

La modelización de la morfología verbal bribri

Modeling the verbal morphology of Bribri

Sofía Flores-Solórzano

Universidad de Costa Rica, Sede del Atlántico, 30501 Turrialba, Costa Rica
sofia.flores.s@gmail.com

Resumen: La modelización computacional de la morfología verbal bribri (lengua de la estirpe Chibchense hablada en el sudeste de Costa Rica) es posible mediante el uso de la tecnología de los transductores de estados finitos (FST). En este trabajo se desarrolla un analizador y generador morfológico a partir de las descripciones gramaticales existentes. Se han definido 11 paradigmas verbales correspondientes a 6 conjugaciones orales y 5 conjugaciones nasales, además de una serie de reglas de alteración fonológica y ortográfica. El sistema soporta casos complejos como los verbos intrínsecamente medios, defectivos o que presenten supletivismo en la raíz, así como la concatenación de sufijos y direccionales.

Palabras clave: Bribri, lenguas indígenas, morfología, análisis morfológico, transductor de estados finitos

Abstract: Computational modeling of the verbal morphology of Bribri (Chibchan family language currently spoken in southeastern Costa Rica) is made possible through the use of finite state transducers (FST). A morphological analyzer and generator have been developed from the existing grammatical descriptions. A total of 11 verbal paradigms corresponding to 6 oral conjugations and 5 nasal conjugations were defined and implemented in the transducer along with a number of phonological and orthographic alteration rules. The system covers complex cases such as middle verbs, defective verbs and suppletivism, as well as the concatenation of suffixes and directionals.

Keywords: Bribri, indigenous languages, morphology, morphological analysis, finite state transducer

1 Introducción

El bribri es una lengua de la estirpe chibchense que hablan en la actualidad aproximadamente unos 7.000 indígenas en el sudeste de Costa Rica (Instituto Nacional de Estadística y Censos, 2013). Hasta la fecha, es probablemente la lengua indígena del país más estudiada y descrita. Se encuentra por lo tanto en una situación favorable para llevar a cabo una investigación aplicada en el campo de la lingüística computacional o el procesamiento del lenguaje natural. Precisamente el analizador morfológico que presentamos constituye un primer intento de modelizar computacionalmente la morfología bribri¹.

El sistema de escritura bribri emplea una serie de diacríticos para representar la tonalidad y la nasalidad; para su introducción se utilizó la herramienta Teclado Chibcha (Flores Solórzano, 2010) que consta de un mapa del teclado y una fuente Unicode que apila adecuadamente los diacríticos, estos se encuentran soportados en el analizador.

El lexicón fue elaborado con el formalismo *Lexc* (*Finite-State Lexicon Compiler*, Compilador Léxico de Estados Finitos.). Este es un lenguaje de programación de tipo declarativo de alto nivel que se usa específicamente para la elaboración de lexicones de lenguas naturales. La sintaxis de este lenguaje está diseñada para facilitar la definición de la estructura morfológica y el tratamiento de algunas irregularidades y ambigüedades ortográficas (Beesley y Karttunen, 2003).

La formulación y ordenamiento de las reglas en cascada, que contienen las alteraciones fonológicas y ortográficas necesarias, se realizó mediante el formalismo denominado *Replace Rules*, reglas de sustitución, desarrollado por Beesley y Karttunen (2003). Las reglas de sustitución son parecidas a las reglas de reescritura utilizadas por la fonología tradicional, incluidas en los trabajos de Chomsky y Halle (1968). Son básicamente expresiones regulares que utilizan unos operadores básicos; su notación es ideal para que los lingüistas puedan definir relaciones complejas de estados finitos de una manera bastante cómoda y familiar (Beesley y Karttu-

¹La herramienta se encuentra disponible para ser evaluada en la URL <http://morphology.bribri.net>.

nen, 2003: 132–137).

Finalmente, con la herramienta *Foma* (Hulden, 2009) realizamos la composición del lexicón y las reglas de sustitución en una única red de estados finitos denominada transductor léxico. El transductor contiene la información morfológica de la lengua bribri: morfemas léxicos, derivativos, inflexivos, alteraciones, infijaciones, supletivismo, interdigitación, composición, etc. Funciona tanto para el análisis como para la generación, en otras palabras, es bidireccional.

