

The Annotation of Compound Suffixation Structure of Quechua Verbs

Maximiliano Duran^(✉)

Université Franche-Comté, Besançon, France
duran_maximiliano@yahoo.fr

Abstract. In the Quechua language we find less than 1400 simple verbs. But, the language has several strategies to increase this verb lexicon generating new verbs by derivation, making use of a particular set of suffixes. First we construct a Boolean matrix showing the valid compounds of the suffixes doing this job. Then we show how we have programmed certain number of morpho-syntactic NooJ grammars to generate the corresponding compound verbs (over 43160). As a result we present a dictionary of the lexicalized compound verbs, including their Spanish and French translations. We have applied both the dictionary and some grammars for automatically annotate a Quechua text obtaining near 90 % of successful matches.

Keywords: Quechua · Compound suffixation · NooJ grammars · Compound verbs · Quechua morphology · Verb morphology · NLP applications · Verb inflection and derivation

1 Introduction

The inventory of Quechua verbs, carried out on the existing paper dictionaries, shows us a lexicon of less than 1400 simple verbs.

In a previous article¹ we have reported that we have been able to increase this list to near 2000 by the addition of some hundreds of verbs, obtained them by parsing some NooJ grammars on our corpus in which they were imbedded in the form of derivations using compound suffixations.

For example the form *asirichiy* appears in the corpus translated as: “make him smile”. We notice that it contains the compound suffix *-ri-chi-* which can be analyzed as follows. *chi*: factitive, make someone do something *ri*: dynamism, to start doing the action defined by the verb. The remaining morpheme *asi-* is the lemma of the quechua verb to laugh. Thus *asiy*: to laugh has been derived by the suffixes *-ri-*, *-chi-* to give the new verb to smile.

Let us see another example: the form *rantikuy* appears translated in the corpus as: to sell. We notice that it contains the suffix *-ku-*. It can be analyzed as follows: The suffix *-ku-*: auto benefic, has induced on the original lemma *ranti-* to buy (acquiring something), a change of semantic field into to sell (get rid of something).

¹ Morphological and syntactic grammars for recognition of verbal lemmas in Quechua. To appear in Proceedings of the 2014 International NooJ Conference. Sassari. Italy.

2 Generation of Quechua Verbs

We might be inclined to think that the fact of having less than 1400 simple verbs could have been a handicap for the production of any kind of extensive literature like French for instance, which has several thousands of verbs. For this language, Dubois, Jean et Dubois-Charlier, Françoise (D&D) have inventoried more than 25000 entries in their dictionary² «Dictionnaire électronique des verbes (français)»

But, Quechua presents a remarkable strategy for generating new verbs by derivation of the simple ones as we have just seen. For this, it makes use of a set of 26 interposition suffixes IPS³. To illustrate this, let us take the simple verb *llamkay* (to work), which is formed by the verbal lemma *llamka-* and the infinitive suffix *-y*. Interposing the suffix *-isi-* between them we obtain the derived verb *llamka-isi-y* (to help someone to work). Joining to the same lemma the suffix *-chka-* (the action is being executed, it's similar to the role of the progressive particle *ing* in English), we obtain the new verb *llamka-chka-y* (to keep working). The parsing of the NooJ grammar of Fig. 1 or its algebraic expression $V_SIP1_INF^4$ on the set of 1400 verbs generates 364000 compound Quechua verbs. Besides the examples, some of these are lexicalized well known verbs like the ones appearing in Table 1.

We might be inclined to think that, the fact of having less than 1400 simple verbs could have been a handicap for the production of any kind of extensive literature like French for instance, which has several thousands of verbs. For this language, Dubois, Jean et Dubois-Charlier, Françoise (D&D) have inventoried more than 25000 entries in their dictionary⁵ «Dictionnaire électronique des verbes (français)»

But, Quechua presents a remarkable strategy for generating new verbs by derivation of the simple ones as we are going to see. For that, it makes use of a set of 26 interposition suffixes IPS⁵. To illustrate this, let us take the simple verb *llamkay* (to work), which is formed by the verbal lemma *llamka-* and the infinitive suffix *-y*. Interposing the suffix *-isi-* between them we obtain the derived verb *llamka-isi-y* (to help someone to work). Joining to the same lemma the suffix *-chka-* (the action is currently executed, it is similar to the role of the progressive particle *ing* in English), we obtain the new verb *llamka-chka-y* (to keep on working). The parsing of the NooJ grammar of Fig. 1 or its algebraic expression $V_SIP1_INF^6$ on the set of 1400 verbs generates 36 4000 compound

² Jean Dubois et Françoise Dubois-Charlier (2007).

