

# Verificación ortográfica de formas verbais e secuencias de pronomes enclíticos en lingua galega

Miguel Anxo Solla Portela  
Universidade de Vigo  
miguel.solla@uvigo.es

## Resumo

Descrición das melloras no comportamento do verificador ortográfico MySpell/Hunspell ante formas verbais en lingua galega, con arquivos de dicionario e de afixos que se elaboraron a partir da versión para o galego dispoñible baixo os termos da licenza GNU GPL, versión 2, aos que se lles modificou a estrutura para que, en cada persoa gramatical da flexión, admita un paradigma diferente de posibles secuencias des pronomes persoais enclíticos á forma verbal consonte a información que se extraeu do *Vocabulario ortográfico da lingua galega* sobre o réxime de construción sintáctica das formas verbais.

## 1. Introducción

Despois de realizar probas cos ficheiros para a lingua galega dispoñibles baixo os termos da licenza [GNU GPL, versión 2](#), para o verificador ortográfico MySpell/Hunspell (que poden utilizar, entre outros, o paquete de ofimática [OpenOffice](#), o verificador para o navegador [Firefox](#), o xestor de correo electrónico [Thunderbird](#), editores de textos como [gedit](#) ou [AbiWord](#), programas de localización de software como o [Poedit](#), [Lokalize](#), [WordForge](#)...); calquera se pode decatarse de que non verifica adecuadamente certas formas verbais, sobre todo formas rizotónicas de verbos con pouco corpo fónico, e de que non reconece formas lingüísticas moi expresivas con máis de dous pronomes enclíticos tras a forma verbal. Foron precisamente estas circunstancias as que motivaron o interese por esculcar o arquivo de sufixos dispoñible no [sitio web do Centro de Referencia e Servizos de Software Libre Mancomún](#) co fin de examinar a dificultade de engadir máis posibilidades de concorrencias de formas verbais con pronomes enclíticos e ver de mellorar a flexión de certas formas. Despois dalgunhas probas, decatámonos enseguida de que non era difícil ampliar as combinacións sintagmáticas de pronomes enclíticos; no entanto, o deseño do arquivo de afixos ía xerar e admitir moitas máis formas agramaticais das que xa coñecía (formas flexivas de primeira e segunda persoa, singular e plural, con secuencias de enclíticos que comezasen por un pronome reflexivo de terceira persoa, pronomes enclíticos acusativos con verbos que non admiten unha construción transitiva...) de se ampliar o paradigma de pronomes enclíticos sen restricións. Respecto da revisión da flexión verbal, obtivéronse bos resultados mais, en certos casos, resultou insuficiente a revisión do arquivo de afixos e fíxose necesario engadir lle alomorfos con acentos gráficos na raíz ao arquivo de dicionario en certos casos. No momento en que se modificou o arquivo de dicionario, xermolou a idea

de impoñer desde este arquivo limitacións no tipo de construción sintáctica de cada lema para restrinxir ou ampliar as secuencias de pronomes enclíticos que se engadisen ás formas flexionadas.

## 2. Elaboración

### 2.1 A información sintáctica

A información sobre o tipo de construción, pese a non ser sistemática<sup>1</sup>, parece abonda para este propósito e está dispoñible en formato electrónico no portal da Real Academia Galega no apartado dedicado ao [Vocabulario Ortográfico da Lingua Galega \(VOLG\)](#). Con este propósito, con data do 14/12/2008 realizouse unha extracción do VOLG mediante buscas por clases de palabra (verbos) que contivesen información sobre o tipo de construción sintáctica (6.738 lemas). Filtráronse as formas non toleradas<sup>2</sup>, que se reutilizaron para compoñer regras de substitución que o verificador emprega para ofrecer suxestións nas secuencias que non considera correctas.

### 2.2 Paradigmas de pronomes persoais enclíticos

A elaboración dos paradigmas de pronomes partiu do paradigma máis extenso (o dunha forma verbal en terceira persoa, singular ou plural, en construción transitiva) con 22 regras para a creación de secuencias de pronomes enclíticos monosílabos e 353 regras para secuencias de pronomes enclíticos polisílabos<sup>3</sup> que chegan a aglutinar ata tres dativos e un acusativo (*trouxéronchemellela*).

