

Semi-Automatic construction of a Quechua-Spanish dictionary

Maximiliano Duran
Université de Franche-Comté
maximiliano.duran@yahoo.fr

Max Silberztein
Université de Franche-Comté
max.silberztein@gmail.com

Abstract

This paper presents a set of linguistic resources that describes Quechua verbs. We first present a dictionary of 1,444 fundamental Quechua verbs, associated with morpho-syntactic grammars to formalize their inflection and their derivations, that can be used to produce over 2,777,000 conjugated Quechua derived verbal forms. We aligned this list of Quechua verbal forms with the corresponding Spanish dictionary that contains 618,000 conjugated verbal forms, thus producing both a Spanish to Quechua and a Quechua to Spanish dictionary.

1 Introduction

Quechua is an endangered language, spoken by over 6 million people in Peru, Equator, Bolivia, and Argentina. Quechua has an important characteristic: its verbal morphology is robust and highly productive. Using the NooJ linguistic development platform (see Silberztein 2016),¹ we have constructed a dictionary of 1,444 Quechua verbs and formalized the corresponding inflectional and derivational morphological systems.

Quechua morphological system is based on the use of suffixes, which we have classified into four fundamental sets, see (Duran, 2014):

¹ NooJ is a free and open-source linguistic development platform linguists use to formalize various linguistic phenomena, from orthographic to semantic. NooJ allows linguists to construct electronic dictionaries, as well as regular, context-free, context-sensitive and unrestricted formal grammars. NooJ contains tools to edit, test, debug and maintain large-coverage linguistic resources. There are linguistic resources in the NooJ format for over 30 natural languages, see <https://nooj.univ-fcomte.fr>.

- Conjugation endings suffixes include two subsets: the Indefinite Tense Endings ITE = {*ni, nki, n, nchik, niku, nkichik, nqaku*}, and the Future Tense Endings FTE = {*saq, nki, nqa, sunchik, saqku, nkichik, nqaku*}.
- Interposition Suffixes (IPS) placed after the verb radical and the infinitive suffix “y” or the conjugation endings. IPS suffixes are divided into two very productive subsets:
- Derivation Inter Position Suffixes IPSd = {*chaku, chi, chka, ykacha, ykachi, ykamu, ykapu, ykari, yku, ysi, kacha, kamu, kapu, ku, lla, mpu, mu, naya, pa, paya, pu, raya, ri, rpari, rqu, ru, tamu*} (27), which will generate new verbs by derivation. For example:

asiy [to laugh] → *asiriy* [to smile]
maskay [to search] → *maskapay* [to inquire]
rimay [to talk] → *rimaykuy* [to greet]
samay [to rest] → *samarquy* [to bivouac].

- Postposition Suffixes (PPS) placed after the conjugation endings, The Postposition Suffixes (PPS), placed after the conjugation endings, allow to express modalities: declarative, interrogative, imperative, emphasis, exclamation, negative or affirmative, passive, active or conditional: PPS = {*-ch, -chá, -chik, -chiki, -chu, -chu(?), -chusina, -m, -mi, -má, -man, -ña, -pas, -puni, -qa, -raq, -s, -si, -taq, -yá*}
- Non Derivation Inter Position Suffixes IPSd, serving to express the Past Tenses (*rqa, sqa, ra*) and some moods Reflexive (*wa*), Subjunctive (*sqa*), and the Gerund

(spa, pti), IPSd = { pti, rqa, spa, sqa, stin, wa}

- Verb Nominalizing Suffixes (VNS) derive a verb into a noun: VNS = {-i, -na, -q, -sqa }.

