

La morfología verbal del español y la generación automática

Juan Rafael Zamorano Mansilla
 Universidad Complutense de Madrid
 juanrafaelz@terra.es

Resumen: El propósito de este artículo es, en primer lugar, llamar la atención sobre un problema concerniente a la generación automática del español: la extensa morfología del verbo crea un problema de almacenamiento de léxico, ya que introducir una a una todas las posibles variaciones morfológicas de cada elemento léxico es una tarea ardua y pesada. Se hace, pues, necesario la especificación de recursos para la generación de todas las posibles variaciones a partir de un mínimo de información contenida en el léxico, con lo cual la introducción de elementos léxicos verbales se hace mucho más simple. La especificación de estos recursos constituye la parte central del presente artículo. Dichos recursos han sido diseñados para el entorno de desarrollo de gramáticas KMPL, que en la actualidad está siendo desarrollado en la Universidad de Bremen. Esto significa que las especificaciones computacionales de estos recursos son válidas sólo para KPML o un entorno similar que utilice una aproximación sistémico-funcional. No obstante, la parte más abstracta del artículo, la que trata sobre la clasificación y organización de los verbos para un almacenamiento léxico más económico, puede ser aplicable a cualquier sistema de generación.

Palabras clave: Morfología verbal, español, generación.

Abstract: The aim of this paper is, firstly, to focus on a problem about automatic generation in Spanish: the long verbal morphology causes a problem about the lexicon storage, because of storing each morphological variation of a term. Therefore, it is necessary the specification of tools to generate all possible variations of a term, from some information stored in the lexicon. The specification of this kind of tools is the main aim of this paper. These tools have been designed for the development toolkit for KMPL grammars, which is presently developed in the University of Bremen. It means that the computational specifications of these tools are only proper for KPML.

Keywords: Verbal Morphology, Spanish, generation.

1 *El problema de la morfología del verbo en español*

El problema planteado por la morfología del verbo español es el siguiente: cada tiempo contiene 3 formas distintas para cada persona (primera, segunda y tercera), y a su vez cada persona puede aparecer en singular o plural. Si consideramos que el español cuenta con 7 tiempos simples, una multiplicación nos dice que cada verbo en español presenta 42 formas distintas. A éstas hemos de sumar las 2 formas del Imperativo y las 3 formas no personales (infinitivo, gerundio y participio).

En principio existen dos formas de tratar esta variación:

- a) Prescindir de recursos morfológicos e incluir las 47 formas en el archivo del léxico del generador. No sólo es una tarea larga, costosa y pesada, sino que también hace crecer de manera excesiva el tamaño de los archivos de vocabulario.
- b) Proporcionar recursos morfológicos que permitan generar, a partir de la raíz del verbo, las 47 formas.

La segunda postura es sin duda más aconsejable, ya que el tiempo empleado en la creación de dichos recursos morfológicos

permite un ahorro considerable en tiempo y espacio a la hora de introducir nuevos elementos léxicos. Por desgracia la conjugación del verbo español ha sido afectada por fenómenos fonéticos que la han alejado de la regularidad del latín. Esto quiere decir que no es suficiente con incluir la raíz verbal en el léxico y posteriormente añadir a dicha raíz las terminaciones apropiadas que dicten los recursos morfológicos.

La propuesta que aquí se hace para el tratamiento de la flexión verbal se halla más cerca de la segunda postura, si bien la peculiaridad del español requiere un grado importante de complejidad en la organización de los recursos morfológicos y la interacción de éstos con la información léxica. No obstante, antes de entrar en los detalles de esta complejidad es conveniente examinar los cambios que afectan a la flexión del verbo para comprender mejor la naturaleza de los problemas que plantea.

2 *Los cambios morfológicos del verbo español*

La flexión del verbo español es la típica de las lenguas Indo-Europeas: a la raíz se añaden terminaciones que contienen la información de persona, número y tiempo. Estas terminaciones varían según un parámetro preciso: el modelo de conjugación o paradigma que sigue el verbo. Así, la terminación de primera persona del plural del Presente de Indicativo no es igual en los verbos de la primera conjugación (*am-áis*), la segunda (*tem-éis*) o la tercera (*part-ís*).