Actualmente existen transductores léxicos y *taggers* (etiquetadores) para casi todas las lenguas de prestigio. También se han desarrollado analizadores morfológicos para un limitado número de lenguas indígenas de América; Mager et al. (2018) citan el quechua (Rios, 2010), el mapudungun (Chandía, 2012), el mohawk (Assini, 2013), el plains cree (Snoek et al., 2014), el odawa (Bowers et al., 2017), entre otras. De acuerdo con Mager et al. (2018), el modelo de estados finitos es el más extendido entre los casos que estos investigadores documentan.

En el siguiente apartado describiremos en detalle el modelado de un paradigma verbal bribri en un transductor de estados finitos.

2 La morfología verbal bribri

Uno de los principales retos de la investigación ha sido modelar el sistema verbal bribri. Dicha modelización se ha basado principalmente en las descripciones de Constenla, Elizondo, y Pereira (1998); Margery Peña (1982) y Jara (2013). Asimismo, hemos contado con la consultoría de los maestros bribris Julio Morales Campos y Franklin Morales Morales, ambos tienen una amplia experiencia en la escritura y enseñanza de su lengua. A continuación nos referiremos a los elementos más esenciales del sistema verbal bribri, y posteriormente a su modelización computacional.

El **tema verbal** se obtiene apartando la consonante glotal (saltillo) o el tono descendente de las formas del perfecto improspectivo activo² (Tabla 1). El perfecto improspectivo es la base del tema verbal y es la única forma que permite predecir las formas de la voz media de la conjugación. Por ejemplo, *apàka'* 'narré, visité' es el perfecto improspectivo activo del verbo *apakók* 'narrar, visitar', cuyo tema verbal es *apàka* (la forma improspectiva sin el saltillo).

El sistema verbal bribri se divide en **verbos orales** y **verbos nasales**. La última vocal del tema del verbo es su vocal temática. Los verbos orales son

Perfecto Improspectivo	Tema
<i>apaka'</i>	<i>apaka-</i> narrar
<i>të'</i>	<i>të-</i> cortar, punzar
<i>ali'</i>	<i>ali-</i> cocinar
<i>yulö'</i>	<i>yulö-</i> buscar
<i>u'</i>	<i>u-</i> moler
<i>ñq'</i>	<i>ñq-</i> comer alimentos suaves
<i>tse'</i>	<i>tse-</i> llevar, traer
<i>ini'</i>	<i>ini-</i> jugar
<i>tó</i>	<i>tó-</i> comprar, costar, valer
<i>ku'</i>	<i>ku-</i> encontrar, jalar
<i>sú</i>	<i>sú-</i> ver
<i>wö'ík</i>	<i>wö'ík-</i> soplar
<i>báts</i>	<i>báts-</i> unir
<i>apëit</i>	<i>apëit-</i> prestar, alquilar

Tabla 1: Obtención del tema verbal bribri

aquellos cuyo tema termina en una vocal oral y los verbos nasales, en una vocal nasal. Los temas verbales que terminan en consonante, como *wö'ík* 'soplar', *báts* 'unir' o *apëit* 'prestar', constituyen un paradigma aparte y generalmente son tratados como verbos orales, aunque hay excepciones; por ejemplo, hemos observado que el tema verbal *bi-kéits* 'pensar', el cual termina en consonante, es tratado por algunos hablantes como un verbo oral, i.e. *bikéitsók*³, mientras que por otros como nasal *bi-kéitsuk*⁴.

Los verbos orales se dividen en 6 conjugaciones determinadas por la vocal temática oral o la consonante final del tema, a saber: **-a**, **-ë**, **-i**, **-ö**, **-u** y **-Cons**. Los verbos nasales se subdividen a su vez en 5 conjugaciones según su vocal temática: **-a**, **-e**, **-i**, **-o** y **u**.

La **voz** es la categoría más importante del sistema verbal bribri. Los verbos transitivos e intransitivos suelen presentar **voz activa** y **voz media**. La voz media se usa generalmente para excluir la presencia de un agente. Hay algunos verbos que se refieren a procesos que carecen de conjugación activa y solo se conjugan en la voz media, son verbos defectivos medios i.e. *kiànuk* 'querer', *tsikìnik* 'nacer', *sènik* 'vivir' etc.

Hay al menos **cinco modos**: indicativo, imperativo, de finalidad, optativo y adversativo. El único modo que presenta oposiciones de aspecto y tiempo es el modo indicativo.

En bribri **el aspecto** es más determinante que el tiempo. Es importante indicar si la acción se ha completado (aspecto perfectivo) o si aún se está desarrollando, va a desarrollarse en el futuro o pue-

²Jara (2013) lo denomina perfecto remoto.

³El morfema *-ók* indica el infinitivo de los verbos orales.

