


# Quechua Module for NooJ Multilingual Linguistic Resources for MT

Maximiliano Duran 

Université de Franche-Comté, Besançon, France  
duran\_maximiliano@yahoo.fr

**Abstract.** The aim of this paper is to present the content of the first version of a Quechua linguistic module for NooJ and to describe its main components. They include several electronic dictionaries of verbs, nouns, and other POSs. The article also describes an ongoing work in the elaboration of a large set of paradigms which formalize Quechua inflectional and derivational morphology as well as morphological and syntactic grammars. We present the morphological codes used in the dictionaries, some graphs representing the numerals, the derivation and finally a demo text.

**Keywords:** Quechua · Quechua for NooJ · Quechua linguistic module · Quechua-French bilingual resources · Morphological grammars · Inflectional and derivational paradigms · Electronic dictionaries · Quechua simple verbs

## 1 Introduction

The linguistic resources presented here are intended to support the description of phenomena occurring in Quechua-French and Quechua-Spanish projects.

NooJ4Qu is a set of linguistic resources developed with NooJ for the automatic processing of Quechua. We follow the structure of the NooJ modules developed in the last ten years for various languages (Barreiro 2008 for Portuguese; Bogacki 2008 for Polish; Chadjipapa et al. 2008 and Gavrilidou et al. 2008 for Greek; Georganta and Papadopoulo 2012 for Ancient Greek; Aoughlis et al. 2014 for Tamazight; Dobrovolic 2014 for Slovene). It is composed of a demo text and contains two types of elements necessary for the linguistic analysis of texts. The first one comprises some dictionaries of lemmas which can operate independently. The second one is a system containing an important number of local grammars describing inflections and derivations obtained using nominal and verbal suffixes. It also includes a local grammar describing numerals ranging from 1 to 999999, a grammar that recognizes and annotates the dates and a grammar for the recognition and annotation of future forms.

It is not the purpose of this module to give users exhaustive dictionaries and complete libraries of local grammars but rather to show how Quechua, with its very rich morphology and specific peculiarities fits in NooJ's linguistic development environment.

## 2 The Main Components of NooJ4QU

The lexical analysis of a text makes use of the following components launched simultaneously:

- Dictionaries of nouns, adjectives and non-verbal POSs,
- A dictionary of Quechua verbs,
- A library of morphosyntactic grammars and of inflectional and derivational rules.

The files describing the inflection of the inflected entries are:

- Ngrammars.flx corresponding to the compiled dictionary Nquechua.nod,
- Vgrammars.flx corresponding to Vquechua.nod,
- A Quechua text containing 8 tales.

## 3 The Noun Electronic Dictionary

The electronic dictionary of simple nouns contains 1,472 entries. It does not include proper nouns. All of the entries have an inflectional paradigm assigned to them, represented by FLX=. For instance the entry unit *wasi* (house) inflects according to the paradigm class NVOCAL, thus the entry appears: *wasi,N+FR*=“maison”+FLX=NVOCAL. Inflectional paradigms are standard pattern models or prototypes based on morphological suffixation rules. These rules cover variation in number, diminutives and superlatives, verbalizations, cases.

Quechua is a highly inflectional language:

- *llaqta-cha*                      « small town »
- *kusillu-hina*                    « resembles to a monkey »
- *Paris-ña*                        « it's Paris already »
- *rima-sqa-y-man-ta*            « in what it concerns my speech »

The category of case is represented by the following set of suffixes {-*hina*, -*kama*, -*man*, -*manta*, -*nta*, -*ninta*, -*ninka*, -*nka*, -*p*, -*pa*, -*pi*, -*paq*, -*pura*, -*rayku*, -*ta*, -*wan*, -*y!*, -*niy!*}. The verbalization of a noun is obtained by one or more of the following derivation suffixes: {-*y*, -*yay*, -*chay*}

### 3.1 Nominal Inflections

There is another subset which is classified as enclitics: {-*ña*, -*raq*, -*puni*, }

The thousands of nominal forms that can be obtained by the inflectional paradigms are generated by the following set of nominal suffixes and their combinations in different layers:

SUF\_N = {*-ch* , *-chá* , *-cha* , *-chik* , *-chiki* , *-chu* , *-chu?* , *-hina* , *-kama* , *-kuna* , *-lla* , *-má* , *-man* , *-manta* , *-...* , *-pa* , *-paq* , *-pas* , *-pi* , *-poss(7v+7c)* , *-puni* , *-pura* , *-qa* , *-rayku* , *-raq* , *-ri* , *-s* , *-si* , *-sapa* , *-su* , *-ta* , *-taq* , *-wan* , *-y!* , *-ya!* , *-yá* , *-yupa* , *-yuq*} (the signs *?* , *!* are not part of the suffix, they are put only to indicate the presence of intonation).

The tag *poss (7v+7c)* represents the two sets of possessive suffixes:

- *poss (7v)*={*-y* , *-yki* , *-n* , *-nchik* , *-yku* , *-ykichik* , *-nku*} applied to nouns with a vowel ending.
- *poss (7c)*={*-niy* , *-niyki* , *-nin* , *-ninchik* , *-niyku* , *-niykichik* , *-ninku*} applied to nouns with a consonant ending.

They give us the following paradigms of possessive forms:

POSSV = *i/POS+s+1* | *iki/POS+s+2* | *iku/POS+PEX+1* | *nku/POS+p+3*;  
 POSSAIC = *nin/POSAI+s+3* | *ninchik/POSAI+PIN+1* | *niikichik/POSAI+p+2*;

The morphosyntactic grammars of dimension 1 (only a single suffix intervene) for the inflection of nouns is shown in Fig. 1.

