

GRAMATICAS MORFO-SINTACTICAS PARA EL RECONOCIMIENTO AUTOMATICO DE RAICES VERBALES EN QUECHUA

Maximiliano Duran
Université de Franche-Comté, CRIT, Besançon, France
LIG, UGA, Grenoble, France
duran_maximiliano@yahoo.fr

Resumen.

Dado que el quechua es una lengua en peligro, tal como lo expresa la UNESCO, se pretende en este trabajo avanzar en la creación de una plataforma de recursos lingüísticos, de modo de poder contribuir a su supervivencia y desarrollo. En este artículo, se presenta parte del algoritmo que utiliza las estructuras flexivas y derivativas de los verbos en quechua que permiten reconocer raíces verbales de un texto. Para ello nos valemos de las reglas morfo-sintácticas de aglutinación de los sufijos verbales de derivación y flexión. Se muestra cómo, con la ayuda del programa informático llamado NooJ, se pueden formalizar estos procesos. En particular: los paradigmas que rigen estructuras flexivas así como las gramáticas correspondientes. La finalidad es obtener nuevos verbos y generar flexiones o formas conjugadas de un solo verbo. Aquí se trabajó con un corpus, basado en escritos antiguos, para identificar automáticamente todas esas formas verbales, y extraer de entre ellos las raíces verbales correspondientes. De ese modo se pudo ‘descubrir’ doscientos noventa y ocho verbos ‘nuevos’, que se agregaron al léxico de verbos simples del quechua.

Palabras clave:

Tiempo indefinido, conjugación formalizada, sufijos inter-posicionales, sufijos-pos posicionales.

MORPHO-SYNTACTIC GRAMMARS FOR THE AUTOMATIC RETRIEVAL OF VERBAL ROOTS IN QUECHUA

Abstract

Given that Quechua is an endangered language, as expressed by UNESCO, this paper aims to advance in the creation of a platform of linguistic resources, in order to contribute to its survival and development. In this article, we present part of the algorithm that uses the inflectional and derivational structures of Quechua verbs to recognize verb roots in a text. For this purpose, we use the morpho-syntactic rules of agglutination of derivational and inflectional verb suffixes. It is shown how, with the help of the computer program called NooJ, these processes can be formalized. In particular: the paradigms governing inflectional structures as well as the corresponding grammars. The aim is to obtain new verbs and to generate inflections or conjugated forms of a single verb. Here we worked with a corpus, based on ancient writings, to automatically identify all these verb forms, and extract from them the corresponding verb roots. Two hundred and ninety-eight 'new' verbs were thus 'discovered' and added to the lexicon of simple Quechua verbs.

Key words:

Indefinite time, formalized conjugation, inter-position suffixes, post-position suffixes.

1. Introducción. Motivación para el proyecto¹.

¹Este trabajo se basa en dos artículos: Formalizing Quechua verb inflection. Formalizing Natural Languages with NooJ 2013 International Conference Cambridge Scholars New Castle. England. 2014 y Morphological

El idioma quechua fue el idioma oficial de la civilización Inca. Se originó en los Andes centrales del Perú alrededor de la primera mitad del primer milenio de la era actual y muy probablemente, fue la lengua de la civilización Huari, asentada en Ayacucho. En 2009, la UNESCO la declaró lengua en peligro. Con este trabajo, procuramos contribuir a su supervivencia y su desarrollo. Nuestro proyecto a largo plazo es construir una plataforma de recursos lingüísticos para el procesamiento automático de textos en quechua. El vocabulario de verbos simples en los diccionarios impresos es relativamente modesto². Sin embargo, sabemos que en los manuscritos y documentos antiguos, se hallan bastantes verbos ocultos o dados por perdidos. Cómo encontrarlos automáticamente para que puedan incrementar el vocabulario verbal, es uno de nuestros propósitos. Nuestro primer paso consistió en construir y formalizar un corpus de base. El segundo paso fue construir un diccionario electrónico francés-quechua³ basado en los veinticinco mil verbos franceses de Dubois & Dubois-Charlier (2007), utilizando los formalismos de la plataforma lingüística NooJ.⁴

2. El Corpus

En nuestro corpus, hemos incluido los siguientes documentos, que datan de inicios del siglo XVI y XVII, y contiene un total de 67900 tokens:

- González Holguin, Diego, 1608, *Vocabulario de la Lengua General de todo el Perú llamada Lengua Qquichua o Lengua del Inca*.
- Santo Thomas, Domingo de, 1560, *Lexicon, o vocabulario de la lengua general del Perú*
- Francisco de AVILA's, 1598? *Dioses y hombres de Huarochiri*. Una narración de ritos y de leyendas Quechua coleccionadas por Francisco de Avila

Previamente, se estandarizó la ortografía de estos textos aplicando el alfabeto oficial de la variante Ayacucho del Quechua.

3. Formas verbales flexionadas

Una forma verbal flexionada típica del quechua tiene la siguiente estructura:

V+ SIP + DESINENCIA TI + SPP

Lo explicamos como sigue:

V: verbo, lema verbal

SIP⁵: sufijo interpuesto, existen treinta y uno,

DESINENCIA TI⁶: es el conjunto de siete sufijos que marcan la persona en el tiempo indefinido TI (se comportan como puntos fijos durante las flexiones),

and Syntactic grammars for recognition of verbal lemmas in Quechua. Formalising Natural Languages with NooJ 2014. Monti J., Silberztein M., Monteleone M., Di Buono M. Eds. Cambridge Scholars Publishing.