Otro rasgo típico de las lenguas Indo-Europeas es que la raíz verbal puede cambiar. También en este caso lo hace según un parámetro determinado: el tiempo. Es habitual que los verbos tengan una raíz para el presente y otra distinta para el pasado. Esto puede observarse en latín, en las lenguas germánicas, el griego o el ruso.

Ahora bien, el principal problema del español es que, como consecuencia de fenómenos fonéticos acontecidos a lo largo de su historia, ha roto estas dos tendencias. Las terminaciones, que desde luego varían según la conjugación, varían también según criterios que nada tienen que ver con ella. Así, la terminación de la primera persona del singular del Presente de Indicativo es *-o* (*am-o*), aunque hay verbos que toman *-oy* (*est-oy*) e incluso *-e* (*h-e*, *s-é*). El Pretérito Indefinido presenta terminaciones

diferentes según el tipo de conjugación, pero también posee unas terminaciones para verbos de pretérito fuerte que son independientes del tipo de conjugación. La tercera persona del plural del Pretérito Indefinido de los verbos fuertes es en ocasiones *-eron* (*dij-eron*), en ocasiones *-ieron* (*tuv-ieron*), independientemente de la conjugación a la que pertenezca el verbo. Y la lista podría prolongarse todavía más.

La segunda tendencia (los cambios de raíz van a asociados a ciertos tiempos) también ha sido rota en español. En nuestro idioma conviven 2 o incluso 3 raíces diferentes dentro del mismo tiempo. Tomemos el ejemplo del verbo *poder*. En el Presente de Indicativo la raíz puede ser *pod-*, como en *pod-emos*, mientras que con otra persona y número la raíz es *pued-*, como en *pued-es*. Estas variaciones de la raíz no son caprichosas. Como ya he dicho, tienen un origen fonético, por lo que presentan cierta regularidad en su distribución. De hecho, es posible crear modelos de conjugación que se superponen a las clásicas primera, segunda y tercera conjugación, y ambos deben ser utilizados para el almacenamiento de léxico en los generadores.

De estos hechos podemos concluir que un programa que pretenda generar formas verbales en español debe:

- a) incluir información sobre los tipos de conjugación posibles según el tipo de variación en la raíz.
- b) incluir información sobre los tipos de conjugación posibles según el tipo de terminación.
- c) armonizar ambos tipos de información con los modelos de conjugación clásicos: primera, segunda y tercera conjugación.
- d) permitir que los parámetros de tiempo, persona y número tengan influencia sobre la elección de la raíz, no sólo sobre la terminación.

3 *Especificación de recursos para la generación de formas verbales*

En esta sección comenzaré por presentar la estructura básica válida para los verbos regulares, a la cual incorporaré los cambios necesarios para dar cabida también a los verbos irregulares. La terminología y los especificaciones computacionales empleadas pertenecen a la gramática sistémico-funcional tal y como son utilizadas por el entorno de

desarrollo de gramáticas KPML (Bateman y Matthiessen 1991, Bateman 1997, Reiter y Dale 2000), la gramática computacional NIGEL (Mann y Matthiessen 1985, Matthiessen 1985) y el generador PENMAN (Mann 1983).

3.1 Estructura simple para verbos regulares

Si todos los verbos del español fuesen regulares, los recursos que a continuación voy a presentar serían suficientes para generar toda la variación morfológica del verbo español.

Dentro de la gramática, en el nivel de generación de palabras (*word-rank*) y dependiendo de la opción *Verbo*, encontramos los sistemas que aparecen en la Tabla 1.