Some verbs accept only one nominalization suffix, whereas others may take two, three, or all four nominalization suffixes. For example:

imikuy → *mikui* [meal]
namikuy → *mikuna* [feed]
qmikuy → *mikuq* [dinner]
sqamikuy → *mikusqa* [what was eaten]

2 Conjugation

The basic indefinite tense Quechua conjugation scheme consists of the agglutination of the verbal root and the ITE suffix. Named as indefinite, it may simultaneously express a present, a past tenses, or a usual activity. The following table presents the conjugation of a verb in the Indefinite tense:

Pronoun	SP-EN	Verb root	ITE
<i>ñoqa</i>	Yo I	VR	<i>ni</i>
<i>qam</i>	Tu You		<i>nki</i>
<i>pay</i>	El(la) He, she		<i>n</i>
<i>ñoqanchik</i>	Nosotros We (inc.)		<i>nchik</i>
<i>ñoqaiku</i>	Nosotros We (ex.)		<i>niku</i>
<i>qamkuna</i>	Ustedes You		<i>nkichik</i>
<i>paykuna</i>	Ellos(IIas) They		<i>nku</i>

Table 1: Indefinite Tense

In the same way, the conjugation paradigm that corresponds to the future tense is shown in Table 2:

Pronoun	SP-EN	Verb root	FTE
<i>ñoqa</i>	Yo I	VR	<i>saq</i>
<i>qam</i>	Tu You		<i>nki</i>
<i>pay</i>	El(la) He, she		<i>nqa</i>
<i>ñoqanchik</i>	Nosotros We (inc.)		<i>sunchik</i>
<i>ñoqaiku</i>	Nosotros We (ex.)		<i>saqku</i>
<i>qamkuna</i>	Ustedes You		<i>nkichik</i>
<i>paykuna</i>	Ellos(IIas) They		<i>nqaku</i>

Table 2: Future Tense

As of now, we have listed 1,444 Quechua fundamental verbs, and have associated each of

them with their corresponding conjugation paradigm.

In Quechua, many combinations of tenses, aspects, and circumstances result from the agglutination of one, two, or three IPS. For example:

mikuchkani [I am eating]
mikurqani [I have eaten]
mikuchkarqani [I was eating]
mikurani [I ate]
mikunaimi karqa [I was going to eat]
mikusaq [I will eat]
mikurqusaq [I am going to eat quickly]

3 Derivation

Fundamental verbs can be agglutinated with one or more suffixes to produce new verb meanings. Next are a few examples of fundamental verbs agglutinated with one IPS suffix:

asiy [to laugh] + *ri* → *asiriy* [to smile]
rimay [to talk] + *yku* → *rimaykuy* [to greet]
ripuy [to go away] + *ku* → *ripukuy* [to move]
suyay [to wait] + *yku* → *suyaykuy* [to wait patiently]

Most derived verbs are listed as entries in traditional dictionaries. For instance, mono-suffixed derived verbs are represented as lexical entries, as the following ones:

asipayay [to make fun of], derived from *asiy* [to laugh]
maskakuy [to search for oneself], derived from *maskay* [to search]
rakinakuy [to divorce], derived from *rakiy* [to split]
rimanakuy [to dialogue], derived from *rimay* [to talk]

but many derived verbs, such as the ones derived from the verb *rimay* [to talk], are not listed in any dictionary: *rimaymanay*, *rimatamuy*, *rimaruy*, *rimapuy*, *rimarquy*, *rimarpariy*. Our project is to recognize them and produce their Spanish translation automatically.

4 Formalization with NooJ

The NooJ platform contains tools that allow linguists to develop various types of grammars, from the orthographic to the semantic levels. NooJ inflectional grammars can be regular or context-free grammars. In its simplest form, a NooJ

regular grammar is just a disjunction of (suffix, property) pairs in the following form:

```
Paradigm = suffix/properties |
suffix/properties | ...;
```

Typically, a suffix is added to the lexical entry to produce an inflected form, which is then associated with some linguistic property. For example, the following paradigm WALK describes the inflection of all the English verbs that take an “s” in the Present, third person singular, “ing” in the Gerund, and “ed” in the Preterit and the Past Participle:

```
WALK = <E>/Infinitive |
s/Present+3+singular | ing/Gerund |
ed/Preterit | ed/PastParticiple;
```

The first term represents the fact that if one adds the empty string (noted <E> in NooJ) to a lexical entry, one produces its infinitive form. The second term represents the fact that if one adds an “s” to the lexical entry, one produces a form in the Present tense, 3rd person, singular.

