

Slash y los fenómenos de extracción múltiple del español

J.G.L.Escribano
Universidad de Oviedo

Resumen

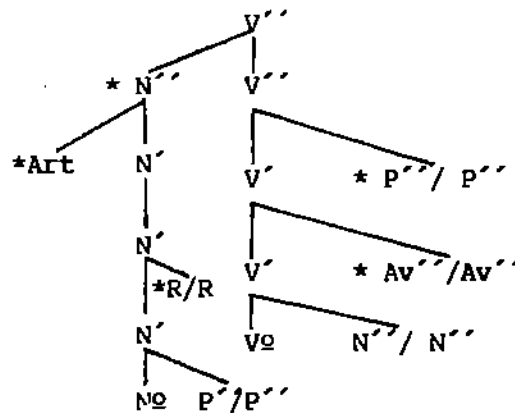
Partiendo de la observación de la existencia en español de casos de extracción múltiple con dependencias en muchos casos entrecruzadas se intenta demostrar que, en su formulación actual, el mecanismo SLASH, junto con los principios HFC,FFP,CAP, STM 1, etc, que rigen su funcionamiento, no basta para computarlas eficientemente, y asimismo que ninguna de las modificaciones sugeridas por Gazdar et al. es técnicamente viable como solución a estos casos, que, asumiendo los presupuestos actuales de la GPSG, resultarían intratables salvo que se recurriera a procedimientos ad hoc de identificación de antecedentes y "gaps" cuyo coste computacional sería elevado y convertiría a la GPSG en un sistema no asimilable a las gramáticas CF. Dada la complejidad y redundancia que el tratamiento en términos de SLASH produce en la gramática y su impotencia ante ese tipo de fenómenos, parece inevitable concluir que SLASH no es la respuesta al problema y la GPSG no ofrece una alternativa preferible al mecanismo antecedente-traza tan rigurosamente desarrollado en la teoría GB de Chomsky.

J.L.G. Escribano

Universidad de Oviedo

De la interacción del mecanismo SLASH con la meta-regla STM 1 y los principios HFC, FFP y CAP que rigen la instanciación de rasgos en la GPSG (cf. Gazdar et al. (1985), asp. cap. 5) se derivan dos predicciones cruciales: 1) que, en general, es imposible "extraer" (= instanciar SLASH en) los constituyentes no dominados por el X' que domina inmediatamente a X, salvo los casos de N''/N'' que por el principio de unificación asociado a FFP generan huecos parasitarios. En (1) ninguno de los nodos que llevan asterisco puede contener <SLASH,v> :

(1)



En efecto, si <SLASH> fuera instanciado en alguno de esos constituyentes FFP lo exigiría también en su nodo materno y HFC en el núcleo (H, en lo sucesivo) de éste, que a su vez lo transmitiría (por HFC) hacia abajo a todos los nodos frásicos dominados por H que tengan carácter de núcleo hasta llegar al H' situado inmediatamente encima de la categoría léxica |SUBCAT,n|, que por no ser frásica no puede instanciar <SLASH>. H' debería heredar entonces <SLASH,v> de alguno de sus constituyentes inmediatos, con lo cual se generaría bajo H' un hueco ilegítimo de tipo X/X, donde X debe ser un valor unificable con el asignado a <SLASH> en el nodo Y / X situado más arriba. Solapándose con el efecto de HFC y FFP, STM 1 actúa únicamente sobre reglas ID léxicas, que generan X' pero no sus especificadores, ni sus modificadores, con lo cual el valor <NULO> sólo aparece en los argumentos de X0 (con exclusión del sujeto) y en los posibles huecos parasitarios asociados a ellos. De acuerdo con HFC, FFP y STM 1, entonces, la predicción es que será imposible "extraer" ningún otro tipo de constituyentes.

La segunda predicción, 2), es que si <SLASH,v> es una función monádica y rigen CAP y el principio de unificación, debe ser imposible extraer más de un constituyente, o, equivalentemente, <SLASH> debe recibir un sólo argumento, por tanto una categoría

como *A/B,C,... es imposible por definición (cf. Maling & Zaenen (1982), Chung & McCloskey (1983)).