Verbo (insert Root) (insert Ending)
FORMA-VERBAL
Presente-Indicativo
Presente-Subjuntivo
Pretérito-Indefinido
Pretérito-Imperfecto
Pretérito-Imperfecto-Subjuntivo
Futuro
Condicional
CONJUGACION
Primera-conjugación
Segunda-conjugación
Tercera-conjugación
PERSONA
Primera-persona
Segunda-persona
Tercera-persona
NUMERO
Singular
Plural

Tabla 1: Sistemas dependientes del rasgo *Verbo*

Los sistemas, en mayúsculas, contienen opciones que se oponen entre sí significativamente. Estas opciones, con mayúscula inicial, reciben el nombre de rasgos y pueden llevar asociadas órdenes sintácticas¹ (entre paréntesis) o conducir a su vez a nuevos

¹ El término *sintácticas* aparece aquí utilizado en sentido amplio, al estilo de la gramática sistémico-funcional, para referirse a todo lo que tiene que ver con el eje sintagmático en oposición al eje paradigmático del lenguaje.

sistemas. En este caso, *Verbo* es un rasgo que conduce a los sistemas FORMA-VERBAL, CONJUGACIÓN, PERSONA y NÚMERO, además de contener dos órdenes, (*insert Root*) y (*insert Ending*), que indican que deben introducirse dos funciones sintácticas. Posteriormente se especificará qué “sustancia” realizará dichas funciones. En el caso de *Root*, la parte de la gramática que decide sobre su realización es la región de transitividad, no la morfología. Supongamos que la palabra a generar es cualquiera de las formas del verbo *amar*. Dicha palabra aparece especificada de la siguiente forma en el archivo de léxico.

```
(LEXICAL-ITEM
:NAME AMAR
:SPELLING "amar"
:FEATURES (LOVE-VERB REACTION-
VERB EXPERIENCE-VERB MIDDLE-VERB FIRST-
CONJUGATION)
:PROPERTIES ((raíz "am"))
)
```

La región de transitividad decide el tipo de verbo que va a generarse según su semántica, por lo que selecciona el elemento léxico guiándose por la información que aparece en la ranura (*slot*) FEATURES. Una vez que en la región de transitividad se ha seleccionado un verbo con los requisitos semánticos MIDDLE-VERB, EXPERIENCE-VERB, LOVE-VERB, etc, se especifica que la función *Root* debe ser realizada por la sustancia que aparece en el léxico junto a la etiqueta *raíz*, o sea, *am*.

La decisión sobre qué sustancia realizará la función *Ending* sí es competencia de los sistemas de morfología verbal expuestos en la tabla 1, ya que la terminación del verbo depende de factores como el tiempo, la persona, el número y la conjugación.

Para poner en relación todos los rasgos que aparecen en la tabla 1 hay que recurrir a los portales (*gates*). En esencia, un portal es un sistema como los de la tabla 1, con la salvedad de que contienen un único rasgo.

Pero veamos con un ejemplo concreto cómo funcionan los portales. Supongamos que deseamos generar la segunda persona de plural de Presente de Indicativo del verbo *amar*, del que ya tenemos la raíz.

En el caso de *amar*, un verbo de la primera conjugación, la terminación para dicho tiempo, persona y número es *-áis*. Así pues el portal requerido sería como sigue:

```
(GATE
:NAME 2ndplPresentIndicative
:INPUTS (AND Presente-Indicativo
          Primera-conjugación
          Segunda-persona
          Plural)
:OUTPUTS ((1.0 2ndplPresentIndicative
           (preselect-substance ending
            2ndplPresentIndicative)))
:REGION verb
)
```

En la ranura INPUTS especificamos las condiciones de entrada, es decir, aquellos rasgos que deben ser seleccionados para acceder a este portal (Presente-Indicativo, Primera-conjugación, Segunda-persona, Plural). En OUTPUTS se especifica la información sintáctica (preselect-substance ending 2ndplPresentIndicative), que en este caso establece que la función *Ending* será realizada por un elemento con el nombre *2ndplPresenIndicative*, que por supuesto ha sido introducido en el léxico previamente de la siguiente forma:

```
(LEXICAL-ITEM
:NAME 2ndplPresenIndicative
:SPELLING "áis"
:FEATURES (2ndplPresenIndicative)
)
```

Así pues, el resultado final es el que muestra la Figura 1.

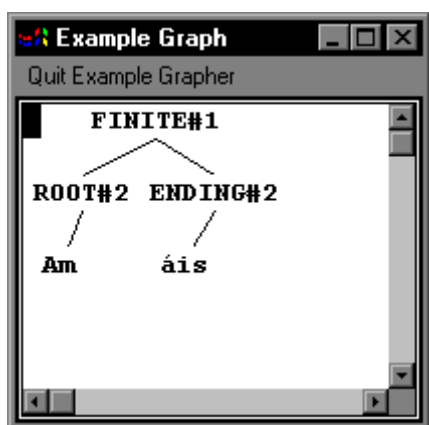


Figura 1

Siguiendo el ejemplo, sólo tenemos que diseñar portales como el de arriba para los distintos tiempos, conjugaciones, personas y números.