² González Holguin (1608), Itier (2001), Perroud (1970) (menos de mil quinientas entradas).

³ Debido al hecho de realizar mi trabajo en Fancia desde hace décadas. La versión Español-Quechua la estamos también trabajando.

⁴ NooJ es una plataforma lingüística realizada por Max Silberztein (2015).

⁵ Sufijos verbales interposicionales SIP= (*chaku, chi, chka, ykacha, ykachi, ykamu, ykapu, ykari, yku, ysi, kacha, kamu, kapu, ku, lla, mpu, mu, naya, pa, paya, pu, ra, raya, ri, rpari, rqu, ru, tamu*). Estos sufijos permiten generar nuevos verbos por derivación. Para obtener la totalidad de los SIP debemos agregar (*na, pti, rqa, sqa, spa, wa*).

⁶ Sufijos verbales del tiempo indefinido TI= (-ni, -nki, -n, -nchik, -niku, -nkichik, nku). Lo llamamos 'Tiempo Indefinido' por lo siguiente: 1°. Colocan en 'tiempo presente' lo que decimos (*kunan rimani/ yo hablo ahora*), 2°. Colocan en 'tiempo pasado' lo que decimos (*qayna-punchau rimani/ yo hablé ayer*), 3°. Colocan en 'modo habitual' lo que decimos (*sapa punchau rimani/ yo hablo todos los días*), mana yachanchikchu imapuni pachapim 'ni' churawasjanchikta.

SPP⁷: sufijo pospuesto, existen diecinueve.

Recordemos que el quechua es una lengua polisintética. Por ejemplo, la una oración larga en español como la que sigue: ‘*Tenemos que hacer el trabajo dejando de lado todo lo demás*’ se convierte en *llamkananchikraqmi*, es decir: está representada por una forma verbal (una palabra) en quechua. Repasemos el comportamiento de algunos de estos sufijos y sus combinaciones en las siguientes flexiones del verbo *qallariy* ‘comenzar’:

Lema verbal: *qallari-*

qallari-nchik ‘comenzamos’, *-nchik* es el sufijo que marca el PR+p+1 (primera persona plural, inclusivo, en presente)

qallari-chka-nchik ‘estamos comenzando’

qallari-ysi-chka-nchik ‘estamos ayudando a empezar a alguien’

qallari-ysi-chka-nchik-ña ‘ya estamos ayudando a empezar a alguien a’

qallari-ysi-chka-nchik-ña-taq ‘y sin embargo ya le(s) estamos ayudando a empezar’

La desinencia personal *nchik* permanece fija durante las combinaciones en que intervienen los SIP y SPP. La morfología del quechua está dominada por este tipo de aglutinaciones de sufijos verbales, nominales, adverbiales o adjetivales.

4. Enfoque matricial de las combinaciones de sufijos verbales

Lo que llamamos el tiempo presente es en realidad, un tiempo indefinido (TI). Por un lado, sitúa el enunciado en el momento en que se produce este enunciado, pero, por otro lado, puede situarlo en un momento en el que el enunciado ya se ha producido y también expresa un modo habitual. La conjugación de las tres personas singulares tiene la siguiente estructura:

ñoqa (yo) lema +**NI**

qam (tú) lema + **NKI**

pay (él, ella) lema +**N**

Para el futuro tenemos el esquema:

ñoqa (yo) lema +**SAQ**

qam (tú) lema + **NKI**

pay, (él, ella) lema +**NJA**

La estructura del tiempo presente juega un papel crucial en la construcción de las conjugaciones de los otros tiempos. Es una especie de punto fijo alrededor del que se construye toda la topología flexiva basada en la combinación de los sufijos (tiempos, modos, aspectos, etc.). Por ejemplo, el pretérito perfecto simple se obtiene tomando esa estructura TI e interponiendo el sufijo SIP *-rqa*, entre el lema verbal y la terminación de persona. Tenemos entonces:

Presente o TI

taki-ni / yo canto

taki-nki / tú cantas

taki-n / él canta

Pretérito

taki-rqa-ni / yo canté

taki-rqa-nki / tú cantaste

taki-rqa-n / él cantó

De acuerdo con la morfología del verbo quechua, podemos construir combinaciones de dos, tres o cuatro sufijos SIP y SPP. Es muy productiva en cuanto a la flexión. Para obtener el conjunto completo de estas combinaciones que sean sintácticamente correctas, primero construimos

⁷ Sufijos verbales posposicionales SPP= (*ch, chá, chiki, chu?, chun, chusina, má, man, m, mi, ña, pas, puni, qa, raq, s, si, taq, yá*).

manualmente una matriz de dos entradas que tiene como primera fila y primera columna todos los sufijos interposicionales SIP en un caso y, en otro caso, el conjunto de los SPP.