Suffixes may contain stack operators to process words, such as (for “Backspace”) to delete the letter at the top of the stack. For instance, the following rule represents the paradigm used to inflect the word “man”:

```
MAN = <E>/singular | <B2>en/plural;
```

If one deletes the two last letters of the lexical entry “man” and then adds the suffix “en”, one produces the plural form “men”.

NooJ contains also “linguistic” operators specific to each language, such as the Spanish operator <Á> (add accent), the French operator <D> (duplicate letter), the Semitic operator <F> (definalize/finalize last consonant), etc.

NooJ contains a handy operator specific for Quechua: <D> which, applied to certain inanimate nouns, duplicates them to generate a new noun bearing a superlative content, such as in the following examples:

```
rumi [stone] + <D> → rumirumi [stony]
sacha [tree] + <D> → sachasacha [forest]
aqu [sand] + <D> → aquaqu [sandy]
```

The following rule formalizes the Quechua Indefinite Tense IT paradigm, described in Table 1 of suffixes:

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

The linguistic codes correspond to the following properties: +1, +2, +3: first, second, or third person; +s: singular; +p: plural; +pin: inclusive plural; +pex: exclusive plural. In NooJ Context-Free grammars, a rule can refer to another rule, using its name prefixed with the colon character (“:”). Rule names are the equivalent of the auxiliary symbols in generative grammars. For example, rule IPSIPS is defined as:

```
IPSIPS = <B> (chi/FACT mu/ACENT
| chi/FACT pu/APT |
chaku/DVAL :CHKA/PROG |
chi/FACT chi/FACT);
```

Here, :CHKA refers to the paradigm **CHKA**. The Quechua dictionary uses a list of 27 semantic features (FACT, ACENT, APT, PROG, etc.). For instance: FACT (factitive), ACENT (towards oneself), PROG (Progressive), etc.

The following rule formalizes verbal derivations that use one suffix:

```
V_SIP1_INF = <B> (:CHAKU | :CHI |
:CHKA | :YKACHA | :YKACHI | :YMANA |
:YKAMU | :YKAPU | :YKARI | :YKU |
:YSI | :KACHA | :KAMU | :KAPU | :KU
| :LLAV | :MU | :NAYA | :PAV | :PAYA
| :PU | :RAYAV | :RIV | :RPARI |
:RQU | :RU | :TAMU) y/V;
```

Applied to the verb *rimay* [to talk], NooJ produces 27 derived verbs automatically, including the following ones:

```
rimachiy,V+FLX=V_SIP1_INF
rimarpariy,V+FLX=V_SIP1_INF
rimaykuy,V+FLX=V_SIP1_INF
```

These derived verbs correspond to the following English translations:

```
rimachiy [make him talk]
rimarpariy [go to talk to all of them]
rimaykuy [say your greetings]
```

4.1 Multi-suffixes inflections

Quechua verbs can be derived using two IPS suffixes. The Boolean matrix of table 3, presented in (Duran 2017), represents 240 valid combinations of IPS suffixes.

SIP_TR	CHAKU	CHI	CHKA	YKACHA	YKACHI	YKAMU	YKAPU	YKARI	YKU	YSI	KACHA	KAMU	KAPU	KU	L
CHAKU	0	1	1	0	1	1	1	1	1	1	1	0	0	0	0
CHI	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1
CHKA	0	0	0	0	1	1	1	1	0	1	0	0	0	0	0
YKACHA	1	1	1	0	0	0	0	0	0	1	0	1	0	1	1
YKACHI	0	0	1	0	0	0	0	0	0	1	0	1	1	1	1
YKAMU	0	0	1	1	1	0	0	1	0	1	0	0	0	0	0
YKAPU	0	0	1	1	1	1	0	1	0	1	0	0	0	0	0
YKARI	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0
YKU	0	0	1	1	1	1	1	1	0	1	0	0	0	1	1
YSI	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0
KACHA	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1
KAMU	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1
KAPU	0	1	1	1	1	1	1	0	1	1	0	0	0	0	0
KU	0	0	1	0	0	1	1	0	1	0	0	0	1	1	1