Esta estructura básica con sus múltiples portales permite generar verbos regulares de

forma eficaz. En el archivo de léxico sólo hay que introducir la raíz del verbo. También hay que introducir las distintas terminaciones, pero una vez introducidas serán empleadas por todos los verbos, por lo que el ahorro de tiempo y espacio es considerable.

La cuestión es ahora qué cambios deben introducirse en esta estructura para dar cabida a los verbos irregulares igualmente.

3.2 Estructura compleja para todo tipo de verbos

Como ya he expuesto más arriba, el principal problema del modelo simplificado de la sección anterior es que:

a) no permite la existencia de distintas terminaciones para el mismo significado, o dicho de otra forma, sólo prevé la variación en las terminaciones motivada por una variación en la persona, el número o/y el tiempo.

b) no prevé la existencia de cambios de raíz.

El primer problema es relativamente sencillo de solucionar. Hay que incluir un nuevo sistema dependiente de las combinaciones de persona, número y tiempo que son realizadas por más de una terminación. Este sistema deberá reemplazar a los portales que vimos más arriba.

Por ejemplo, ya hemos visto que la primera persona del singular del Presente de Indicativo puede ser realizada por la terminación *-o*, pero también por *-oy* o *-e*. Así pues, un sistema que fuese sensible a esta variación no motivada por ningún factor como el tiempo, la persona o el número, sería como sigue.

```
(SYSTEM
:NAME 1stsgPresentIndicative
:INPUTS (AND Presente-Indicativo
          Primera-persona
          Singular)
:OUTPUTS ((0.33333334 o-Present-ending
           (preselect-substance ending o-
            PresentEnding)
           (0.33333334 oy-Present-ending
            (preselect-substance ending oy-
             PresentEnding)
           (0.33333334 e-Present-ending
            (preselect-substance ending e-
             PresentEnding)))
:CHOOSE 1stsgPresentType
:REGION verb
)
```

Una vez más, dentro de INPUTS podemos observar los rasgos que deben haber sido escogidos para que el programa acceda a este sistema (Presente-Indicativo, Primera-persona, Singular). Dentro de OUTPUTS tenemos las tres

posibles opciones: una terminación en *-o*, en *-oy*, o en *-e*. Entre paréntesis se hallan las reglas sintácticas, que son iguales en los tres casos: *Ending* debe ser realizado por un elemento léxico que se encuentra en el archivo de vocabulario bajo la etiqueta *o-PresenEnding*, *oy-PresenEnding*, o bien *e-PresenEnding* según el caso. La elección la realiza el sistema basándose en información contenida en la entrada léxica del verbo. Por ejemplo, el verbo *estar* pertenece al grupo de verbos cuya primera persona del singular de Presente de Indicativo acaba en *-oy*. Este detalle debe haber sido incluido al introducir *estar* en el archivo de léxico, en la ranura FEATURES.

```
(LEXICAL-ITEM
:NAME ESTAR
:SPELLING "estar"
:FEATURES (RELATIONAL-BE-VERB
RELATIONAL-VERB MIDDLE-VERB FIRST-
CONJUGATION OY-1STSGPRESENT)
:PROPERTIES ((raiz "est"))
)
```

Ahora bien, puesto que los verbos que utilizan esta terminación (o la terminación en *-e*) constituyen una minoría, no sería práctico incluir en todos los verbos regulares una información equivalente del tipo O-1STSGPRESENT, para informar de que dichos verbos tienen una terminación en *-o*. En vez de eso, el programa busca dentro de la ranura FEATURES del elemento léxico si existe información a este respecto. En caso contrario, el programa asume que el verbo es regular y elige por defecto la opción de terminación en *-o*.