Esa concepción restrictiva del mecanismo SLASH asegura la aplicabilidad del CAP y brinda un atractivo, sencillo y elegante procedimiento de identificación de "gaps" sin necesidad de recurrir a reglas transformacionales como MOVE X. Asimismo da cuenta automáticamente de ciertos efectos ATB, nace superfluas ciertas restricciones inicialmente descubiertas por Ross (cf. Ross (1986)) como las denominadas "Coordinate Structure Constraint" y "Left Branch Condition", genera uno de los efectos ECP (o |COMP t|) más importante, y, en fin, permite prescindir también del denominado "WH Island Constraint" (o de una de las aplicaciones del Principio de Subyacencia), al impedir la extracción de cualquier elemento WH|+/-Q| de una cláusula que ya tenga otro de esos elementos en la posición COMP.

Sin embargo, las lenguas naturales ofrecen por doquier hechos que contradicen esas dos predicciones cruciales del mecanismo SLASH, los principios que rigen su funcionamiento y por tanto la esencia de la GPSG. En primer lugar, el tándem constituido por SLASH, HFC, FFP, CAP y STM 1 excluye como agramaticales múltiples posibilidades de extracción de constituyentes que se dan con toda naturalidad. En el sintagma nominal (N^{\max}), por ejemplo, es posible extraer, contra lo que predice la GPSG, frases preposicionales (P^{\max}) y cláusulas de relativo en casos de "extraposición" como los de (2):

- (2) a. Kiki traía un bañador el otro día de esos transparentes con el que nos puso a todos negros.
- b. Conocí a una chica hace un par de días en una discoteca que resultó ser inspectora de Hacienda.

Asimismo, en lo que respecta al V^{\max} , nada impide extraer todo tipo de complementos circunstanciales no subcategorizados, como prueban los ejemplos de (3):

- (3) a. ¿Cuándo crees tú que vendrá Paco a Gijón?
- b. Antes de finales de julio no creo que venga.
- c. ¿En qué fecha leyó Pepe su tesis?
- d. Antes del 80 no creo que la haya leído.
- e. ¿Por qué llegas tan tarde?
- f. ¿Dónde podría comprar un comecocos? ...etc.

Sin embargo, en cualquier análisis \bar{X} esos elementos corresponden a posiciones en las que STM 1 no actuaría y que, por los citados principios HFC, FFP, etc., rechazan SLASH y, según la teoría GPSG, deben ser inextraíbles.

En cuanto a la imposibilidad de extracciones múltiples que también predicen esos mismos principios junto con el criterio de unificabilidad y FCR 20 (cf. Gazdar et al. (1985)), ejemplos como los de (4-12) dejan bien a las claras que se trata de una predicción falsa:

- (4) De Gazdar, ¿cuándo crees tú que sería conveniente que hablásemos?
- (5) a. La única persona a quien nunca sé qué regalar es mi padre.
- b. Esas son cosas de las que no sé por qué he de ocuparme yo.

- (6) Juan el dinero a su mujer ahora; cómo se lo envía?
- (7) De "pársers", ¿con quién hablé yo hace poco de la politécnica que me dijo que estaba construyendo uno?
- (8) Ana puede tocar, y el pequeño tararear, las Suites Inglesas enteras de memoria sin fallar una nota.
- (9) Ana toma todos los días el café completamente frío.
- (10) Dí todavía ayer dinero al chico para sus gastos.
- (11) Mandó de muy malos modos a Pepe salir.
- (12) Por las mañanas viene Juan.

En (4) hay extracción evidente de un "tópico" y un elemento +QU de la misma cláusula; en (5) la doble extracción afecta a un relativizador y un interrogativo; en (6) se extraen varios tópicos y un adverbio +QU de modo; en (7) hay extracción múltiple y con dependencias entrecruzadas, ya que en el lado izquierdo ha habido extracción de un argumento de hablé pero, al mismo tiempo, de un segundo argumento +QU de ese mismo verbo, con quién..., asimismo extraído hacia la izquierda, se extraen hacia la derecha dos modificadores del N^{\max} cuyo núcleo es quién, una frase P^{\max} y una cláusula de relativo; en (8) parece inevitable postular no menos de cuatro extracciones, i.e., N^{\max} , A^{\max} , P^{\max} y otra P^{\max} ; en (9), (10) y (11), las reglas ID nos obligan a postular dos extracciones en cada una; otro tanto ocurre en (12) si se asume para el español una regla LP del tipo $N'' < V''$.