³ IPS = (*chi, chka, ikacha, ikachi, ikamu, ikapu, ikari, iku, isi, kacha, kamu, kapu, ku, lla, mpu, mu, naya, pa, paya, pu, raya, ri, rpari, rqu, ru, tamu*).

⁴ $V_SIP1_INF = \langle B \rangle (\text{chi/FACT} \backslash \text{chka/PROG} \backslash \text{ikacha/DISP} \backslash \text{likachi/POL1} \backslash \text{likamu/PREAT} \backslash \text{ikapu/SOIT3} \backslash \text{likari/PONC} \backslash \text{liku/COURT} \backslash \text{lisi/COLL} \backslash \text{lkacha/ARO} \backslash \text{lkamu/AOL} \backslash \text{kapu/RAS} \backslash \text{ku/AUBE} \backslash \text{lla/POL1} \backslash \text{mpu/INSP} \backslash \text{mu/ACENT} \backslash \text{naya/ENV} \backslash \text{pa/PEAU} \backslash \text{paya/FREQ} \backslash \text{pu/APT} \backslash \text{raya/DUR} \backslash \text{ri/DYN} \backslash \text{rpari/ASUR} \backslash \text{rqu/PAPT} \backslash \text{ru/PRES} \backslash \text{tamu/AEP}) (\text{y/INF});$

⁵ IPS = (*chi, chka, ikacha, ikachi, ikamu, ikapu, ikari, iku, isi, kacha, kamu, kapu, ku, lla, mpu, mu, naya, pa, paya, pu, raya, ri, rpari, rqu, ru, tamu*).

⁶ $V_SIP1_INF = \langle B \rangle (\text{chi/FACT} \backslash \text{chka/PROG} \backslash \text{ikacha/DISP} \backslash \text{likachi/POL1} \backslash \text{likamu/PREAT} \backslash \text{ikapu/SOIT3} \backslash \text{likari/PONC} \backslash \text{liku/COURT} \backslash \text{lisi/COLL} \backslash \text{lkacha/ARO} \backslash \text{lkamu/AOL} \backslash \text{kapu/RAS} \backslash \text{ku/AUBE} \backslash \text{lla/POL1} \backslash \text{mpu/INSP} \backslash \text{mu/ACENT} \backslash \text{naya/ENV} \backslash \text{pa/PEAU} \backslash \text{paya/FREQ} \backslash \text{pu/APT} \backslash \text{raya/DUR} \backslash \text{ri/DYN} \backslash \text{rpari/ASUR} \backslash \text{rqu/PAPT} \backslash \text{ru/PRES} \backslash \text{tamu/AEP}) (\text{y/INF});$

Quechua verbs. Some of these are lexicalized well known verbs like the ones appearing in the following table.

Table 1. Lexicalized compound Quechua verbs

Lexicalized V	English V	V Lemma	IPS
<i>rimaikuy</i>	greet	<i>rima</i> (talk)	<i>-iku-</i>
<i>amichiy</i>	bore	<i>ami</i> (tired of N)	<i>-chi-</i>
<i>atipay</i>	win	<i>ati</i> (can)	<i>-pa-</i>
<i>aiqiryi</i>	start fleeing	<i>ayqi</i> (flee)	<i>-ri-</i>
<i>aysariy</i>	tow	<i>aysa</i> (carry)	<i>-ri-</i>

But most of them are relatively unknown ones as we will see soon.

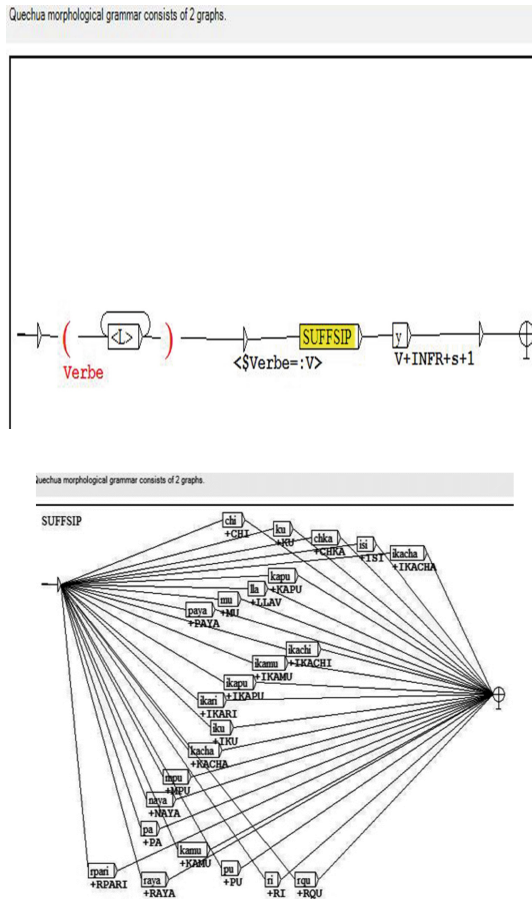


Fig. 1. The NooJ grammar that generates one dimension compound verbs using the 26 inter positioned suffixes IPS