No hay más que construir los sistemas adecuados, siguiendo el ejemplo, para todos aquellos casos en los que encontramos varias terminaciones que no están motivadas por la elección de tiempo, persona, número o conjugación. Tal es el caso del Imperativo, a veces con *-e* final (*sube*), a veces sin ella (*ven*); la tercera persona del plural del Pretérito Indefinido, a veces en *-eron* (*dijeron*), a veces en *-ieron* (*vinieron*); o el Futuro Simple, a veces con las terminaciones completas (*amaré*, *temeré*, *subiré*), a veces con las terminaciones contractas (*podré*, *querré*), por citar unos ejemplos. En cualquier caso es importante que siempreelijamos una de las terminaciones como la opción por defecto, pues de este modo evitaremos tener que incluir información en el léxico para un buen número de verbos, como ya he indicado antes.

El segundo problema, los cambios de raíz del verbo, es algo más complejo en su resolución. Una aproximación clásica al problema es la de crear submodelos de conjugación, dentro de los tres modelos principales (primera, segunda, tercera conjugación). Uno puede tomar un verbo (*poder*, por ejemplo) y crear un paradigma (modelo *poder*) bajo el que se agrupan todos los verbos que experimentan los mismos cambios de raíz que dicho verbo. El problema es que de este modo es posible establecer un número considerable de modelos. El Esbozo de la RAE (RAE 1991) describe 19 tipos básicos de alteración en la raíz, si bien muchos verbos combinan más de uno de esos tipos; Andrés Bello (Bello 1847) establece 13 tipos con múltiples subdivisiones cada uno; y Leonardo Gómez Torrego (Gómez Torrego 2000), sin duda el más completo y exhaustivo, descubre hasta 57 paradigmas.

La aproximación que propongo en este artículo es ligeramente diferente y presenta notables ventajas. Se basa en la observación de que, si bien es cierto que muchos verbos españoles presentan algún tipo de irregularidad, no es menos cierto que todos presentan regularidad en alguna parte de la conjugación. El verbo *ser*, sin ir más lejos, uno de los más irregulares de nuestra lengua, posee un Futuro Simple de Indicativo absolutamente regular. Y el verbo *haber*, también altamente irregular, tiene un Pretérito Imperfecto regular por completo.

Mi propuesta es establecer paradigmas de irregularidad para cada tiempo de forma aislada, y tratarlos independientemente del resto de los tiempos. En las clasificaciones tradicionales, el verbo *poder*, por ejemplo era mostrado como modelo de un tipo de irregularidad no sólo en el Presente de Indicativo, sino también en el Pretérito Indefinido, el Presente de Subjuntivo, etc. El número de verbos que se comportan exactamente igual que *poder* a lo largo de toda la conjugación es bastante reducido, mientras que son muchos los que se comportan igual que él en el Presente de Indicativo, por ejemplo. Si establecemos modelos de irregularidad que se aplican únicamente a un tiempo y no a la conjugación al completo, obtenemos 5 modelos en el tiempo más irregular (Presente de Subjuntivo), aunque la mayoría de los tiempos sólo tienen 2. Las ventajas de tener que especificar una media de 3 modelos por cada

tiempo en vez de más de 50 para el total de la conjugación son apreciables, especialmente a la hora de especificar los recursos gramaticales encargados de la morfología, pues tendremos que crear un menor número de sistemas y de menor complejidad.

Pero pasemos a ver en detalle un ejemplo de esto. Concentrémonos en el Presente de Indicativo, uno de los tiempos más sujeto a irregularidades de toda la conjugación. Si clasificamos todos los verbos del español tomando en consideración únicamente los cambios en el Presente de Indicativo, obtenemos 4 modelos:

- 1) los verbos regulares, que usan la misma raíz en todas las personas.
- 2) los verbos del tipo *pedir*, que usan una raíz para las 3 primeras formas del singular y la tercera del plural y otra para la primera y segunda personas del plural.
- 3) los verbos del tipo *tener*, que usan una raíz para la primera persona del singular, otra para la segunda y tercera del singular y la tercera del plural, y una tercera raíz para la primera y segunda personas del plural.
- 4) los verbos del tipo *poner*, que usan una raíz para la primera persona del singular y otra para el resto de personas.