	CHI	CHKA	IKACHA	IKACHI	IKAMU	IKAPU	IKARI	IKU	ISI	KACHA	KAMU	KAPU	KU	LLAV	MU	NAV
CHI	0	1	1	1	1	1	0	1	1	0	1	1	1	1	1	4
CHKA	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
IKACHA	1	1	0	0	1	1	0	0	1	0	0	0	1	1	1	0
IKACHI	1	1	0	0	1	1	0	1	1	0	0	0	1	1	1	0
IKAMU	1	1	0	0	1	0	0	0	1	0	0	1	1	1	0	0
IKAPU	1	1	0	1	0	0	0	1	0	0	1	1	1	1	1	0
IKARI	1	1	0	1	1	0	0	1	0	0	1	1	1	1	1	0
IKU	1	1	0	0	1	0	0	0	1	0	0	1	1	1	0	0
ISI	1	1	0	1	0	0	0	1	0	0	1	1	1	1	1	0
KACHA	1	1	0	1	1	1	0	1	0	0	0	1	1	1	1	1
KAMU	1	1	1	1	1	0	0	1	0	0	0	0	0	1	1	1
KAPU	1	1	1	1	1	0	0	1	0	0	0	0	0	1	1	1
KU	0	1	0	0	1	1	0	1	0	0	0	1	0	1	0	1
LLAV	1	1	1	1	1	0	0	1	1	0	0	0	0	0	1	1
MU	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1
NAV	0	0	0	0	0	0	0	0	0	0	4	0	0	2	0	0
KAMU	1	1	1	0	0	0	1	0	0	0	0	0	0	1	0	0
NAKU	0	1	1	0	0	0	1	1	0	0	0	0	0	1	1	0
NAYA	1	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0
PAYA	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1
PU	0	1	0	0	1	1	0	1	0	0	0	0	1	0	0	0
RA	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
RAYA	1	1	0	0	0	0	0	1	1	0	0	0	1	1	1	1

Tabla 1 Matriz de combinaciones bidimensionales de sufijos interposición interposicionales SIP

Allí llenamos las novecientas sesenta y un celdas con 0 o 1 para los SIP. El valor '1' significa 'combinación gramaticalmente válida' y '0' significa 'no válido'. Por ejemplo,

- la celda correspondiente al punto (*chi*, *chka*) como coordenadas en esta matriz lleva '1', porque la combinación *-chichka-* es compatible y puede aglutinarse a la raíz *taki-* del verbo '*cantar*' *takiy* para obtener la conjugación:

taki-chichka-ni 'Le estoy haciendo cantar',
taki-chichka-nki 'Le estás haciendo cantar',
taki-chichka-n 'Le está haciendo cantar', etc.

- la celda (*kacha*, *ku*), que también lleva '1', tendrá la combinación *-kachaku-*, *taki-kachaku-y* /*mantenerse cantando repetidamente o canturrear para sí mismo*, que se conjuga como sigue:

takikachakuni/ canturreo para mí mismo;
takikachakunki/ canturreas para tí mismo;
takikachakun/ canturrea para él mismo, etc.

El total de '1' encontrados en la matriz binaria de SIP es trescientos seis.

Luego buscamos las combinaciones válidas de tres SIP. En la Tabla 2, aparece la tabla que contiene la matriz correspondiente. Tiene como primera fila el conjunto de treinta y un SIP y como primera columna las trescientas seis combinaciones binarias válidas que acabamos de obtener. De las ocho mil doscientas sesenta y dos celdas, tres mil y una contienen el valor '1', correspondiente a las combinaciones ternarias SIP válidas como las que siguen:

*-ñachusinam -ñapaschá -ñapaschik -ñapaschu -ñapasmi -ñataqsi -punichusinam -puniñach -
puniñachá -puniñachik -puniñachu? -puniñachu -puniñachusina -puniñamá -puniñam -puniñapas -
puniñas*

SEMIBINOMIAL	CHAKU	CHI	CHKA	YKACHA	YKACHI	YKAMU	YKAPU	YKARI	YKUI	YSI	KACHA	KAMU	KAPU	KU	LLAV	MPU	NU	NAYA	PA	PAYA	PU	RAYA	RI	RPARI	RQU	RU	TAMU		
CHAKUCHI	0	0	1	0	1	1	1	1	1	1	0	1	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	1
CHAKUCHKA	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CHAKUYKACHI	0	0	1	0	0	1	1	0	1	0	1	0	1	1	1	0	1	0	1	1	1	1	0	1	1	1	1	1	
CHAKUYKAMU	0	0	1	1	1	1	1	0	1	0	1	1	0	1	1	0	1	0	1	1	0	1	0	1	1	1	1	1	
CHAKUYKAPU	0	1	1	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	
CHAKUYKARI	0	1	1	0	0	0	0	0	1	1	0	0	0	1	1	0	1	0	0	1	1	0	0	1	1	1	1	0	
CHAKUYKUI	0	1	1	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	0	1	1	0	0	1	1	1	1	0	
CHAKUYSI	0	1	1	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	0	
CHAKUKACHA	0	1	1	0	0	1	1	0	1	0	0	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1	
CHAKULLAV	0	1	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
CHAKURPARI	0	1	1	0	0	1	1	0	1	0	1	0	1	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	
CHAKURQU	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
CHAKURU	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
CHAKUTAMU	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
CHICHKA	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CHICHKACHI	0	0	1	0	0	1	1	1	0	1	0	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	
CHICHKAMU	0	0	1	1	1	1	0	1	1	0	1	1	0	0	1	1	0	0	1	1	0	1	0	0	1	1	1	1	
CHICHKAPU	0	0	1	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	
CHICHKARI	0	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	0	1	0	0	1	1	0	0	1	1	1	0	
CHICHKUI	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	1	0	0	1	1	0	0	1	1	1	0	
CHICHYSI	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	
CHICHKAMU	0	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	
CHICHKU	0	0	1	0	0	0	1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	
CHICHLLAV	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
CHICHMU	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
CHICHPU	0	0	1	0	0	1	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	