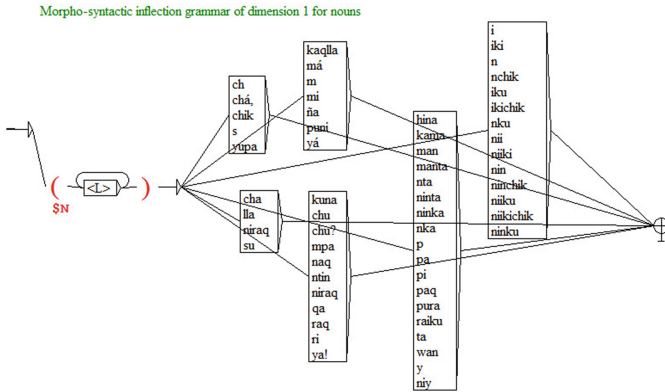


Fig. 1. Local grammar of one-dimension nominal inflection

For the combination of two nominal suffixes, we have developed an algorithm (See Duran 2013a which allows us how to obtain 517 valid agglutinations, some of which appear in the following paradigm:

N\_V\_2=:DCHACH I:DCHACHÁ I:DCHACHUI I:DCHACHU I:DCHACHIKI I:DCHAM  
I:DCHAGEPA ... I :DCHALLA I:GEPACHUN I:GEPACHIK I:GEPACHU  
I:GEPAGEPA I:GEPAKAMA I:GEPAKUNA I:GEPALLA I:GEPAMAA I:GEPAMAN ;

For the three-layer combinations of nominal suffixes we have obtained 2,108 valid ones. And for the four-layer of four suffixes we obtain less than 1,800.

Nominal inflections like: *wasi-cha-lla-ikichik-paq-hina-raq*, “it is something that would look nice in your house” are generated by paradigms containing six agglutinated suffixes like:

CHALLA\_1 = <E>/N+s | :CHALLAIKICHIKPAQHINARAQ;  
CHALLAIKICHIKPAQHINARAQ= :CHA I:LLA I:IKICHIK I:PAQ I:HINA I:RAQ;

Grammatical gender is not proper to Quechua. Natural gender is differentiated in lemmas, e.g. *maqta* boy, *pasña* girl or by noun phrases, e.g. *urqu allqu* male dog, *china allqu* female dog. Most of the kinship forms distinguish sex of both persons involved in the relationship:

man’s brother	<i>wawqi</i>	woman’s brother	<i>turi</i> .
man’s sister	<i>pani</i>	woman’s sister	<i>ñaña</i> .

### 3.2 Properties of the Noun Dictionary

According to their linguistic attributes words integrate different hierarchical ontology classes and subclasses. For many entries we have also provided syntactic-semantic properties. (Remark: most of the examples come from dictionaries QU-FR ou QU-SP (Itier 2011; Perroud 1970) and, in some cases we give their English translation). For instance:

1. QU *allqu* (FR *chien*) is classified as a common noun (imapa sutin IS, in QU), vertebrate animal (wasa tulluyuq quñi yawarniyuq WTQY, in QU) mammal (ñuñuq Ñ, in QU).
2. QU *pacha* (EN *cloth*) is classified as common noun (IS, in QU), *pachakunapaq* (PA, in QU), clothing, soft thing made of fabric, leather, and so on.
3. QU *llaqta* (EN *city*) is classified as a common noun (imapas sutin IS, in QU), proper name (runa kaqla sutiuyuq RKS in QU) denoting a geographic (allpaman qatiq, AQ in QU) place, geographical entity, and geographical location.

Let us comment the properties accompanying some entries of the sample presented in Fig. 2.

```

# NooJ V5
# Language is: qu
allqu, N+FR="chien"+FLX=NVOCAL+DRV=YAY+DRV=CHAY+NC+MAMIPH
ayni, N+FR="coopération"+FLX=NVOCAL+DRV=YY+NC
chakra, N+FR="terrain agricole"+FLX=NVOCAL+DRV=YAY+NC+GEO
Inka ñan, N+FR="chemin de l'Inca"+FLX=NCONSO+NPROP+UNAMB
kullu, N+FR="tronc"+FLX=NVOCAL+DRV=YAY+NC
laiqa, N+FR="sorcier"+FLX=NVOCAL+DRV=YY+DRV=YAY+NC
llaqta, N+FR="ville"+FLX=NVOCAL+DRV=YY+DRV=CHAY+DRV=YAY+NC+GEO
mama, N+FR="mère"+FLX=NVOCAL+DRV=YAY+NC
maqta, NM+FR="garçon"+FLX=NVOCAL+DRV=YAY+NC
ñan, N+FR="chemin"+FLX=NCONSO+DRV=YAY+DRV=CHAY+NC
pacha, N+FR="la terre"+FLX=NVOCAL+DRV=YAY+NC+GEO
pacha, N+FR="le temps"+FLX=NVOCAL+NC+TP
pacha, N+FR="vêtement"+FLX=NVOCAL+DRV=YAY+NC+VMT
paukar, N+FR="jardin"+FLX=NCONSO+DRV=YAY+DRV=CHAY+NC
purun, NA+FR="désert"+FLX=NCONSO+DRV=CHAY+DRV=YAY+NC
quillur, N+FR="étoile"+FLX=NCONSO+NC
sipas, N+FR="jeune fille"+FLX=NCONSO+DRV=YAY+NC
waman, N+FR="faucon"+FLX=NCONSO+NC+VTBRSANGCH+AVE
wasi, N+FR="maison"+FLX=NVOCAL+DRV=YAY+DRV=CHAY+NC
yupi, N+FR="trace"+FLX=NVOCAL+DRV=CHAY+NC
yura, N+FR="arbuste"+FLX=NVOCAL+DRV=CHAY+NC