⁴*-ík* indica el infinitivo de los verbos nasales.

Tema improspectivo transitivo ali- ‘cocinar’					
Voz activa					
Imperfecto primero	<i>al-</i>	<i>è</i> →	Imperfecto habitual	<i>alè-</i>	<i>ke</i>
Infinitivo	<i>al-</i>	<i>ók</i>	Imperfecto habitual negativo	<i>alè-</i>	<i>ku</i>
Modo imperativo	<i>al-</i>	<i>ó</i>	Imperfecto potencial	<i>alè-</i>	<i>mi</i>
Modo imperativo negativo	<i>al-</i>	<i>ar</i>	Imperfecto futuro	<i>alè-</i>	<i>dá ~ rá</i>
Modo de finalidad	<i>al-</i>	<i>ó</i>	Imperfecto futuro negativo	<i>alè-</i>	<i>pa</i>
Modo optativo	<i>al-</i>	<i>a'ku</i>			
Modo adversativo	<i>al-</i>	<i>a'</i>			
Perfecto prospectivo	<i>al-</i>	<i>é</i>			
Perfecto improspectivo	<i>al-</i>	<i>i'</i>			
		↓			
Voz Media					
Imperfecto primero	<i>alì-</i>	<i>r</i> →	Imperfecto habitual	<i>alìr-</i>	<i>ke</i>
Infinitivo	<i>alì-</i>	<i>nuk</i>	Imperfecto habitual negativo	<i>alìr-</i>	<i>ku</i>
Perfecto prospectivo	<i>alì-</i>	<i>na</i>	Imperfecto potencial	<i>alìr-</i>	<i>mi</i>
Perfecto improspectivo	<i>alì-</i>	<i>ne</i>	Imperfecto futuro	<i>alìr-</i>	<i>dá</i>
Modo de finalidad	<i>alì-</i>	<i>nó</i>	Imperfecto futuro negativo	<i>alìr-</i>	<i>pa</i>
			Forma anterior	<i>alìr-</i>	<i>ule</i>

Tabla 2: Conjugación de los verbos con vocal temática -i

de desarrollarse y está incompleta (aspecto imperfectivo). Hay dos formas perfectivas: el perfecto improspectivo y el perfecto prospectivo. Y cinco formas imperfectivas: el imperfecto primero, el imperfecto habitual, el imperfecto potencial, el imperfecto futuro y la forma anterior. Los verbos bribris carecen de morfemas que especifiquen exclusivamente el tiempo. Las formas perfectivas se refieren generalmente al pasado remoto y al presente; y las formas imperfectivas se pueden referir al pasado, al presente y al futuro. En las traducciones al español se debe apelar al contexto para decidir muchas veces qué tiempo utilizar.

2.1 Paradigma de los verbos terminados en -i

Por razones de espacio no nos referiremos a todas las conjugaciones del sistema verbal bribri. Nuestra intención es explicar la morfotaxis de los verbos regulares mediante un paradigma verbal: la conjugación de los verbos transitivos terminados en -i, y así mostrar su implementación en un transductor de estados finitos.

Los verbos con la vocal temática -i se conjugan en su mayoría de manera regular (Tabla 2). Obsérvese como la vocal temática -i cae o desaparece en la voz activa, y se manifiesta solo en las formas medias y en el perfecto improspectivo. Este comportamiento de la vocal temática es típico de las conjugaciones regulares del sistema verbal bribri, y por consiguiente, de la conjugación de los verbos en -i⁵, como se observa en los siguientes ejemplos:

⁵Sin embargo, es importante tener en cuenta que hay al-

alók (ali') ‘cocinar’, *ichàkók* (ichàki') ‘preguntar’, *tsakók* (tsaki') ‘reventar algo’, *inúk* (inì')’, etc. La primera forma es el infinitivo y la forma entre paréntesis, el perfecto improspectivo.

En la conjugación de la voz activa se agregan una serie de desinencias (-è, -ók, -ó, -ar, -ó, -a'ku, -a', -é) a la raíz del tema verbal (Tabla 2). Entendemos por raíz verbal el tema de verbo sin la vocal temática.

Asimismo, nótese que a partir del imperfecto primero activo, se derivan todas las demás formas imperfectivas activas mediante la concatenación de las correspondientes terminaciones (-ke, -ku, -mi, -dá ~ -rá, -pa).

En la voz media, las terminaciones medias (-r, -nuk, -na, -nó) se agregan al tema verbal –la raíz verbal más la vocal temática–, y la vocal temática cambia de tono bajo a tono alto o descendente.

