

## Detección automática de chilenismos verbales a partir de reglas morfosintácticas. Resultados preliminares

### *Automatic detection of verbal chilenismos using morphosyntactic rules. First results*

**Walter A. Koza, Pedro Alfaro Faccio**  
Pontificia Universidad Católica de Valparaíso  
Av. El Bosque 1290, Viña del Mar, Chile  
walter.koza@ucv.cl

**Ricardo Martínez Gamboa**  
U. Diego Portales/U. de La Frontera  
Vergara 240, Santiago, Chile  
ricardomartinezg@gmail.com

**Resumen:** En el presente artículo, se describen las tareas realizadas para el desarrollo de un extractor automático de verbos diferenciales del español chileno mediante la aplicación de reglas de lenguaje natural. A partir de este objetivo, se procedió a la modelización de características léxicas, morfológicas y sintácticas de estas expresiones, la cual permitió la posterior implantación en máquina. En primer lugar, se clasificaron los chilenismos verbales en cuatro tipos, según su uso registrado en los diccionarios y su naturaleza sintáctica: puros, puros pronominales, de sentido y de sentido pronominales. En segundo lugar, se establecieron reglas sintácticas para el reconocimiento automático. En el trabajo computacional se utilizaron las herramientas Smorph y Módulo Post Smorph, que trabajan en bloque a base de reglas del lenguaje natural. Este método se probó en un corpus compuesto por 5.194 *tweets*, producidos por usuarios chilenos, logrando 85,54% de precisión, 96.16% de cobertura y 90,53% de medida f. Los resultados permiten validar el método propuesto, aunque se detectaron algunas limitaciones y detecciones erróneas, lo que implicaría la necesidad de especificación de algunas reglas y la creación de nuevas, tanto para la detección automática como para el filtrado de etiquetados erróneos.

**Palabras clave:** Chilenismo verbal, detección automática, reglas morfosintácticas, Smorph, MPS.

**Abstract:** In this paper, the tasks made for obtaining an automatic extractor for verbal chilenismos using natural language rules are described. With this objective, a formalization of lexical, morphological and syntactic features was made, for a subsequent computational implementation. Firstly, verbal chilenismos were classified in four kinds, according to the use registered in the dictionaries and syntactic features: pure, pure-clitic, of sense, and of sense-clitic. Secondly, syntactic rules were established for the automatic recognition. Smorph and Post Smorph Module were used in the computational work, both use natural language rules. The method was tested in a corpus composed by 5194 *tweets* produced in Chile, obtaining 85.54% of precision, 96.16% of coverage, and 90.53% of F-measure. The results show that this method is able for this kind of work, all the same, some limitations and mistakes were detected and more specific and new rules are necessary for the recognition task and for filtering wrong tagged.

**Keywords:** Verbal chilenismo, automatic detection, morphosyntactic rules, Smorph, MPS.

## 1 Introducción

Los diccionarios diferenciales son obras lexicográficas que buscan codificar una variedad nacional de una lengua a través de un proceso de estandarización. Particularmente en Chile se han elaborado varios de estos diccionarios. Más allá de la relevancia de estas obras, las metodologías que se han solido utilizar podrían considerarse insuficientes al momento de enfrentar grandes bases de datos de lenguaje natural, en la medida en que ocupan principalmente técnicas manuales en las que priman el contraste entre obras, las entrevistas a hablantes y el juicio de expertos. En este sentido, contar con una herramienta automática que permita la recopilación de voces y la construcción de lexicones sería de gran ayuda para los lexicógrafos.

En este marco, se propone una modelización del chilenismo verbal a través de una descripción morfosintáctica, que permita su implantación en máquina con el fin de establecer un método de detección automática basada en lenguaje natural. Mediante la concreción de este objetivo se pretende aportar a las tareas de extracción de información, aplicadas a la lexicografía diferencial, y proponer una modelización de estructuras morfosintácticas.

Así, el trabajo consistió en la elaboración de un diccionario electrónico con los lemas univerbales del DUECh, a los que se les asignó un modelo de acuerdo a sus especificidades morfológicas. Esto permitió contar con una herramienta que generara chilenismos verbales en las diversas formas flexivas y que los pudiera detectar en textos de lenguaje natural.

