

UN TRATAMIENTO FUNCIONAL DE LA SÍNTESIS (EN TRADUCCIÓN AUTOMÁTICA)

J. Carlos Ruiz Antón

Sección de Lingüística General
Universitat Jaume I (Castelló)
Email: ruiz@fil.uji.es

1. Resumen

Este artículo investiga las posibilidades de emplear las ideas de la *Functional Grammar* (FG) como base de un generador de oraciones (tal como se entiende, por ejemplo, en traducción automática). Este enfoque combina las líneas maestras de la FG (esquemas funcionales de orden, reglas de colocación y reglas de expresión, fundamentalmente) con el procedimiento de unificación, de amplio uso en lingüística computacional. El sistema resultante es bastante más declarativo y monotónico que otras implementaciones de la FG.

El generador en cuestión (que llamaremos GFU: *Generador Funcional de Unificación*) se encuentra integrado en un sistema experimental de traducción automática, diseñado y utilizado como herramienta didáctica para cursos de lingüística computacional. El sistema utiliza una versión de LFG como analizador, y un sistema simple de traducción de rasgos en la fase de transferencia.

Esta comunicación describe en detalle el generador. La sección 2 traza una breve descripción de la teoría de la FG, especialmente las reglas y principios que emplea para producir la forma superficial de las oraciones. La sección 3 expone algunos detalles de la implementación de GFU, cuyo funcionamiento se examina con un ejemplo en la sección 4.

2. El modelo de la FG

La FG es una teoría funcional de la sintaxis y la semántica de las lenguas naturales. En esta sección describiremos únicamente los aspectos del modelo que conciernen a la generación de oraciones; para un conocimiento más completo el lector deberá consultar la bibliografía especializada.

La FG representa la semántica de las oraciones mediante fórmulas predicativas bastante abstractas, donde los participantes en la acción (argumentos y adjuntos) están señalados por su función semántica y pragmática. Por ejemplo, la representación de la frase *el agricultor plantó las zanahorias* sería:

$$(1) [\text{Past } e_i \text{ plantar}_v (d \ 1 : \text{agricultor}_N(x_1))_{\text{AgTop}} (d \ m \ x_2 : \text{zanahoria}_N(x_2))_{\text{GoFoc}} (e_i)]$$

Esta predicación (algo simplificada) describe un acontecimiento e_i situado en tiempo pasado (Past). Este acontecimiento es una acción descrita por el predicado verbal *plantar*, con un agente (Ag) y un objetivo (Go). El agente es un individuo x_1 , definido (d), perteneciente a la clase nominal *agricultor*, mientras que el objetivo es un conjunto múltiple (m) y definido de individuos de la clase nominal *zanahoria*. Obsérvese que la fórmula especifica también la función pragmática de los términos: el agente es también el tópico, y el objetivo está marcado como foco.

Como puede verse, las fórmulas predicativas se componen fundamentalmente de predicados (*plantar*, *agricultor*, *patata*) y de operadores (temporales, cuantificadores, de definitud, etc.).

La proyección entre estas fórmulas predicativas y la forma superficial de las oraciones tiene lugar en una serie de etapas:

- Asignación de funciones sintácticas (sujeto y objeto) entre los constituyentes de la fórmula.
- Aplicación de una serie de reglas de expresión, que especifican las formas que adoptan los constituyentes, con arreglo a su función sintáctica y a los operadores presentes en la fórmula. En la práctica estas reglas se ocupan de insertar elementos léxicos que corresponden a operadores, y de tratar morfológicamente las palabras de la oración.
- Ordenación de los constituyentes resultantes. El ordenamiento se obtiene por la aplicación de unas reglas de colocación (*placement rules*) que asignan a los constituyentes una cierta posición dentro de un esquema funcional de ordenamiento (*functional pattern*).
- Procedimientos morfológicos de *sandhi*, que dependen del orden relativo de los elementos de una oración. (por ejemplo, el que en inglés selecciona la forma *an* del artículo indeterminado cuando la palabra siguiente comienza por vocal).