1 Vid. Álvarez e Xove, 2002, p. 239.

2 Vid. [Estrutura das entradas](#), no sitio web do VOLG.

3 Os paradigmas de pronomes persoais enclíticos polisílabos tratan de representar, respecto da orde, o que se dispón sobre as secuencias de clíticos en Álvarez e Xove, 2002, p. 570-571.

A partir deste paradigma obtivéronse, por unha banda, o paradigma da terceira persoa da construción intransitiva tras eliminar todas as secuencias que contiñan pronomes enclíticos marcados morfoloxicamente como acusativos e as que contiñan o pronome reflexivo de terceira persoa; e, por outra banda, os paradigmas das demais persoas da construción transitiva tras eliminar de cada paradigma as secuencias que contiñan o reflexivo de terceira persoa ou o pronome reflexivo respectivo da persoa da flexión. A partir do paradigma para unha forma flexionada en terceira persoa en construción intransitiva obtivéronse os paradigmas para as demais persoas na construción intransitiva mediante a eliminación das secuencias que contiñan o pronome reflexivo de terceira persoa ou o pronome reflexivo respectivo da persoa da flexión; e, ademais, os paradigmas da construción pronominal tras eliminar as formas que non contiñan o reflexivo correspondente a cada persoa acompañado ou non de formas de dativo doutras persoas.

### 2.3 Revisión morfolóxica e asociación cos paradigmas dos pronomes enclíticos

Os paradigmas da flexión parten da revisión morfolóxica dos paradigmas preexistentes para cada conxugación, que se triplicaron (un para cadansúa construción sintáctica) e nos que se relacionou cada regra de creación dun sufixo co seu correspondente paradigma consonte as súas posibilidades sintagmáticas (combinacións con enclíticos monosílabos e polisílabos, acusativos de P3 tras -r, -s ou ditongo decrecente...) e consonte a persoa gramatical da flexión.

### 2.4 Paradigmas de formas verbais impersoais e defectivas

Os paradigmas das formas impersoais reducíronse á flexión en terceira persoa. Inclúese a terceira persoa de plural para recoller usos metafóricos con suxeito en plural do tipo *choven chuzos, tronaban os canóns*.

Os paradigmas dos verbos defectivos compuxéronse, cando non coincidían cos dos verbos impersoais, de acordo coas limitacións flexivas que precisaban.

Para o verbo *decir* (forma tolerada no VOLG, que remite a *dicir*) creouse un paradigma que inclúe as formas de infinitivo, xerundio, e P4 e P5 do presente de indicativo.

### 2.5 Elaboración do dicionario

Para os lemas do dicionario empregouse a selección dos termos que resultaron de eliminar os termos non tolerados do extracto do VOLG. Para os alomorfos destes lemas, aproveitáronse, cando existían, os alomorfos da versión anterior. Os demais introducíronse manualmente. A relación cos afixos

responde tamén, en parte, coa da versión anterior que se editou, manualmente nalgunhas ocasións e coa axuda dunha folla de cálculo noutras, para adaptalo ás mudanzas que se introduciron durante a revisión morfolóxica, aos lemas novos, ás variacións de paradigmas, á asociación coa súa construción, á anotación morfolóxica...

Cómpre subliñar que os verbos que se consideraron susceptibles de aparecer como auxiliares en perífrases verbais (*acabar, acostumar, andar, botar, cesar, chegar, comezar, continuar, dar, deixar, empezar, estar, levar, parar, pasar, principiar, rematar, terminar, tornar, coller, deber, haber, poder, poñer, pór, ser, ter, volver, ir, seguire vir*), manipuláronse para que, independentemente da información sintáctica que reciban, admitan a posibilidade de engadir enclíticos de calquera dos tres tipos de construcións.

Os lemas do arquivo de dicionario que non se corresponden con formas verbais coinciden coas formas preexistentes.

## 3. Comportamento

Fronte á versión anterior, prevese un paradigma diferente para que cada persoa gramatical da flexión poida responder a posibilidades combinatorias específicas (deste xeito, evítase, por exemplo, que o reflexivo de terceira persoa poida acompañar formas flexionadas doutras persoas gramaticais: *\*cómose, \*coméchesse, \*andádesse, \*collémosse...*). As posibles secuencias de pronomes enclíticos varían, ademais, segundo o réxime de construción.