Desde una perspectiva teórica, se establecieron cuatro tipos de chilenismos verbales: (1) puros ('achorar'), (2) puros pronominales ('chacrearse'), (3) de sentido ('afilarse') y (4) de sentido pronominales ('hacerla'). Posteriormente, se elaboraron reglas de carácter sintáctico a partir de la combinación entre pronombres clíticos, verbos y otras estructuras, como, por ejemplo, sintagmas preposicionales.

Para el trabajo computacional, se recurrió a los *software* Smorph (Ait-Mokthar, 1998) y Módulo Post Smorph, MPS (Abacci, 1999), que trabajan en bloque. Smorph es un analizador y generador textual que, en una sola etapa, realiza la segmentación, lematización y análisis morfológico. MPS, por su parte,

toma como entrada el *output* de Smorph y, a través de reglas de recomposición, descomposición y correspondencia, analiza la cadena de lemas resultante del análisis morfológico.

Este método fue probado en un corpus compuesto por 5.194 *tweets*, producidos en Chile, provistos por AnalITIC<sup>1</sup>.

## 2 Caracterización del chilenismo verbal

### 2.1 Acerca del chilenismo

Como se mencionó, este trabajo tiene por objetivo desarrollar un método de detección automática de chilenismos verbales -o verbos diferenciales chilenos-, a partir de reglas de lenguaje natural. Se ha adoptado una definición de chilenismo verbal que se corresponda con sus características lingüísticas y que, a la vez, sea acorde con el trabajo computacional de detección automática. En este sentido, se considera que un chilenismo es toda palabra (o expresión) de uso documentado en el habla de Chile, cuyo lema: (i) no está registrado en el DRAE ('funar'); (ii) el DRAE lo registra como chilenismo ('vitriñar'); y (iii) si bien el lema se encuentra en el DRAE, en Chile, se utiliza para expresar un significado distinto al descrito en dicho diccionario ('pinchar').

Otra de las características de estas expresiones es que tienden a estar registradas en diccionarios diferenciales. Este tipo de trabajos constituyen obras lexicográficas que buscan codificar los significados de una variedad nacional de una lengua (Chávez, 2010). La obra más importante de este tipo en Chile es el *Diccionario ejemplificado de chilenismos y de otros usos diferenciales del español de Chile, DECh*, primer diccionario diferencial chileno de corte científico (Morales, 1984). En esta obra Morales (1984: XXXI) procura incluir entre sus voces:

*todo uso o acepción que, además de emplearse o de haberse empleado alguna vez en Chile, no perteneciera al empleo común o general, es decir, no dialectal, tal como lo registra la Real Academia en sus diccionarios oficiales.*

Con este propósito, el procedimiento para determinar las voces que no corresponden al español general consistió en contrastar ítems léxicos de uso documentado en Chile que no estuviesen incorporados en el DRAE (Sáez,

<sup>1</sup> [www.analitic.cl](http://www.analitic.cl)

2011). Para ello el autor estableció diferentes tipos de diferencialidad, indicados por abreviaciones o signos específicos :

- [N] no incorporada en el DRAE
- [\*] desplazamiento gramatical
- [f] cambio fónico
- [g] cambio gráfico
- [l] cambio lexemático
- [c] cambio de la extensión semántica
- [C] cambio en la comprensión (metáforas, metonimias)

Gracias a esta caracterización, las diferencialidades fueron abordadas de manera más detallada que la simple no incorporación al DRAE, al tiempo en que se profundizó en la naturaleza de las voces diferenciales de la variedad del español de Chile.

En 2010 la Academia Chilena de la Lengua publica su propio diccionario diferencial de chilenismos, el *Diccionario de uso del español de Chile, DUECh*, cuya construcción sigue los principios que se utilizaron para el DECh. En efecto, Matus (2010: 4) ha señalado que en este diccionario

*para verificar esta diferencialidad dialectal se ha empleado una batería de contrastividad constituida por un conjunto de diccionarios que contienen léxico general, corpus electrónicos, buscadores (como Google) y encuestas aplicadas a informantes. Esta batería ha sido aplicada rigurosamente a cada una de las unidades léxicas y para cada una de sus acepciones.*

Si bien, el DUECh no indica los tipos de diferencialidad y, en términos metodológicos, no existe información del modo en que los ítems léxicos fueron seleccionados (Sáez, 2011), esta obra constituye el referente más actualizado y exhaustivo de los usos dialectales en Chile.