La forma anterior y los imperfectos medios también se derivan del imperfecto primero medio, al cual se le concatenan las desinencias imperfectivas y la desinencia de la forma anterior (-ule).

gunos verbos cuyo tema termina en -i que mantienen la vocal temática en la raíz del tema verbal, es decir, no se produce la caída de la vocal, i.e. *kiók* (ki') ‘llamar’, *biók* (bi') ‘escarbar’, *tuúk* (tuí) ‘anochecer’. En este último ejemplo, *tuúk* (tuí), la conjugación es nasal a pesar de que el tema verbal sea oral. Como se ejemplifica en los casos expuestos, parece que si el tema es monosilábico y termina en una vocal cerrada anterior -i, hay una clara tendencia a conservar la vocal en la conjugación activa.

Declaración de los símbolos multi-carácter (etiquetas)	
Lexicon Root	Verbos ; Sustantivos ;
Lexicon Verbos	
morfema <i>upper:lower</i>	<i>continuum class</i> ; ! comentario

Tabla 3: Estructura de un archivo *Lexc*

2.2 Implementación de los verbos terminados en -i en un transductor de estados finitos

Como señalamos en la Introducción, el lexicón se elaboró con el Compilador Léxico de Estados Finitos (*Lexc*). La estructura de un archivo *Lexc* tiene tres secciones: a) La declaración de los símbolos multi-carácter o etiquetas; b) una declaración obligatoria que se denomina *Lexicon Root*, donde se enumeran las principales clases del lexicón (generalmente corresponden a las categorías gramaticales, i.e. verbos, sustantivos, adjetivos, pronombres etc.); y c) las diferentes clases del lexicón, *Lexicon classes*, con la descripción morfológica (Tabla 3).

La declaración de los símbolos multi-carácter es opcional y en ella se enumeran las etiquetas que se han definido en las clases del lexicón. La declaración es útil para que el transductor interprete los símbolos multi-carácter precisamente como símbolos o etiquetas y no como cadenas de símbolos, o sea, que interprete un símbolo como +Sust como un único símbolo multi-carácter y no como cinco símbolos: + S u s t.

La declaración *Lexicon Root* es obligatoria. Debajo de ésta se enumeran las principales clases del lexicón. Los nombres de las clases son de libre elección, pero deben finalizar con un punto y coma (;). Las clases enumeradas en esta sección no son las únicas clases presentes en el lexicón sino las que se encuentran en el punto de inicio de la red.

Las clases se introducen con la palabra LEXICON seguida del nombre de la clase. Debajo se describen las diferentes bases léxicas o morfemas; un morfema por línea con su clase continua (*continuum class*) o en su defecto, el símbolo numeral o almohadilla (#) que indica que la cadena no va a admitir más morfemas. Después un punto y coma (;) obligatorio que indica el fin de la cadena. Optativamente se puede añadir un comentario que se indica con el signo de exclamación (!). En nuestro caso, usamos el comentario para indicar la traducción al español de las bases léxicas o morfemas.

En el siguiente fragmento del Lexicón, escrito en *Lexc*, mostramos la modelización de la conjugación de los verbos de tema no monosilábico terminados en -i en un transductor léxico de estados finitos (ejemplo 1). Por razones de espacio no presentamos la clase continua de los direccionales y otros sufijos que pueden presentar los verbos bribris. En figura 1 se observa la visualización del transductor.

- (1) Multichar_Symbols + V + Imp1Tran è + Imp2 ke + Imp2Neg ku + ImpPot mi + ImpFut dâ + ImpFutNeg pa + Inf ók + ModoImp ó + ModoImpNeg ar + ModoFin ó + ModoOpt a'ku + ModoAdvers a' + Prosp é + PerfImp + VozMedia uk a e + Imp1 ó + Anterior ule

LEXICON Root
Verbos ;

LEXICON Verbos
al -i ; ! cocinar tr.