Verbo (insert Root) (Insert Ending)
FORMA-VERBAL
Presente-Indicativo
TIPO-DE-PRESENTE
Regular (preselect-substance-as-stem Root raíz)
Tipo-pedir
Tipo-tener
Tipo-poner
Presente-Subjuntivo
Pretérito-Indefinido
Pretérito-Imperfecto
Pretérito-Imperfecto-Subjuntivo
Futuro
Condicional

Tabla 2: Sistema para la representación de los 4 tipos de Presente de Indicativo

Esto queda reflejado en la gramática de la siguiente forma: en primer lugar, puesto que ahora tenemos 4 tipos distintos de Presente de Indicativo, hemos de incluir un sistema con las 4 opciones correspondientes dependiente del

rasgo *Presente-Indicativo*, como se aprecia en la Tabla 2.

Se observará que el único rasgo que contiene reglas sintácticas es *Regular*, debido a que los verbos regulares emplean una única raíz para todas la personas. Los otros tres rasgos, al utilizar distintos tipos de raíz según la persona y el número, no pueden especificar nada sobre la raíz a utilizar sin combinar esta información previamente con la de persona y número, por lo que no llevan asociadas reglas sintácticas de momento. Estas aparecen, una vez más, en los portales, sistemas capaces de combinar la información de tiempo, número, persona, conjugación, tiempo y tipo de presente.

Veamos, pues, los dos portales necesarios para los verbos con un Presente del tipo *pedir*.

```
(GATE
:NAME raíz1
:INPUTS (AND Presente-Indicativo
Tipo-pedir
(OR
(AND Plural
Tercera-persona)
(AND Singular
(OR Primera-persona
Segunda-persona
Tercera-persona)))
:OUTPUTS ((1.0 Raíz1
(preselect-substance-as-stem Root
raíz1)))
:REGION:verb
)

-----

(GATE
:NAME raíz2
:INPUTS (AND Presente-Indicativo
Tipo-pedir
(AND Plural
(OR Segunda-persona
Primera-persona)))
:OUTPUTS ((1.0 Raíz2
(preselect-substance-as-stem Root
raíz2)))
:REGION:verb
)
```

El primer portal es el encargado de generar las formas de primera, segunda y tercera persona del singular, además de la tercera del plural. La razón para esta distribución es que, como hemos visto, los verbos del tipo *pedir* utilizan una raíz para estas personas y otra para las restantes. En la ranura INPUTS de ambos portales se especifican los rasgos que deben haber sido seleccionados previamente para acceder al portal. El comando AND indica que un rasgo debe estar presente, mientras que OR sólo exige la presencia de uno entre varios rasgos. En la ranura OUTPUTS se encuentra la información sobre el tipo de raíz a seleccionar. Podrá observarse que en el primer portal se

identifica la función *Root* con la forma Raíz1, mientras que en el segunda portal la identificación es con Raíz2. Evidentemente, Raíz1 y Raíz2 son etiquetas que hacen referencia a formas incluidas en el archivo de léxico. El verbo *pedir*, por ejemplo, tendría esta especificación léxica.

```
(LEXICAL-ITEM
:NAME                PEDIR
:SPELLING             "pedir"
:FEATURES             (TELLING SYMBOLIC-
                      VERB EFFECTIVE-VERB THIRD-CONJUGATION)
:PROPERTIES           ((raiz1 "ped")
                      (raiz2 "pid"))
)
```

En el caso de verbos del tipo *tener*, los que presentan una mayor irregularidad en el Presente, debemos incluir 3 raíces:

```
(LEXICAL-ITEM
:NAME                TENER
:SPELLING             "tener"
:FEATURES             (POSSESSION-VERB
                      RELATIONAL-VERB MIDDLE-VERB SECOND-
                      CONJUGATION)
:PROPERTIES           ((raiz1 "ten")
                      (raiz2 "tien") (raiz3 "teng"))
)
```

Del mismo modo sólo hay que continuar construyendo portales como los ya mostrados, cambiando las condiciones de entrada en INPUTS para reflejar las asociaciones entre personas y número con la raíz correspondiente según los 4 modelos expuestos.