### 3.1 Paradigma dunha construción transitiva

A forma flexiva non acepta, na secuencia de pronomes enclíticos, o reflexivo da mesma persoa gramatical na primeira posición da secuencia, salvo na terceira persoa, de singular e de plural, para a formación das construcións impersoal activa e impersoal pasiva.

O verificador admite: *retíñanllelas, mantíñana, mantíñasnola, tráiolle, produciuse un erro, advirteselles ás persoas interesadas...* e rexeita *\*saltácheste a clase, \*bebinme o leite, \*mercástesvos cadanseu vestido...* A detección deste tipo de castelanismos sintácticos non queda resolta na terceira persoa, de singular e de plural, pois a aparición do pronome reflexivo forma construcións impersoais.

## Arquivo de dicionario

## Arquivo de sufixos

Raíz

**Versión previa**

+ sufixo flexivo

[+ pronome/s persoal/is enclítico/s]

**Modificación**

Raíz con información sobre o tipo de construción sintáctica que admite o verbo

+ sufixo flexivo segundo o modelo de conxugación: flexión completa, impersoal ou defectiva

[+ pronome/s persoal/is enclítico/s: paradigmas diferentes segundo a persoa gramatical da forma flexiva e segundo o tipo de construción sintáctica do verbo]

Existen tamén outras limitacións para este comportamento: se a forma verbal ou un termo homógrafo (ou unha forma verbal susceptible de auxiliar a outra nunha perífrase) admiten tamén o réxime de construción pronominal, o verificador non pode diferenciar o uso concreto e vai permitir tamén as secuencias de pronomes enclíticos prevista para esta construción.

un uso en primeira ou en terceira persoa de singular do copretérito do verbo cantar, de tal xeito que *cantábame unha canción*, vaise validar independentemente de que responda a un uso indebido en primeira persoa, porque o verificador identifica cun suxeito en terceira persoa de singular, debido a que a secuencia de enclíticos comeza polo reflexivo de primeira persoa, que non forma parte do paradigma de secuencias de pronomes enclíticos previsto para a forma en primeira persoa.

### 3.2 Paradigma dunha construción intransitiva

A forma flexiva non acepta, na secuencia de pronomes enclíticos, o reflexivo da mesma persoa gramatical da forma flexiva na primeira posición da secuencia, salvo na terceira persoa, de singular e de plural, para a formación da construción impersoal intransitiva (agás os verbos impersoais, que tampouco admiten o reflexivo na terceira persoa, pois non precisan marcar sintacticamente a indeterminación do suxeito); tampouco admite na secuencia un pronome enclítico acusativo marcado morfoloxicamente (P2 / P3 / P6).

Deste xeito, o verificador admite usos como *acontecéuchemelles*, *abóndalles*, *concorriase con frecuencia*, *aquí vívese ben...* e non admite *\*aconteceuna*, *\*morreuno*, *\*abondádelas*, *\*finóuchemellela*, *\*névase*. Con todo, existen limitacións, coma na construción transitiva, para este comportamento: se a forma verbal ou un termo homógrafo (ou unha forma verbal susceptible de auxiliar a outra nunha perífrase) admiten tamén o réxime de construción pronominal, o verificador non pode diferenciar o uso concreto e vai permitir tamén as secuencias de pronomes enclíticos prevista para esta construción.

Ademais, tanto en construcións transitivas coma en construcións intransitivas, as formas flexivas con sincretismo da persoa gramatical inclúen as posibilidades sintagmáticas de todas as persoas que representan: *cantaba unha canción* pode responder a

### 3.3 Paradigma dunha construción pronominal

A forma flexiva só acepta secuencias de pronomes enclíticos nas que estea presente o reflexivo da mesma persoa gramatical (ou formas marcadas morfoloxicamente como dativo + reflexivo) e non admite na secuencia un pronome enclítico acusativo marcado morfoloxicamente (P2 / P3 / P6)

O verificador admite usos como *desentendéronse*, *resentiuselle*, *entrecruzámoschenoslle*, *ativéstešllesvos...*, pero non admite *\*desentendéronlle*, *\*resentiulle*, *\*entrecruzámoschelle*, *\*ativéstešlles*, *\*arrepuxémonola...* As limitacións deste comportamento son similares ás que se expuxeron para os paradigmas de construcións transitivas e intransitivas.