3. GFU como generador de índole funcional

3.1 Representaciones de entrada

Las representaciones de contenido de las que parte el generador GFU no se corresponden estrictamente con las fórmulas predicativas de la FG descritas en la sección anterior. En su lugar, el sistema utiliza estructuras funcionales, bastante similares a las que se emplean en formalismos de unificación como LFG (v. Bresnan 1982) o FUG (v. Kay 1985). Estas estructuras son producidas por un *parser* descrito en Ruiz, Abaitua y Zubizarreta (1991).

Las estructuras funcionales son conjuntos de pares atributo-valor (*rasgos*), donde la posibilidad de disponer de valores complejos permite prescindir (con ventaja) de los clásicos árboles de estructura de frase. Además, las estructuras funcionales pueden transmitir, *grosso modo*, la misma información que los esquemas predicativos de la FG. Véase, como ejemplo, la estructura funcional correspondiente a la fórmula predicativa (1):

(2) *El agricultor plantó las zanahorias*

```
[ pred=plantar,
  voz=activa,
  tiempo=pasado,
  A1 = [ pred=agricultor,
        def=+,
        num=sing,
        prag=top ],
  A2 = [ pred=zanahoria,
        def=+
        num=pl,
        prag=foco ]
]
```

Los operadores están representados aquí como rasgos simples, mientras que los términos de la predicación aparecen como rasgos complejos. En lugar de los papeles semánticos característicos de la FG (agente, objetivo, destinatario, etc) hemos preferido utilizar etiquetas semánticas más abstractas: primer argumento (A₁) y segundo argumento (A₂). Muchas reglas gramaticales se pueden formular de manera más simple en términos de estas funciones abstractas que mediante papeles semánticos.

3.2 Arquitectura de GFU

GFU esta formado por una serie de módulos, tal como muestra el siguiente cuadro:

1. Asignación de función sintáctica
2. Inserción léxica
3. Ordenamiento de constituyentes
4. Morfología
 - a. Flexión
 - b. Sandhi

Se observará que la estructura de GFU se diferencia de la estrategia típica de la FG (v. *supra*) en que las reglas de expresión se encuentran desdobladas ahora en dos tipos de reglas: reglas de inserción léxica (que se aplican antes de las reglas de colocación de constituyentes) y reglas que tratan la flexión de las palabras (que intervienen después). De esta forma las reglas flexivas y las reglas de *sandhi* se pueden agrupar, de forma más natural, en el mismo módulo morfológico.

GFU es un sistema fundamentalmente declarativo: sus reglas se escriben en un formato especial, y se convierten automáticamente en términos de PROLOG cuando se cargan las gramáticas. De esta forma se refuerza la modularidad del sistema, al mantener separados el conocimiento lingüístico (expresado en diversos tipos de reglas), y los algoritmos de generación, que están invisibles al usuario. Esta es una diferencia substancial con otras implementaciones de la FG: así, las reglas del sistema ProfGlot (Dik 1992) están escritas directamente como programas de PROLOG.

3.3 Tipos de reglas en GFU

Al margen de las reglas morfológicas (que por falta de espacio no abordaremos aquí), los tipos de reglas de que se sirve GFU son las siguientes:

A. Reglas de asignación de función sintáctica

Estas reglas determinan los constituyentes que ocupan las funciones sintácticas de sujeto y objeto en cada cláusula. Así, la regla (3a) selecciona el segundo argumento como sujeto si la estructura funcional contiene el rasgo 'pasiva'. En caso contrario, la regla (3b) opta por el primer argumento.

- (3)
- a. $A_2 > \text{subj} / [\text{voz}=\text{pasiva}]$.
 - b. $A_1 > \text{subj}$.