2.1 Combinations of Two Interposed Suffixes

The Quechua grammar allows agglutinations of IPS and consequently to obtain more new verbs. We can have combinations of two or more of them. For instance the combination *-chka-isi-* which can be added to the lemma *llamka-* to obtain the new verb *llamka-chka-isi-y* (to keep helping someone to work). However the permutated combination **-isi-chka-* is not grammatically correct. To answer the question of which two-fold combinations are grammatical, we have built manually the matrix of Fig. 2 based on field work. There, the valid combination is noted by 1 and the invalid ones by 0.

Matrix binaire avec suffixes interposés																					
	CHI	CHKA	IKACHA	IKACHI	IKAMU	IKAPU	IKARI	IKU	ISI	KACHA	KAMU	KAPU	KU	LLAV	MPU	MU	NAYA	PA	PAYA	PU	RAYA
CHI	1	1	0	1	1	1	0	1	1	0	1	0	1	1	0	1	0	0	0	1	0
CHKA	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
IKACHA	1	1	0	0	0	0	0	0	1	0	1	0	1	1	0	1	0	0	0	1	0
IKACHI	0	1	0	0	1	1	1	0	1	0	1	1	1	1	0	1	0	0	1	1	1
IKAMU	0	1	1	1	0	1	1	0	1	1	0	0	0	1	0	0	1	0	1	0	1
IKAPU	1	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	1	0	1	0	1
IKARI	1	1	0	0	0	0	0	1	1	0	0	0	1	1	0	1	0	0	1	1	0
IKU	1	1	0	0	0	0	0	0	1	0	0	1	1	1	0	1	0	0	1	1	0
ISI	1	1	0	0	0	0	1	1	0	0	0	0	1	1	1	1	0	0	1	1	1
KACHA	1	1	0	0	1	1	0	1	0	0	1	1	1	1	0	1	1	0	1	1	1
KAMU	0	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	1	0	1	0	1
KAPU	1	1	1	1	1	1	0	1	1	0	0	0	0	1	0	1	0	0	0	0	1
KU	0	1	0	0	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	1
LLAV	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1
MPU	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
MU	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
NAYA	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1
PA	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	1	1	0	0	0	1
PAYA	1	1	1	0	0	0	1	1	1	1	1	1	1	1	0	1	1	0	0	1	1
PU	0	1	0	1	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
RAYA	1	1	1	1	1	1	0	0	1	0	0	0	1	1	0	1	0	0	0	0	1
RI	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1	1	0	0	1	1
RPARI	1	1	0	1	1	1	0	1	1	0	1	1	1	1	0	1	0	0	0	0	1

Fig. 2. Boolean matrix of two-fold IPS combinations

We have verified that at least 292 are grammatically correct new verb generators. Here is a sample of the resulting agglutination of 2 dim IPS compounds:

- CHIMU = :CHI :MU;
- CHIPU = :CHI :PU;
- CHKAIKAMU = :CHKA :IKAMU;
- CHKAIKAPU = :CHKA :IKAPU;
- IKURQU = :IKU :RQU;
- ISICHI = :ISI :CHI;

Which allows us to write Nooj grammars to generate new 2-fold IPS compound verbs:

```
V_SIP2_INF = <B> (:CHICHI |:CHICHKA |:CHIIKACHI |:CHIIKAMU |:CHIIKAPU |:CHII-
KARI |:CHIIKU |:CHIIISI |:CHIKAMU |:CHIKU |:CHILLAV |:CHIMU |:CHIPU |:CHIRPARI
|:CHITAMU |:CHKAIKACHI ... |:IKAMUCHKA |:IKAMUIKACHA |:IKAMUIKACHI
|:IKAMUIKAPU |:IKAMUIKARI |:IKAMUISI |:IKAMUKACHA |:IKAMUKU |:IKAMULLAV
|:IKAMUNAYA |:IKAMUPAYA)y/INF;
```