Table 3: Boolean Matrix of IPS combinations

The rule that describes the corresponding combinations of two IPS suffixes is implemented in NooJ as follows:²

SIP2_TR_INF = (:CHAKUCHKA | :CHICHI | :CHICKA | :KURQU | :LLAVCHKA | :LLAVYKACHI | :NAYACHKA | :NAYALLAV | :PAYAYSI | :PAYAKACHA | :PAYAKAMU | :PAYAKAPU | :PAYAKU | :PAYALLAV | :PAYAMU | :RAMU | :RICHI | :RICHKA | :RIYKACHI | :RPARILLAV | :RPARIMU | :TAMUCHKA | :TAMUYKACHI | :TAMUYKAMU | :YKURQU) y/V;

When applied to any fundamental verb, the valid combination of two IPS suffixes will produce 240 new verbs. The following is an extract of the dictionary of verbs derived from the verb *rimay* [to talk]:

rimachimuy, *rimay*, V+FACT+ACENT+INF
rimarparirquy, *rimay*, V+ASUR+PAPT+INF
rimarparimuy, *rimay*, V+ASUR+ACENT+INF
 ...

The corresponding translations are:

rimachimuy [make him talk]
rimarparimuy [go to talk to all of them]
rimaykurquy [say your greetings rapidly]

4.2 Automatic verbal inflection

Quechua Transitive verbs are conjugated according to the following rule:

V_TR_CONJO = :PR | :F | :IP | :RQA_PR_PLU | :C | :RQU_PR_PLU | :SPA_PRM2_PLU |

² For the detailed description of all the paradigms (CHA, KU, CHKA, CHAKUCHKA..etc) appearing in the article in capital letters see Duran(2017).

:RU_PR_PLU | :PTIPRM1QA | :F_ÑA | :PTI_PRM2_PLU | :V_SIP0_STIN | :V_RECIP_CONJO;

Following is the rule used to conjugate intransitive verbs:

V_ITR_CONJO = :PR | :F | :IP | :RQA | :C | :RQU_PR | :SPA_PRM2_PLU | :SPA | :RU_PR | :RA_PR | :RQU_F | :PASSA | :SQA_PRM2_PLU | :RUPRM1MAN | :RQUPRM1MAN | :PTIPRM1QA | :F_ÑA | :PTI_PRM2_PLU | :NAV_PRM2 | :NAV_PRM2_MM | :SQA_PR;

Following is a rule used to conjugate pronominal verbs:

V_PRON_CONJO = :PR_PRON | :F_PRON | :IP_PRON | :RQA_PR_PLU_PRON | :C_PRON | :RQU_PR_PLU_PRON | :SPA_PRON | :SPA_PRM2_PLU_PRON | :RA_PR_PRON | :PTI_PRM2_PRON | :NAV_PRM2_PRON | :NAV_PRM2_MM_PRON | :V_STIN_PRON;

Following is the rule used to conjugate impersonal verbs:

V_IMP_CONJO = :PR_IMP | :F_IMP | :I_IMP | :RQA_IMP | :C_IMP | :SPA | :SPA_PRM2_IMP | :PTIPRM1QA_IMP | :RQU_IMP | :SPA | :F_ÑA_IMP | :STIN_IMP | :PTI_PRM2_IMP | :NAV_PRM2_IMP | :SQA_SS_IMP;

All rules can be applied to a verb derived by adding one or more IPS suffixes to a fundamental verb such as *asiriy* [to smile]. NooJ can then produce the list of all conjugated forms of all the verbs derived using one or more IPS suffixes.