Cabe destacar que es en este contexto en que surge la motivación para el presente trabajo, pues se busca contribuir desde la extracción automática de información a la elaboración de este tipo de lexicones. A tales efectos, se elaboró un diccionario electrónico para la detección automática a partir de la lista de lemas verbales presentes en el DUECh. Posteriormente, se le asignó un modelo específico a cada uno de ellos, de acuerdo con sus características morfosintácticas.

## 2.2 Morfología del chilenismo verbal

En el DUECh se registran 944 casos de chilenismos verbales. Estos se distribuyen en las terminaciones “ar” (629 casos, 66,6%), “er” (10 casos, 1,1%), “ir” (13 casos, 1,5%) y los restantes se distribuyen en modelos pronominales (289 casos, 30,6%), tales como “arse”, “earla” o “árselas”. La forma lexicogenética verbal más productiva del español de Chile es la terminación “ear” (329 casos, 34,85%), tal como ya documentara Morales, Quiroz y Mayorga (1969). Esta es, además, la menos documentada por el DRAE, dado que se incluyen en los diccionarios de la Academia solo el 24,3% de estos verbos. Otro aspecto relevante consiste en que una de las maneras más productivas del español de Chile para formar verbos nuevos es la que, de acuerdo con Morales y Quiroz (1983), corresponde a un desplazamiento de tipo gramatical, en este caso, la pronominalización de un verbo ya documentado en el español general, como, por ejemplo “agarrar(la)”. En general estos verbos forman un complejo transitivo interno en que se desplaza gramaticalmente la función verbal desde lo intransitivo o lo transitivo (externo) a formas en que se internaliza el Objeto Directo. En estos casos prácticamente la totalidad de las formas no clíticas están presentes en el DRAE.

A partir de esta descripción, se elaboraron modelos morfológicos y se establecieron cuatro categorías a partir de características morfosintácticas:

1. Puros: se trata de verbos cuyo uso se da casi de manera exclusiva en Chile. La mayoría de estos no está registrado en el DRAE (‘marquetear’) o este último los clasifica como chilenismos (‘lolear’).
2. Puros pronominales: además de tener un uso exclusivo en Chile, estos verbos van acompañados de clíticos (‘enyegüecerse’).
3. De sentido: verbos que, si bien sus lemas están incluidos en el DRAE, poseen un uso particular en Chile. Este es el caso de ‘pinchar’, que significa ‘tener a una relación sentimental sin compromiso de exclusividad ni vínculo legal o religioso’.
4. De sentido pronominales: al igual que los de sentido, su lema está registrado en el DRAE, no obstante, para conformar una expresión propia de Chile, se deben combinar con clíticos (‘podérsela’, ‘casarse’).

Esta clasificación obedece a dos motivos. Por un lado, se pretende establecer una clasificación que contemple la naturaleza de los chilenismos verbales y, por otro, busca ser análoga al trabajo computacional.

El método fue probado en un corpus textual compuesto por 5.194 *tweets*, producidos en Chile entre el 22 y el 28 de noviembre de 2013 con el hashtag #Falabella, provistos por AnaliTIC.

En la sección siguiente, se presenta la implantación en máquina realizada a partir de la categorización propuesta.

### 3 Metodología

A fin de corroborar la descripción morfológica de los verbos diferenciales chilenos, se llevó a cabo una modelización de dicha descripción para, posteriormente, realizar una implantación en máquina y, así, generar la conjugación; con ellos, finalmente, se pretende detectar estas expresiones en textos de lenguaje natural.

Para el trabajo informático, se recurrió a las herramientas, que trabajan en bloque, Smorph (Ait Mokhtar, 1998) y Módulo Post Smorph, MPS (Abacci, 1999).