	A	B	C
2931	TAMUIKAPURA= :TAMU :IKAPU :RA;		
2932	TAMURQUCHKA= :TAMU :RQU :CHKA;		
2933	TAMURQUISI= :TAMU :RQU :ISI;		
2934	TAMURQULLAV= :TAMU :RQU :LLAV;		
2935	TAMURACHI= :TAMU :RA :CHI;		
2936	TAMURAPU= :TAMU :RA :PU;		
2937	RACHICHKA= :RA :CHI :CHKA;		
2938	RACHIIKACHI= :RA :CHI :IKACHI;		
2939	RACHIIKAMU= :RA :CHI :IKAMU;		
2940	RACHIIKAPU= :RA :CHI :IKAPU;		
2941	RACHIIKARI= :RA :CHI :IKARI;		
2942	RACHIIKU= :RA :CHI :IKU;		
2943	RACHIIISI= :RA :CHI :ISI;		
2944	RACHIKAMU= :RA :CHI :KAMU;		
2945	RACHIKU= :RA :CHI :KU;		
2946	RACHILLAV= :RA :CHI :LLAV;		
2947	RACHIMU= :RA :CHI :MU;		
2948	RACHIPU= :RA :CHI :PU;		
2949	RACHIRPARI= :RA :CHI :RPARI;		
2950	RACHITAMU= :RA :CHI :TAMU;		
2951	RAMUCHKA= :RA :MU :CHKA;		
2952	RAMULLAV= :RA :MU :LLAV;		

Fig. 3. 3-dimension grammatical verb-generators

Moreover, these combinations are capable of generating three fold agglutinations by adding again one IPS. The respective Boolean matrix contains 7592 entries, but not all are grammatical. Manual verification yields only 2952 “1’s”, i.e. grammatically correct compounds. Figure 3 shows a sample of the last 22 of this list.

V_IPS4
chi-iku-na-lla
ku-lla-chka-rqa
chi-isi-mu-chka
 Agglutinations of five dimensions
 V_IPS5
chi-ku-na-lla-pti
chi-ku-lla-wa-pti
chi-isi-mu-chka-pti

3 The Compound Verbs

Agglutinating these grammatical compound suffixes to the 1400 simple verb lemmas we obtain 43160 grammatically correct compound verbs. In Fig. 4 we present a sample of them.

	A	B	C	D	E	F	G	H
43141	yuyapuy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_APT	INF	
43142	yuyapayay	yuyay	V	TR	FR="rappele	FLX=V_SIP1_FREQ	INF	
43143	yuyapay	yuyay	V	TR	FR="rappele	FLX=V_SIP1_PEAU	INF	
43144	yuyanayay	yuyay	V	TR	FR="rappele	FLX=V_SIP1_ENV	INF	
43145	yuyamuy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_ACENT	INF	
43146	yuyampuy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_INSP	INF	
43147	yuyallay	yuyay	V	TR	FR="rappele	FLX=V_SIP1_POL1	INF	
43148	yuyakuy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_AUBE	INF	
43149	yuyakapuy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_RAS	INF	
43150	yuyakamuy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_AOL	INF	
43151	yuyakachay	yuyay	V	TR	FR="rappele	FLX=V_SIP1_ARO	INF	
43152	yuyaisiy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_COLL	INF	
43153	yuyaikuy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_COURT	INF	
43154	yuyaikariy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_PONC	INF	
43155	yuyaikapuy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_SOIT3	INF	
43156	yuyaikamuy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_PREAT	INF	
43157	yuyaikachiy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_POL1	INF	
43158	yuyaikachay	yuyay	V	TR	FR="rappele	FLX=V_SIP1_DISP	INF	
43159	yuyachkay	yuyay	V	TR	FR="rappele	FLX=V_SIP1_PROG	INF	
43160	yuyachiy	yuyay	V	TR	FR="rappele	FLX=V_SIP1_FACT	INF	

Fig. 4. A sample of the generated dictionary of 43000 compound verbs

4 The Semantics of the Agglutinations

But, grammatically correct forms do not necessarily mean meaningful forms. For instance what is the precise meaning of the verb *llamkarachitamuy?* or *tiyarachitamuy*, where *llamka-*: to work and *tiya-*: to sit are the lemmas and *rachitamuy* is the valid combination of the suffixes *-ra-*, *-chi-*, *-tamu-*.

Table 2. Neologisms proposed instead of loans in use

Generated compound verb	Originating simple verb	neologism	may avoid the Spanish loan word
aysariy, V + FR = “re morquer, traîner un petit bout de chemin”	aysay : to pull, to drag	aysariy: to tow	rimulkay < remorcar (to tow)
yapariy, V + FR = “raj outer un peu de quelque chose”	yapay : to add	yapariy: to award a prise	premiay < premiar (to award a prise)
kausariy, V + FR = “r evenir à la vie, revenir à soi”	kausay : to live	kausariy: to revive	resusitay < resucitar (to revive)
qipariy, V + FR = “se retarder un peu lentement”	qipay : to delay a little	qipariy : to get delayed	tardiyary < tardarse (to get delayed)
utiriy, V + FR = “devenir fou”	utiy : to go crazy	utiriy : to go crazy	lukuyay < alocarse (to go crazy)