Entrar en los detalles de los sistemas dependientes del resto de tiempos simples es una tarea que excede los límites de espacio de este artículo. Pero la siguiente tabla muestra de manera resumida los tipos de irregularidad que he establecido para cada tiempo, representados como opciones dentro de sistemas dependientes de los tiempos correspondientes.

Verbo (insert Root) (Insert Ending)
FORMA-VERBAL
Presente-Indicativo TIPO-DE-PRESENTE Regular (preselect-substance-as-stem Root raíz) Tipo-pedir [222112] Tipo-tener [322112] Tipo-poner [311111]
Presente-Subjuntivo TIPO-DE-PRESENTE-SUBJUNTIVO Regular (preselect-substance-as-stem Root raíz) Tipo-pedir (preselect-substance-as-stem Root raíz2) Tipo-acertar [222112] Tipo-tener (preselect-substance-as-stem Root raíz3) Tipo-sentir [222552]
Pretérito-Indefinido TIPO-DE-PRETÉRITO-INDEFINIDO Regular (preselect-substance-as-stem Root raíz) Tipo-pedir [112112] Tipo-tener (preselect-substance-as-stem Root raíz5)
Pretérito-Imperfecto TIPO-DE-PRETÉRITO-IMPERFECTO Regular (preselect-substance-as-stem Root raíz) Tipo-ser (preselect-substance-as-stem Root raíz4)
Pretérito-Imperfecto-Subjuntivo TIPO-DE-PRETÉRITO-IMPERFECTO-S. Regular (preselect-substance-as-stem Root raíz) Tipo-pedir (preselect-substance-as-stem Root raíz2) Tipo-tener (preselect-substance-as-stem Root raíz5) Tipo-sentir (preselect-substance-as-stem Root raíz3)
Futuro TIPO-DE-FUTURO Regular (preselect-substance-as-stem Root raíz) Tipo-tener (preselect-substance-as-stem Root raíz6)
Condicional TIPO-DE-CONDICIONAL Regular (preselect-substance-as-stem Root raíz) Tipo-tener (preselect-substance-as-stem Root raíz6)

Tabla 3: Sistemas dependientes de los tiempos para la especificación de los tipos de irregularidad

Es de destacar que muchos verbos, dentro de su irregularidad, mantienen la misma raíz para todas las personas de un tiempo. De ahí que

muchos de los rasgos de la tabla 3 lleven asociadas órdenes sintácticas sin necesidad de aguardar a conocer información sobre la persona y el número. En los casos en los que exista variación de la raíz según la persona y número, estas órdenes sintácticas deben aparecer en los portales correspondientes, y no asociadas directamente al rasgo.

Obsérvese que en la tabla se hace mención a varias raíces numeradas (raíz, raíz2, raíz3, raíz4, etc). Los números del 1 al 3 están reservados para las raíces que aparecen en el Presente. El 4 es para aquellos verbos excepcionales que poseen una raíz exclusiva para el Pretérito Imperfecto de Indicativo. El número 5 se reserva para las raíces exclusivas del Pretérito Indefinido, mientras que el 6 es para la raíz del Futuro de Indicativo.

En aquellos modelos de conjugación que presentan variación de la raíz según la persona, he incluido entre corchetes la alternancia de ésta siguiendo el orden tradicional de las personas y números. Por ejemplo, 322112 significa que en este modelo se emplea la raíz 3 para la primera persona del singular, la raíz 2 para la segunda y tercera persona del singular más la tercera del plural, y la raíz 1 para la primera y segunda persona del plural.

Una vez visto el esquema de los recursos gramaticales, veámos ejemplos de especificaciones léxicas como prueba del ahorro que supone este método de generación. El verbo *hacer*, uno de los más irregulares, quedaría del siguiente modo.

```
(LEXICAL-ITEM
:NAME HACER
:SPELLING "hacer"
:FEATURES (CREATION-VERB DO-
VERB EFFECTIVE-VERB PRESENTE-TIPO-PONER
FUTURO-TIPO-TENER PRESENTE-SUBJUNTIVO-
TIPO-TENER INDEFINIDO-TIPO-TENER SECOND-
CONJUGATION)
:PROPERTIES ((raiz1 "hac")
(raiz3 "hag") (raiz5 "hic") (raiz6 "ha"))
)
```

En PROPERTIES hemos incluido las 4 raíces diferentes del verbo, mientras que en FEATURES he subrayado la información requerida por la gramática. Obsérvese que, puesto que este verbo presenta irregularidades en 4 tiempos, hemos especificado el modelo que sigue en cada uno de esos tiempos de manera individual. Por otro lado, el programa asume que aquellos tiempos sobre los que no se especifica nada son regulares.