Gazdar et al. (1985) p.81 reconocen que el mecanismo SLASH podría necesitar modificaciones que implicarían un debilitamiento de la hipótesis inicial en caso de que fuera necesario computar dependencias múltiples como las descubiertas por Engdahl en las lenguas escandinavas. Sugieren que SLASH podría dar cuenta de esos fenómenos o bien admitiendo como valores listas de categorías, o posiblemente un "stack", o bien permitiendo argumentos del tipo A/B (i.e., configuraciones del tipo $X/|A/B|$) o, en fin, recurriendo a "some other modification". Es evidente, sin embargo, que tanto ellos como Klenk (1987) consideran esa "modificación" como algo trivial, como una simple cuestión "de ejecución" que no merece la pena desarrollar.

Sin embargo, qué solución se dé a ese problema dista mucho de ser una cuestión trivial, ya que de ella depende la viabilidad del enfoque SLASH de la GPSG ante una gran cantidad de fenómenos como los de arriba. Nótese que si SLASH no es capaz de dar cuenta de esas dependencias la tesis de monoestratalidad se derrumba inmediatamente, ECPO se convierte en un principio vacío, el formato ID+LP se revela incapaz de generar gramáticas explícitas y en consecuencia la esencia de la GPSG se desintegra.¹

Pues bien, mi opinión es que ninguna de las modificaciones tenuemente sugeridas por Gazdar et al. es en realidad técnicamente viable ante casos como los que nos conciernen. Recurrir a un "stack" que legitime categorías del tipo X/A-B-C... es, ante todo, conceptualmente repugnante si <SLASH> es una función monádica y el CAP ha de mantenerse (y, naturalmente, de esos presupuestos se deriva lo que de eficaz, sencillo y matemáticamente elegante tiene el mecanismo SLASH), pero además ni si-