```

**Fig. 2.** A sample of the dictionary of nouns

1. *allqu* is marked (N) for noun; it inflects according to the morphological paradigm NVOCAL (FLX=NVOCAL) for nouns ending in a vowel, the mark FR for French gives the French translation of the entry: “chien” (dog). We introduce some semantic properties as: common name (NC in FR), mammal animal (MAMIF in FR mammifère), and can derivate into verbs by the suffixes *-yay* and *-chay*;
2. *chakra* a noun (N) which inflects according to the morphological paradigm NVOCAL (FLX=NVOCAL) for nouns ending in a vowel, corresponding to the FR noun “terrain agricole” (cultivated field), with semantic properties as common name (NC in FR), a geographical emplacement (GEO in FR), and can derivate into verb by the suffix *-yay*;
3. *Inca ñan* a noun (N) which inflects according to the morphological paradigm NCONSO (FLX=NCONSO) for nouns ending in a consonant, a proper noun (NPROP) corresponding to the FR noun Chemin de l’Inca (Inca Trail), a multi word unit defined as unambiguous UNAMB;
4. *pacha* a noun (N) which inflects according to the morphological paradigm NVO-CAL (FLX=NVOCAL) for nouns ending in a vowel, corresponding to the FR noun “la Terre” (the earth), with semantic properties as common name (NC), a geographical emplacement (GEO), and can derivate into a verb by the suffix *yay*; which differs from the following homonym;
5. *pacha* a noun (N) which inflects according to the morphological paradigm NVO-CAL (FLX=NVOCAL) for nouns ending in a vowel, corresponding to the FR noun “le temps” (time), with semantic properties as common name (NC), concerning the time (TP), but does not derivate; this form has the following third homonym:

6. *pacha* a noun (N) which inflects according to the morphological paradigm NVO-CAL (FLX=NVOCAL) for nouns ending in a vowel, corresponding to the FR noun “vêtement” (cloth), with semantic properties as common name (NC), relative to clothing (VTM), and can derivate into a verb by the suffix *yay*.

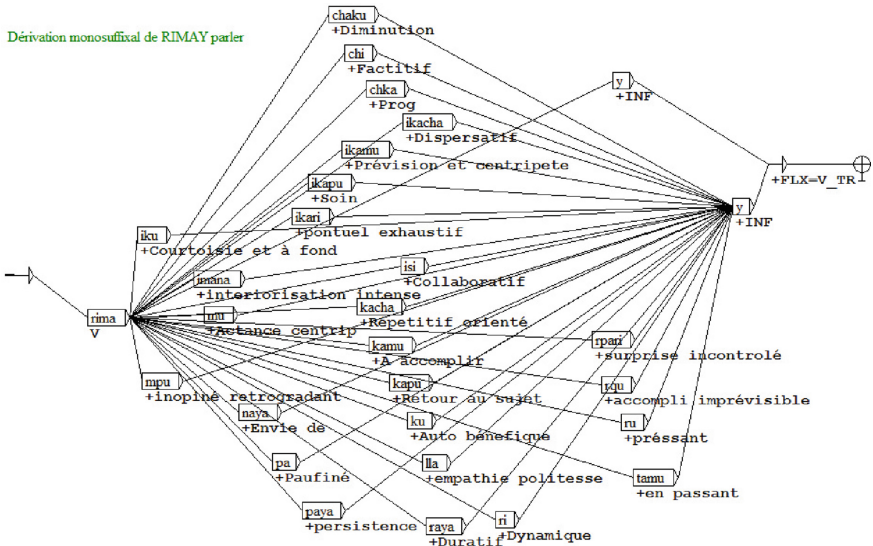


Fig. 3. Graphical grammar for derivation of a transitive verb

Let us note that one QU entry corresponds to one FR transfer, which allows obtaining disambiguated words. For instance, we see that *pacha*=temps (time) and also *pacha*=vêtement (clothing), the first one corresponds to the time (TP) and the second one defines a concrete thing made of fabric or leather or another material which serves to wear. So we try to have the disambiguation at the dictionary level by enriching the lexicon with the addition of syntactic-semantic properties; this is why we try to have as many entries as there are meanings for the same word in the source language. The compiled Nquechua.nod dictionary is able to recognize more than one million inflected forms deriving from 1500 lemmas.

#### 4 The Electronic Dictionary of Quechua Verbs

The electronic dictionary of simple verbs contains **1,472 entries**. It does not include compound verbs, neither phrasal verbs, which are presented in another dictionary. Each verb has an inflectional paradigm assigned to it. For instance the entry unit *rimay* (to talk) inflects according to the paradigm class V\_TR, thus the entry becomes: *rimay*,V +FR=“parler”+FLX=V\_TR whose structure is relatively complex as we will see.

V\_TR contains among other things, the paradigms of conjugation in present and future tenses: PR and FUT:

PR=<B>(ni/PR+s+1|nki/PR+s+2|n/PR+s+3|nchik/PR+pin+1|nkichik/PR+p+2|nku/PR+p+3|niku/PR+pex+1);

Example: *taki-ni* I sing, *taki-nki* you sing *taki-n* he sings

FUT = <B>(saq/F+s+1 | nki/F+s+2 | nqa/F+s+3 | saqku/F+pex+1 | sunchik/F+Pin+1 | nkichik/F+p+2 | nqaku/F+p+3);

Example: *taki-saq* I will sing, *taki-nki* you will sing *taki-n* he will sing

It also contains some syntactic and semantic information, like the two main classes of verbs:

Transitive (TR): *rimay* to talk, Intransitive (ITR): *mikuy* to eat.

The intransitive class is quite small. It contains less than one hundred stems. The class of impersonal verbs (INP) includes those relating to weather: *paray* to rain; *lastay* to snow.

As we showed in a previous work (Duran 2013 and 2014), a typical Quechua inflected verbal form has the following structure:

<V><IPS><PR ENDING><PPS>, where:

- V: Verb stem,
- IPS<sup>1</sup>: Interposed suffixes (placed between the lemma and the ending),
- PPS<sup>2</sup>: Post-posed suffix (placed after the ending),
- PR ENDING<sup>3</sup>: is the set of seven present tense personal endings (which will behave as fixed points during the inflections).