Table 3. Semantic values for suffixes IPS

#Suffixes inter posées entre le lemme verbal et la désinence de personne
chi,SIP+FACT+"VS=invite FACT_2 autorise FACT_3 pousse l'objet du verbe à réaliser l'action
chka,SIP+PROG+"VS=en train déréaliser l'action
ikacha,SIP+DISP+"VS=en désordre DISP_2 désorienté réalise l'action
ikachi,SIP+POLI+"VS=poliment POLI_2 concrètement POLI_3 précisément POLI_4 courtoisement
ikamu,SIP+PREAT+"VS=vers le sujet PREAT_2 en prévoyant PREAT_3 attentionné PREAT_4 de haut en bas
ikapu,SIP+SOIN3+"VS=avec attention SOIN3_2 soigneusement SOIN3_3 concernant un tiers SOIN3_4 recommençant
ikari,SIP+APRP+"VS= ponctuelle et rapidement APRP_2 à la hâte mais avec précision réalise l'action
iku,SIP+COURT+"VS=courtoisement COURT_2 soigneusement COURT_3 amicalement COURT_4 vers le sujet
isi,SIP+COLL+"VS=collabore COLL_2 aide réalise l'action
kacha,SIP+ARO+"VS=collabore ARO_2 aide réalise l'action
kamu,SIP+AAR+"VS=aller àréaliser l'action
kapu,SIP+RAS+"VS=auto bénéfécie réalise l'action
ku,SIP+AUBE+"VS=se responsabilisant AUBE_2 affectueusement AUBE_3 s'impliquant réalise l'action
lla,SIP+POLI+"VS=gentiment POLI_2 poliment POLI_3 doucement POLI_4 emphatiquement
mu,SIP+ACENT+"VS=se déplaçant ACENT_2 sous contrainte réalise l'action
mpu,SIP+INSP+"VS=inoipinément réalise l'action
na,SIP+OBL+"VS=obligé POLI_2 potentiellementréaliser l'action
naya,SIP+ENV+"VS=envié de ENV_2 souhaiteréaliser l'action
pa,SIP+PEAU+"VS=peaufine PEAU_2 réitère PEAU_3 réalise l'action
paya,SIP+FREQ+"VS=répétition Fréquente FREQ_2 persiste réalise l'action
pu,SIP+APT+"VS=en substitution APT_2 en action centrifuge réalise l'action
ra,SIP+PASS+"VS=avais réalise l'action
raya,SIP+DUR+"VS=demeure un temps pour réalise l'action
zi,SIP+DYN+"VS=commence à DYN_2 recommence à réalise l'action
rpari,SIP+ASUR+"VS=action surprise ASUR_2 impulsivement réalise l'action
rqu,SIP+PAFP+"VS=accompli en peu de temps APT_2 en action centrifuge réalise l'action"
ru,SIP+PRES+"VS=pressantele_sujetPRES_len_peu_de_tempsPRES_2 de_façon_pressante réalise_l_action
tamu,SIP+AEP+"VS=en passant réalise_l_action"

What are the actual meanings of this amazing quantity of the generated new verbs? Are they really currently used by the native speakers? Which ones are really meaningful in the language?

Many are certainly candidates to become neologisms like in the following table:

But many others seem not to have a plausible meaning.

We are aware that the only way to answer these questions is by hand verification on the field, nevertheless to ease this task we have written some NooJ grammars which give us, as a first step, the corresponding annotations of the suffixes contained in the verbal form, like in Fig. 5. Then, it proposes automatically the glossed translation. For this, we

```
#use grammaire_Verbs_SIP_INF.nof
#
kuyaramullay,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+ACENT+POLI+INF
kuyaramuchkay,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+ACENT+PROG+INF
kuyarachitamuy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+AEP+INF
kuyarachirpariy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+ASUR+INF
kuyarachipuy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+APT+INF
kuyarachimuy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+ACENT+INF
kuyarachillay,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+POLI+INF
kuyarachikuy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+AUBE+INF
kuyarachikamuy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+AOL+INF
kuyarachiisiy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+COLL+INF
kuyarachiikuy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+ACENT+INF
kuyarachiikariy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+PONC+INF
kuyarachiikapuy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+SOIN3+INF
kuyarachiikamuy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+PREAT+INF
kuyarachiikachiy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+POLI+INF
kuyarachichkay,kuyay,V+FR="to love"+FLX=V_SIP3_INF+PASS+FACT+PROG+INF
kuyatamurapuy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+AEP+PASS+APT+INF
kuyatamurachiy,kuyay,V+FR="to love"+FLX=V_SIP3_INF+AEP+PASS+FACT+INF
```

Fig. 5. Annotated 3-dim 2931 verbs derived from the verb to love

have first inventoried the IPS suffixes and their corresponding main semantic values as it appears in Table 3.