One can automatically obtain all the conjugated forms of the 1,444 fundamental verbs, as well as their derived verbs, using one or two IPS suffixes. By combining the inflectional and the derivational grammar rules and applying them to the dictionary of fundamental Quechua verbs, NooJ has generated 2,777,418 verbal forms (see Figure 1), including for the verb *rimay* [to talk]:

0 IPS: *rimanqa* [he will talk]
 1 IPS: *rimaykurqanchik* [we saluted]
 2 IPS: *rimapayaykunku* [they are arguing]

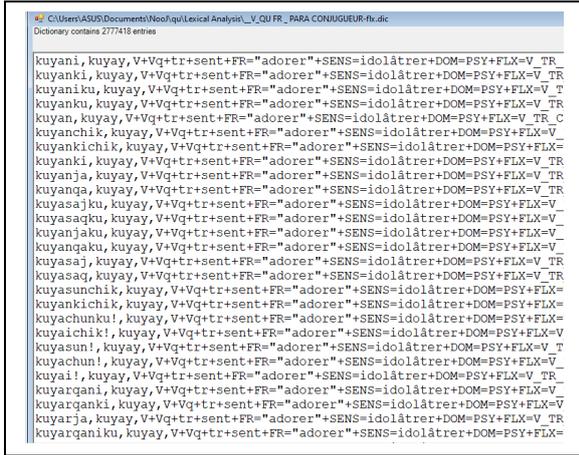


Figure 1: NooJ produced 2,777,418 conjugated forms of simple and derived verbs

4.3 Postposed suffixes

The second set of suffixes that generate a different kind of inflected forms are the Post Posed Suffixes (PPS) which appear after the Indefinite Tense endings (ITE), or the Future Tense Endings (FTE), e.g.:

rimaniña [I already talked]

Here the PPS suffix “*ña*” appears after the ITE suffix “*ni*”.

rimasaqraq [I will talk before]

Here the PPS suffix “*raq*” appears after the FTE suffix “*saq*”. The following paradigm formalizes the use of one PPS:

PPS1 = :CHAAV | :CHV | :CHIKV |
 :CHIKIV | :CHUIV | :CHUNV |
 :CHUSINAV | :MAAV | :MV | :ÑAV |
 :PASV | :PUNIV | :QAV | :RAQV |
 :SV | :TAQV | :YAAV;

When applied to the verb *rimay* [to talk], this paradigm generates 17 verbal forms including the following ones:

rimankichu [you did not talk]
rimanmi [he talked indeed]
rimanjakus [they say that they will talk]
rimanmi [I did talk]

Verbs can accept one, two or three PPS suffixes, e.g.:

2 PPS: *rimankiraqchu?* [Will you still talk?]

3 PPS: *rimanchikmanñataq* [What about if we should talk?]

4.4 Mixed suffix agglutination

When we analyze some type of conjugated verb forms, it’s interesting to remark that the ITE and the FTE suffixes behave as fixed points, around which the two sets of suffixes IPS and PPS appear agglutinated to produce a general verbal form for the Indefinite Tense, e.g.:

llamkachkanraqmi [He is still working]

or, for the Future Tense, e.g.:

llamkachkanqaraqmi [He will be still working]

In the first verbal form, the IPS suffix “*chka*” is positioned between the verb root *llamka* [work] and the verbal ending “*n*” mark of the third singular person in the indefinite tense, and the two PPS suffixes *raq* and *mi* located after the ending. These verbal forms occur actually very frequently in daily communication.

Following are more examples of verbal forms that contain ITE endings used as fixed points:

mikullachkankuña [fortunately, they are already eating]

Here, there are two IPS: *lla* and *chka*, and one PPS: *ña*.

pukllachichkaniraqmi [I am making him play]

Here, there are two IPS *chi* and *chka* and two PPS *raq* and *mi*.

mikuchiyykuchkanillaraqmi [I am carefully helping him to eat before anything else]

We have implemented grammar rules for these mixed constructions. For instance, the following one represents the insertion of one or two IPS and one PPS:

V_MIX21 = :SIP2_PR_V :SPP1_PR_V |
 :SIP2_PR_C :SPP1_PR_C;

For a detailed description of each paradigm, see Duran (2017). Based on our dictionary of 1,444 verbs and the inflectional and derivational grammars, NooJ can recognize any Quechua verbal inflected and derived verbal form,

lemmatize and analyze it, and reciprocally produce 2,777,418 Quechua inflected and derived verbal forms.