Examples:

*rima-nki*, ‘you talk’, present tense 1+s

*rima-ri-nki*, ‘you start talking’, the IPS suffix *-ri* is interposed

*rima-nki-man*, ‘you should talk’, the PPS suffix *-man* is post posed

*rima-ri-nki-man*, ‘you should perhaps start talking’, the IPS suffix *-ri* and the PPS suffix *-man* are mixed.

In these examples, each class of suffix intervenes in the inflection only once at both sides of the ending. But, the Quechua grammar allows having several layers of IPS and PPS. (For a more detailed explanation on the construction of valid agglutinations of IPS and PPS suffixes and their mixed patterns see Duran 2014).

<sup>1</sup> IPS = {chaku, chi, chka, ykacha, ykachi, ykamu, ykapu, ykari, yku, ysi, kacha, kamu, kapu, ku, lla, mpu, mu, na, naya, pa, paya, pti, pu, ra, raya, ri, rpari, rqa, rqu, ru, spa, sqa, stin, tamu, wa}.

<sup>2</sup> PPS = {ch, chaa, chik, chiki, chu(?), chu, chusina, má, man, m, mi, ña, pas, puni, qa, raq,ri, sí, s, taq, yá}.

<sup>3</sup> PR ENDING = {-ni, -nki, -n, -nchik, -niku, -nkichik, -nku}.

In fact, considered separately we will have 231 combinations of two PPS, 132 combinations of three PPS and 92 combinations of four valid combinations of PPS. Here are some samples of the corresponding paradigms:

```
SVPP2= (CHUSINAVMV MANVCHAAV MANVCHIKV MANVCHIKIV
ANVQAVMANVCHUIV MANVRAQV MANVSIV MANVTAQV
MANVYAVMANVYAAV ÑAVCHV ÑAVCHAAV ÑAVCHIKIV ÑAVCHUIV
...)
SVPP3 = (MANVCHUSINAVMV MANVÑAVCHV MANVÑAVCHAAV
MANVÑAVCHIKV MANVÑAVCHIKIV MANVÑAVCHUIV MANVÑAVCHUNV
MANVÑAVCHUSINAV RAQVTAQVYAAV TAQVCHUSINAVMV...)
```

On the other hand, a very large number of inflected forms include a mixture of one or more layers of IPS and one or more PPS, we call them mixed inflections MIFLX.

For instance, in the form *rima-ri-lla-nki-man-raq* ‘I think you should talk before’, we have two suffixes IPS (*-ri*, *-lla*) and two PPS (*-man*, *-raq*). Besides, in the form *miku-cha-ku-na-lla-n-paq* ‘hoping that she will kindly eat it », we have four suffixes IPS (*cha*, *ku*, *na*, *lla*) and one suffix SPP (*paq*) (4 + 1).

In order to generate automatically all these forms, we have constructed a series of mixed paradigms which we symbolize by V\_MIXn. We present the details for V\_MIX1 which contains 2\*2 = 4 sub-grammars (2 four IPS consisting in one layer, ending in a vowel or a consonant, and 2 for the SPP including one layer placed after the endings, ending in a vowel or a consonant):

```
V_MIX1=(:(SIP1_PR_V):(SPP1_V)|(:(SIP1_PR_C):(SPP1_C)
|(:(SIP1_PR_M_V):(SPP1_V)|(:(SIP1_PR_M_C):(SPP1_C);
```

Thus, summarizing, one of the dominant paradigms for the inflection of a verb contains all these grammars:

```
V_TR= :V_SPP|:V_TR_SIP|:V_CONJ|:IMP; where
V_PP=:SPP1_C|:SPP1_F|:SPP2;
V_TR_SIP = :SIP1_G|:SIP2_G;
SIP1_G = <E>/INF|:SIP1|:V_SIP1_INF|:SIP1_N; and so on.
```

#### 4.1 Verb-Verb Derivations: Generating New ULAVs

Among the SIP, we find a sub-set of 27 suffixes which derivate a verb into another verb. This strategy allows Quechua to generate new ULAV<sup>4</sup>. These new forms can be conjugated as if they were simple verbs. We name this set “SIP\_DRV”.

<sup>4</sup> ULAV: ‘unités linguistiques atomiques verbales conjugables’ or ‘Conjugable verbal atomic linguistic units’, after Silberstein (2016).



```

# NooJ V5
# Dictionary
rimaruy, rimay, V+FR="parler"+FLX=V_SIP1_PRES_1_réalise_l'action_de_façon_
rimarquy, rimay, V+FR="parler"+FLX=V_SIP1_PAPT_1_accompli_l'action_en_peu_
rimarpariy, rimay, V+FR="parler"+FLX=V_SIP1_ASUR_1_action_surprise_sans_hé:
rimapayay, rimay, V+FR="parler"+FLX=V_SIP1_FREQ_1_répétition_fréquente+INF
rimapay, rimay, V+FR="parler"+FLX=V_SIP1_PEAU_1_peaufine_l'action_PEAU+INF
rimanayay, rimay, V+FR="parler"+FLX=V_SIP1_ENV_1_envié_de_souhaite_de+INF
rimamuy, rimay, V+FR="parler"+FLX=V_SIP1_ACT_1_actance_centripète_d'approx:
rimakuy, rimay, V+FR="parler"+FLX=V_SIP1_AUBE_1_actance_se_responsabilisan:
rimakapuy, rimay, V+FR="parler"+FLX=V_SIP1_RAS_1_auto_bénéfice_réalise_l'act:
rimakamuy, rimay, V+FR="parler"+FLX=V_SIP1_AAR_1_aller_à_réaliser_l'action+IN:
rimakachay, rimay, V+FR="parler"+FLX=V_SIP1_ARO_1_action_répété_orienté+INF
rimaisiy, rimay, V+FR="parler"+FLX=V_SIP1_COLL_1_aide_à_réaliser_l'action+:
rimaikuy, rimay, V+FR="parler"+FLX=V_SIP1_COURT_1_courtoisement_soigneusem:
rimaikariy, rimay, V+FR="parler"+FLX=V_SIP1_APPR_1_ponctuelle_et_rapidemen:
rimaikapuy, rimay, V+FR="parler"+FLX=V_SIP1_SOIN3_1_avec_attention+INF
rimaikamuy, rimay, V+FR="parler"+FLX=V_SIP1_PREAT_1_vers_le_sujet+INF
rimaikachiy, rimay, V+FR="parler"+FLX=V_SIP1_POLI_1_poliment+INF
rimachkay, rimay, V+FR="parler"+FLX=V_SIP1_PROG1_en_train_deréaliser_l'act:
rimachiy, rimay, V+FR="parler"+FLX=V_SIP1_FACT_1_assiste_aide+INF