Where for the first one CHI, we have three factitive values (in English and French for this suffix but only in French for the rest)

FACT_1 : the subject aids, helps le_sujet_assiste, aide

FACT_2 : the subject invites, authorizes, incites le_sujet invite, autorise, incite

FACT_3 : the subject forces, commands a third party to do the action; le_sujet oblige, commande à un tiers à réaliser l'action.

5 Proposing Automatic Transducers from Quechua Compound Verbs into French

Using Table 2 we have written some NooJ grammars to annotate the 3-dim 2931 verbs derived from the verb to love *kuway* as we see in the sample of Fig. 5.

We have searched plausible meanings for the generated compound verbs by applying on the annotated forms the semantic values of Table 2. We show some results of this approach in Fig. 6 for the derivations of the verb *rimay* to talk:

```
#use grammaire_Verbs_26feb15.nof
#
rimatamurquy, rimay, V+FR="parler"+FLX=V_SIP2_INF
+le_sujet_en_passant_réalise_l'action
+le_sujet_accompli_en_peu_de_temps
+APT_2_en_action_centrifuge_réalise_l'action+INF
rimatamuikapuy, rimay, V+FR="parler"+FLX=V_SIP2_INF
+le_sujet_en_passant_réalise_l'action
+le_sujet_avec_attention_SOIN3_2_soigneusement
+SOIN3_3_concernant_un_tiers_SOIN3
+4_recommençant_l'action_interrompue_réalise_l'action+INF
rimatamuikamay, rimay, V+FR="parler"+FLX=V_SIP2_INF
+le_sujet_en_passant_réalise_l'action+le_sujet_vers_le_sujet
+PREAT_2_en_prévoyant_PREAT_3_attentionné
+PREAT_4_de_haut_en_bas_réalise_l'action+INF
rimatamuikachiy, rimay, V+FR="parler"+FLX=V_SIP2_INF
+le_sujet_en_passant_réalise_l'action+le_sujet_poliment
+POLI_2_concrètement_POLI_3_précisément
+POLI_4_courtoisement_réalise_l'action+INF
rimatamuikachay, rimay, V+FR="parler"+FLX=V_SIP2_INF
+le_sujet_en_passant_réalise_l'action+le_sujet_en_désordre
+DISP_2_desorienté_réalise_l'action+INF
rimatamuchkay, rimay, V+FR="parler"+FLX=V_SIP2_INF
+le_sujet_en_passant_réalise_l'action
+le_sujet_en_train_de_réaliser_l'action+INF
rimatamuchiy, rimay, V+FR="parler"+FLX=V_SIP2_INF
+le_sujet_en_passant_réalise_l'action+le_sujet_invite
+FACT_2_autorise_FACT_3_pousse_l'objet_du_verbe_à_réaliser_l'action+INF
rimarullay, rimay, V+FR="parler"+FLX=V_SIP2_INF
+le_sujet_en_peu_de_temps
+PRES_2_de_façon_pressante_réalise_l'action+le_sujet_gentiment
```

Fig. 6. Glossed meanings for some compound verbs derived from *rimay* to talk

After verification of the pertinence of these glossed outputs we may propose the possible meaning for the compound verb, like in the following examples:

*ayqiriy, ayqiy, V + FR = "échapper" + FLX = V_SIP_INF + le_sujet_commence_à_D
YN_2_recommence_à_réalise_l'action + INF*

aisariy, aisay, V + FR = “tirer” + FLX = V_SIP_INF + le_sujet_commence_à_DYN_2_recommence_à_réalise_l_action + INF

which could be interpreted as: the subject starts towing something, and so *aisariy* should be: to tow, as it has been actually lexicalized.

rimaikuy, rimay, V + FR = “parler” + FLX = V_SIP_INF+le_sujet_courtoisement_COURT_2_soigneusement_COURT_3_amicalement_COURT_4_vers_le_sujet_réalise_l_action + INF

which means (the subject) talks someone courteously, carefully, friendly, which could in fact has been lexicalized as: to greet

These meanings may be opposed to the existing lexicalized entries that we have gathered out of our corpus. We see that for the three first ones, they match well:

aiqiriy, V + FR = “commencer à fuir, entreprendre un retrait” + SP = “comenzar a huir, emprender la retirada” + FLX = V_TR

rimaikuy, V + FR = “adresar la parole à qqn avec courtoisie” + SP = “dirigir la palabra a alguien atentamente” + FLX = V_TR

amichiy, V + FR = “faire qqn s’ennuyer” + SP = “hacer aburrir a alguien” + FLX = V_TR

aiqiriy, V + FR = “se retirer lentement à une petite distance” + SP = “retirarse lentamente a pequeña distancia” + FLX = V_TR

asiriy, V + FR = “sourire” + SP = “sonreír” + FLX = V_TR

asnariy, V + FR = “commencer à sentir (la viande)” + SP = “comenzar a oler (carne)” + FLX = V_TR