Liste de combinaciones ternarias SIP válidas	CHAKU	CHI	CHKA	YKACHA	YKACHI	YKAMU	YKAPU	YKARI	YKUI	YSI	KACHA	KAMU	KAPU	KU	LLAV	MPU	NU	NAYA	PA	PAYA	PU	RAYA	RI	RPARI	RQU	RU	TAMU		
CHAKUCHI	0	0	1	0	1	1	1	1	1	1	0	1	0	1	0	1	0	0	0	1	0	0	1	0	0	1	0	0	1
CHAKUCHKA	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CHAKUYKACHI	0	0	1	0	0	1	1	0	1	0	1	0	1	1	1	0	1	0	0	1	1	1	1	0	1	1	1	1	
CHAKUYKAMU	0	0	1	1	1	1	1	0	1	0	1	1	0	1	1	0	1	0	1	1	0	1	0	1	1	1	1	1	
CHAKUYKAPU	0	1	1	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	
CHAKUYKARI	0	1	1	0	0	0	0	0	0	1	1	0	0	0	1	1	0	1	0	0	1	1	0	0	1	1	1	0	
CHAKUYKUI	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	0	1	1	0	0	1	1	1	0	
CHAKUYSI	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	0	
CHAKUKACHA	0	1	1	0	0	1	1	0	1	0	0	1	1	1	1	0	1	1	0	1	1	1	1	1	1	1	1	1	
CHAKULLAV	0	1	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CHAKURPARI	0	1	1	0	0	1	1	0	1	0	1	0	1	1	1	0	1	0	0	0	1	0	0	0	0	0	0	0	
CHAKURQU	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
CHAKURU	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
CHAKUTAMU	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
CHICHKA	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CHICHKACHI	0	0	1	0	0	1	1	1	0	1	0	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	
CHICHKAMU	0	0	1	1	1	1	1	0	1	0	1	1	0	0	1	1	0	0	1	1	0	1	0	0	1	1	1	1	
CHICHKAPU	0	0	1	1	1	1	0	0	1	0	1	1	0	0	1	1	0	1	1	0	1	1	1	1	1	1	1	1	
CHICHKARI	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	
CHICHKUI	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	
CHICHYSI	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	
CHICHKAMU	0	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	
CHICHKU	0	0	1	0	0	0	1	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	
CHICHLLAV	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
CHICHMU	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
CHICHPU	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0

Tabla 2 Combinaciones ternarias SIP válidas

5. Sufijos posposicionales SPP

Estos sufijos son los que se colocan después de la desinencia verbal del TI o del futuro, como en los siguientes ejemplos:

- rima-nki-man / deberías hablar*
- rima-nki-man-pas / además, deberías hablar*
- rima-nki -man-pas-cha / además quizás deberías hablar*
- rima-n - man-ña-taq / temo que él hable*

La matriz de combinaciones binarias de SPP contiene cincuenta y seis combinaciones compatibles (celdas con valor ‘1’), como se muestra en la Tabla 3.

	CH	CHAA	CHIK	CHUI	CHUN	CHUSIN/MAA	M	MMI	ÑA	PAS	PUNI
CH	0	0	0	0	0	0	0	0	0	0	0
CHAA	0	0	0	0	0	0	0	0	0	0	0
CHIK	0	0	0	0	0	0	0	0	0	0	0
CHUI	0	0	0	0	0	0	0	0	0	0	0
CHUN	0	0	0	0	0	0	0	0	0	0	0
CHUSIN	0	0	0	0	0	0	1	0	0	0	0
MAA	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0
MMI	0	0	0	0	0	0	0	0	0	0	0
ÑA	1	1	1	1	1	1	1	0	0	1	0
PAS	1	1	1	0	1	0	0	1	0	0	0
PUNI	1	1	1	1	1	1	1	0	1	1	0
QA	0	0	0	0	0	0	0	0	0	0	0
RAQ	0	1	1	1	1	1	1	0	1	0	1
S	0	0	0	0	0	0	0	0	0	0	0

Tabla 3 Matriz de combinaciones binarias compatibles de sufijos posposicionales.

Aquí también, el valor “1” correspondiente al punto (*ña, m*) indica que la combinación *-ñam* es gramaticalmente válida. De forma similar a los de SIP, se obtuvo la matriz de combinaciones terciarias de SPP, que tiene como primera columna las cincuenta y seis combinaciones binarias compatibles encontradas y como primera fila el vector SPP. El resultado contiene ochenta elementos no nulos, como se muestra en la Tabla 4.

rima-nchik-ña-m, que ya hemos mencionado (*'m'* solo si sigue a una vocal, *'mi'* si sigue a una consonante o *'ai'*).