```

**Fig. 4.** ULAVs obtained by derivation of the verb *rimay* (to talk)

SIP\_DRV= {-chaku, -chi, -chka, -ykacha, -ykachi, -yakamu, -ykapu, -ykari, -yku, -ysi, -kacha, -kamu, -kapu, -ku, -lla, -mpu, -mu, -naya, -pa, -paya, -pu, -raya, -ri, -rpari, -rqu, -ru, -tamu}

For example, the NooJ grammar of Fig. 3. generates all the new ULAVs for the verb *rimay* shown in Fig. 4.

We can obtain combinations of two, three or four IPS suffixes. We may obtain in this way more than 2,000 derived forms for a single verb. They look like the ones appearing in the following sample which includes different layers.

```

rimarquy, rimay, V+FR="parler"+FLX=V_SIP_INF+PAPT+INF (talk thoroughly)
rimarpariy, rimay, V+FR="parler"+FLX=V_SIP_INF+ASUR+INF (talk to each one with details)
rimapayaikachaikamuy, rimay, V+FR="parler"+FLX=V_SIP3_INF+FREQ+ARO+PREAT +
riariikapuchkay, rimay, V+FR="parler"+FLX=V_SIP3_INF+DYN+SOIN3+PROG+INF
rimaisimullay, rimay, V+FR="parler"+FLX=V_SIP3_INF +COLL+ACENT+POL1+INF
rimaisimuchkay, rimay, V+FR="parler"+FLX=V_SIP3_INF+COLL+ACENT+PROG+INF

```

It is an important question to know the meaning of these generated forms. Are they only theoretical forms with no tangible sense or are they used currently in the daily conversations? Many are currently used and are meaningful for the users but their translation needs long descriptions. For instance, we have the translation of the entry:

```
rimariikapuchkay, rimay, V+FR="parler"+FLX=V_SIP3_INF+DYN+SOIN3+PROG+INF
```

as follows: ‘a kind demand to start talking in behalf of a third person’.

On the other hand, if we attempt to get an automatic translation based in the meaning of each suffix participating in the combination, we may use the ‘meaning value’ of each of the suffixes noted in the entry, and put them in a grammatical order as follows:

<i>rima</i>	<i>ri</i>	<i>ikapu</i>	<i>chka</i>	<i>y</i>	
talk		DYN	SOIN3	PROG	INF

The noted symbols carry on the following meanings<sup>5</sup>:

DYN	“the action begins”
SOIN3	“a demand to do something in behalf of a third, with care”
PROG	Progressive (the action is taking place at present)
INF	Infinitive

Thus, ‘a demand + to do the action of talking + done at present + executed gently in behalf of a third person + to start now’ (should be equivalent to) ‘a kind demand to start talking in behalf of a third’, which seems to be the case.

The complete translation of the new ULAVs is a real challenge for NLP and MT research in Quechua.

## 4.2 Nominalization

The nominalization suffixes are placed at the end of the verbal lemma. (For details on their construction, see Duran 2013b) They are generated by the following set of suffixes:

S\_NV= {*y*, *na*, *q*, *sqá*}.

As we can see in the following examples, certain verbal stems can be nominalized by more than one of them.

<i>tiyana</i>	<\$V,N+Instrumental>
<i>tiyaq</i>	<\$V,N+ Agentif >
<i>rimana</i>	<\$V,N+Instrumental> (an object for talking: microphone)
<i>rimai</i>	<\$V,N+Substantif> (the talk)
<i>rimaq</i>	<\$V,N+Agentif> (the speaker)
<i>rimasqa</i>	<\$V,N+Participe> (the speech)

These forms can be generated by the grammar shown in Fig. 5:

<sup>5</sup> A condensate dictionary of ‘enunciation moods’ of all the IPS suffixes in form of a table can be found in (Duran 2016).