B. Reglas de inserción léxica

El cometido de estas reglas es añadir elementos léxicos dada la presencia de determinados rasgos gramaticales o semánticos en la estructura funcional. Por ejemplo, en (4a), si la estructura de rasgos de entrada contiene polaridad negativa, la posición Neg se ocupa con la partícula *not*, insertándose al mismo tiempo un auxiliar *do*. La regla (4b) también es necesaria: se limita a insertar Neg cuando ya existe un auxiliar en la cláusula. Por su parte, (4c) lexicaliza el rasgo 'pasiva' como el verbo auxiliar *be*, que rige forma de participio pasivo (PAP) en el verbo siguiente.

- (4)
- a. $\text{pol}=\text{neg} \rightarrow [\text{aux}_1=\text{v}:\text{do}+\text{INF}, \text{Neg}=\text{not}]$.
 - b. $\text{pol}=\text{neg} \rightarrow [\text{Neg}=\text{not}]$.
 - c. $\text{voz}=\text{pasiva} \rightarrow [\text{aux}=\text{v}:\text{be}+\text{PAP}]$.

C. Reglas de colocación

Las gramáticas de GFU deben especificar los esquemas funcionales para cada dominio sintáctico (cláusula, sintagma nominal, etc.) y las reglas de colocación correspondientes. Por ejemplo, (5) es un esquema funcional (simplificado) que detalla la estructura de la cláusula en inglés:

(5) R P₁ Adv₀ V_f Neg S Adv₁ V_i CL Adv₂ O MOD SUB

Donde R es la posición de los relatores, como *that*, *because*, etc; P₁ es la posición inicial, reservada en la FG para constituyentes de especial relevancia pragmática, como por ejemplo, el tópico (v. Dik 1989: §17.2); Adv₀-Adv₂ son diversas posiciones adverbiales; V_f es la posición de los verbos en forma personal (*finite verbs*); Neg es la partícula negativa (*not*); S es sujeto; V_i es para los verbos en forma no personal; CL corresponde a los clíticos pronominales; O es la posición del objeto; MOD se reserva para otros complementos y adjuntos; SUB es la posición normal de las cláusulas subordinadas complejas (completivas y adverbiales).

En el esquema, cada posición sólo puede ser ocupada por un constituyente, excepto V_f y MOD, que admiten más de un ocupante.

Las reglas de colocación utilizadas por el generador son de la forma A → B, donde A se refiere a un constituyente, y B indica la posición que le corresponde dentro del esquema del dominio. La especificación del constituyente puede incluir condiciones que restrinjan la aplicabilidad de la regla. Véanse, por ejemplo, algunas reglas para la cláusula del inglés:

- (6) a. X = sn[qmor='Q'] → P₁.
 b. X = sn[pragm=top] → P₁.
 c. subj → P1.
 d. subj → S.
 e. arg1 → MOD=sn[prep=p:by].
 f. arg2= sn[pro=clit] → CL.
 g. arg2 → O.
 h. arg3 → MOD.
 i. adj → MOD.

Algunas de estas reglas necesitan una aclaración: las reglas (6a-b) colocan en la posición P₁ los constituyentes interrogativos (con el rasgo qmor='Q') y los sintagmas nominales con la función pragmática de tópico. La regla (6e) sitúa en MOD el primer argumento de las cláusulas pasivas (y además le añade la preposición *by*). Por su parte, (6f) coloca los clíticos en la posición CL.

Para completar esta visión global de las reglas de colocación en inglés, añadiremos varias de las reglas que sitúan constituyentes en las posiciones verbales y adverbiales (más abajo se verá su uso en un ejemplo).

- (7) a. aux → V_f.
 b. pred → V_f.

- c. pred \rightarrow V₁.
- d. freq \rightarrow Adv₁ & [V₁=aux].
- e. freq \rightarrow Adv₀.

Estas reglas sitúan el auxiliar en la posición de V₁ (7a), y al predicado, alternativamente, en las posiciones V₁ y V_i (7b-c). El adverbio de frecuencia va a Adv₁ si la estructura funcional ya contiene un auxiliar (7d), o a Adv₀ en los demás casos (7e).