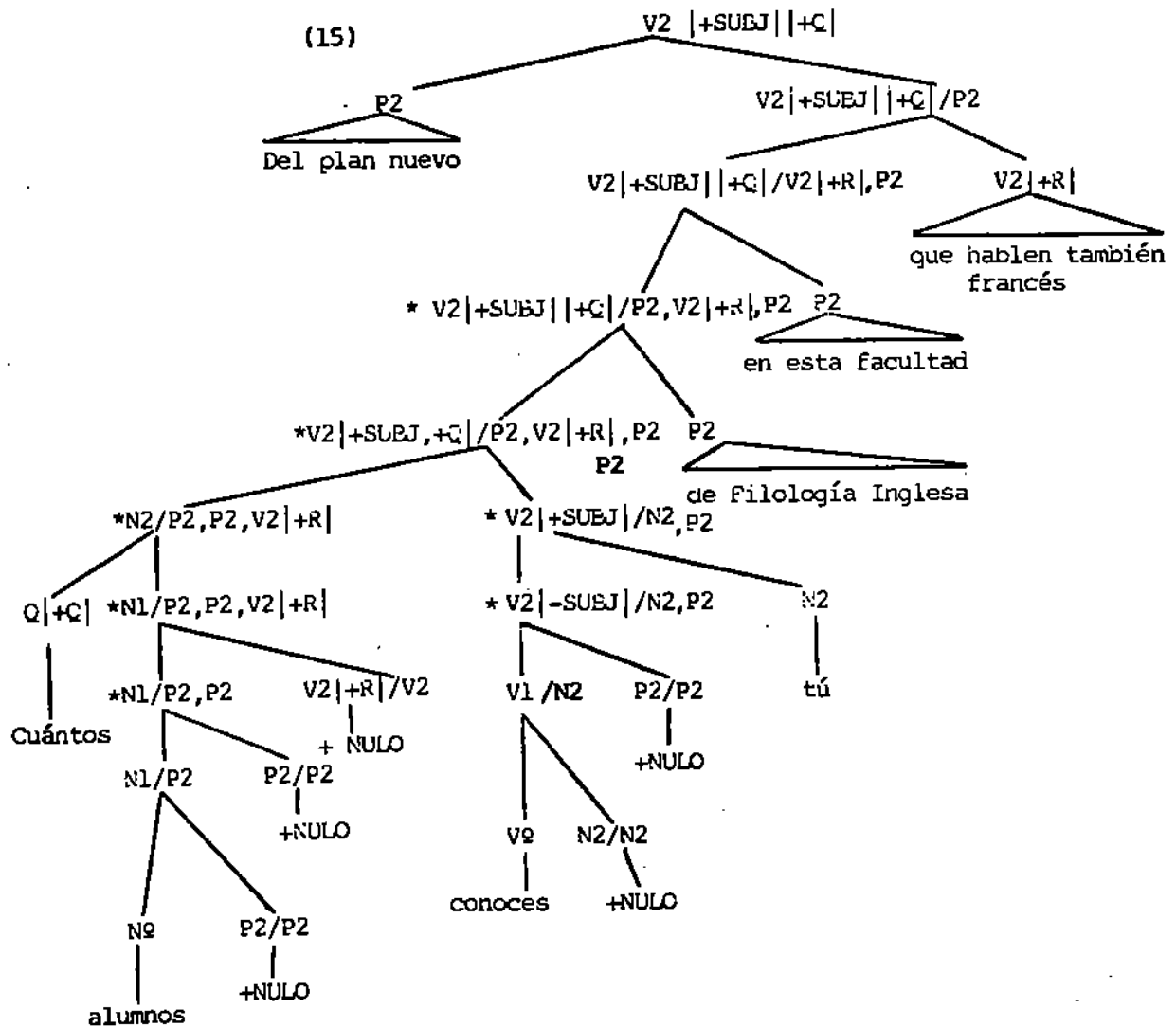
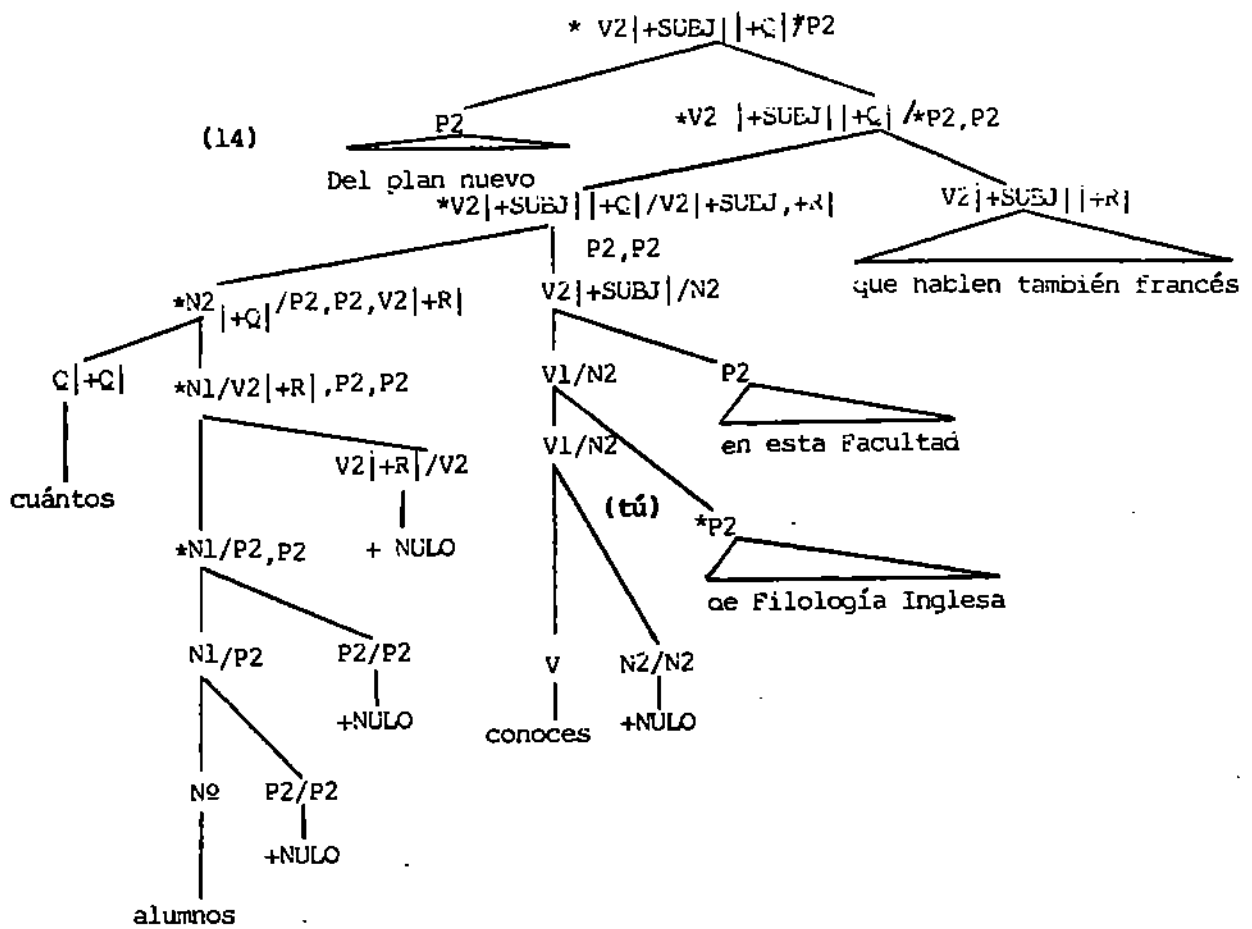
quiera parece técnicamente factible salvo de un modo extremadamente ad hoc y nada extrapolable. Piénsese en que un "stack" requiere un principio de ordenación (first in - first out, last in - first out, ...etc.). Sin embargo, en los casos de extracción que comentamos no hay ningún modo sintácticamente no arbitrario de ordenar los huecos acumulados en una categoría. Prácticamente en cualquiera de los ejemplos de (4-12) tan arbitrario es decir que "a un X al que le falta A le ha sido extraído además un B" ,i.e., $|X/A| / B$, como que "a un X al que faltaba B le ha sido extraído A". Pero, claro está, sin un principio estructural de ordenación, ese hipotético "stack" tendría que descansar en algún procedimiento de codesignación ad hoc de los huecos y sus antecedentes cuyo coste computacional es una incógnita y que en todo caso no resulta precisamente elegante ni esclarecedor en cuanto engranaje de una teoría de las lenguas naturales.