LEXICON -i
i TemaImprospectivo ;
i:0 TemaVozActivaOral ;

LEXICON TemaVozActivaOral
+ V:0 TiemposVozActivaOral ;

LEXICON TiemposVozActivaOral
+ Imp1Tran:è TiemposImperfectoPrimero ;
+ Inf:ók # ;
+ ModoFin:ó # ;
+ Prosp:é # ;
+ ModoOpt:a'ku # ;
+ ModoAdvers:a' # ;
+ ModoImp:ó # ;
+ ModoImpNeg:ar # ;

LEXICON TemaImprospectivo
+ V:0 Improspectivo ;

LEXICON Improspectivo
+ VozMedia:d TiemposVozMedia ;
+ PerfImp:' # ;

LEXICON TiemposVozMedia
+ Prosp:a # ;
+ PerfImp:e # ;
+ Inf:uk # ;
+ Imp1:0 TiempoAnterior ;
+ ModoFin:ó # ;

LEXICON TiempoAnterior
+ Anterior:ule # ;
0 TiemposImperfectoPrimero ;

LEXICON TiemposImperfectoPrimero	
+ Imp2:ke	# ;
+ Imp2Neg:ku	# ;
+ ImpFut:dâ	# ;
+ ImpFutNeg:pa	# ;
+ ImpPot:mi	# ;
0	# ;

De acuerdo con la morfotáctica definida en el ejemplo 1, la base o raíz verbal⁶, en este caso, *al*, tiene como clase continua la clase *-i* (correspondiente a la vocal temática). A su vez la clase *-i* presenta dos recorridos posibles: si la cadena de entrada termina con la vocal *-i*, su clase continua es la clase *TemaImprospectivo*; si la entrada no lleva la vocal temática (de ahí el símbolo 0), le corresponde la clase *TemaVozActivaOral*.

En las clase *TemaVozActivaOral*, la cadena de entrada se contrasta con el elemento “+V”, símbolo multi-carácter que se encuentra en el lado superior del red, y concatena este símbolo en el lado superior del recorrido.

En la clase *TiemposVozActivaOral*, la cadena de entrada se vuelve a contrastar con el lado inferior de la red, donde se encuentran los morfemas del imperfecto primero transitivo, el infinitivo, el modo de finalidad, el prospectivo, el modo optativo, el modo adversativo, el modo imperativo y el modo imperativo negativo. En caso de existir correspondencia, se concatena el símbolo correspondiente en el lado superior. Por razones prácticas, hemos llamado a estas clase *TiemposVozActivaOral*, aunque somos conscientes de que la mayoría de los morfemas son de modo.

Tras concatenar alguna de las formas léxicas identificadas con los símbolos multi-carácter +Imp1Tran, +Inf, +ModoFin, +Prosp, +ModoOpt, +ModoAdvers, +ModoImp o +ModoImpNeg, todas las cadenas finalizan el recorrido, como lo indica el símbolo almohadilla, con excepción de la cadena que presente *è* en el lado inferior, que continúa hacia la clase siguiente: *TiemposImperfectoPrimero*.

Por otro lado, si en la forma superficial de la cadena de entrada está presente la vocal temática, a la cadena le corresponde la morfotáctica de las clases *TemaImprospectivo*, *Improspectivo* y *TiemposVozMedia*.

Finalmente, apuntar que *-d* es el morfema de la voz media bribri. En el nivel fonológico /d/ tiene dos alófonos en distribución complementaria, [r] y [n]. Mediante tres reglas de alteración aplicadas en paralelo, hemos modelado la realización del fonema /d/:

(2) define Alófonos

⁶La raíz verbal es el tema verbal sin la vocal temática.

d (->) r +V tono V' V _ [.#. Vo Cons] ,,
d (->) n V tono V _ [Vn .#. Vn] ,, d (->) n Vn tono _ .#. ;

De acuerdo con la primera regla del ejemplo 2, *d* puede optativamente realizarse como *r* antes de una vocal con o sin tono o saltillo y seguida de posición final de palabra o vocal oral o consonante. Según la segunda regla, *d* se realiza como nasal, *n*, antes de una vocal con o sin tono seguida de una vocal nasal que puede estar en posición final. La tercera regla indica que la realización nasal de /d/ puede darse antes de una vocal nasal con tono y seguida de posición final.

3 Resultados

La composición del lexicón junto con las reglas de alteración produce un transductor de estados finitos compuesto por 10.609 estados y 38.515 arcos. El análisis morfológico automático (*apply up*) de la conjugación del verbo *alók* en dicho transductor arroja los siguientes resultados:

- (3) alè
ali + V + Imp1Tran
+ Adv [duda]
- alók
ali + V + Inf
- aló
ali + V + ModoImp
- alar
ali + V + ModoImpNeg
- aló
ali + V + ModoFin
- ala'ku
ali + V + ModoOpt
- ala'
ala' + Sust
ali + V + ModoAdvers
- alé
ali + V + Prosp
- alèke
ali + V + Imp1Tran + Imp2
- alèku
ali + V + Imp1Tran + Imp2Neg
- alèmi
ali + V + Imp1Tran + Suf [incoativo]
ali + V + Imp1Tran + Dir