	CHA	CHAA	CHIK	CHUCH	CHUI	CHUN	CHUS	CHUSIN/MAA	M	MMI	ÑA	PAS	PUNI
MANVCHUS	0	0	0	0	0	0	0	0	0	ÑAPASM	0	0	0
MANVMMI	0	0	0	0	0	0	0	0	0	0	0	0	0
MANVÑA	0	0	0	0	0	0	0	0	0	0	0	0	0
MANVPAS	ÑATAQCHA	ÑATAQCHAA	ATAQCHITAQCHU	0	ATAQCHU	0	TAQCHUSATAQMA	0	ATAQMNATAQÑ	0	ATAQPUI	0	0
MANVQA	0	0	0	0	0	0	0	0	0	0	0	0	0
MANVRAQ	PASCHCHA	PASCHCHAA	ASCHCHISCHCHU	ASCHCHIASCHCHU	ASCHCHU	ASCHCHU	ASCHMA	0	ASCHMV	0	ASCHPA	0	0
MANVSSI	0	0	0	0	0	0	0	0	0	0	0	0	0
MANVTAQ	0	0	0	0	0	0	0	0	0	0	0	0	0
MANVYAA	0	0	0	0	0	0	0	0	0	0	0	0	0
ÑACHUSINA	0	0	0	0	0	0	0	0	0	PUNIMM	0	0	0
ÑAMAA	0	0	0	0	0	0	0	0	0	0	0	0	0
ÑAM	0	0	0	0	0	0	0	0	0	0	0	0	0
ÑAMMI	0	0	0	0	0	0	0	0	0	0	0	0	0
ÑAPAS	PUNIRAQCHA	UNIRAQCHA	NIRAQCI	0	0	0	0	0	0	0	0	0	0
ÑAS	0	0	0	0	0	0	0	0	0	0	0	0	0

Tabla 4 Vista parcial de la matriz de valores de combinaciones ternarias compatibles de sufijos posposición.

SVPP3= (MANV-CHUSINAV-MV, MANV-ÑAV-CHV, MANV-ÑAV-CHAAV, MANV-ÑAV-CHIKV, MANV-ÑAV-CHIKIV, MANV-ÑAV-CHUIV, MANV-ÑAV-CHUNV, MANV-ÑAV-CHUSINAV, RAQV-TAQV-YAAV, TAQV-CHUSINAV-MV...)

Siguiendo el mismo método, obtenemos la matriz de combinaciones compatibles de cuatro sufijos posposicionados. Algunas de las combinaciones válidas resultantes se enumeran a continuación:

-manñapaschá, -puniraqpaschá, -puniraqpaschiki, -puniraqpaschu -raqpuniñachu, -raqpuniñachus, -raqpuniñachusina, -manñapaschiki, -manñapaschun

Estamos trabajando en la matriz de combinaciones de cinco y seis SPP.

6. Combinaciones mixtas SIP y SPP

Estas son formas que contienen uno o varios sufijos SIP y uno o varios sufijos SPP. Debemos resaltar el hecho de que las desinencias TI permanecen invariables. Las reglas de la morfosintaxis quechua permiten estas mezclas con el objeto de obtener un gran número de formas verbales conteniendo aglutinaciones mixtas de sufijos inter y posposicionales. Por ejemplo, en la forma verbal siguiente tenemos dos sufijos SIP, *ku* y *lla*, que preceden a la desinencia *ni* de la primera persona singular. Ej: *tapu-ku-lla-ni / yo pregunto respetuosamente.*

Es posible aglutinarle sufijos SPP después de la desinencia *ni*. Si agregamos el sufijo *raq* obtendremos la forma verbal siguiente:

tapu-ku-lla-ni-raq / antes, yo pregunto respetuosamente.

Emplearemos el símbolo MIX21 para señalar una forma verbal mixta conteniendo dos sufijos SIP y un SPP, MIX22 para señalar una forma conteniendo dos sufijos SIP y dos, SPP, etc.

miku-ku-rqa-n-pas-chiki «pareciera que él ha comido también» (MIX22)

miku-cha-ku-na-lla-n-paq «para que él pueda comer tranquilamente» (MIX41)

miku-ku-lla-chka-rqa-ni-taq «nótese que yo estaba comiendo discretamente» (MIX41)

qaya-cha-ku-lla-wa-pti-yki-pas «aún cuando tú casi me llamaras gentilmente» (MIX51)
(construir sobre la estructura modificada TI 1)

qaya-cha-mu-chka-lla-yki-man-raq-pas-hina «es como si yo estuviera llamándote delicadamente» (MIX44) (construir sobre la estructura modificada TI1)

qaya-cha-mu-chka-lla-n-man-raq-pas-hina-s «da la apariencia de que él te estuviera llamando con delicadeza» (MIX45).

En general, V_MIXij (i=1...4 y j=1...5), indica una forma conteniendo 'i' sufijos SIP y j sufijos SPP, con un máximo de cuatro para los SIP y cinco para los SPP.

7. Programando las gramáticas NooJ correspondientes

Aplicamos estos resultados para programar los paradigmas correspondientes a las diferentes formas conjugadas. Elegimos la descripción en forma algebraica de la flexión, siguiendo criterios similares a los de K. Bogacki (2008), en lugar del enfoque gráfico para describir las gramáticas. He aquí algunos ejemplos, cuyos detalles se pueden encontrar en Durán (2014, 2016):

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

```
VERBEAY = <E>/INF|:CHU |:PRESENT |:FUT |:RQA |:PREASS |:CHKAASS
|:IMP |:COND |:PPL |:PTIC |:presenCHU |:progCHU|:impeCHU |:FUTCHU
|:iptiiCHU |:imanCHU |:sqaCHU |:ptiijaCHU|:GDYN |:STINCHU |:TA
|:SQAIKI|:TRTS1 |:WANCHU |:TRDE1 |:DE1PCHU |:TRDE3 |:DE3PCHU
|:TRTA2 |:accustfTA |:WANKICHU |:TRDE1CHU |:TRTS1CHU |:PIDF2
|:PIDF2CHU |:PIDF2 |:PICTR |:PICTRAC |:SPA |:GSPA |:IPI;
```

Usando el diccionario de verbos simples, NooJ generará las formas verbales anotadas correspondientes. A continuación, tenemos una pequeña lista de entradas para el verbo *takiy* / *cantar*.

```
takinichu,takiy,V+FR="chanter"+FLX=conjugVERBES+s+1+NEG
takichkanikuchu,takiy,V+FR="chanter"+FLX=conjugVERBES+pex+1+NEG
takichkankuchu,takiy,V+FR="chanter"+FLX=conjugVERBES+p+3+NEG
```

Hasta el momento, hemos programado más de doscientos paradigmas. En el futuro, completaremos el estudio para los casos de combinaciones de más de tres sufijos.