Se se toma como exemplo unha forma verbal que admita os tres réximes de construción, obsérvase que, mesmo con todos os posibles paradigmas, o verificador nunca vai admitir unha construción pronominal cunha secuencia de enclíticos que conteña un pronome marcado morfoloxicamente como acusativo: *\*batémonolos*, *\*botámonola*. Admite usos como *había colonia e botámola*, *había colonia e botamos*, *había colonia e botámonos*, pero non admite *había colonia e \*botámonola*<sup>4</sup>.

<sup>4</sup> Álvarez e Xove, 2002, p. 556.

### 3.4 As suxestións de substitución

Cando se acadou o comportamento que se vén de describir, observouse que as regras para as suxestións de termos cando verificador atopa unha forma que non reconece melloraran sensiblemente, xa que se eliminaran a formas agramaticais na flexión.

As formas verbais que o VOLG inclúe como formas non toleradas introducíronse como regras de substitución con resultados moi satisfactorios:

```
$ hunspell -m -d gl_ES -i utf-8 -a
@(#) International Ispell Version 3.2.06
(but really Hunspell 1.2.6)1.2.6

olvidouna

& olvidouna 1 0: esqueceuna
```

## 4. Particularidades de Hunspell

A documentación, en lingua inglesa, e as instrucións de descarga e instalación da programa de verificación ortográfica están dispoñibles [no seu sitio web](#).

### 4.1 A recursividade

O verificador MySpell/Hunspell permite un envío de relacións desde o arquivo de dicionario (lema e alomorfos) cara ao arquivo de afixos (que xera a súa flexión), e unicamente outro envío desde cada unha destas formas flexivas cara a ese mesmo arquivo de afixos (co que se xera a secuencia de enclíticos).

Esta limitación na recursividade impón que os paradigmas de pronomes persoais enclíticos sexan tan analíticos, debido a que non resulta posible segmentar as secuencias de pronomes enclíticos en unidades palabra, que sería unha descrición máis precisa do comportamento lingüístico.

### 4.2 A segunda forma do artigo e a guionización

Non se obtiveron resultados satisfactorios nas probas que se fixeron para verificar o alomorfo do artigo determinado (-la, -las, -lo, -los): a documentación de Hunspell xa advirte de que se pode empregar a instrución WORDCHARS - para que o verificador, unicamente nun terminal, non divida as palabras con esta grafía (unha solución deste tipo faise precisa en lingua galega, debido que a sílaba do artigo afecta para a acentuación gráfica na combinación de formas verbais que rematan en -r ou -s co alomorfo do artigo). Cos sufixos que se probaron obtivéronse bos resultados co analizador morfolóxico de Hunspell executándose nun terminal, no entanto, produciron efectos non desexados en todos os programas de edición de texto cos que se experimentou, xa que todos dividen en palabras diferentes a secuencia que

haxa antes do guión respecto da que figure a continuación do guión pese a que existan regras de sufixación que o inclúan.

Co exemplo *cóme-lo caldo*, o verificador non vai atopar a secuencia *cóme* na flexión do lema *comer*, porque a súa acentuación responde á ligazón co alomorfo do artigo, que precisa da grafía con guión, mais os editores de texto van interpretar dúas unidades palabra diferenciadas que se segmentan co trazo.

Como froito destas probas e a raíz da documentación que figura [wiki de Mancomún](#) comezouse tamén a elaboración dun arquivo de guionización acorde coa silabación en lingua galega, que se inclúe co ficheiro de afixos e co dicionario, e que interpreta xa algunhas características propias do galego, pero que aínda precisa de moitas melloras.

Cómpre ter en conta que este arquivo impón regras de segmentación silábica para o uso do guión ao final de liña e que non o empregan todas as aplicacións que utilizan o verificador, senón que os programas que o xestionan adoitan ter un xeito particular de facelo. As probas realizáronse co motor, propio, do OpenOffice 3.