5 Translating Spanish to Quechua

Because there are linguistic resources for many languages that have been formalized with the same NooJ platform³ using the same formalism, it becomes relatively easy to align them and thus construct bilingual lexical resources.

5.1 The Spanish-Quechua Verb Dictionary

To produce a Quechua translation of any Spanish verbal form, we need both the Quechua dictionary and a Spanish dictionary.

There are few available open-source Spanish electronic dictionaries, and even fewer bilingual Spanish to Quechua (SP-QU) dictionaries: The dictionary used by A. Rios (2015) for Spanish utilizes a set of part of speech tags from the Royal Academy of Spanish and, for the Quechua language, a set of tags corresponding to the Quechua Cuzco variant.

The NooJ website (<https://nooj.univ-fcomte.fr>) offers an electronic version of the RAE (Real Academia Española) dictionary implemented by (Fuentes, Gupta, 2014). It contains over 61,000 entries, including about 11,000 simple and compound verbs.

Using the tools offered by NooJ, we were able to align the Quechua list of fundamental and derived verbs with this list of 11,000 Spanish verbs, thus producing a bilingual Spanish-Quechua dictionary, see Figure 2.

```

caminar, V+mov+QU="puriy"+FLX=AMAR
caminar, V+mov+QU="riy"+FLX=AMAR
caminar, V+tr+QU="ichiy"+FLX=AMAR
canalizar, V+tr+QU="yaku-fianta ruray"+FLX=TRAZAR
canalizar, V+tr+QU="yarjachay"+FLX=TRAZAR
cancelar, V+tr+QU="jopitay"+FLX=AMAR
cancelar, V+tr+QU="jopoy"+FLX=AMAR
canjear, V+tr+QU="yamkiy"+FLX=AMAR
cansar, V+itr+Tec+Psy+QU="amiruy"+SENS=fatigar+FLX=AMAR
cansar, V+itr+Tec+Psy+QU="amiy"+SENS=fatigar+FLX=AMAR

```

Figure 2: SP-QU Electronic Dictionary

In this bilingual dictionary:

- Each entry is associated with its part of speech category (“V” for verbs), some

³ To download linguistic resources in the NooJ format for over 30 languages, see: <https://nooj.univ-fcomte.fr/resources.html>.

syntactic and/or semantic property (e.g., “+tr” for transitive verbs) and an inflectional paradigm (e.g., “+FLX=AMAR” for the conjugation paradigm AMAR).

- Each entry is associated with its Quechua translation (property “+QU”).

This dictionary contains simple verbs as well as multiword verbs such as:

andar a ciegas → +QU=*taplaykachay* [to tangle]
andar a paso lento → +QU=*purichakuy* [to walk slowly]

There are many cases of polysemy, both on the Spanish and on the Quechua sides. Some Spanish verbs may have more than one translation. For example, *caminar* [to walk] is associated with three Quechua translations: *puriy*, *riy*, *ichiy*.

5.2 Translating conjugated forms

This SP-QU dictionary only contains the lemma, i.e., the infinitive form of each Spanish verb. However, to develop a Machine Translation system, we need to associate each Spanish conjugated form with its corresponding form in Quechua. Although there is no perfect correspondence between Spanish and Quechua tenses, moods and aspects, we entered the closest Quechua form for each Spanish form. Table 4 presents an extract of the resulting dictionary.

Abreviación y su descripción en SP (Español) y QU (quechua)			
Equivalencias simbólicas esenciales entre SP y QU			
SP	QU	Flexión en SP	Valor en QU
inf	INF	inf infinitivo	INF infinitivo
ppio	GER1	ppio participio indefinido	GER1 gerundio flexionado
ger	GER	ger gerundio	GER gerundio -spa simultáneo
pres	PR	pres presente	PR presente
ppp	PPA	ppp pasado pluscuam perfecto	PPA -sga PR mikusqaani
pps	PS	pps pasado simple (cantaste)	PS pasado simple
fut	F	fut futuro	F futuro
cond	C	cond condicional	C condicional
ppi	PASRA	ppi pret. perfecto indef (cantabas)	PASRA pasa. realizado (takirani)
pasmr	PASRQU	pasmr pasado modo rápido	PASRQU pasado modo rápido
subi	SUBJ	subi subjuntivo	SUBJ subjuntivo spaga
subjmr	SUBJMR	subjmr subj. modo rápido rqusga	SUBJMR subj. modo rápido rqusga
imp	IP	imp imperativo	IP imperativo
pp	PPI	pp pretérito perfecto	PPI pretérito perfecto rqa
1+s	1+s	1+s primera persona singular	1+s primera persona singular
2+s	2+s	2+s segunda persona singular	2+s segunda persona singular