La segunda posibilidad sugerida por Gazdar et al. (loc.cit) y vislumbrada también por Piera (1985) p. 682, i.e., permitir que <SLASH> tome como argumentos categorías del tipo A/B, es conceptualmente aceptable dados los presupuestos de la GPSG, pero plantea problemas técnicos que la hacen inviable y en cualquier caso insuficiente. Los más obvios son los siguientes: a) ¿Cómo computar los huecos de una configuración $X/|A/B|$ cuando A y B son elementos no adyacentes y están dominados por constituyentes diferentes? b) ¿Cómo ordenar de un modo no arbitrario A, B, C, ... en una configuración $X/A, B, C, \dots$ cuando de un constituyente extraído se extraen a su vez, no uno, sino dos o más constituyentes no dominados por un solo nodo? c) ¿Cómo satisfacer el CAP y reducir un hueco B cuando el antecedente B extraído está en una posición tal que no "comanda" a A/B? d) ¿Por qué procedimiento puede asegurarse la correspondencia correcta uno a uno entre antecedentes y huecos homocategoriales múltiples? e) ¿Cómo evitar que sean unificados valores homocategoriales no unificables, con la consiguiente pérdida de argumentos y transgresión de los requisitos de interpretabilidad? f) ¿Cómo delimitar un conjunto finito de reglas $A \rightarrow A/B, C \quad B, C ?$, y, en fin, g) ¿Qué reglas semánticas van a interpretar en una gramática monoestratal las sorprendentes estructuras que una gramática de inspiración GPSG se vería obligada a asignar a casos como (5) o (7) supra?

Los fenómenos que uno encuentra a todas horas en el estudio de las lenguas naturales son lo suficientemente complejos como para plantear de hecho todos esos problemas, entre otros. A modo de demostración sucinta de que esos problemas no son artificiales analizaré tan sólo un ejemplo significativo, el de (13) (el lector puede comprobar fácilmente que cualquier ejemplo de (4-12) o similar plantea en mayor o menor grado dificultades del mismo tenor).

(13) Del Plan Nuevo, ¿a cuántos alumnos conoces tú de Filología Inglesa en esta Facultad que hablen también francés?

