

# Alternancias morfografémicas y paradigmas irregulares en la morfología flexiva verbal del gallego

Javier Gómez Guinovart  
Universidade de Vigo  
jgomez@uvigo.es

José Luis Aguirre Moreno  
Universidade de Vigo  
jlagui@uvigo.es

## Resumen

En esta comunicación presentamos los aspectos lingüísticos y computacionales de una implementación de los paradigmas regulares con alternancias morfografémicas y de los paradigmas irregulares de la morfología flexiva del verbo gallego en el formalismo DATR de representación del conocimiento léxico. Mediante este lenguaje formal, proponemos una representación tipificada de los elementos léxicos verbales del gallego, utilizando estructuras de rasgos morfológicos y redes léxicas con herencia.

## 1. Introducción

Este trabajo presenta una implementación de la morfología verbal regular e irregular del gallego, dentro del marco teórico lingüístico desarrollado en [Álvarez et al., 1986]. En [Gómez Guinovart y Aguirre, 1998] ofrecemos una visión global del diseño del componente léxico elaborado para los paradigmas verbales completamente regulares, es decir, sin alternancias morfografémicas. El objetivo de esta comunicación consiste precisamente en mostrar la generalización de este diseño a los paradigmas regulares con alternancias morfografémicas y a los paradigmas verbales irregulares.

La implementación aquí propuesta se basa en la tipificación de las formas verbales del gallego, utilizando estructuras de rasgos morfológicos y redes léxicas con herencia especificadas en DATR ([Evans y Gazdar, 1996]), un lenguaje formal determinista, simple y declarativo diseñado para la representación de la información lingüística asociada con los elementos léxicos y que permite definir redes con herencia múltiple no monotónica mediante ecuaciones atributo-valor.

En tanto que lenguaje formal de programación lingüística, DATR no presupone ninguna teoría concreta del lenguaje, ni está limitado a la representación de un único nivel de análisis, siendo así un formalismo conveniente tanto para implementaciones de amplia cobertura lingüística, como para la comprobación automatizada de la consistencia de una teoría gramatical o de sus predicciones. En este sentido, conviene destacar que existen descripciones en DATR para una amplia gama de fenómenos lingüísticos, desde la morfología flexiva [Jenkins, 1990; Gazdar, 1992; Gómez Guinovart y Aguirre, 1998], hasta la semántica léxica [Kilgarriff, 1993], la morfofonología [Cahill, 1993a], la morfoprosodia [Reinhard y Gibbon, 1991] y la sintaxis [Evans et al. 1995, Carroll et al., 1998; Gómez Guinovart, 1998].

También cabe resaltar que el motor de inferencia de DATR se puede implementar en Prolog —como en la implementación original de R. Evans (Univ. de Sussex) o en QDATR de J. Kilbury (Univ. de Duesseldorf)— y en C —como en ZDATR de C. Schillo y D. Gibbon (Univ. de Bielefeld)— de manera sencilla y eficiente, lo que facilita la incorporación de los léxicos desarrollados en DATR en sistemas de procesamiento del lenguaje natural más amplios [Andry et al., 1992; Cahill, 1993b; Carroll et al., 1998].

## 2. Marco general

De acuerdo con [Álvarez et al.], los morfemas constituyentes de las formas verbales son raíz (R), vocal temática (VT), sufijo modo-temporal (SMT) y sufijo de número-persona (SNP). Se distinguen tres clases flexivas regulares (C1-C3), seis personas (P1-P6) y catorce tiempos: indicativo presente (IPr), indicativo pretérito (IP), indicativo copretérito (ICP), indicativo antepretérito (IAP), indicativo pospretérito (IPP), indicativo futuro (IF), subjuntivo presente (SPr),

subjuntivo pretérito (SP), subjuntivo futuro (SF), imperativo (Imp), infinitivo (Inf), infinitivo conjugado (IC), gerundio (Xer) y participio (Part).

