

La conjugación verbal: propuesta desde un sistema de traducción por ordenador.

Pilar Salamanca
Dpto de Lingüística
Facultad de Filosofía y Letras
Universidad Autónoma de Madrid

Resumen

Todo trabajo en que se involucre la Lingüística Informática requiere no sólo profundos fundamentos lingüísticos, sino también una adecuada formalización de los fenómenos de la lengua.

Pero no siempre es fácil. Con bastante frecuencia, el comportamiento de la lengua resulta difícil de constreñir en un conjunto de reglas.

Concretamente, la conjugación verbal en español presenta una considerable complejidad: por una parte, existen diferentes paradigmas de conjugación, formas compuestas, numerosos tiempos, varios modos, desinencias de diversos tipos, etc.; y, por otro lado, nos encontramos distintos tipos de irregularidades, algunos relacionados con cambios en la raíz, otros con las partículas flexivas.

Todo ello complica el tratamiento que deba darse del fenómeno en el nivel morfológico del procesamiento de la lengua, en nuestro caso, el nivel EMS (eurotra morphological structure) del proyecto EUROTRA de traducción por ordenador.

La conjugación verbal: propuesta desde un sistema de traducción por ordenador

Todo trabajo en que se involucre la Lingüística Informática requiere no sólo profundos fundamentos lingüísticos, sino también una adecuada formalización de los fenómenos de la lengua.

Pero no siempre es fácil. Con bastante frecuencia, el comportamiento de la lengua resulta difícil de constreñir en un conjunto de reglas. Veámoslo con un ejemplo.

La conjugación verbal en español presenta una considerable complejidad: por una parte, existen diferentes paradigmas de conjugación, formas compuestas, numerosos tiempos, varios modos, desinencias de diversos tipos, etc.; y, por otro lado, nos encontramos distintas clases de irregularidades, algunas relacionadas con cambios en la raíz, otros con las partículas flexivas (1).

Todo ello complica el tratamiento que deba darse del fenómeno en el nivel morfológico del procesamiento de la lengua, en nuestro caso, el nivel EMS (EUROTRA MORPHOLOGICAL STRUCTURE) del Proyecto EUROTRA de traducción por ordenador (2). En EMS, nos movemos en los límites de la palabra (3), por lo que los tiempos compuestos se formarán en el siguiente nivel (ECS -EUROTRA CONFIGURATIONAL STRUCTURE-) mediante la combinación de los elementos correspondientes. El motivo que nos impide tratar en EMS los tiempos compuestos como palabras únicas es la intercalación de elementos entre el auxiliar y el participio, caso poco frecuente en español, pero posible. Si bien nos parece preferible que los elementos que acompañen al verbo (e.g., los adverbios) respeten la continuidad entre ambas formas, tendremos en cuenta el caso contrario por la posibilidad de su aparición en los textos. De modo que, de momento, sólo nos ocuparemos del análisis de las formas simples.

El verbo en español puede segmentarse, desde un enfoque meramente teórico, de acuerdo con el siguiente esquema:

| 1 | 2 | 3 | 4 |
|-------|----------------|----------------|----------------|
| Raíz | Vocal temática | Tiempo-Aspecto | Número-Persona |
| com- | -e- | -ría- | -mos |
| tem- | -i- | -ste- | -is |
| part- | -i- | -ó- | Ø |

Hay que puntualizar que lo que hacemos realmente no es segmentar. EUROTRA no toma como unidad la palabra, sino el carácter. Los caracteres se agrupan registrándose en los niveles básicos las cadenas legales reconocidas por el prototipo y registradas en el diccionario. Lo que hacemos en la etapa de análisis no es, pues, segmentar, sino establecer los mecanismos de combinación de esas cadenas aceptadas previamente. Éstas, dicho sea de paso, no siempre coincidirán con morfemas en el sentido clásico. El procedimiento que emplearemos para decidir qué cadenas son legales será la segmentación, pero lo que el prototipo hace no es segmentar sino agrupar cadenas para formar palabras.

En cuanto al análisis del verbo, es lógico que tendremos que aislar siempre la raíz, puesto que varía en cada caso. Es la parte de la palabra que aporta la información semántica.

En cuanto a los componentes 2, 3 y 4, no existe un único alomorfo y, en ocasiones, la alomorfia se produce incluso en la

flexión de un mismo tiempo dentro de un mismo paradigma de conjugación.

Uno de esos alomorfos es [Ø]. El problema es que, al menos en el análisis, no podemos tratar lo que no esté gráficamente representado en el texto que se introduce para ser traducido. Por ello, ninguno de los componentes de una palabra podrá ser un elemento cero ni siquiera en alternancia con otro alomorfo. Ello nos obliga a alejarnos de la descripción gramatical ortodoxa para conseguir una formalización implementable.