#### 3.1 Smorph

Smorph es un analizador y generador textual que, en una única etapa, realiza segmentación, lematización y análisis morfológico. Se trata de una herramienta declarativa, en la cual la información lingüística está separada de la maquinaria algorítmica, lo que permite que se la pueda adaptar tanto a cualquier lengua como a cualquier variedad lingüística -por ejemplo, en este caso, al español chileno. En este programa se declaran cinco tipos de informaciones: (1) Códigos Ascii, (2) Entradas, (3) Modelos, (4) Terminaciones y (5) Rasgos.

Los códigos Ascii refieren a la notación específica de Smorph, por lo que no ha sido intervenida para este experimento. Describimos las demás informaciones a continuación.

##### 3.1.1 Entradas

Las entradas constituyen el diccionario lingüístico en el que las expresiones (palabras) tienen la posibilidad de aparecer. En este archivo, la información se declara de tres maneras posibles:

- A partir de los lemas con la indicación precisa del modelo morfológico que siguen (1)
- Directamente con la indicación de los rasgos morfológicos (2)
- Con la indicación de categoría gramatical y la información considerada pertinente por el usuario (3)

(1) penquear	@vch1
(2) lo	/clac .
(3) de	/prepde .
con	/prepron .

En el caso de 'penquear' ('reprender') se presenta el lema que se expresa convencionalmente con la forma infinitiva, tal como ocurre en los diccionarios comunes. Es decir, 'penquear' es el lema que representa al grupo de verbos 'penqueo', 'penqueas', 'penquea', 'penqueamos', 'penqueáis', 'penquean', 'penqueé', 'penqueaste', etc. En el caso de (2), no se recurre a ningún modelo, sino que solo se señala el carácter de pronombre clítico acusativo mediante la expresión 'clac'. En el caso de las preposiciones (3), fue necesario destacar cada una de ellas, por lo que, además de la etiqueta 'prep', de preposición, se le adicionó la preposición misma. En esta ocasión, se establecieron modelos morfológicos y morfosintácticos para los verbos chilenos puros, diferenciándose aquellos que eran pronominales. Así, por ejemplo, 'penquear' se considera un chilenismo puro cuando remite a 'reprender', pero cuando se combina con un clítico reflexivo ('me penqueé'), el significado alude a embriagarse. Para diferenciar ambos significados, en el archivo entradas, el lema 'penquear' aparece dos veces. De este modo, se distingue el uso con pronombre clítico del que no lo requiere.

(4) penquear	@vch1
penquear	@vchpron1

Adicionalmente, es necesario señalar que se asignaron etiquetas especiales con información morfológica y sintáctica para los verbos de sentido pronominales, según el pronombre que requieren para convertirse en chilenismo y, en caso de corresponder, la preposición pertinente. Por ejemplo, 'comer', cuando alude a 'tener relaciones sexuales', se combina con un clítico reflexivo, más la preposición 'a', más un sintagma nominal ('Juan se come a la vecina'). Similar

comportamiento tiene ‘hacer’ (‘tener algo como objeto frecuente de acción’, ejemplo, ‘le hace al canto’). Para estos verbos, se creó la etiqueta ‘vchpronrefa’ (verbo chileno pronominal preposición ‘a’).

Para la presente investigación, se utilizó la lista de entradas correspondientes a verbos, nombres, adjetivos, adverbios, preposiciones, siglas y marcadores discursivos desarrollado por el equipo Infosur<sup>2</sup> de la Universidad Nacional de Rosario, Argentina. A este archivo se le adicionaron los verbos incluidos en el DUECh.

De este, se han extraído 960 unidades univocales que corresponden a 483 verbos definidos como intransitivos y 477 verbos definidos como transitivos.

### 3.1.2 Modelos

En los modelos, se consigna la estructura morfológica. Los modelos se introducen a través del símbolo @, que indica el lugar en que va la forma básica o raíz a la que se concatenan las terminaciones. En el ejemplo, se muestra un fragmento para el modelo 1 de verbos chilenos regulares de la primera conjugación.

@vch1	-2
+o	vch/pres/ind/1a/sg/c1/r
+as	vch/pres/ind/2a/sg/c1/r
+ás	vch/pres/ind/2a/sg/c1/r
+ai	vch/pres/ind/2a/sg/c1/r/ch
+a	vch/pres/ind/3a/sg/c1/r
+amos	vch/pres/ind/1a/pl/c1/r
+áis	vch/pres/ind/2a/pl/c1/r
+an	vch/pres/ind/3a/pl/c1/r
+aba	vch/imp/ind/1a/sg/c1/r
+abas	vch/imp/ind/2a/sg/c1/r
+abai	vch/imp/ind/2a/sg/c1/r/ch
+aba	vch/imp/ind/3a/sg/c1/r
(...)	