En líneas generales, la formalización en DATR desarrollada se compone de cinco módulos. En el primer módulo, común a todos los verbos, se declaran los diversos morfos para los morfemas VT, SMT y SNP. Por ejemplo, en la figura 1, puede observarse la declaración del morfo de SMT para IAP, y de los alomorfos (condicionados por la clase flexiva C) de SMT para ICP.

```
<mor smt iap> == ra
<mor smt icpc1> == ba
<mor smt icpc2c3> == a
```

Figura 1

En el segundo módulo se establece (mediante ecuaciones *camino == valor*) la configuración morfológica de las 65 formas que forman el paradigma verbal completamente regular en cada una de las tres clases flexivas. Los *caminos* o cadenas de atributos representan cada una de las formas verbales, a las que se les asigna su correspondiente valor o composición morfológica. Por ejemplo, la declaración DATR recogida en la figura 2 describe la forma de la primera persona del presente de indicativo. Su valor es la concatenación de los 4 morfemas que constituyen esa forma: (1) La raíz del verbo que corresponda<sup>1</sup>. (2) La vocal temática, representada por el morfo cero, alomorfo de este morfema en las tres clases flexivas. (3) El sufijo modo-temporal de IPr, también realizado como morfo cero. (4) El sufijo de número-persona para P1, representado por su segundo alomorfo, descrito como alo2.

<sup>1</sup> El valor de un atributo global, representado en DATR entre comillas (como "<mor r>" en la figura 2) es el valor que tiene ese atributo en el nodo original desde el que se realiza la consulta. Esta manera de compartir atributos entre diversos nodos se denomina *herencia global*.

```
<mor ipr p1> == "<mor r>"
<mor vt clc2c3alo0> <mor smt ipr>
<mor snp p1 alo2>
```

Figura 2

La descripción de las formas verbales se distribuye a través de una red de nodos. Se pretende agrupar los rasgos comunes, de modo que cada lexema contenga todas sus formas verbales correctamente descritas y de la manera menos redundante posible. Cada clase de lexema (o tipo) que obtenemos mediante esta distribución es un nodo. La distribución y jerarquía de los distintos nodos quedan reflejadas en la figura 3.

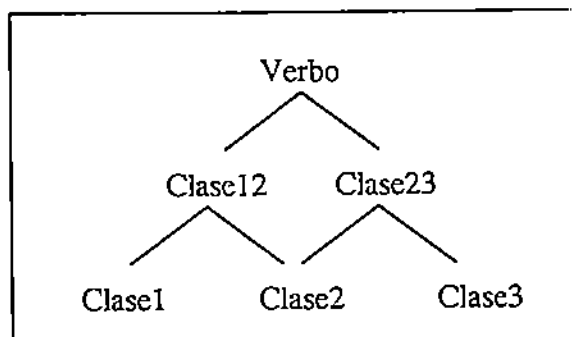


Figura 3

En el nodo *Verbo* se incluyen los rasgos que tienen en común todas las formas verbales regulares. Por ejemplo, todos los verbos regulares poseen la misma configuración morfológica en P1 del IPr (R + VT morfo Ø + SMT morfo Ø + SNP morfo o), por lo que su descripción se recoge en el nodo superior de la jerarquía, siendo heredada así por todos los nodos subordinados. P2, P3 y P6 del IPr, en cambio, poseen una configuración en C1 y C2, y otra en C3, por lo que su descripción se recoge, respectivamente, en los nodos *Clase12* y *Clase3*. En *Clase3* se describen los rasgos comunes a los verbos de C3, mientras que en *Clase12* se agrupan las características compartidas por C1 y C2, siendo sus rasgos heredados por los nodos *Clase1* y *Clase2*.

### 3. Paradigmas regulares con alternancias morfográficas

En el tercer módulo se describen los verbos que, siguiendo el paradigma regular en la mayoría de sus formas, presentan alternancias morfográficas en la raíz o en la vocal temática. Se dividen en tres grupos, a partir de la clasificación que aparece en [Álvarez et al.], complementada con las indicaciones de [Real Academia Galega e Instituto da Lingua Galega, 1995]: (1) Variaciones en la vocal radical. (2) Variaciones gráficas en la raíz condicionadas por la distribución de grafemas, y variaciones gráficas en la raíz y en la vocal temática de los verbos en *-uir*. (3) Paradigmas irregulares en el tema de presente.