En la generación sí podríamos introducir elementos cero, incluyéndolos como conjuntos de rasgos en el diccionario como los demás elementos del mismo y borrándolos en alguno de los niveles básicos de la generación, antes de llegar al texto en la lengua meta.

En el modelo que proponemos, hay distintos tipos de desinencias. En las formas no personales, distinguimos lo que habitualmente se llama vocal temática de la partícula característica de cada forma (-r, para el infinitivo; -ndo, para el gerundio; -do, para el participio). Este análisis se ajustaría aproximadamente a la descripción teórica habitual:

```
:b:
regla_nopers = {unid_léx=X, cat=palabra, clas_pal=v,
                forma_hipercategórica=Y, conjugado=no}
                [{unid_léx=X, cat=raíz, clas_pal=v,
                  conjugación=C},
                 {cat=voc_tem, conjugación=C},
                 {cat=des, forma_hipercategórica=Y}].

:l:

cantar = {unid_léx=cantar, cat=raíz, clas_pal=v, cadena=cant,
          conjugación=regl}.
r = {unid_léx=r, cat=des, cadena=r,
     forma_hipercategórica=infinitivo}.
ndo = {unid_léx=ndo, cat=des, cadena=ndo,
       forma_hipercategórica=gerundio}.
do = {unid_léx=do, cat=des, cadena=do,
      forma_hipercategórica=participio}.
a = {unid_léx=a, cat=voc_tem, cadena=a, conjugación=1}.
```

En EUROTRA, como en otros sistemas, la información se distribuye, dentro de cada nivel, entre el diccionario o lexicón y la gramática. Existen también traductores de unos niveles a otros, pero no nos ocuparemos de ello en esta ocasión. En el interior de la gramática, las reglas más importantes son las **reglas de construcción de estructuras**, también llamadas reglas :b:. Estas reglas construyen estructuras y en el análisis amplían los árboles hacia arriba desde los niveles inferiores. Tales árboles agrupan todos los elementos que componen el texto (lingüísticos y de formato) según relaciones de dominancia y precedencia. Asimismo, recogen en cada nivel la información que suministra la aplicación de las distintas reglas. En este caso, la combinación de cadenas básicas formará las palabras añadiéndoles la información morfológica pertinente (número, persona, etc.). En el árbol, las palabras resultantes dominarán a las cadenas básicas.

En la formalización correspondiente, el primer objeto entre llaves es la cabeza o nodo resultante de la aplicación de la regla. A continuación del corchete, encontramos los nodos hijos a partir de los cuales se ha formado el nodo madre. Los hijos son elementos léxicos que deben contrastarse en el diccionario, formado por las reglas :l:, reescritas a continuación de las

reglas de construcción de estructuras.

En este caso, por la aplicación de la regla, obtendríamos cualquiera de las formas no personales del verbo **cantar** con la información correspondiente (unidad léxica a la que pertenece, clase de palabra, etc.). Las variables (expresadas por medio de mayúsculas) sirven para hacer coincidir los valores de pares de distintos nodos: así, la raíz y la vocal temática deben corresponder a la misma conjugación.

Para las formas personales (siempre en el ámbito de los tiempos simples), el reconocimiento de componentes no se corresponde tan estrictamente con la Gramática Teórica. La regla sería la siguiente:

```
:b:
regla_tiempo = {unid_léx=X, cat=palabra, clas_pal=v, tiempo=Z,
                persona=P, num=N, modo=J, conjugado=sí}
                [{unid_léx=X, cat=raíz, clas_pal=v, conjugación=C}
                 {{{cat=des, conjugación=C, par_inf=sí},
                  {cat=des, conjugación=C, par_inf=sí,
                    tiempo=Z, persona=P, num=N, modo=J}}};
                 {{cat=des, conjugación=C, persona=P, num=N,
                   tiempo=Z, modo=J, par_inf=no}}).
```

```
:l:
cantar = {unid_léx=cantar, cat=raíz, clas_pal=v,
          conjugación=regl}.
ar = {unid_léx=ar, cat=des, cadena=ar;as;ás;ár, conjugación=1,
      par_inf=sí}.
a = {unid_léx=a, cat=des, persona=1, num=sing, tiempo=imperfecto,
     modo=subjuntivo, par_inf=sí}.
```

Con las reglas anteriores, podemos construir la primera persona del imperfecto de subjuntivo del verbo 'cantar', pues los objetos léxicos de las reglas :l: se unifican con los nodos hijos de la regla :b:.