Un importante problema que afecta a la implementación del generador es determinar la secuencia de aplicación de las reglas de ordenamiento. El método preferido en varias de las implementaciones de la FG (Connolly 1983, 1991; Bakker 1989; Dik 1992) es una ordenación no estricta siguiendo de izquierda a derecha las expectativas del esquema. De seguir esta estrategia, en nuestro esquema de cláusula se aplicarían primero las reglas que colocan los relatores (R), luego las reglas que sitúan constituyentes en P₁, y así sucesivamente.

Nuestra opción es distinta: consiste en aplicar las reglas de colocación a partir de los rasgos, según el orden (irrelevante) en que aparezcan en su estructura funcional. Esta aplicación producirá una estructura de rasgos parcial, que deberá unificar con el producto de la aplicación sucesiva de las reglas de colocación sobre los demás rasgos de la estructura. De esta manera, la unificación impide que una posición pueda estar ocupada por más de un constituyente, con excepción de las posiciones abiertas, como MOD o V_i (v. *supra*).

4. Un ejemplo

Veamos ahora un ejemplo de cómo tiene lugar la generación de una frase en GFU. Tomaremos como punto de partida la estructura funcional correspondiente a la oración pasiva *El discurso fue interrumpido a menudo*:

```
(8)
[ pred=interrumpir,
  voz=pasiva,
  tiempo=pasado,
  A2= [ pred=discurso,
        def=+,
        num=sing,
        prag=top ],
        freq = [ pred= a-menudo ]
  ]
```

Esta es la estructura producida por el *parser* de castellano. Su transferencia a inglés es directa, dando lugar a la siguiente estructura funcional:

```
(9)
[ pred=v:interrupt,
  voz=pasiva,
  tiempo=pasado,
  A2= [ pred=n:speech,
        def=+,
        num=sing,
        prag=top ],
        freq = [ pred= adv:often ]
  ]
```

Esta estructura funcional queda ligeramente modificada al asignar la función sintáctica de sujeto al término A_j . (cf. la regla 3a).

A continuación entran en liza las reglas de inserción, que insertan el auxiliar *be* por la presencia del rasgo 'pasiva' (cf. regla 4c), y el artículo *the* dada la presencia del rasgo 'definido':

- (10) [pred=v:interrupt,
voz=pasiva,
aux = be+PAP
tiempo=pasado,
subj= [pred=n:speech,
def=+,
det=the
num=sing,
pragm=top],
freq = [pred=adv:often]
]

A partir de este punto intervienen las reglas de colocación, activando el esquema de cláusula. Según éste, el rasgo 'pred' puede ocupar (de acuerdo con las reglas 7b-c) las posiciones V_f y V_i . Por el momento recibe V_f y la estructura parcial resultante es

- (11) [$V_f = v:interrupt$]

Sin embargo, el rasgo 'aux' sólo puede ocupar la posición V_f , lo que se contradice con esta estructura parcial. Por tanto, se impone probar la otra alternativa de colocación de 'pred' (V_i). El resultado es entonces:

- (12) [$V_f = aux:be+PAP,$
 $V_i = v:interrupt$]

Le llega ahora el turno al rasgo temporal 'pasado'. Como no hay declarada ninguna regla que lo procese, se ignora, y se continúa con el rasgo 'subj'. Las reglas (6c) y (6d) prevén las posiciones P1 y S (en este orden). Aceptando por el momento la alternativa P1 (y pasando por alto la trivial linealización del sintagma nominal) tenemos ahora

- (13) [$V_f = aux:be+PAP,$
 $V_i = v:interrupt,$
 $P_1 = [det=the,$
 head = n:speech]
]

Ya únicamente resta por situar el rasgo 'freq'. En este caso se aplica la regla (7d) que lo coloca en Adv_1 si el verbo que ocupa V_f es un auxiliar (en otro caso, la posición hubiera sido Adv_0 , según 7e). Así, la estructura resultante es

- (14) [$V_f = aux:be+PAP,$
 $V_i = v:interrupt,$
 $P_1 = [det=the,$
 head=n:speech],
 $Adv_1 = [pred=adv:often]$
]

Ordenando estas posiciones de acuerdo con el esquema (5) (esto es: $P_1 V_f Adv_1 V_i$) y aplicando las reglas morfológicas se obtiene la secuencia *The speech was often interrupted*.