Dentro de cada grupo, se han diferenciado diversos tipos léxicos, en función de sus características morfológicas, agrupados en tres categorías (figura 4):

(3.1) variaciones en la vocal radical ([Álvarez et al.: 305-310]),

(3.2) variaciones gráficas en la raíz ([Real Academia Galega e Instituto da Lingua Galega 1995: §16.2]), y

(3.3) paradigmas irregulares en el tema de presente ([Álvarez et al.: 338-339]).

#### 3.1. Variaciones en la vocal radical

- Tipo léxico C1vocra-*iu*. Algunos verbos acabados en *-iar*, *-uar* de C1 [307] (como *afiar* "afilar" y *recuar* "recular") presentan acento gráfico en la vocal radical en sus formas rítonas (*afío-afiamos*, *recúo-recuamos*).
- Tipo léxico C3vocra-*e1*. Verbos de C3 (como *advertir* y *vestir*) con vocal radical *-i-* en las formas rítonas y *-e-* en las arítonas [308] (*advirto-advertimos*, *visto-vestimos*).
- Tipo léxico C3vocra-*e2*. Verbos de C3 (como *servir* y *mentir*) con vocal radical *-e-* en las formas arítonas, e *-i-* en las formas rítonas, con la excepción de P2, P3 y P6 del IPr donde vuelve a aparecer *-e-* [308-309] (*sirvo-serves-servimos*, *mintomentes-mentimos*).

- Tipo léxico C3vocra-*u*. Verbos de C3 (como *fuixir* "huir" o *cuspir* "escupir") con vocal radical *-o-* en P2, P3 y P6 de IPr y *-u-* en todas las otras formas (*fuixu-fuixes*, *cuspo-cospes*).

#### 3.2. Variaciones gráficas en la raíz

Esta categoría incluye variaciones gráficas en la raíz condicionadas por la distribución de grafemas, y variaciones gráficas en la raíz y en la vocal temática de los verbos en *-uir*.

- Tipo léxico C1raizen-*gu*. Verbos de C1 con raíz acabada en *-gu* (como *minguar* "menguar"). Presentan diéresis ante los sufijos flexivos encabezados por *-e* (*minguo-mingüe*).
- Tipo léxico C2raizen-*qgc*. Verbos de C2 con raíz terminada en una de estas consonantes (como *verquer* "verter", *erguer* "erguir" o *quecer* "calentar"). Presentan las alteraciones gráficas *c/qu*, *g/gu* y *z/c* regulares en la lengua (*verco-verques*, *ergo-ergues*, *quezo-queces*).
- Tipo léxico C3raizen-*qgc*. Verbos de C3 de raíz acabada en estas consonantes (como *delinquir*, *distinguir* o *traducir*), con alternancias gráficas semejantes a las del tipo léxico anterior (*delinco-delinques*, *distingo-distingues*, *traduzo-traduces*).
- Tipo léxico C1raizen-*czg*. Verbos de C1 de raíz acabada en estas consonantes (como *sacar*, *cazar* o *pegar*), con las alternancias gráficas regulares semejantes a las de los dos tipos anteriores (*saco-saque*, *cazo-cace*, *pego-pegue*).
- Tipo léxico C3en-*uir*. Verbos de C3 acabados en *-uir* (como *incluir*), con alternancia debida a la tilde o diéresis empleada para indicar el hiato en la raíz o en la vocal temática (*inclúo-incluimos*, *incluíamos-incluamos*).