La forma verbal conjugada se obtiene del proceso que permite reconocer, en primer lugar, la raíz. No habría inconveniente para que una raíz que tuviera más de un alomorfo, se ajustara a la regla, con tal de que el verbo fuera regular en cuanto al paradigma flexivo. En segundo lugar, podemos encontrar alternativamente la presencia de una cadena característica del futuro y condicional de indicativo, e imperfecto de subjuntivo, seguida de las correspondientes desinencias. Dicha cadena estaría formada por la llamada vocal temática y parte de la marca de tiempo-aspecto. El rasgo 'par_inf=sí' es necesario para que sólo se combinen con la secuencia de futuro-condicional-imperfecto las desinencias correspondientes. Las demás formas se componen con las desinencias marcadas con 'inf_par=no'.

De este modo, el reconocimiento y combinación de los formantes permite crear la correspondiente forma verbal simple.

Otra tarea importante es el tratamiento de los verbos irregulares, aunque, para no extendernos en formalismos, no haremos más que indicar su planteamiento. Algunas de las irregularidades afectan sólo a la alomorfía de la raíz. Generalmente, se deben a razones de tipo fonológico (volver, vuelva; herir, hiera, por ejemplo). Estos cambios se solucionan, como queda dicho, señalando en el diccionario las posibles variantes mediante la notación de alternancia: ;. En todo lo demás, estos verbos podrían ajustarse al paradigma verbal que les corresponda.

Por ejemplo, el verbo **dormir** tiene varios alomorfos para la raíz, pero su comportamiento flexivo es regular:

PRESENTE DE INDICATIVO

duerm - o - Ø - Ø
duerm - e - Ø - s
duerm - e - Ø - Ø
dorm - i - Ø - mos
dorm - Ø - Ø - is
duerm - e - Ø - n

(4)

Otro grupo de verbos adopta modelos flexivos distintos de las tres conjugaciones conocidas tradicionalmente. A veces, esos modelos son relativamente rentables y aparecen en más de un caso aislado de la lengua. Una posibilidad es incluir tales grupos de desinencias en la gramática como paradigmas flexivos. En los casos extremadamente infrecuentes, podrían incluirse todas las formas conjugadas en el diccionario con la información gramatical ya incorporada desde el lexicón. Pero en el momento de redactar estas líneas, aún no se ha adoptado ninguna decisión definitiva al respecto.

Conclusión

Una primera conclusión que podemos extraer de esta presentación es la distancia que media a veces entre las reglas formuladas por el lingüista teórico y por quien necesita escribir una gramática implementable en un lenguaje de programación. Estos resultados pueden extrañar e incluso escandalizar a quienes estudian la lengua desde una óptica quizá más ortodoxa, pero, sin duda y por desgracia, menos aplicable.

Por ello, la segunda conclusión, evidente para todos los que trabajamos en tareas de Lingüística Informática, pero no demasiado habitual en la formación universitaria, es la imperiosa necesidad de completar los estudios teórico-descriptivos de la lengua con una formalización precisa de los datos obtenidos. Las diversas experiencias en el terreno no hacen sino apoyar esta afirmación.

NOTAS

(1) Piénsese en la diferencia que existe entre la conjugación verbal de una lengua romance como el español frente a la sencillez del modelo inglés, pongamos por caso.

Para una visión detallada de la cuestión, pueden consultarse las numerosas gramáticas de que disponemos, en particular, las que mencionamos en las referencias bibliográficas.

(2) Véase Arnold y des Tombe (1987).

(3) En este caso, nos referimos a palabra como secuencia de texto separada en la escritura por dos blancos. Si bien desde el punto teórico esto puede ser discutible, para el tratamiento de la lengua escrita es una definición idónea.

(4) Me ajusto a la segmentación propuesta por Stockwell, Bowen y Martín (1965).

BIBLIOGRAFIA

- Alcina Franch, J. y Blecua, J. M. (1975): **Gramática española**. Esplugas de Llobregat, Ariel.
- Alsina, R. (1976): **Todos los verbos castellanos conjugados**. Barcelona, Teide.
- Arnold, D. y Tombe, L. des (1987): "Basic theory and methodology in Eurotra", en S. Nirenburg, **Machine Translation**. Cambridge, Cambridge University Press.
- Marcos Marín, F. (1984): **Curso de gramática española**. Madrid, Cincel.
- R. A. E. (1977): **Esbozo de una nueva gramática de la lengua española**. Madrid, Espasa-Calpe.
- Stockwell, R., Bowen, J. y Martin, J. (1966): **The Grammatical Structures of English and Spanish**. Chicago, The University of Chicago Press.