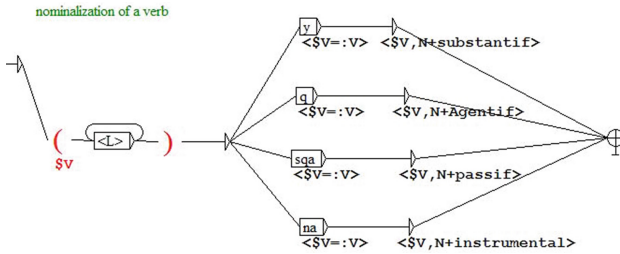


Fig. 5. Nominalization grammar

Gathering all this information concerning the verbs, we have built the dictionary of verbs contained in the Quechua Module for NooJ. We can see in a sample in Fig. 6:

```
# NooJ V5
# Dictionary
# Language is: qu
achay,V+TR+FR="chauffer"+FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO+DRV=NAS:NVOCAL+DRV=SQAS:NVOCAL
achikyay,V+TR+FR="faire jour"+FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO+DRV=SQAS:NVOCAL
away,V+TR+FR="tisser"+FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO+DRV=NAS:NVOCAL+DRV=SQAS:NVOCAL
ayniy,V+TR+FR="aider"+FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO
aypay,V+TR+FR="atteindre"+FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO+DRV=NAS:NVOCAL+DRV=SQAS:NVOCAL
ayuyqay,V+TR+FR="enrôler"+FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO+DRV=SQAS:NVOCAL
ayuy,V+TR+FR="faire adultère(pour un homme)" +FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO+DRV=SQAS:NVO
aywiy,V+TR+FR="agiter" +FLX=V TR+DRV=IS:NCONSO+DRV=NAS:NVOCAL+DRV=SQAS:NVOCAL
chakchay,V+TR+FR="mâcher des feuilles de coca"+FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO
chakiy,V+TR+FR="sécher"+FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO+DRV=NAS:NVOCAL+DRV=SQAS:NVOCAL
chakmay,V+TR+FR="casser la terre"+FLX=V TR+DRV=IS:NCONSO+DRV=NAS:NVOCAL+DRV=SQAS:NVOCAL
chakmay,V+TR+FR="mettre en qachère"+FLX=V TR+DRV=QS:NCONSO+DRV=NAS:NVOCAL+DRV=SQAS:NVOCAL
chakmay,V+TR+FR="préparer la terre"+FLX=V TR+DRV=IS:NCONSO+DRV=NAS:NVOCAL+DRV=SQAS:NVOCAL
challay,V+TR+FR="arroser"+FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO+DRV=SQAS:NVOCAL
challay,V+TR+FR="asperger" +FLX=V TR+DRV=IS:NCONSO+DRV=NAS:NVOCAL+DRV=SQAS:NVOCAL
chamqay,V+TR+FR="jeter"+FLX=V TR+DRV=IS:NCONSO+DRV=QS:NCONSO+DRV=NAS:NVOCAL+DRV=SQAS:NVOCAL
```

Fig. 6. The verb dictionary of the module

## 5 Dictionaries for Other POSs

We have included, in the module, a dictionary of adjectives and another of other POS containing adverbs, pronouns and numerals. The set of suffixes that allows inflecting each category are given separately. For example the following sets correspond to adjectives and pronouns:

- SUF\_A\_V = {-ch, -chá, -cha,-chik, -chiki, -chu, -chu?, - hina, -kama, -kuna, -lla, -má..., -raq, -ri, -s, -su, -ta, -taq, -wan, -yá, -yupa}, adjectival suffixes.
- SUF\_PRO\_V = {-ch, -chá, -,chik, -chiki, -chu, -chu?, -hina, -kama, -lla, -má, -man..., -qa, -rayku, -raq, -ri, -s, -ta, -taq, -wan, -yá, -yupa}, pronominal suffixes applied to pronouns ending in vowels.
- SUF\_PRO\_C = (-chá, -chik, -chiki, -chu, -chu?,-hina, -kama, -kuna, -lla, -má, -man, -manta..., -ri, -si, -ta, -taq, -wan, -yupa), pronominal suffixes applied to pronouns ending in consonants.

For instance, here is a sample of paradigm for adjectives ending in a vowel:

```
A_V_1 = <E>|:CH |:CHAA |:CHIK |:CHIKI |:CHUN |:CHUI |:DCHAI:GEP |:GEP  
|:HINA|:KAMA, ..., |:POSV_v |:POSV_c |:PUNI |:PURA |:QA |:RAYKU |:RAQ |:RI  
|:SSIV |:TA |:TAQ |:WAN |:YAA |:YUPA;
```

This will generate 56 inflected one-dimension forms for the adjective *taksa* (small) as follows:

```
taksach,taksa,A+FR=« petit »+ FLX=A_V_1+DPRO  
taksachá,taksa,A+FR=« petit »+ FLX=A_V_1+DPRO  
taksacha,taksa,A+FR=« petit »+ FLX=A_V_1+DPRO
```

As for the case of nouns, Quechua allows agglutination of adjectival suffixes up to seven layers. The following sample shows forms with two suffixes:

```
sumaqkamachá,sumaq,A+FR="beau"+FLX=A_C_2+LIM+DPRO  
sumaqkamalla,sumaq,A+FR="beau"+FLX=A_C_2+LIM+ISO  
sumaqkunachá,sumaq,A+FR="beau"+FLX=A_C_2+PLU+DPRO
```

Applying similar methods to nouns and verbs has driven to hundreds of paradigms, for the inflection of these categories containing two or more agglutinated suffixes. Figure 7 shows a sample of the dictionary of pronouns.

---

```
# NooJ V5  
# Dictionary  
ñuqa,PRO+FR="je moi"+FLX=FLXPROVOY  
qam,PRO+FR="tu toi"+FLX=PROPLU  
pai,PRO+FR="il elle lui"+FLX=+FLX=FLXPROCON  
pay,PRO+FR="il elle lui"+FLX=FLXPROCON  
ñuqanchik,PRO+FR="nous, (tous)"+FLX=PROPLU  
ñuqaiku,PRO+FR="nous, (pas vous)"+FLX=PROPLU  
kikii,PRO+FR="moi même"+FLX=FLXPROVOY  
kikiiki,PRO+FR="toi même"+FLX=FLXPROVOY  
kikin,PRO+FR="soi même"+FLX=FLXPROCON  
kikinichik,PRO+FR="nous même"+FLX=FLXPROCON  
kikiiku,PRO+FR="nous même t'excluant"+FLX=FLXPROVOY  
kikiikichik,PRO+FR="vous même pluriel"+FLX=PROPLU  
kikinku,PRO+FR="eux même"+FLX=FLXPROVOY  
ñoqaiku,PRO+FR="nous autres"+FLX=FLXPROVOY  
qamkuna,PRO+FR="vous pluriel"+FLX=FLXPROVOY  
juk,PRO+FR="otro"+FLX=FLXPROCON  
huk,PRO+FR="otro"+FLX=FLXPROCON  
quk,PRO+FR="otro"+FLX=FLXPROCON  
chai,PRO+FR="cela"+FLX=FLXPROCON  
chay,PRO+FR="otro"+FLX=FLXPROCON  
wakin,PRO+FR="quelques uns"+FLX=FLXPROCON  
llapan,PRO+FR="tous toutes"+FLX=FLXPROCON
```

---

Fig. 7. The Dictionary of pronouns

## 6 Some Additional Grammars to the Library

We present, as an illustration, three graphical grammars contained in the Quechua NooJ Module: Pachayupay.nog (the dates in QU), NumQU.nog (the numbers en QU) and Future.nog (the future in QU).