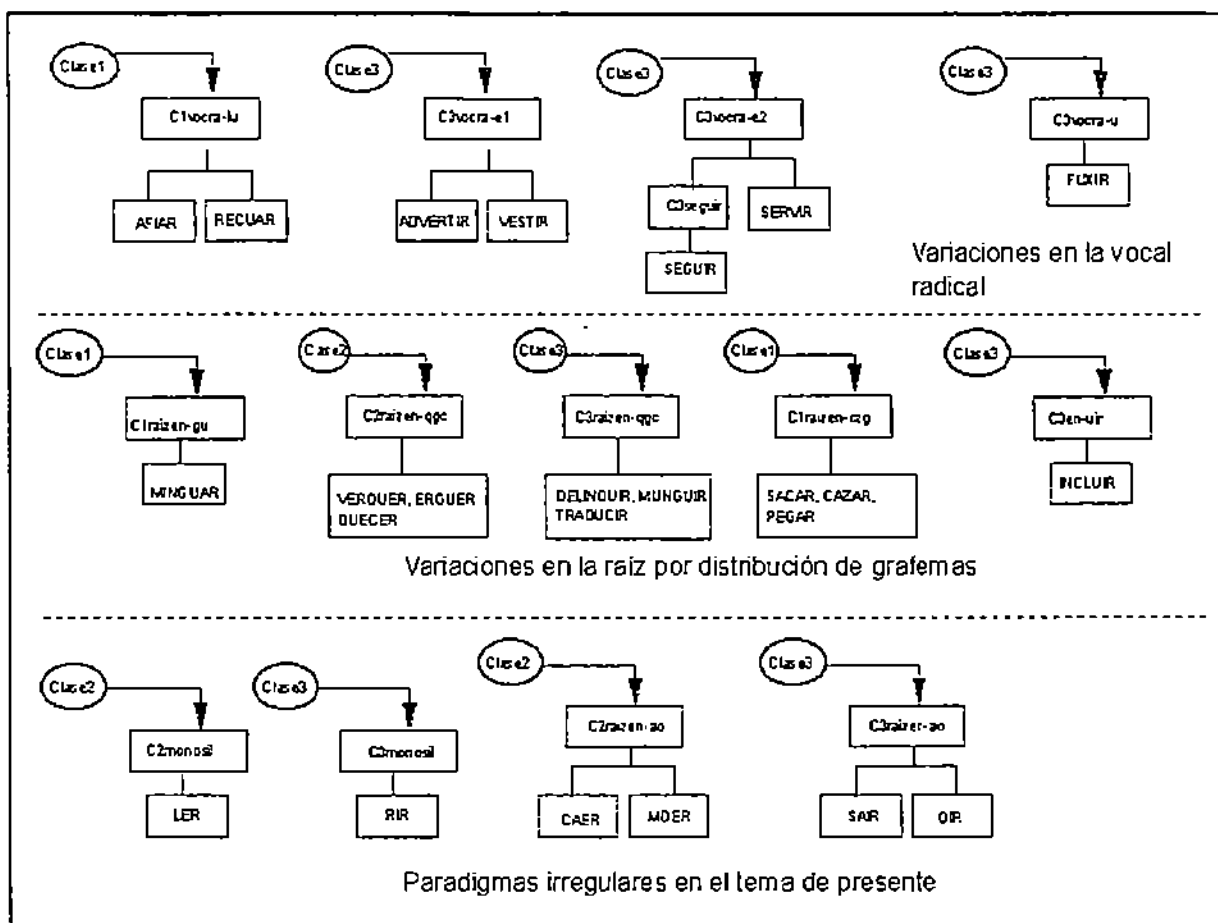


Figura 4

### 3.3. Paradigmas irregulares en el tema de presente

- Tipo léxico C{2|3}monosil. Verbos de C2 y C3 (como *ler* "leer" y *rir* "reír") en los que "cómpre ter en conta algúns encontros vocálicos que na fala e na escritura (...) se resolven en contracción" ([RAG-ILG: §16.2]). Como esta contracción se produce en casi todas las formas, en esta implementación se describen las excepciones en el nodo del tipo léxico, y el resto se conjuga de manera regular, considerando *efi* como VT.
- Tipo léxico C{2|3}raizen-ao. Verbos de C2 y C3 con raíz acabada en *-al-o* (como *caer* y *moer* "moler") que "intercalan un *-i-* sen valor morfolóxico entre a raíz e a desinencia, cando esta empeza por *o* ou por *a*" ([RAG-ILG: §16.2]). En la implementación presentada, esta *-i-* de *caio*

("caigo") se considera un alomorfo de VT, complementario de *-í-* (*caíches* "caíste") y de *-i-* (*caíamos* "caíamos").