Table 4: Verbal properties in Spanish and Quechua

Using NooJ to produce all inflected forms corresponding to the 11,000 Spanish verbal entries, we produced a list SP-QU of over 618,000 Spanish verbal forms associated with their Quechua translation. Then, using simple SQL operations, we reversed the SP-QU dictionary and produced the reverse Quechua to Spanish dictionary. Figure 3 shows an extract of it:

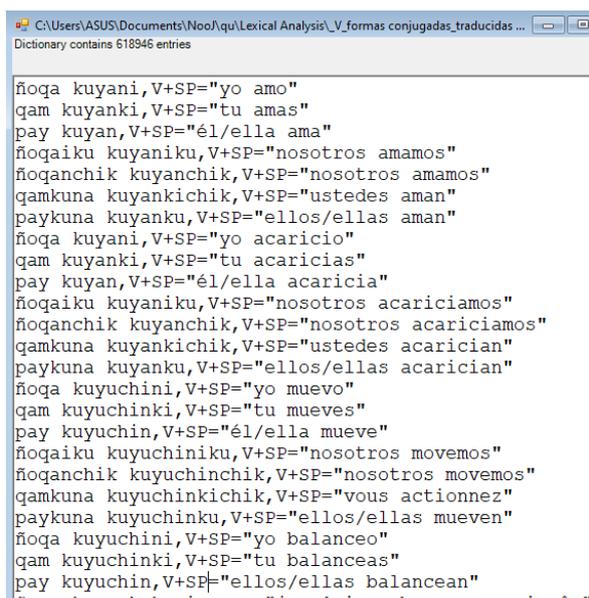


Figure 3: QU-SP Electronic dictionary.

This list constitutes the most comprehensive bilingual available lexical resource: it currently contains over 618,000 simple and compound lexical entries translated into Spanish, each entry being associated with their respective inflectional and derivational paradigm.

6 Evaluation

We have used NooJ to apply this Quechua-Spanish dictionary to a corpus consisting of ten story tales in Quechua (see references), totaling 16,874 words. Applying the query <V> (to extract all Quechua verbal forms) produced a concordance of 227 entries, out of 234 really present in the text, i.e., a 97% recall. Out of these 227 entries, 2 words were incorrectly recognized as verbs: *llunpay* [too much] and *patanta* [the board], i.e., a 99% precision. It must be noted though that these two word forms could actually be verbal forms, in other syntactic contexts: we will develop and apply NooJ local syntactic grammars to disambiguate them automatically.

7 Conclusion and perspectives

Using the NooJ development environment platform, we have implemented a set of resources that formalizes the morphology of the Quechua verbal system. These resources contain a dictionary of fundamental Quechua verbs (about 1,444 entries), a derivational grammar that formalizes the derivations of each fundamental verb using one, two or three interposition suffixes,

and an inflectional grammar that describes the conjugation of all fundamental and derived verbs. Based on these formalized resources, NooJ produced over two million Quechua verbal forms automatically. We aligned the resulting list with the list of Spanish conjugated verbal forms, resulting in a Spanish to Quechua database that associates any Spanish conjugated verb with its Quechua translation. Finally, using simple SQL operations, we reversed the database and produced a dictionary that associates 618,946 Quechua derived inflected verbal forms with their Spanish translation.