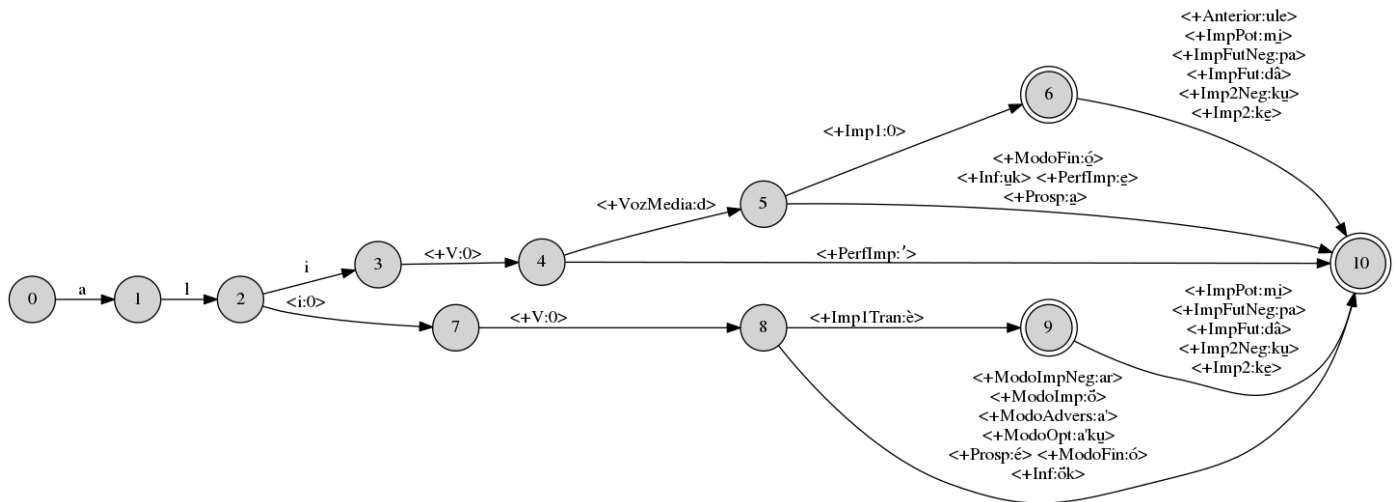


Figura 1: Ejemplo de la modelización en una red de estados finitos de tipo transductor de la conjugación del verbo *alók* ‘cocinar’ (Gráfico generado automáticamente con *Foma* y el paquete *GraphViz*)

ali + V + Imp1Tran + ImpPot

alèdâ

ali + V + Imp1Tran + ImpFut

alèrà

ali + V + Imp1Tran + ImpFut

alèpa

ali + V + Imp1Tran + ImpFutNeg

alìr

ali + V + VozMedia + Imp1

alìnuḱ

ali + V + VozMedia + Inf

alìna

ali + V + VozMedia + Prosp

alìne

ali + V + VozMedia + PerfImp
 ali + V + PerfImp + Suf[reiterativo]
 ali' + Sust + Suf[iterativo]

alìnó

ali + V + VozMedia + ModoFin

alìrke

ali + V + VozMedia + Imp1 + Imp2

alìrku

ali + V + VozMedia + Imp1 + Imp2Neg

alìrmi

ali + V + VozMedia + Imp1 + Suf[incoativo]
 ali + V + VozMedia + Imp1 + Dir
 ali + V + VozMedia + Imp1 + ImpPot

alìrdâ

ali + V + VozMedia + Imp1 + ImpFut

alìrpa

ali + V + VozMedia + Imp1 + ImpFutNeg

alìrule

ali + V + VozMedia + Imp1 + Anterior

El analizador nos devuelve la cadena que se encuentra en el lado superior de la red (*upper-side*) más una serie de etiquetas que se componen de un operador “+” y una categoría gramatical. Como puede comprobarse, las “etiquetas” contienen la información morfosintáctica.

Los resultados del análisis automático del ejemplo 3 son satisfactorios, pues todas las formas léxicas –aquellas que se encuentran en el lado superior de la red– existen en la lengua bribri. En el ejemplo 3, la mayoría de los casos presentan el tema verbal, *ali*, más el símbolo multi-carácter +V, que indica que se trata de una forma verbal. A continuación, dependiendo de la forma de entrada, pueden concatenarse otros símbolos multi-carácter, que indican el aspecto, modo y tiempo verbal.