La figura 4 ilustra la sección de la red de nodos relativa a este tercer módulo, con los tipos léxicos correspondientes a cada tipo de alternancia y las clases léxicas regulares de las que heredan. Por ejemplo, los nodos correspondientes a los lexemas *AFIAR* y *RECUAR* heredan del nodo *Clase1* la descripción morfológica de todas sus formas flexivas, excepto de aquellas que están expresamente descritas en el nodo *C{vocra}-iu*, que tienen prioridad sobre las anteriores en virtud del mecanismo de herencia no monotónica de DATR.

Para la generación de las formas verbales con alternancias morfografémicas en la raíz, se emplea el recurso de especificar en la entrada léxica los distintos alomorfos del radical del verbo (*R*, *R2*, *R3*..., *Rn*). Por ejemplo, en la entrada léxica de *advertir*, se especifican los alomorfos de la raíz *advert-* y *advirt-*,

```

C3vocra-e1:  % advertir, vestir ... %
<> = Clase3
<mor ipr p1> = "<mor r2>" <mor vt clc2c3alo0> <mor smt ipr> <mor snp p1 alo2>
<mor ipr $p236> = "<mor r2>" <mor vt clc2c3alo1> <mor smt ipr> <mor snp $p236>
<mor spr $pla6> = "<mor r2>" <mor vt clc2c3alo0> <mor smt sprc2c3> <mor snp $pla6>
<mor imp p2> = "<mor r2>" <mor vt clc2c3alo1> <mor smt imp> <mor snp p2 alo3>

ADVERTIR:
<> = C3vocra-e1
<mor r> = advert
<mor r2> = advirt.

```

Figura 5

representados en sus formas flexivas *advirto* "advierdo" y *advertimos*. En la figura 5 se recoge el nodo ADVERTIR, correspondiente a esta entrada léxica, y la descripción de las formas flexivas que contienen la segunda raíz, agrupadas en el nodo C3vocra-e1.

Mediante esta descripción de ADVERTIR, el lexema construye sus formas verbales siguiendo el patrón de los verbos regulares de la Clase3, con la excepción de las formas agrupadas en C3vocra-e1, que se componen con la raíz alternativa *advirt-*.

En general, DATR permite describir los tipos léxicos de manera no redundante. Aun así, algunos tipos léxicos de este módulo deben adoptar idénticas descripciones, por ejemplo, en el caso de algunos fenómenos de alternancia que se dan de igual manera en las distintas conjugaciones: C2raizen-qgc y C3raizen-qgc describen las mismas formas del mismo modo, pero deben mantenerse como nodos independientes para heredar de sus respectivas clases regulares (Clase2 y Clase3).

Por último, aunque en la mayor parte de las ocasiones hemos seguido los criterios de clasificación establecidos en [Álvarez et al.], en determinados casos se han tenido que adoptar soluciones diferentes. Por ejemplo, la descripción de [Álvarez et al.] del verbo *seguir* y sus derivados es fonémica, mientras que la de esta implementación es grafémica. Por ello hemos debido establecer una triple alternancia en la raíz, con los alomorfos *segu-*, *sigu-* y *sig-* (*seguirá*, *sigues*, *sigamos*). La variante con *-i-* es *sig-* en todas las formas en las que aparece, excepto en P2 de Imp (*sigue*). Esta forma hace necesaria una tercera raíz *sigu-*, y un nuevo nodo C3seguir con la descripción de P2 de Imp, nuevo nodo que (como se observa en la

figura 4) hereda el resto de características morfológicas de C3vocra-e2.

#### 4. Paradigmas verbales irregulares

El cuarto módulo de la implementación implementa los paradigmas verbales irregulares. En gallego, las irregularidades afectan casi exclusivamente al tema verbal (raíz o vocal temática) ([Álvarez et al.: 318]). Además no se dan de manera tan arbitraria como pudiera parecer. La mayor parte de los verbos considerados irregulares presentan un tema de perfecto (319) con el que se forman los tiempos IP, IAP, SP y SF. En nuestra implementación en DATR, este tema se genera mediante la especificación de un alomorfo de la raíz (R2) en la entrada léxica y de un alomorfo de la vocal temática.