5. Conclusión

Hemos esbozado un nuevo concepto de generación que aprovecha las posibilidades de las reglas de orden de la FG. Este enfoque hace uso de esquemas funcionales de orden y de reglas de colocación, y nos parece preferible (tanto desde un punto de vista teórico como metodológico) a otros tratamientos basados en series de transformaciones (como por ejemplo MacCord 1989), ya que facilita un tratamiento declarativo y fácilmente extensible de las gramáticas de generación.

Referencias

- BAKKER, DIK (1989) «A formalism for functional grammar expression rules». En Connolly, J.H. y S. Dik (eds.) *Functional grammar and the computer*. Dordrecht, Foris, págs. 45-63.
- BRESNAN, JOAN (ed.) (1982) *The mental representation of grammatical relations*. Cambridge (Mass.), MIT Press.
- CONNOLLY, JOHN H. (1983) «Placement rules and syntactic templates». En S. C. Dik (ed.) *Advances in functional grammar*, Dordrecht: Foris, págs. 247-266.
- CONNOLLY, JOHN H. (1991) *Constituent order in functional grammar: synchronic and diachronic perspectives*. Dordrecht, Foris.
- DAHL, VICTORIA (1985) «Hiding complexity from the casual writer of parsers». En Dahl, V. y P. Saint-Dizier (eds.) *Natural language understanding and logic programming*. (Vol. I), Elsevier, págs. 1-19.
- DIK, SIMON (1989) *The theory of functional grammar. part I: the structure of the clause*. Dordrecht, Foris.
- DIK, SIMON (1991) «Functional grammar». En F. G. Droste y J. E. Joseph (eds.) *Linguistic theory and grammatical description*. Amsterdam, John Benjamins, págs. 247-274.
- DIK, SIMON (1992) *Functional grammar in PROLOG*. Berlín/Nueva York, Mouton de Gruyter.
- HUTCHINS, W. JOHN y HAROLD SOMERS (1992) *An introduction to machine translation*. Londres, Academic Press.
- KAY, MARTIN (1985) «Functional unification grammar: a formalism for machine translation». *COLING-84*, Stanford University (California), págs. 75-78.
- MCCORD, MICHAEL C. (1989): «Design of LMT: a Prolog-based machine translation system». *Computational Linguistics* 15.1, págs. 33-52.
- PERLMUTTER, DAVID (1981) «Functional grammar and relational grammar: points of convergence and divergence». En T. Hoekstra, H. van der Hulst y M. Moortgat (eds.) *Perspectives on Functional Grammar*. Dordrecht, Foris, págs. 319-352
- ROHRER, CHRISTIAN (1986) «Linguistic bases for machine translation». *11th International Conf. on Computational Linguistics (COLING)*, Bonn, págs. 353-355
- RUIZ ANTÓN, JUAN C., JOSEBA ABAITUA y J. RAMÓN ZUBIZARRETA (1991) «Un compilador de LFG y su aplicación al euskera». *Procesamiento del Lenguaje Natural* 9, págs. 177-191
- SHIEBER, STUART (1986) *An introduction to unification-based approaches to grammar*. Center for the Study of Language and Information. Stanford University.
- SIEWIERSKA, ANNA (1991) *Functional grammar*. Londres, Routledge
- TUCKER, ALAN B. (1987) «Current strategies in machine translation research and development». En Nirenburg, S. (ed.) *Machine translation: theoretical and methodological issues*. Cambridge University Press, págs. 22-41