### 6.1 The Dates and Numerals in Quechua

Recognition and annotation of the dates is done by the following grammar (Fig. 8):

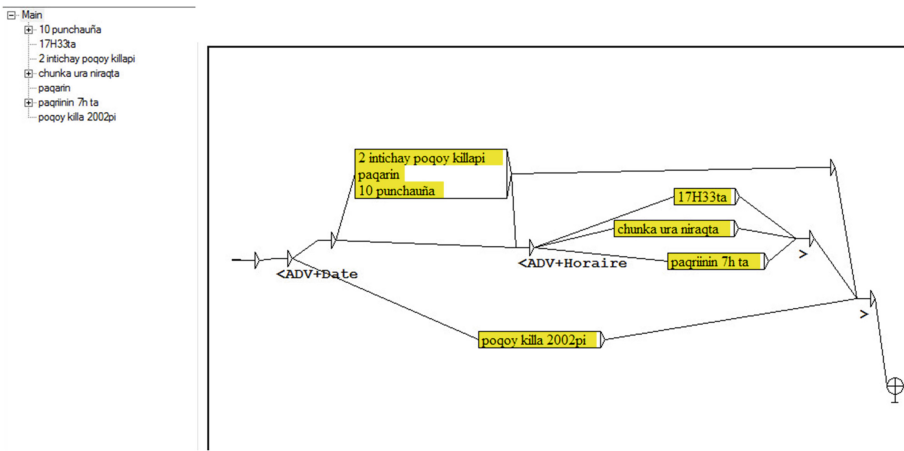


Fig. 8. Pachayupay.nog grammar (the dates in QU)

Figure 9 annotates the numeric determinants in full letters for the numbers going from 1 to 999999: “huk” to “isqun pachak isqun chunka isqunniyuq waranqa isqun pachak isqun chunka isqunniyuq”.

dnumQU

This graph annotates the numeric determinants in full letters, ex. from 1 to 999999 "huk" to "isqun pachak isqun chunka isqunniyuq waranqa isqun pachak isqun chunka isqunniyuq"

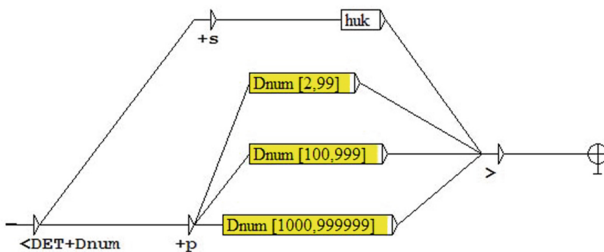


Fig. 9. Annotation of numeric determinants in full letters

## 6.2 Recognition Grammar for the Future Tense

(See Fig. 10)

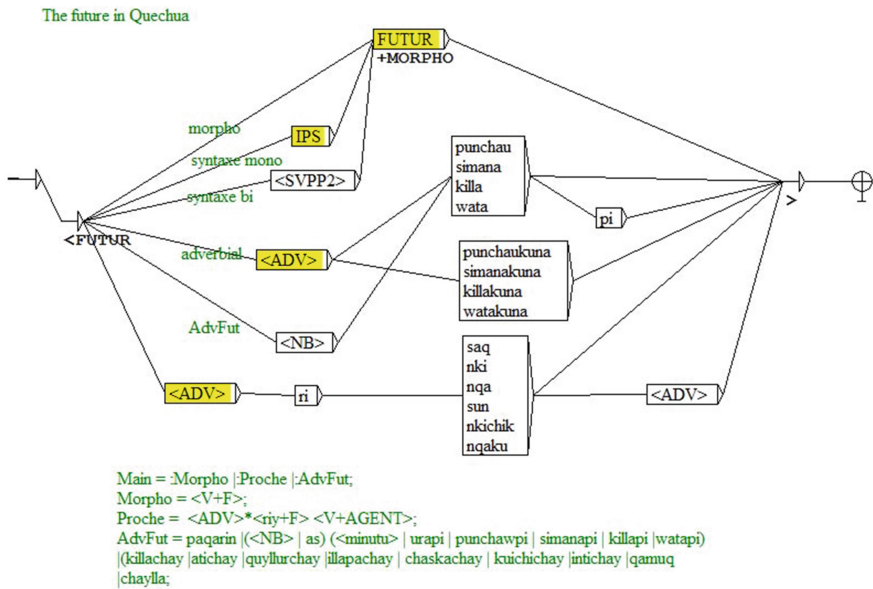


Fig. 10. Future Quechua.nog grammar (The future in QU)

## 7 Linguistic Analysis

The linguistic resources that we have presented can be applied to analyze texts. Figure 11 shows an extract of the results of the concordance query  $<N>+<V+POL1><ADV>$  applied to a text of Quechua tales. It recognizes and annotates 1,320 forms.