Esto se lee de la siguiente manera, primero se indica el número de caracteres que se extrae al lema. Eso significa que a un verbo como ‘lolear’ se le quita ‘ar’ y se va combinando con las diferentes desinencias correspondientes, con las variaciones de persona, número, tiempo y modo.

A cada uno de ellos se le asignó el modelo correspondiente de acuerdo a sus particularidades de regularidad, por ejemplo, en el caso de ‘huevear’, al ser un verbo

regular se le asignó el modelo de verbos 1. En cambio, a un verbo del tipo ‘acolloncar’, se le asignaron dos modelos, uno correspondiente a la forma regular ‘acollonc-’ (vch10) y otro, a la irregular ‘acollonqu-’ (vch11). Además, cabe destacar que en los modelos se incluyó la variación de segunda persona del singular del español chileno, para expresiones como ‘penqueai’.

### 3.1.3 Terminaciones

Se trata de una serie de caracteres que expresan un rasgo o un conjunto de rasgos. En las terminaciones se incluyen, entre otros aspectos, las desinencias verbales. Vale aclarar que Smorph permite la inclusión de lo que se ha denominado ‘terminaciones distinguidas’ (Aït-Mokhtar y Lázaro, 1995). Estas consisten en los finales de palabras que permiten determinar la categoría gramatical, son similares a la noción de sufijo aunque pueden diferir en algunos casos. Así, por ejemplo, se sabe que toda palabra terminada en –ción es un nombre femenino singular o que la terminación –ó es propia de un verbo en pretérito perfecto simple, de la tercera persona del modo indicativo. Las terminaciones distinguidas permiten detectar aquellas palabras que no estén incluidas en el archivo de entradas, tales como los neologismos. En esta ocasión, se cargó ‘ó’ como terminación distinguida de la siguiente manera:

ó v/3/sg/perf/ind

### 3.1.4 Rasgos

Para construir los modelos, se recurre a rasgos morfológico-sintácticos y, en esta ocasión, a información léxica presente en los diccionarios diferenciales. Por ejemplo, se tienen: EMS (etiqueta morfosintáctica), que incluye los valores ‘n’ (nombre), ‘adj’ (adjetivo), ‘v’ (verbo), ‘vch’ (verbo chileno), ‘vchpron’ (verbo chileno pronominal), ‘cl’ (clítico), ‘prep’ (preposición), ‘adv’ (adverbio).

A partir de estas cuatro informaciones, Smorph realiza su análisis. La figura 1 muestra un ejemplo de *tweet* que luego será analizado:

RT @RadarInformador: quieres sapear a tu vecina cuando se saca la ropa? helicoptero con control a 15 lukas falabella

Figura 1: Ejemplo extraído del corpus

<sup>2</sup> www.infosurrevista.com.ar.

A partir del ejemplo de la figura 1, Smorph da como resultado un archivo con la información asignada a cada uno de sus constituyentes. La tabla 1 muestra este contenido de modo esquemático, destacándose en negrita el verbo:

'RT'.	['RT', mi].
'@RadarInformador'.	['@RadarInformador', mi].
'.'	['2p', 'EMS', 'dosp'].
'quieres'.	['querer', 'EMS', 'v', 'EMS', 'ind', 'PERS', '2a', 'NUM', 'sg', 'TPO', 'pres', 'TR', 'hi', 'TDIAL', 'est'].
'sapear'.	<b>['sapear', 'EMS', 'vch', 'EMS', 'infin', 'TR', 'r', 'TC', 'c1']</b> .
'a'.	['a', 'EMS', 'prep'].
'tu'.	['tu', 'EMS', 'det', 'TDET', 'pos'].
'vecina'.	['vecino', 'EMS', 'adj', 'GEN', 'fem', 'NUM', 'sg'], ['vecino', 'EMS', 'nom', 'GEN', 'fem', 'NUM', 'sg'].
'cuando'.	['cuando', 'EMS', 'rel'].
'se'.	['lo', 'EMS', 'cl', 'TPCRF', 'rflse'].
'saca'.	['sacar', 'EMS', 'v', 'EMS', 'ind', 'PERS', '3a', 'NUM', 'sg', 'TPO', 'pres', 'TR', 'r', 'TC', 'c1', 'TDIAL', 'estpi'].
'la'.	['el', 'EMS', 'det', 'TDET', 'art'], ['lo', 'EMS', 'cl', 'TPCL', 'nrfl'].
'ropa'.	['ropa', 'EMS', 'nom', 'GEN', 'fem', 'NUM', 'sg'].
'?'.	['nif', 'EMS', 'pun'].
'helicoptero'.	['helicoptero', mi].
'con'.	['con', 'EMS', 'prep'].
'control'.	['control', 'EMS', 'nom', 'GEN', 'NUM', 'sg'].
'a'.	['a', 'EMS', 'prep'].
'15'.	['num', 'EMS', 'numer'].
'lukas'.	['lukas', mi].
'falabella'.	['falabella', mi].

Tabla 1: Esquema de datos de salida de Smorph. Ejemplo extraído del corpus

### 3.2 Módulo Post Smorph, MPS

MPS tiene como *input* la salida de Smorph y, a partir de reglas de recomposición, descomposición y correspondencia, declaradas por el usuario, analiza la cadena de lemas resultante del análisis morfológico. Con este programa, se elaboraron reglas sintácticas para

la combinación de verbos chilenos con y sin pronombres.

Las fuentes declarativas de MPS están constituidas por un único tipo de archivo, rcm.txt, que incluye un listado de reglas que especifican cadenas posibles de lemas con una sintaxis informatizada. Las reglas pueden ser de tres tipos: (1) recomposición:  $D + N = SN$ ; (2) descomposición:  $Contracc = P + D$ ; y (3) correspondencia:  $Art = D$ .

En el presente trabajo se recurrió a reglas de reagrupamiento, de las etiquetas 'cl' (clítico: 'me', 'se', etc.), los verbos clasificados como vchr y ciertas preposiciones. Algunas de las combinaciones fueron las siguientes:

Reglas	Ejemplos
cl + cl + vchr = chilenuismo	'me la rebusco'
clref + vchr + a + SN = chilenuismo	'se come a la vecina'
cldat + vchr + a + SN = chilenuismo	'le hace al canto'
vchr (en forma infinitiva) + cl = chilenuismo	'enyegüecerse'

Tabla 2: Reglas para MPS y ejemplos de chilenuismos

A partir de este tipo de reglas, MPS logró detectar chilenuismos del modo en que muestra la figura 2.

```
'@fdoverdugo'.['@fdoverdugo', mi]. '':[ '2p', 'EMS', 'dosp']. '#Falabella'.['#Falabella', mi]. 'te jode'. ['te joder', 'EMS', 'chil']. 'con'. ['con', 'EMS', 'prepcon']. 'CtaCte'.
```

Figura 2: Información de salida de MPS.

Este método fue aplicado a un corpus compuesto por 5.194 *tweets*, producidos en Chile con el *hashtag* #Falabella, entre el 22 y el 28 de noviembre de 2013.

## 4 Resultados

El corpus contenía 443 chilenuismos verbales, de los cuales el método propuesto fue capaz de detectar 426. Asimismo, se detectaron de forma errónea 72 verbos. A partir de allí, se determinó una precisión de 85,54%, una cobertura de 96,16% y una medida f de 90,53%.

En la tabla 3, se muestran los chilenismos clasificados de acuerdo con la propuesta presentada, junto con las cantidades obtenidas.