Además de este tema de perfecto, todos los paradigmas irregulares presentan algunas alternancias "más difícilmente sistematizables" (325) en determinadas formas del presente. La mayoría de ellos, sin embargo, utiliza en el SP un tercer alomorfo de la raíz, implementado en DATR como R3.

En [Álvarez et al.: 318-326], los verbos irregulares se agrupan mediante dos criterios independientes: por un lado, las propiedades de su tema de perfecto y, por otro, las características de sus irregularidades en los tiempos de presente (IPr y SPPr). A partir de la descripción morfológica de [Álvarez et al.], la clasificación empleada en esta implementación agrupa los paradigmas irregulares en dos grandes grupos (figura 6), tratando de capturar —en la medida de lo posible— las generalidades lingüísticas reconocibles en su morfografía,

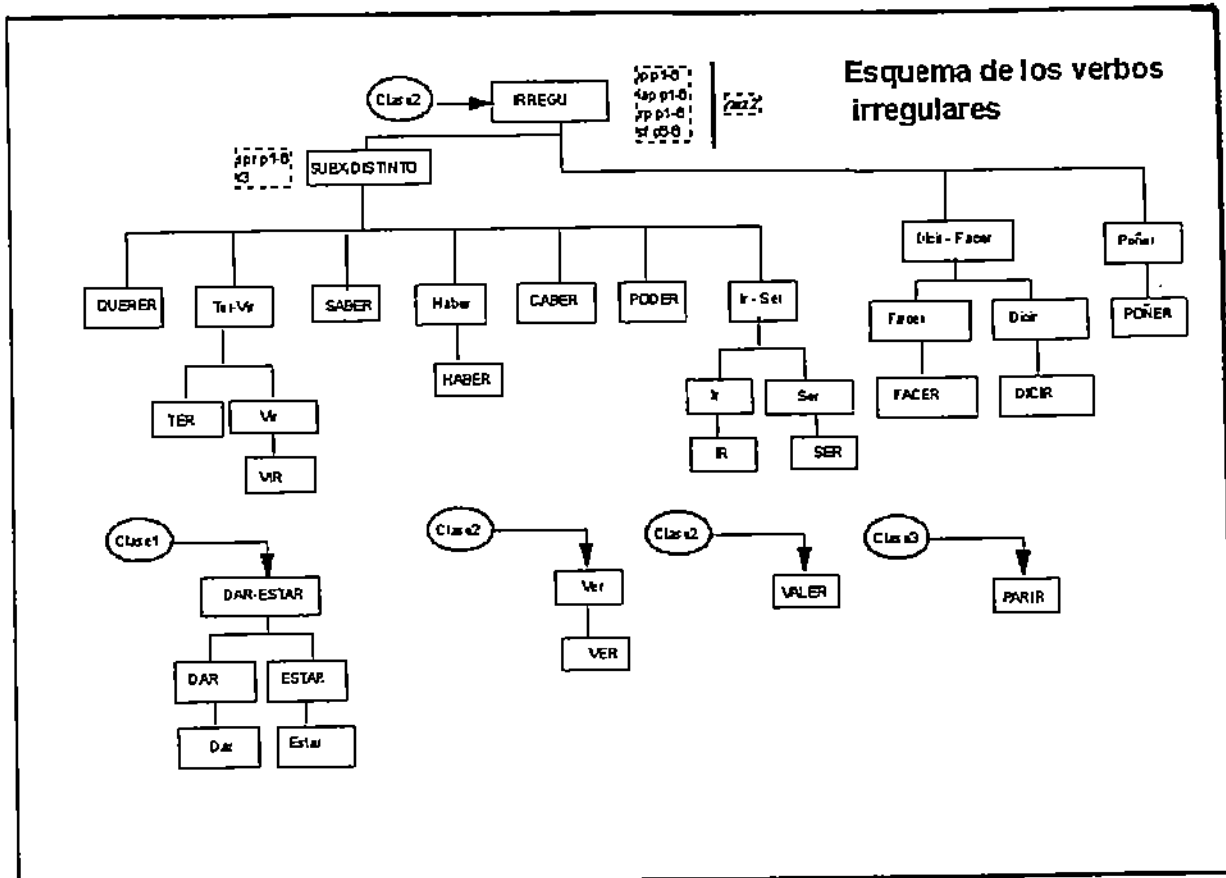


Figura 6

con el objetivo de describir el comportamiento flexivo de los verbos irregulares de manera tipificada.