A su vez, un verbo regular, como *temer*, no necesitaría más que las siguientes especificaciones.

```
(LEXICAL-ITEM
:NAME TEMER
:SPELLING "temer"
:FEATURES (FEAR-VERB REACTION-
VERB EXPERIENCE-VERB MIDDLE-VERB SECOND-
CONJUGATION)
:PROPERTIES ((raiz "tem"))
)
```

Es decir, dentro de FEATURES sólo se ha indicado la conjugación. Y la única raíz necesaria aparece en PROPERTIES.

4 Resumen y conclusiones

En este artículo he mostrado el problema planteado por la extensa variación morfológica del verbo español a la hora de almacenar léxico en los generadores automáticos. Tras hacer un repaso de las principales irregularidades de la conjugación, he mostrado una clasificación de éstas basada en la creación de modelos de conjugación por cada tiempo individual, en vez del tradicional modelo de conjugación completa. He defendido las ventajas de esta clasificación, a saber: por un lado, el número de modelos resultantes es menor, por lo que los recursos morfológicos a especificar son menos complicados; además, reconoce el hecho de que incluso el más irregular de los verbos tiene tiempos regulares.

Finalmente, he mostrado cómo gracias a la creación de recursos morfológicos dentro de la gramática, la cantidad de información que debemos incluir en la entrada léxica de un verbo es mucho menor. Un máximo de 5 raíces y otros tantos rasgos referentes a la conjugación y el modelo de variación a seguir en los tiempos que sean irregulares bastan para proporcionar al generador la información necesaria para generar la conjugación completa de cualquier verbo.

En mi opinión el trabajo que supone especificar todos los recursos morfológicos queda compensado por el ahorro de espacio y tiempo a la hora de introducir léxico. No olvidemos que el léxico es, en principio, ilimitado, mientras que los recursos morfológicos, una vez creados, son utilizados de forma recurrente por todos los verbos.

Creo también que esta especificación morfológica no sólo supone un avance de tipo técnico, sino que desde un punto de vista lingüístico es más correcto que la inclusión de

las distintas formas del verbo una a una en el léxico.

Bibliografía

- Bateman, J. 1997. "Enabling technology for multilingual natural language generation: the KPML development environment". *Journal of Natural Language Engineering*, 31.:15-55.
- Bateman, J.A. y Matthiessen, C.M.I.M 1991. *Text Generation and Systemic-Functional Linguistics. Experiences from English and Japanese*. London: Pinter Publishers.
- Bello, A. 1847. *Gramática de la lengua castellana destinada al uso de los americanos*, Santiago de Chile. Cito edición de 1980, Madrid: EDAF.
- Gómez Torrego, L. 2000. *Gramática didáctica del español*. Madrid: SM.
- Mann, W.C. 1983. "An overview of the PENMAN text generation system". Technical Report ISI/RR-83-114, USC/Information Sciences Institute, Marina del Rey, California.
- Mann, W.C. y Matthiessen, C.M.I.M. 1985. "Demonstration of the Nigel text generation computer program". En J.D. Benson y W.S. Greaves, ed., *Systemic Perspectives on Discourse*, Volumen I. Norwood, New Jersey: Ablex.
- Matthiessen, C.M.I.M. 1985. "The systemic framework in text generation: Nigel". En J.D. Benson y W.S. Greaves, ed., *Systemic Perspectives on Discourse*, Volumen I. Norwood, New Jersey: Ablex.
- Real Academia Española. Comisión de Gramática 1991. *Esbozo de una nueva gramática de la lengua española*. Madrid: Espasa-Calpe.
- Reiter, E. y Dale, R. 2000. *Building Natural Language Generation Systems*. Cambridge: CUP.