Table 4. Automatic glossed translation compared to lexicalized entries

Automatic glossed translation	Lexicalized entries translated into FR
<i>rimariy, rimay,</i> V + FR = “parler” + FLX = V_SIP_INF + le_sujet_commence_à_DYN_2_recommence_à_réalise_l_action + INF	<i>rimariy, V</i> + FR = “commencer à parler” + EN = “to start talking” + FLX = V_TR
<i>rimapamuy, rimay,</i> V + FR = “parler” + FLX = V_SIP_INF + le_sujet_peaufine_PEAU_2_réitère_PEAU_3_réalise_l_action + le_sujet_se_déplaçant_ACENT_2_sous_contra-inte_réalise_l_action + INF	<i>rimapamuy, V</i> + FR = “parler à la place de quelqu’un” + EN = “to talk in behalf of someone” + FLX = V_TR
<i>rimaikachay, rimay,</i> V + FR = “parler” + FLX = V_SIP_INF + le_sujet_en_désordre_DISP_2_desorienté_réalise_l_action + INF	<i>rimaikachay, V</i> + FR = “parler constamment, sans arrêt” + EN = “to talk constantly” + FLX = V_TR
<i>rimanayay, rimay,</i> V + FR = “parler” + FLX = V_SIP_INF + le_sujet_envié_de_ENV_2_souhaite_réaliser_l_action + INF	<i>rimanayay, V</i> + FR = “avoir envie de parler” + EN = “to have the desire to talk” + FLX = V_TR

yaikuriy,V + FR = “entrer un peu, un moment, et aussi, entrer en étant de passage” + SP = “entrar un poco, y también: entrar estando de paso” + FLX = V_TR
 In the next table we show some more comparisons for other verbs (Table 4).

6 Results

As a result of this hand verifications carried out on some hundreds of cases, we have elaborated a trilingual dictionary (Qu, Fr, Sp) of Quechua compound verbs. It contains 1600 entries which can be added to our 1400-simple verbs lexicon. It includes their Spanish and French translations. We present below a sample of the entries of this dictionary (Fig. 7):

huqariy,V+FR=" lever, hausser, ramasser"+ SP="levantar, alzar, recoger"+FLXV_TR
qapariy,V+FR=" crier"+SP=" gritar"+FLXV_TR
paqariy,V+FR=" commencer à faire jour, lever du jour, naissance du jour"+SP="amanecer, nacer"+FLXV_TR
anchichiy,V+FR=" aimer quelqu'un"+SP="quejarse a alguien"+FLXV_TR
aiqichiy,V+FR=" laisser échapper"+SP="dejar escapar"+FLXV_TR
aysachiy,V+FR=" tirer avec force"+SP="tirar con fuerza"+FLXV_TR
akllachiy,V+FR="ordonner de choisir ou le permettre"+SP="ordenar escoger, o permitirlo"+FLXV_TR
amichiy,V+FR=" causer du dégoût"+SP="causar asco"+FLXV_TR
allqachiy,V+FR=" entraîner une faute"+SP="causar una falta"+FLXV_TR
anchuchiy,V+FR=" appeler quelqu'un quelque part"+SP="llamar a alguien a un lado"+FLXV_TR
apachiy,V+FR=" commander de porter quelqu'un"+SP="mandar llevar a alguien"+FLXV_TR
asichiy,V+FR=" faire rire"+SP="hacer reír"+FLXV_TR
asirichiy,V+FR=" faire sourire"+SP="hacer sonreír"+FLXV_TR

Fig. 7. A sample of the entries of the dictionary of compound Quechua verbs

7 Text Annotations

With the help of this dictionary and some NooJ grammars like V_SIP1_INF presented before we may automatically annotate a Quechua text. We applied them on a collection of eight Quechua tales. We show in Fig. 8 the annotated correspondences obtained. We have found around 90 % of successful matches, 6 % of partial matches and 4 % are incorrect matches, mainly because of ambiguities.

PAKPAKU CHAYÑAWAN RIMANAKUN

L'hibou et le moineau se parlent

Tutallamanta, manaraq anchata achikiachkaptin, juk sumaq chayñacha, kusikusi **takikuchkarqa** sachap kallmanpi. Jinamanta pakpaku chayaramun llakillaki. Juk kallmapi **rataykuspa kaymanta** wakman **qawaykachan**, mana imanispa.

Chayñacha, kusi ñawinwan **qawaspa, nin**:

- Pakpakucha, imanasqataq **llakichkanki?** **nispa**.