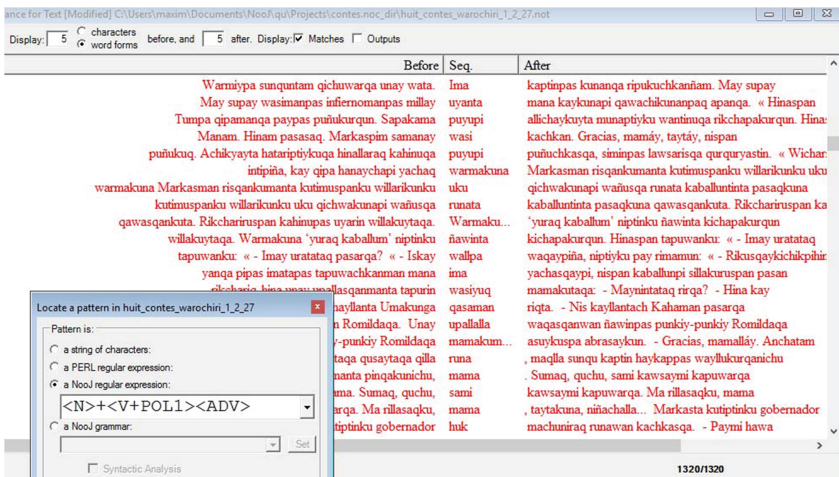


Fig. 11. An extract from the results of the query  $<N>+<V+POL1><ADV>$

## 8 Conclusion and Perspectives

In this paper we have described the first version of the Quechua Module for NooJ. We have seen how the electronic dictionaries are built. We have shown different grammars and how they interact with the dictionaries and the morphological paradigms. We have also seen a simple example of application of these grammars to a text. The simple verbs dictionary was extracted from M. Duran dictionary Quechua-Français (Duran 2009). Our next step will consist in the enhancement of the French-Quechua verb dictionary, by the inclusion of our current research on the electronic bilingual French - Quechua dictionary containing the translation of the 25,000 French verbal senses of Dubois et Dubois-Charlier 2000. On the other hand, we expect to add some syntactic-semantic grammars and disambiguation grammars.

## References

- Aoughlis, F., Nait-Serrad, K., Hamid, A., Ferroudja, A., Said, H.M.: New Tamazight module for NooJ. In: Koeva, S., Mesfar, S., Silberztein, M. (eds.) *Formalising Natural Languages with NooJ 2013. Selected Papers from the NooJ 2013 International Conference*, pp. 13–26. Cambridge Scholars Publishing, Newcastle (2014)
- Barreiro, A.: Port4NooJ: Portuguese linguistic module and bilingual resources for machine translation. In: Blanco, X., Silberztein, M. (eds.) *Proceedings of the 2007 International NooJ Conference*, pp. 19–47. Cambridge Scholars Publishing, Newcastle (2008)
- Bogacki, K.: Polish module for NooJ. In: Blanco, X., Silberztein, M. (eds.) *Proceedings of the 2007 International NooJ Conference*, pp. 48–66. Cambridge Scholars Publishing, Newcastle (2008)
- Chadjipapa, E., Papadopoulou, E., Gavrilidou, Z.: New data in the Greek NooJ module: compounds and proper nouns. In: Blanco, X., Silberztein, M. (eds.) *Proceedings of the 2007 International NooJ Conference*, pp. 96–102. Cambridge Scholars Publishing, Newcastle (2008)
- Duran, M.: *Dictionnaire Quechua-Français-Quechua*, Editions HC, Paris (2009)
- Duran, M.: Formalizing Quechua noun inflection. In: Donabédian, A., Khurshudian, V., Silberztein, M. (eds.) *Formalising Natural Languages with NooJ2013*, pp. 41–50. Cambridge Scholars Publishing, Newcastle (2013)
- Duran, M.: Formalizing Quechua verb inflections. In: Koeva, S., Mesfar, S., Silberztein, M. (eds.) *Formalising Natural Languages with NooJ 2013. Selected Papers from the NooJ 2013 International Conference*, pp. 41–50. Cambridge Scholars Publishing, Newcastle (2014)
- Duran, M.: Morphological and syntactic grammars for the recognition of verbal lemmas in Quechua. In: Monti, J., Silberztein, M., Monteleone, M., di Buono, M.P. (eds.) *Proceedings of the NooJ 2014 International Conference on Formalising Natural Languages with NooJ2014*, pp. 28–36. Cambridge Scholars Publishing, Newcastle (2015)
- Duran, M.: The annotation of compound suffixation structure of Quechua verbs. In: Okrut, T., Hetshevich, Y., Silberztein, M., Stanislavenka, H. (eds.) *NooJ 2015. CCIS*, vol. 607, pp. 29–40. Springer, Heidelberg (2016). doi:[10.1007/978-3-319-42471-2\\_3](https://doi.org/10.1007/978-3-319-42471-2_3)
- Gavrilidou, Z., Chatjipapa, E., Papadopoulou, E.: The new Greek NooJ module: morphosemantic issues. In: Blanco, X., Silberztein, M. (eds.) *Proceedings of the 2007 International NooJ Conference*, pp. 96–102. Cambridge Scholars Publishing, Newcastle (2008)

- Georganta, M., Papadopoulou, E.: Towards an ancient Greek NooJ module. In: Vučković, K., Bekavac, B., Silberstein, M. (eds.) *Formalising Natural Languages with NooJ. Selected Papers from the NooJ 2011 International Conference*, pp. 41–49. Cambridge Scholars Publishing, Newcastle (2011)
- Itier, C.: *Dictionnaire Quechua-Français. L'Asiathèque*, Paris (2011)
- Perroud, P.C.: *Diccionario Castellano kechwa, kechwa castellano. Dialecto de Ayacucho*. Santa Clara Seminario San Alfonso, Peru (1970)
- Silberstein, M.: *Formalizing Natural Languages: The NooJ Approach*. Wiley, London (2016)
- Silberstein, M.: NooJ's dictionaries. In: Vetulani, Z. (ed.) *Proceedings of the 2nd Language and Technology Conference*. Wydawnictwo Poznańskie Sp. z o.o., Poznan (2005)