De este modo, los dos conjuntos principales de verbos irregulares establecidos en la implementación son los siguientes (figura 6):

- (4.1) verbos irregulares de tipo *Irrregu*, y
- (4.2) otros paradigmas verbales irregulares.

#### 4.1. Tipo *Irrregu*

Se trata de un conjunto de verbos irregulares que construyen de la misma forma las formas en las que aparece el tema de perfecto, es decir, IP, IAP, SP y SF. La mayoría de los verbos irregulares que forman parte de este grupo comparten el procedimiento de construcción de SP<sub>r</sub>, por lo que el nodo *Spr-distinto* que lo describe depende directamente del nodo *Irrregu*. Como la gran mayoría de los irregulares de este grupo pertenecen a la segunda conjugación, la categoría hereda por defecto del nodo *Clase2*. Las tres únicas excepciones a esta norma (a saber, *vir* "venir",

*dicir* "decir" e *ir*) se solucionan superponiendo a la herencia por defecto la información morfológica adecuada en los nodos correspondientes a estos tres lexemas verbales. En *VIR* y *DICIR*, la inconsistencia se soluciona simplemente indicando que su conjugación es la tercera, de manera que su VT quede descrita con corrección. En el nodo *IR*, en cambio, debe especificarse la utilización del alomorfo cero para VT, ya que una de sus irregularidades consiste precisamente en la asimilación de la VT con la vocal de la raíz.

#### 4.2. Otros paradigmas irregulares

Verbos agrupados por separado que heredan directamente por defecto de clases regulares. Por ejemplo, el nodo *Dar-Estar* hereda la descripción morfológica de sus formas no irregulares de *Clase1*. El verbo *ver* presenta irregularidades propias, mientras que *valer* y *parir* sólo son irregulares en algunos tiempos de presente, y no en los de tema de perfecto.

La implementación de las irregularidades que atañen al tema de presente se realiza especificando en la entrada léxica los distintos alomorfos de la raíz (R, R2, R3..., Rn), listos para ser utilizados en la descripción de las formas verbales pertinentes. Pueden observarse algunos ejemplos en la tabla 1:

Lexema	R	R2	R3	R4	R5
TER	t	tiv	teñ	tiñ	te
PODER	pod	puid	poid		
DICIR	dic	dix	di	dig	
QUERER	quer	quix	queir		

Tabla 1

Como estrategia general de implementación, se han agrupado los verbos irregulares por sus similitudes flexivas de la manera más amplia posible para evitar la redundancia. Por ejemplo, los verbos *dicir* y *facer* "hacer" construyen muchas formas flexivas de la misma manera. Así, se define un nodo (denominado *Dicir-Facer*) que describe las formas flexivas comunes a ambos verbos, nodo del que heredan sus propiedades los nodos pertenecientes a los tipos léxicos de *Dicir* y *Facer* (véase figura 6). De estos dos últimos nodos, heredarían tanto los nodos lexemáticos correspondientes a los lexemas *DICIR* y *FACER*, como sus verbos derivados (*bendicir*, *contradicir*, *desdicir*, *maldicir*, *afacer*, *desfacer*, *refacer*, *satisfacer*).

Hay que destacar que, en la implementación presentada, todos los paradigmas irregulares heredan una parte de sus propiedades directa o indirectamente de las clases verbales regulares, lo que demuestra su grado de cohesión con el sistema morfológico flexivo de la lengua.