Una gramática general que permita generar automáticamente todas las formas no mixtas de formas verbales en presente, futuro, pasado, y diferentes modos como el subjuntivo, condicional, o los gerundios, etc. para los verbos transitivos, lo obtenemos utilizando estos paradigmas y que lo simbolizamos por V_TR se presenta como sigue:

$$V_TR =:V_SPP | :V_TR_SIP | :V_CONJ_TR | :I_TR^8;$$

Cuando la aplicamos a un verbo transitivo como *mikuy* / *comer*, obtenemos más de siete mil quinientas formas flexionadas como se muestra en la Fig. 1.

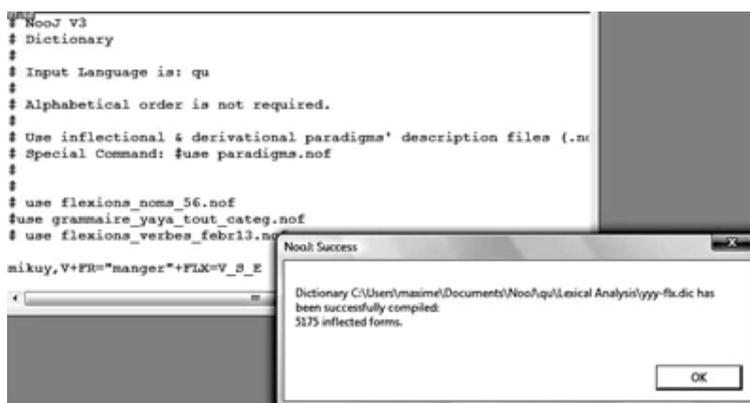


Fig. 1 Formas flexionadas del verbo *mikuy*

Al aplicar esta gramática V_TR, a un conjunto de cuatrocientos verbos quechuas transitivos hemos obtenido 2.749.968 formas flexionadas como se muestra en la Fig. 2.

⁸ Donde se tiene: V_TR: verbo transitivo; V_SPP: paradigma para sufijo verbal posposicional; V_TR_SIP: paradigma para sufijo verbal interposicional para verbos transitivos; V_CONJ_TR: paradigma de conjugación de verbos transitivos; I_TR: paradigma de verbos impersonales transitivos.

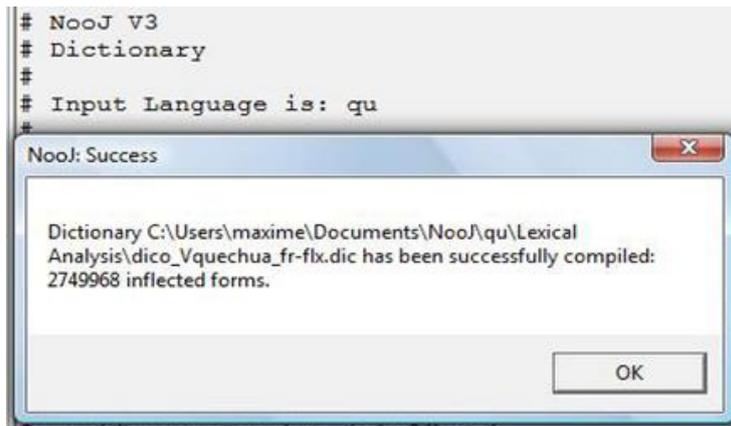


Fig. 2 Formas verbales flexionadas generadas a partir de cuatrocientos verbos transitivos.

Tratando de ampliar la formalización de la morfología verbal, siguiendo un camino similar he construido algunos programas que generan formas verbales mixtas como muestro a continuación:

```
V_MIX12= (:SIP1_TI_V) (:SPP2_V) | (:SIP1_TI_C) (:SPP2_C)
| (:SIP1_TIM_V) (:SPP2_V) | (:SIP1_TIM_C) (:SPP2_C);
V_MIX21= (:SIP2_TI_V) (:SPP1_V) | (:SIP2_TI_C) (:SPP1_C);
V_MIX22= (:SIP2_TI_V) (:SPP2_V) | (:SIP2_TI_C) (:SPP2_C);
```

Los detalles se pueden encontrar en Durán (2017). En la Fig. 3 aparece el total de las formas verbales conjugadas mixtas de *rimay/ hablar*, en el que se incluye un SIP y dos SPP. Sigue después, una muestra de las formas generadas automáticamente por la gramática V_MIX12.