Chilenismos	Total	Detectados	Omitidos	Errores
Puros	50	48	2	0
Puros pro.	11	11	0	0
De sentido	187	178	9	51
De sentido pro.	195	189	6	21

Tabla 3: Resultados generales

A continuación, en la tabla 4, se presentan algunos ejemplos de chilenismos verbales puros hallados en el corpus:

Type	Ejemplo
Funar	['Ayúdame' , 'EMS', 'v+cl']. 'a'. ['a', 'EMS', 'prep']. 'funar' . [ 'funar' , 'EMS', 'chil']. 'a'. ['a', 'EMS', 'prep']. '#Falabella'. ['#Falabella', mi].
Tincar	'Lo'. ['lo', 'EMS', 'art']. 'que'. ['que' , mi]. 'mas'. ['mas' , mi]. 'me'. ['me', 'EMS', 'cl']. 'tincó'. ['tincar', 'EMS', 'chil']. 'del'. ['del', mi]. 'cybermonday'. ['cybermonday', mi]. 'de'. ['de' , mi]. 'falabella'. ['falabella' , mi]. 'fue las'. ['ir las', 'EMS', 'chil']. 'space'. ['space' , mi]. 'bag'. ['bag', mi]. 'xdd' ['xdd', mi]
Huevear	'para'. ['parar', 'EMS', 'prep']. 'el'. ['el', mi]. '""'. ['""' , mi]. 'hueveo'. ['huevear', 'EMS', 'chil']. '""'. ['""' , mi].
Maraquear	'pacos'. ['pacos', mi]. 'maraqueando'. ['maraquear', 'EMS', 'chil'].
Agringar	'flaytes'. ['flaytes', mi]. 'agringados'. ['agringar', 'EMS', 'chil'].
Pitutear	'oooppss'. ['oooppss' , mi ]. 'Pituteando'. ['pitutear', 'EMS', 'chil' ].

Tabla 4: Ejemplos de reconocimiento automático

Un aspecto de interés consiste en que se detectaron dos neologismos a partir de las terminaciones distinguidas: 'loguear' y 'clickear'. Estos candidatos a chilenismos verbales pueden ser sometidos a análisis por parte de los lexicógrafos.

## 5 Conclusiones

A partir de los resultados, puede señalarse que el método propuesto resulta útil y adecuado

para la detección de lo que aquí se ha denominado chilenismos verbales puros, clíticos o no. Esto se debe a que se pudo implantar en máquina los lemas recogidos en el diccionario de chilenismos y se logró conjugarlos mediante la modelización de las estructuras morfológicas que presentan.

No obstante, se detectaron algunos problemas derivados de verbos cargados en el archivo de Smorph como chilenos pronominales, pero que presentan un uso no diferencial. Tal es el caso de 'hacerla', cuando el pronombre 'la' remitía a un Complemento Directo referenciado por este. Este fue uno de los errores de mayor importancia (superior al 50%), debido a la frecuencia de uso no diferencial de verbos como 'hacer' o 'poder'. A fin de poder subsanar este inconveniente, se considera la posibilidad de adicionar métodos estadísticos que permitan reanalizar los datos.

Un segundo problema consiste en la variación grafemática de los verbos. Se observó que verbos como 'huevear', aparecen escritos en el corpus de distintas maneras por los usuarios: 'webear', 'weviar', 'huear', etc. Al respecto cabe señalar que los datos del corpus provienen de un modo de comunicación en el que las prácticas discursivas tienden a ser informales, lo que permite que los hablantes utilicen diversas opciones para escribir una misma palabra. Asimismo, se observa este hecho con mayor frecuencia en los verbos en la forma voseante: 'comís', 'comih', 'comi"', etc. Esto se debe que el poco prestigio del uso de estas formas en la comunicación escrita, no ha permitido que en Chile se estandarice su escritura. A pesar de que se intentó normalizar este hecho a través de la modelización de verbos en diversas posibles formas de escritura, será necesario en trabajos futuros otorgar mayor importancia a este tipo de variaciones y declararlas en el archivo de modelos verbales de Smorph.

Tal como se demostró, es posible incorporar a Smorph un diccionario con información dialectal, en este caso, del habla chilena, específicamente a nivel de morfología verbal. Si bien, mediante este método, no es posible determinar la procedencia del autor (en este caso, el autor del *tweet*), la herramienta permite detectar palabras registradas como propias del español de Chile. Asimismo, al tratarse de una herramienta declarativa, existe la posibilidad de adaptarla para modelizar cualquier variedad

lingüística del español, cargando los diccionarios y los modelos adecuados.