### 5. Representación de las entradas léxicas

Finalmente, en el quinto módulo de la implementación, se especifican los distintos alomorfos de la raíz (R, R2, R3..., Rn) y el paradigma flexivo de cada lexema verbal, de acuerdo con sus características morfológicas, como se ejemplifica en la figura 7.

```

COLLER:
<> == Clase2
<mor r> == coll.

FUXIR:
<> == C3vocra-u
<mor r> == fux
<mor r2> == fox.

PODER:
<> == Spr-distinto
<mor r> == pod
<mor r2> == puid
<mor r3> == poid.

```

Figura 7

Esta información asociada con las entradas léxicas permitirá que los mecanismos de herencia de DATR asignen la configuración morfológica adecuada a cada una de las formas verbales del lexema.

### 6. Disponibilidad de la implementación

Por motivos de espacio, no se ha incluido en la comunicación el código completo. Las personas interesadas en obtener esta implementación en DATR de la morfología verbal regular e irregular del gallego pueden hacerlo a través de Internet en cualquiera de las siguientes dos localizaciones:

- <ftp://ftp.uvigo.es/pub/lingware/morgal.zip>
- <http://www.uvigo.es/webs/h06/webh06/sli/>

### 7. Bibliografía

- ÁLVAREZ, R.; REGUEIRA, X. L.; y MONTEAGUDO, H. (1986), *Gramática galega*, Galaxia, Vigo.
- ANDRY, F.; FRASER, N.; MCGLASHAN, S.; THORNTON, S.; y YOLD, N. (1992), "Making DATR Work for Speech: Lexicon Compilation in SUNDIAL", *Computational Linguistics*, 18 (3), pp. 245-267.
- CAHILL, L. (1993a), "Morphology in the Lexicon", en *Proceedings of the 6th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 87-96.

- CAHILL, L. (1993b), "Some reflections on the conversion of the TIC lexicon into DATR", en T. Briscoe et al. (eds.), *Inheritance, Defaults and the Lexicon*, Cambridge University Press, pp. 47-57.
- CARROLL, J.; NICOLOV, N.; SHAUMYAN, O.; SMETS, M.; y WEIR, D. (1998), "Grammar Compaction and Computation Sharing in Automata-based Parsing", en *Proceedings of the First Workshop on Tabulation in Parsing and Deduction*, pp. 16-25.
- EVANS, R. y GAZDAR, G. (1996), "DATR: A Language for Lexical Knowledge Representation", *Computational Linguistics*, 22 (2), pp. 167-216.
- EVANS, R.; GAZDAR, G.; y WEIR, D. (1995), "Encoding Lexicalized Tree Adjoining Grammars with a Nonmonotonic Inheritance Hierarchy", en *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, pp. 77-84.
- GAZDAR, G. (1992), "Paradigm Function Morphology in DATR", en L. Cahill y R. Coates (eds.), *Sussex Papers in General and Computational Linguistics*, Universidad de Sussex, pp. 43-53.
- GÓMEZ GUINOVART, J. (1998), "Representación del conocimiento léxico en el formalismo DATR: un tratamiento computacional de las alternancias diatéticas", en *Actas del III Congreso de Lingüística General*, Universidad de Salamanca.
- GÓMEZ GUINOVART, J. y AGUIRRE, J. L. (1998), "Representación del conocimiento léxico en el formalismo DATR: un tratamiento computacional de la morfología verbal del gallego", en *Actas del XVI Congreso de la Asociación Española de Lingüística Aplicada*, Universidad de La Rioja.
- JENKINS, E. (1990), "Japanese Verbs in DATR", en R. Evans y G. Gazdar (eds.), *The DATR Papers I*, Universidad de Sussex.
- KILGARRIFF, A. (1993), "Inheriting Verb Alternations", en *Proceedings of the 6th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 213-221.
- REAL ACADEMIA GALEGA e INSTITUTO DA LINGUA GALEGA (1995), *Normas ortográficas e morfolóxicas do idioma galego*, 13ª ed. rev., RAG-ILG, Vigo.
- REINHARD, S.; y GIBBON, D. (1991), "Prosodic Inheritance and Morphological Generalisations", en *Proceedings of the 5th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 131-136.