En algunos casos del ejemplo 3, el análisis arroja más de un resultado e incluso otras categorías gramaticales, como *alè* +Adv[duda] y *ala'* +Sust, correspondientes con el adverbio de duda *alè* ‘tal vez, quizás’⁷ y el sustantivo *ala'* ‘olor’. Esas formas léxicas son homónimas de las formas verbales *alè* ‘cocino’ y *ala'* ‘cocino para molestar a alguien’ (modo adversativo).

Cuando el análisis arroja más de una forma léxi-

⁷Constenla, Elizondo, y Pereira (1998) lo llama partícula de duda.

ca, evidentemente, se trata de casos de homonimia que existen en la lengua y que el desarrollo actual de la herramienta no nos permite desambiguar porque aún no hemos trabajado en el nivel sintáctico; en otras palabras, aunque podemos analizar morfológicamente frases y enunciados, la herramienta no tiene en cuenta el contexto de la frase u oración.

Por cierto, con excepción de los símbolos multicarácter definidos como tales, los diacríticos bribris son tratados por el transductor como símbolos individuales, lo cual permite distinguir y analizar de manera correcta pares bribris que solo se distinguen por el tono o la nasalidad. Por ejemplo, la distinción dentro de los verbos transitivos, entre el imperfecto primero y el perfecto prospectivo, solo se marca por el tono alto y el tono descendente, respectivamente, i.e. *kékará ye' àrros alè* 'yo cocino arroz siempre' y *ye' àrros alé chki* 'yo cociné arroz ayer', esta distinción se encuentra soportada en la herramienta.

Por último, conviene recordar que la escritura bribri no posee un sistema estandarizado; además coexisten al menos tres variantes dialectales, el bribri de Amubri, Coroma y Salitre. Como es común en muchas lenguas indígenas de América, los hablantes de cada variante defienden un sistema de escritura propio que refleje fielmente sus particularidades fonéticas. Así, el verbo *stsók* 'cantar', documentado con -s inicial en Constenla *et al.* (1998) y Margery Peña (1982) –ambos investigadores documentan principalmente los dialectos de Amubri y Salitre– pierde esta s en el dialecto de Coroma, hecho que se refleja en la escritura de esa forma en la variedad de Coroma y posiblemente en la variedad de Amubri, *tsók* 'cantar'.

También el verbo *ijtsók* 'sentir' sufre aféresis de -ij en los tres dialectos, hecho que se refleja en la forma escrita utilizada mayoritariamente, *tsók* 'sentir'; por consiguiente, esta forma es homónima de *tsók* 'cantar'.

En todos estos casos, hemos decidido recoger en el lexicón todas las formas documentadas y utilizadas. Empleamos el mecanismo *upper:lower* del lenguaje *Lexc* para reflejar estos cambios, como se observa en el ejemplo 4:

- (4) sts -ë ; ! cantar intr.
sts:ts -ë ; ! cantar intr.

ijts -ë ; ! sentir, oír tr.
ijts:ts -ë ; ! sentir, oír tr

Tras compilar la gramática, el análisis de *stsók* devuelve una única forma léxica:

- (5) apply up > stsók
stsë + V + Inf

Mientras que el análisis de *tsók* devuelve dos formas léxicas:

- (6) apply up > tsók
ijtsë + V + Inf
stsë + V + Inf

Como puede comprobarse en el ejemplo 6, las dos formas léxicas son válidas para nuestro sistema, y también para el sistema de escritura bribri actual. Asimismo, cuando realizamos el proceso contrario, la forma léxica *stsë + V + Inf* genera (*apply down*) los siguientes resultados:

- (7) apply down > stsë + V + Inf
stsók
tsók

Es decir, dos formas superficiales válidas (ejemplo 7).

4 Conclusiones

El presente trabajo se ha focalizado en modelar el sistema verbal bribri. Lo más destacable a este respecto ha consistido en la elaboración de los paradigmas verbales. La descripción formal que hemos realizado de la morfología verbal bribri nos ha permitido mejorar la comprensión del funcionamiento del sistema verbal y realizar satisfactoriamente su implementación en *Lexc*.

