B>ni/s+1|<B>nki/s+2|<B>n/s+3|<B>nchik/pin+1|<B>nkichik/p+2|
<B>nku/p+3|<B>niku/pex+1

Generates the infinitive and the conjugation of the present tense of any verb (e.g. rimay *to talk*)
rimay,rimay,V+FR="aprendre"+FLX=codePRESENT+s+1
rimani,rimay,V+FR="aprendre"+FLX=codePRESENT+s+1
rimanki,rimay,V+FR="aprendre"+FLX=codePRESENT+s+2
riman,rimay,V+FR="aprendre"+FLX=codePRESENT+s+3
rimanchik,rimay,V+FR="aprendre"+FLX=codePRESENT+pin+1
rimankichik,rimay,V+FR="aprendre"+FLX=codePRESENT+p+2
rimanku,rimay,V+FR="aprendre"+FLX=codePRESENT+p+3
rimaniku,rimay,V+FR="aprendre"+FLX=codePRESENT+pex+1

A little more complex inflections are generated by the paradigm

conjugVERBES = <E>/INF|:CHU |:progCHU |:pasCHU |:futCHU |:impeCHU |:iptiiCHU |:imanCHU |:nominITA|:GSTA;

Below a sample of the 44 conjugated forms generated for the verb **takiy** (***chanter***) *to sing*

takiychu,takiy,V+FR="chanter"+FLX=conjugVERBES+NEG
takichkanikuchu,takiy,V+FR="chanter"+FLX=conjugVERBES+pex+1+NEG
takichkankuchu,takiy,V+FR="chanter"+FLX=conjugVERBES+p+3+NEG
takirqanikuchu,takiy,V+FR="chanter"+FLX=conjugVERBES+PP+pex+1+NEG
takirqanchikchu,takiy,V+FR="chanter"+FLX=conjugVERBES+PP+pin+1+NEG
takirqanchu,takiy,V+FR="chanter"+FLX=conjugVERBES+PP+s+3+NEG

We have added to these, more than 50 paradigms, some of which are very complex, and we applied them to a set of 2000 quechua verbs and generated a dictionary of 680 000 conjugated forms.

Color matching

We have applied the « Color matching sequences in text », within NOOJ's Analyse Linguistique, on a corpus consisting of a short Quechua tale[2]. In the following sample we can see that matching of our inflected dictionaries on the text, gives us the following results (bold dark for the nouns and red for the verbal forms):

Chaytam kunan willasqayki masiiman qina. Uyariwayyari: Ñoqaqa **tutallapim** kausani; **tutallapim** takini llakisqapas, kusisqapas.

[2] Guardia Mayorga (1973). Pakpaku chayñachawan rimanakun The owl's dialogue with the goldfinch.

Manam **pitapas** llakichiita, manchachiita munanichu. Takiillam takikuni, takiita munaspa. **Runakunataq**, imanasqachá, "kay **pakpakup** takinqa **atim**", ninku. Wañuisi qamun **ñoqa** takiptii, manataqmi ñoqá yachanichu ima chai wañuikasqanta. Kausaillatam ñoqaqa yachani. Chaymi llapan **runakuna**, **chai iskay chakipi puriq**, chai puñuysiki **runakuna**, cheqniwan. Ñoqa takikuni **tutakunapi**, paykunataq nin: " Taytallay! , piraq wañunqa, **pakpakum , sachapi** waqachkan" nispa. Chaymi cheqniwanku; chaimi warmakunapas rikuwaspa **rumiwan** choqawan, **sacha kallmapi punchau** puñukuptii. Chaymi ripusaq, **chaiñacha**.
- Manam chaiqa allinchu, nispa, **chaiñacha** kutichin. Imanasqataq mana quk takiiman **takiikita** tikrankichu? Jinallata takiptikiqa, mayta riptikipas, maypi kaptikipas, qinallataqmi **runaqa** cheqnisunki.

We can verify that the nouns and the different conjugated forms have been correctly matched by NOOJ. The non-colored lexemes, showing the absence of matches, correspond to adjectives, adverbs or their inflexions which we will program later.

References

Duran, M. 2009, Diccionario Quechua-Castellano. Editions HC. Paris.

Guardia Mayorga, C. 1973, Gramatica Kechwa,Ediciones los Andes, Lima. Peru.

Silberztein, M. 2003, Nooj Manual. htpp://www.nooj4nlp.net (220 pages updated regularly).