8 References

- Adelaar, W. (1979). Tarma Quechua. Grammar, Texts, Dictionary. In: L'Homme.
- Duran, M. (2014, June). Formalizing Quechua verb inflections. In Formalising Natural Languages with NooJ 2013: Selected Papers from the NooJ 2013 International Conference (p. 41). Cambridge Scholars Publishing.
- Duran, M. (2017). Dictionnaire électronique français-quechua des verbes pour le TAL (Doctoral dissertation, Université Bourgogne Franche-Comté).
- Fuentes, S., Gupta A. (2014). Updated Spanish Module for NooJ. In Formalising Natural Languages with NooJ 2013: Selected papers from the NooJ 2013 International Conference (p. 51). Cambridge Scholars Publishing.
- Gross, M. (1992). Forme d'un dictionnaire électronique, Actes du colloque. "L'environnement traductionnel de Mons". AUELUF-UREF.
- Holguin, D. (1608). Vocabulario de la Lengua General de todo el Perv llamada lengua Qquichua, o del Inca. Corregido y renovado conforme a la propiedad cortesana del Cuzco. Ciudad de los Reyes Lima, Impreso por Francisco del Canto.
- Parker, Gary J. (1973). Derivacion verbal en el quechua de Ancash. Working Paper 25. Lima : Universidad Nacional Mayor de San Marcos, Centro de Investigación de Lingüística Aplicada.
- Perroud, P. (1970). Dictionario castellano kechwa, kechwa castellano. Dialecto de Ayacucho. Santa Clara, Peru. Seminario San Alfonso.
- Rios, A. (2015). A basic language technology toolkit for quechua (Doctoral dissertation, University of Zurich).
- Silberztein, M. (2003-). NooJ Manual. Available at <https://nooj.univ-fcomte.fr>.

Silberstein, M. (2016). Formalizing natural languages: The NooJ approach. John Wiley & Sons.

Soto C. (1976). Diccionario quechua ayacuchochanca, (coll. "Lengua y sociedad", 4), 183 p.

Taylor, G. (1979). Morphologie comparée du verbe quechua : l'expression de l'actance, Première partie : Le sujet, en relations prédicat-actance(s) dans des langues de types divers II, ed. C. Paris. LACITO-documents Eurasie 3. Paris : SELAF, 171-86.

Wallace K. (1988). Parsing Quechua Morphology for Syntactic Analysis. University of California. UCLA. Los Angeles, California.

1ra Edición. IEP, Instituto Francés de Estudios Peruanos. Lima. Perú.

Lexical Resources available on the WEB

<https://www.quechua.org.uk/Eng/Main/>

<https://infolingu.inivmlv.fr/DonneesLinguistiques/Dictionnaires/telechargement.html>

<https://www.cairn.info/revue-langages-2010-3-page-221.htm>

https://www.freelang.com/dictionnaire/quechua_cuzco.php

http://www.lexilogos.com/quechua_dictionnaire.htm

<https://fr.glosbe.com/es/qu/>

<http://aulex.org/qu-es/>

Our corpus of fairy tales

Andahuaylas local journal (2010-2011).

de Avila, Francisco (2012). *Dioses y hombres de Huarochiri*. Narración quechua recogida por Francisco de Avila 1598. Traducción J. M. Arguedas. Lima. Perú 1966. (178p). Edición bilingüe facsimilar.

Guardia Mayorga, Cesar (1973). *Pakpaku chayñawan rimanakun*. Conte (p 383-84) Gramática Kechwa Ediciones Los Andes. Lima Perú.

Lira, Jorge (1990). *Cuentos del Alto Urubamba*. Centro Bartolomé de las Casas, Cuzco Eds.

Meneses, Porfirio (2001). *Sept contes en quechua* (1. Yupinta, 2. Hukpa wasin, Kawsayninchik kutimuptin, 4. Muti maskaypi, 5. Tayta Matias, 6. Warmichaykita waylluy, 7. Chiqnipacha. Edition quechua-français. Auteur C. Itier. Langues et Mondes Eds. Paris.

Oregón Morales, José (1994). *Loro Ccolluchi. Exterminio de loros y otros cuentos*. Lluvia Editores. Lima, 1994.

Taylor, G. (1987). *Ritos y Tradiciones de Huarochiri, Manuscrito Quechua de comienzos del siglo XVII,*