Una de los principales limitaciones de trabajar computacionalmente con lenguas no mayoritarias es la falta de corpora y recursos lingüísticos. En consecuencia, la evaluación del analizador se ha llevado a cabo con el único corpus (Flores Solórzano, 2017) elaborado hasta la fecha, compuesto por textos provenientes del habla cotidiana que fueron grabados y transcritos entre los años 2014 y 2016. La cobertura de este corpus es del 100 %, con excepción de algunos préstamos castellanos. Somos conscientes de la necesidad de aumentar su tamaño y también de la conveniencia de utilizar otros corpus. Parte del trabajo futuro será precisamente aumentar el corpus con más textos provenientes del habla cotidiana, y al mismo tiempo reunir las publicaciones existentes, solicitar los respectivos permisos y derechos de autor, para contar con un corpora sólido que nos permita no solo mejorar la cobertura del analizador, sino desarrollar otras herramientas útiles para contribuir a la documentación y estudio de esta lengua.

Agradecimientos

Esta investigación ha sido posible gracias al financiamiento otorgado por la Vicerrectoría de Investigación de la Universidad de Costa Rica mediante el proyecto B5233.

Simbología

(->) Sustitución optativa.

.#. Marcador de límite de palabra.

- _ Marca la posición contextual del elemento que se sustituye en una regla de sustitución.
- | Operador que indica la unión de dos conjuntos.
- || Marca el contexto en el que se va a aplicar la regla.

+**Adv** Adverbio.

+**Anterior** Forma anterior.

Cons Consonante.

+**Dir** Direccional.

+**Imp1** Imperfecto primero (medio).

+**Imp1Tran** Imperfecto primero transitivo.

+**Imp2** Imperfecto segundo.

+**Imp2Neg** Imperfecto segundo negativo.

+**ImpFut** Imperfecto futuro.

+**ImpFutNeg** Imperfecto futuro negativo.

+**ImpPot** Imperfecto potencial.

+**Inf** Infinitivo.

+**ModoAdvers** Modo adversativo.

+**ModoFin** Modo de finalidad.

+**ModoImp** Modo imperativo.

+**ModoImpNeg** Modo imperativo negativo.

+**ModoOpt** Modo optativo.

+**PerfImp** Perfecto improspectivo.

+**Prosp** Prospectivo.

+**Suf** Sufijo.

+**Sust** Sustantivo.

+**V** Verbo.

Vn Vocal nasal.

Vo Vocal oral.

+**VozMedia** Voz media.

Bibliografía

- Assini, A. A. 2013. Natural language processing and the mohawk language: creating a finite state morphological parser of mohawk formal nouns. Master's thesis.
- Beesley, K. R. y L. Karttunen. 2003. *Finite state morphology*. CSLI (Center of the Study of Language and Information), Stanford, California.
- Bowers, D., A. Arppe, J. Lachler, S. Moshagen, y T. Trosterud. 2017. A morphological parser for odawa. En *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, páginas 1–9.

Chandía, A. 2012. Dunggupeyem: analizador y generador morfológicos para mapudungun.

Chomsky, N. y M. Halle. 1968. *The sound pattern of English*. Harper and Row, New York.

Constenla, A., F. Elizondo, y F. Pereira. 1998. *Curso básico de bribri*. Editorial de la Universidad de Costa Rica, San José.

Flores Solórzano, S. 2010. Teclado chibcha: Un software lingüístico para los sistemas de escritura de las lenguas bribri y cabécar. *Revista de Filología y Lingüística de la Universidad de Costa Rica*, 36(2):155–161.

Flores Solórzano, S. 2017. Corpus oral pandialectal oral de la lengua bribri. Disponible en <http://bribri.net>.

Hulden, M. 2009. Foma: a finite-state compiler and library. En *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, páginas 29–32. Association for Computational Linguistics.

Instituto Nacional de Estadística y Censos. 2013. *X Censo Nacional de Población y VI de Vivienda: Territorios Indígenas*. INEC, San José, Costa Rica, 1 edición, abril.

Jara, C. V. 2013. Morfología verbal de la lengua bribri. *Estudios de Lingüística Chibcha*, 32:95–152.

Mager, M., X. Gutierrez-Vasques, G. Sierra, y I. Meza. 2018. Challenges of language technologies for the indigenous languages of the americas. En *In Proceedings of the 27th International Conference on Computational Linguistics*.

Margery Peña, E. 1982. *Diccionario fraseológico bribri-español, español-bribri*. Editorial Universidad de Costa Rica, San José.

Rios, A. 2010. *Applying Finite-State Techniques to a Native American Language: Quechua*. Ph.D. thesis, Institut für Computerlinguistik, Universität Zurich.

Snoek, C., D. Thunder, K. Lõo, A. Arppe, J. Lachler, S. Moshagen, y T. Trosterud. 2014. Modeling the noun morphology of plains cree. 06.