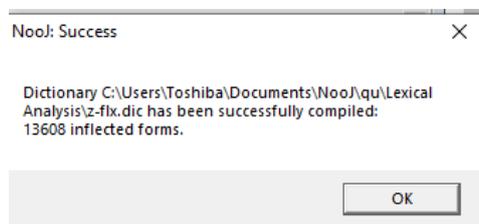


Fig. 3 Formas verbales conjugadas mixtas de *rimay/ hablar* conteniendo un SIP y dos SPP.

rimawankichichusinam,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikmanchá,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikmanmi,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikmanpas,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikmanraq,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikmantaq,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikpunimá,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikpunim,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikraqpas,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikraqpuni,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikraqsi,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikraqtaq,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikraqya,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikraqyá,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichiktaqsi,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichiktaqyá,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikraqchu?,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p

rimawankichikpunichu?,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawankichikñachu?,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+2+p
rimawanmanchá,rimay,V+FR="aimer"+FLX=V_MIX12+TRASL+3+s

8. Encontrar formas verbales y hacer marcado de textos

Si queremos encontrar y marcar las formas verbales conjugadas simples o mixtas en un texto, podemos hacerlo utilizando las gramáticas NooJ correspondientes. Por ejemplo, si buscamos las formas verbales de los verbos transitivos contenidos en el capítulo 2 del manuscrito de Huarochiri recopilado por De Avila (1598), hemos utilizado la gramática:

```
<V+PR>|<V+F>|<V+IP>|<V+PPI>|<V+PAPT>|<V+GER>|<V+GER1>|<V+PASS>  
|<V+RQU_F>|<V+PPA2>|<V+PASSA>|<V+PASS1>|<V+PASS2>|<V+FP>|<V+FA>  
|<V+GER2>|<V+ OBL2>|<V+ OBL >|<V+ SUBI >|<V+ RS1>|<V+GER3>
```

En el resultado obtenido, aplicamos en NooJ, los operadores: Syntactic Analysis>Locate>all occurrences obtenemos una lista parcial ochenta y tres verbos extraídos del texto. Estos pueden ser visualizados y resaltados en color sobre el texto utilizando los operadores: Concordance>Color matching. El resultado lo podemos observar en la Fig. 4, donde se puede ver que los verbos reconocidos han sido coloreados.

9. Reconocimiento extracción y recuperación de verbos perdidos

He aplicado varias operaciones de concordancia sobre nuestro corpus usando operadores como

```
NI_q_extr == Find/ Replace (PERL pattern, ni$ | q$, extract  
lines)  
VOC-ni_q == NI_q_extr (VOC-H_brut)  
VOC-chay_rayay_nayay == Find/ Replace (PERL pattern, chay$|,  
rayay$| nayay$|, extract lines)
```

lo que me permitió extraer, en primer lugar, todas las formas que contienen posibles lemas verbales (5.541 formas).

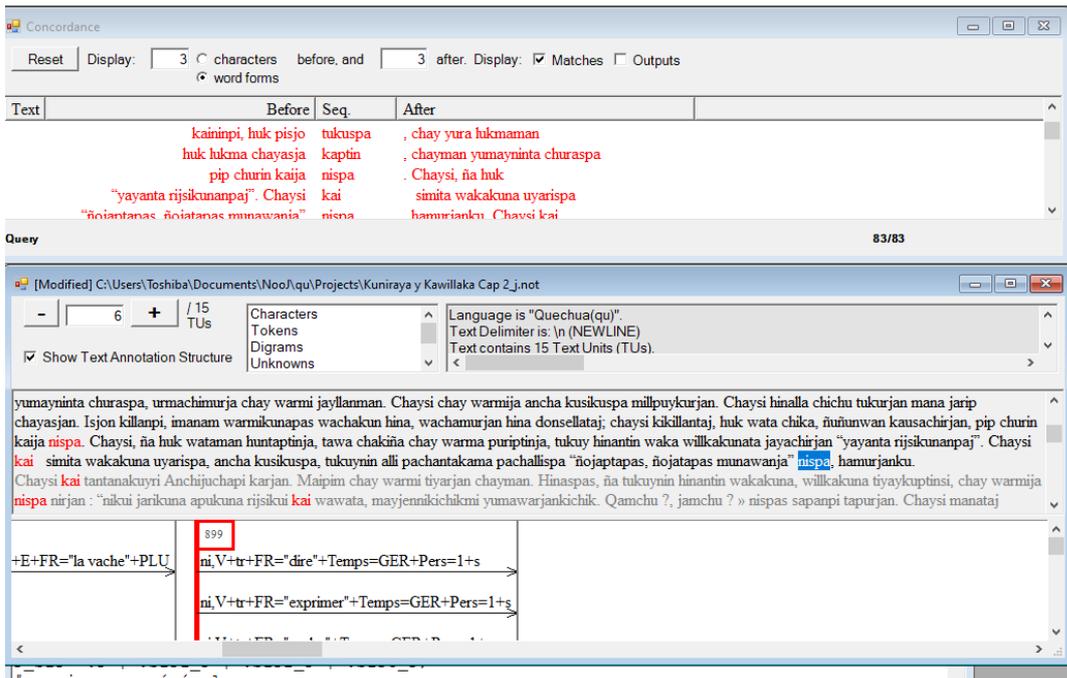


Fig.4 Concordancia y coloración de formas verbales

Después, he podido extraer los lemas verbales inmersos aplicando algunos algoritmos que emplean operadores NooJ. Finalmente, después de un análisis manual, obtuve una lista de doscientos noventa y ocho verbos considerados como verbos 'perdidos' o 'desconocidos'. El rescate de este catálogo de nuevos verbos es un paso importante para la preservación y enriquecimiento del léxico verbal de la lengua. Presentamos una muestra de esos verbos obtenidos en la Fig. 5:

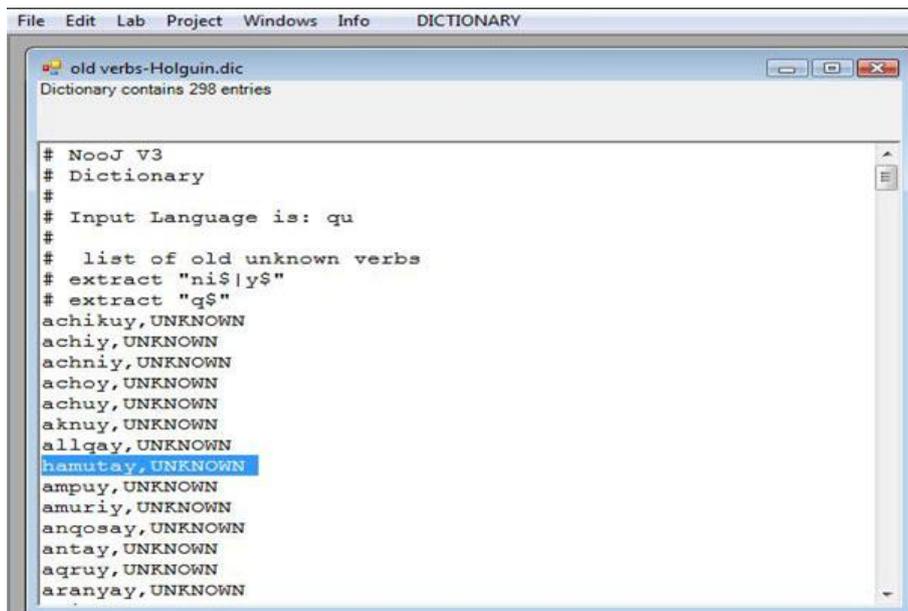


Fig. 5 Extracto de verbos desconocidos hallados en el corpus

qalluykuy/ hacer trampa;
aknay / exhibir;

rampay / guiar a un ciego;
tullpuy / teñir;
utiq / volverse loco;
takuriy / revolucionar;
tokapuy / decorar.

10. Conclusión

He realizado la formalización de los paradigmas que rigen las estructuras flexivas, dependientes de los sufijos posposicionales aglutinados a las desinencias del tiempo indefinido; y de las estructuras derivativas, inducidas por la aglutinación de los sufijos interposicionales, ubicados entre la raíz y la desinencia verbal, de los verbos en quechua. He presentado algunas gramáticas formalizadas, siguiendo los requerimientos de la plataforma lingüística NooJ, que permiten obtener nuevos verbos y también generar miles de flexiones o formas conjugadas simples o mixtas a partir de un solo verbo. He aplicado estas gramáticas formalizadas a nuestro corpus, basado en escritos antiguos, para identificar automáticamente todas esas formas verbales, y extraer de entre ellos las raíces verbales correspondientes. De ese modo, pude ‘descubrir’ doscientos noventa y ocho verbos ‘nuevos’, que agregamos a nuestro léxico de verbos simples del quechua.

11. Referencias

- Bogacki, Krzysztof: *Polish module for NooJ*. In the Proceedings of the 2007 NooJ Conference. Autonomus University of Barcelona. Cambridge Scholars Publishing. Newcastle. (2008)
- De Avila, Francisco, 1598?: *Dioses y hombres de Huarochiri*. Narración Quechua recogida por Francisco de Avila Traducción J. M. Arguedas. Lima. Peru 1966. Edición bilingüe facsimilar. (2012)
- Dubois, Jean et Dubois-Charlier, Françoise (D&D) : *Le verbes français le «dictionnaire électronique des verbes français (DEV)*, 1992 Diffusé à partir de septembre 2007 par MoDyCo dans un format Excel. (2007)
- Durán, Maximiliano: *Diccionario Quechua-Castellano*. Éditions HC. Paris. (2009)
- Durán, Maximiliano : *Formalizing QU verb inflection*. Formalizing Natural Languages with NooJ 2013 International Conference Cambridge Scholars New Castle. England. (2014)
- Durán, Maximiliano: *Morphological and Syntactic grammars for recognition of verbal lemmas in Quechua*. In: *Formalising Natural Languages with NooJ 2015*. Monti J., Silberztein M., Monteleone M., Di Buono M. Eds. Cambridge Scholars Publishing. England.(2016)
- Durán, M.: *Dictionnaire électronique français-quechua des verbes pour le TAL*. Thèse Doctorale. Université de Franche-Comté. Paris (2017).
- Gonçales Holguin, Diego. 1608: *Vocabulario de la Lengua General de todo el Perú llamada Lengua Quichua o del Inca*. Edición y Prólogo de Raúl Porras Barrenechea. Lima, Universidad Nacional Mayor de San Marcos (1952).
- Itier, César: *Dictionnaire Quechua-Français*, Paris. L’Asiathèque. París. (2001)
- Perroud, Pedro Clemente: *Diccionario Castellano - Kechwa Dialecto de Ayacucho*. Lima. Edición (1970)
- Santo Thomas, Domingo de: *Lexicon, o vocabulario de la lengua general del Perú*. Valladolid. Francisco Fernandez de Cordova (1560).
- Silberztein, Max: *NooJ Manual*. <http://www.nooj4nlp.net> (220 pages updated regularly) (2003)
- Silberztein, Max: *Formalizing Natural Languages: The NooJ Approach*. Iste Ediciones. London. (2016)