Pakpakutaq muti ñawinwan **qawarayan** upallalla. Chayñachataq juktawan tapuykun:

- Manachu imatapas **nita munanki?** Chiqñiwánkichu?

- Manam, chayñacha, **cheqñikichu**? Ñoqanchik purapiqa manam cheqñikanchu. **Llakisqam kachani** llaqtaymanta ripunay **kaptin**.

- Imanasqataq **ripunkiri?** **nispa** musuqmanta **tapun, kaymanta** wakman qawaykachaspa:

- Manam runakuna **kuyawanchu**, chaimi **ripuyta munani**.

- Imanasqataq mana **kuyastunkichu?**

- Chaitam **kunan** willasqayki masiman jina. **Uyariwaiyari**: Ñoqaqa tutallapim **kausani**; tutallapim **takini llakisqapas, kuisqapas**. Manam pitapas **llakichiita, manchachiita munanichu**. **Takillam takikuni, takita munaspa**. Runakunataq, imanasqachá, **kay** pakpakup **takinqa** atim", **ninku**. Waiñisi jamun ñoqa **takipti**, manataqmi ñoqa **yachanichu** ima chai waiñiykasqanta. Kausayllatam ñoqaqa **yachani**. Chaimi lllapan runakuna, chai **iskay** chakipi **puriq**, chai puñiyisiki runakuna, **cheqñivan**. Ñoqa **takikuni** tutakunapi, paykunataq **nin**: "Taytallay!, piraq **waiñunqa** pakpakum, sachapi **waqachkan**" **nispa**. Chaimi **cheqñiwanku** chaimi warmakunapas **rikuwaspa** runiwian **choqawan**, sachá kallmapi panchau **puñukupti**. Chaimi **ripusaq**, chayñacha.

- Manam chaiqa allinchu, **nispa**, chayñacha **kutichin**. Imanasqataq mana juk **takiiman** takikita **tikrankichu?** Jinallata **takiptikichu**, mayta **ripikipas**, maypi **kaptikipas**, jinallataqmi runaqa **cheqñisunki**.

Pakpaku umanta kunkan **purupi** chintikuspa, karuneqta **qawaspa**, jamaatapas **kanman jina, kay | ninkun**.

Fig. 8. Recognition of compound verbal forms in one text of the corpus

8 Conclusion

We have studied the key role of inter positioned suffixes IPS, for the generation of new Quechua verbs. After the study of thousands of combinations we have found altogether 3249 valid compounds of up to three IPS suffixes which will generate that amount of new verbs out of a single one. This considerably increases the verb lexicon. In fact parsing the NooJ grammar V_SIP_INF on our dictionary of around 1400 simple verbs gives us 43160 new compound Quechua verbs. With the help of morpho-syntactic NooJ grammars and the use of the semantic annotations corresponding to the IPS suffixes we propose a glossed form in order to figure out the meaning of these verbs.

Perspectives

Increase the compound verb bilingual dictionary.

Improve our grammars to obtain less ambiguous translations

References

1. Bogacki, K., Gwiazdecka, E.: Derivational structure of polish verbs and the expansion of the dictionary. In: Automatic Processing of Various Levels of Linguistic Phenomena. Selected Papers from the NooJ 2011 International Conference, pp. 50–62. Cambridge Scholars Publishing, New Castle upon Tyne (2012)
2. Dubois, J., Dubois-Charlier, F.: Le verbes français (le « dictionnaire électronique des verbes français (DEV). In: 1992 Diffusé à partir de septembre 2007 par MoDyCo dans un format Excel (2007)
3. Duran, M.: Dictionnaire Quechua-Français-Quechua. Éditions HC, Paris (2009)

4. Duran, M.: Morphological and syntactic grammars for recognition of verbal lemmas in quechua. In: Proceedings of the 2014 International Conference and Workshop, Sassari (2014). (at Print)
5. Silberztein, M.: NooJ Manual (2003). www.nooj4nlp.net (220 pages updated regularly)
6. Silberztein, M.: Syntactic parsing with NooJ. In: Proceedings of the NooJ 2009 International Conference and Workshop, 177–190. Centre de Publication Universitaire, Sfax (2010)
7. Silberztein, M.: Automatic transformational analysis and generation. In: Proceedings of the 2010 International NooJ Conference, pp. 221–231. Democritus University Editions, Komotini (2011)
8. Silberztein, M.: Variable unification in nooj V3. In: Automatic Processing of Various Levels of Linguistic Phenomena: Selected Papers from the NooJ 2011 International Conference, pp. 50–62. Cambridge Scholars Publishing, New Castle upon Tyne (2011)
9. Vietri, S.: The annotation of the predicate-argument structure of transfer nouns. In: Formalising Natural Languages with NooJ, pp. 88–99. Cambridge Scholars Publishing, New Castle upon Tyne (2013)