## 4. Conclusións

Os arquivos que se obtiveron tras este traballo están dispoñibles, coa mesma licenza que os arquivos orixinais dos que se partiu, no enlace [http://webs.uvigo.es/miguelsolla/gl\\_ES.zip](http://webs.uvigo.es/miguelsolla/gl_ES.zip).

O comportamento que se describiu parece mellorar sensiblemente a eficacia do verificador e abre as portas para establecer novas regras de substitución que aumenten a súa utilidade.

Os comentarios en cada regra de sufixación, que se empregaron inicialmente para identificar o código con maior precisión, reconvertéronse durante o proceso de revisión en anotación que se adaptou, tamén no arquivo de dicionario, para o analizador morfolóxico de Hunspell.

```
$ hunspell -m -d gl_ES -i utf-8

trouxémoschas

trouxémoschas
st:traer
is:alomorfo traer transitiva
ds:pretérito P4 + enclítico
po:pronome persoal enclítico
is:monosílabo P4 transitiva dativo P2 +
acusativo P3
```

Deste xeito facilítase moito a depuración de comportamentos inesperados e o verificador fornécese dunha ferramenta que se pode estender no futuro para outras clases de palabras. É preciso ter en conta que para poder empregar o analizador é preciso que estean instalados Hunspell e myspell-gl-es no

sistema e que, na actualidade, myspell-gl-es instala uns arquivos herdeiros doutros verificadores, en normativa de mínimos, que se deben substituír manualmente, para empregalos por omisión, ou indicarlle ao analizador morfolóxico de Hunspell cada vez que se use a localización dos ficheiros que se queiran utilizar.

E alén do obxecto destas liñas, cómpre salientar que aínda quedan por facer diferentes *thesauri* que completen as posibilidades do verificador e amplíen a súa eficacia; pero tamén é certo que xa dispoñemos de ferramentas lingüísticas en código aberto que superan moito a simple verificación da ortografía, coma o [Golfiño](#) ou o [Exeria](#), ambos os dous coa análise lingüística do [FreeLing](#).

## 5. Bibliografía

Real Academia Galega e Instituto da Lingua Galega. 2003. *Normas ortográficas e morfolóxicas do idioma galego*, 18ª edición.

Santamarina, Antón e Manuel González González (coord.). 2004. *Vocabulario ortográfico da lingua galega*, Real Academia Galega / Instituto da Lingua Galega.

Álvarez, Rosario e Xosé Xove. 2002. *Gramática da lingua galega*. Editorial Galaxia, Vigo.

Freixeiro Mato, Xosé Ramón. 2000. *Gramática da lingua galega II. Morfosintaxe*. Edicións A Nosa Terra, Vigo, 1ª edición.

Álvarez, Rosario, X. L. Regueira e H. Monteagudo. 1986. *Gramática galega*, Editorial Galaxia, Vigo.

Hermida Gulías, Carme. 2004. *Gramática práctica (morfosintaxe)*. Sotelo Blanco Edicións, Santiago de Compostela.

Hermida, Avelino. 2006. *Conxugación verbal da lingua galega século 21*. Edicións do Cumio / Editorial Galaxia.

González González, Manuel, Carmen García Mateo, Eduardo Rodríguez Banga e Elisa Fernández Rei. 2002. *Diccionario de verbos galegos Laverca*, Edicións Xerais de Galicia, Vigo.

Díaz Regueiro, Manuel. 1992. *Os verbos galegos*. Consellería de Educación e Ordenación Universitaria / Dirección Xeral de Política Lingüística.

Fernández Rei, Francisco. 1991. *Dialectoloxía da lingua galega*. Edicións Xerais de Galicia, Vigo, 2ª edición.

Graña Núñez, Xosé. 1993. *Vacilacións interferencias e outros “pecados” da lingua galega*. Ir Indo Edicións, Vigo.

González Rei, Begoña. 2004. *Ortografía da lingua galega*- Galinova Editorial, A Coruña.

Hermida Gulías, Carme. 2001. *Ortografía práctica*. Sotelo Blanco Edicións, Santiago de Compostela.