En relación con el objetivo de este trabajo -la creación de un extractor automático de verbos diferenciales del español chileno- se puede señalar que, en esta primera etapa, se logró desarrollar un diccionario electrónico que contiene chilenismos verbales puros, pronominales y no pronominales, a cuyos lemas les fueron asignados modelos morfológicos que permiten detectar las posibles flexiones en un corpus.

En el caso de los chilenismos verbales de sentido, se establecieron modelos que, además de los rasgos morfológicos, contienen información de nivel sintáctico, esto es: (i) características de la flexión verbal y (ii) propiedades sintácticas. En (ii) se consignaron los tipos de palabras con las que debía relacionarse el verbo para convertirse en chilenismo. Como ya se mencionó, un verbo de estas características sería ‘comer’ que se combina con un clítico, la preposición ‘a’ y un SN, o bien ‘abanderizar’ (‘simpatizar con una causa’) que también va combinado con un clítico, una preposición (en este caso, ‘con’) y un SN: ‘se abanderiza con una causa perdida’.

Cabe señalar que se hace necesario realizar una exploración exhaustiva del carácter sintáctico de los chilenismos verbales de sentido a fin de obtener información de su comportamiento sintáctico y evaluar la posibilidad de implantarlos de modo informatizado bajo reglas de lenguaje natural.

Otro de los desafíos consiste en detectar locuciones y neologismos verbales que puedan incorporarse al conjunto de verbos chilenos. Para ello, se podría apelar, en algunos casos, a las terminaciones distinguidas de Smorph. No obstante, algunas terminaciones pueden generar ambigüedades. Para ello, otra opción sería apelar al contexto sintáctico que rodea a la expresión neológica.

Por último, se requiere afinar las reglas de detección a fin de evitar etiquetados erróneos como los ya señalados en la sección anterior. Para ello, una opción sería extenderse más allá de los pronombres y preposiciones que requiere el verbo.

Se espera, una vez establecidas las reglas de detección automática, por un lado, contar con una herramienta que ayude a las tareas lexicográficas y, por otro, corroborar las hipótesis lingüísticas acerca de la estructura morfológica y sintáctica del chilenismo verbal.

El trabajo a futuro se organizará en torno a los siguientes ejes: (1) ampliar el corpus y combinar las reglas establecidas con estrategias estadísticas; (2) mejorar la precisión en la detección de chilenismos verbales de sentido mediante una afinación de reglas; (3) elaborar reglas de detección de neologismos verbales; (4) Analizar y detectar automáticamente locuciones verbales chilenas.

### **Bibliografía**

- Abacci, F. 1999. Développement du Module Post-Smorph. Clermont-Fd.: Memoria del DEA de Linguistique et Informatique. Universidad Blaise-Pascal/GRIL.
- Academia Chilena de la Lengua. 2010. *Diccionario de uso del español de Chile*. Santiago, MN Editorial Ltda.
- Ait-Mokthar, S. 1998. SMORPH: Guide d'utilisation. Rapport technique. Clermont-Fd.: Universidad Blaise Pascal/GRIL.
- Ait-Mokthar, S. y Lázaro, M. 1995. Segmentación y análisis morfológico en español utilizando el sistema Smorph. *Procesamiento del lenguaje natural*, 17, 29-41.
- Chávez, S. 2010. Ideas lingüísticas en prólogos de diccionarios diferenciales del español de Chile. Etapa 1875-1928. *Boletín de filología*, XLV(2) 49-69.
- Matus, A. 2010. Un diccionario para la lexicografía clásica chilena. En Morales Pettorino, F. 2010. *Nuevo Diccionario Ejemplificado de Chilenismos. Edición refundida y actualizada*. Suplemento. Valparaíso, Edit. Puntángelos (pp. VII-XIII).
- Morales Pettorino, F. 1984. *Diccionario ejemplificado de chilenismos*. Valparaíso, Academia Superior de Ciencias Pedagógicas de Valparaíso.
- Morales Pettorino, F., Quiroz, O. y Mayorga, D. 1969. Los verbos en -ear en el español de Chile. Santiago, Editorial del pacífico.
- Sáez, L. 2011. El léxico del dialecto chileno: Diccionario de uso del español de Chile DUECh. *Estudios filológicos*, 49, 137-15.