(13) tendría, extrapolando a partir de dos principios \bar{X} que Gazdar et al. parecen suscribir, una estructura como (14), o tal vez (15). Por supuesto ambas infringen HFC, FFP, CAP, FCR 20, STM 1, etc., al igual que las que corresponderían a los casos de (4-12) supra, pero aún ignorando el marco GPSG actual y aceptando las modificaciones de Gazdar et al. como conceptualmente posibles en principio, nótese que al



N^{\max} cuyo H es alumnos le han sido extraídos no menos de tres constituyentes que ni se dominan unos a otros ni son adyacentes (problemas a) y b) supra), y que el propio N^{\max} ha sido extraído a su vez. Por otra parte, si la estructura es (14), como parece razonable, uno de los elementos extraídos del N^{\max} , la P^{\max} de Filología Inglesa, ha "quedado atrás" bajo el V^{\max} , de modo que ni siquiera "comanda" al hueco P^{\max} correspondiente. ¿Cómo se satisfaría el CAP en un caso semejante? (Problema c) supra). Igualmente, ¿cómo determina el "parser" qué hueco P2 debe ser reducido por cada constituyente P2 que va apareciendo en posición de extracción? (problema d) supra). ¿Cómo "averigua" el "parser" que el N2 tú no debe reducir el hueco N2 que lleva $V2|-\text{SUBJ}|$ si no es recurriendo a marcarlo con algún rasgo ad hoc? ¿Cómo evitar que los tres valores P2 sean unificados y se rompa la correspondencia biunívoca entre los antecedentes y sus huecos? Todas esas son preguntas legítimas e inexcusables que deben ser planteadas a quien habla de posibles modificaciones de SLASH como simple cuestión de ejecución. Y aún cabe añadir otras, por ejemplo: ¿cuantos miles o millones de reglas del tipo $X/A, B, C \rightarrow X/A, B \dots C$ como las necesarias para generar (14) o (15) serían necesarias para generar/reconocer exhaustivamente un corpus irrestricto de una lengua natural? O también: ¿qué reglas semánticas tendrían el poder suficiente como para, ante una representación monoestratal como (14) o (15), evitar interpretar P2 Del Plan Nuevo, o P2 de Filología Inglesa, o incluso la cláusula de relativo, como modificadores de un sintagma equivocado? Naturalmente, en ausencia de una hipótesis detallada acerca de la naturaleza de las hipotéticas modificaciones, no es fácil probar concluyentemente la inviabilidad absoluta de un planteamiento. No obstante, sí parece razonable concluir que a) tal y como ha sido desarrollado conceptualmente el mecanismo SLASH de la GPSG hasta ahora, resulta manifiestamente inadecuado para computar dependencias como las que hemos visto; b) ninguna de las vías de modificación vagamente sugeridas por Gazdar et al. conduce en realidad a una solución técnicamente aceptable; y c) que si han de mantenerse los presupuestos que en la actualidad definen el planteamiento GPSG (monoestratalidad, etc.) la única estrategia que cabe entrever como "solución" a esos problemas consiste en recurrir al uso indiscriminado de "marcas" de identificación ad hoc, con lo cual la potencia expresiva de la GPSG se convertiría en indefinida, su coste computacional en incalculable y su capacidad de ofrecer explicaciones sustantivas para los fenómenos lingüísticos en nula. Ciertamente, no parece creíble que una gramática de esas características pueda asimilarse a un sistema CF eficientemente computable ni constituir una alternativa seria a las que están siendo desarrolladas en el marco de la teoría GB de Chomsky.

Notas.

1. En otro trabajo titulado "Una crítica de la GPSG de Gazdar et al." que he enviado para su publicación a la Revista Española de Lingüística he intentado demostrar con una argumentación detallada que es imposible reproducir aquí por razones de espacio esos y otros extremos que arrojan serias dudas sobre la sustantividad de la GPSG.

Referencias.

- Chung, S. & J. McCloskey (1983) : "On the interpretation of certain island facts in GPSG." Linguistic Inquiry, 14(4):704-713.
- Gazdar, G., et al. (1985) : Generalized Phrase Structure Grammar. Oxford: Blackwell.
- Klenk, U. (1987) : "Una gramática de estructura de frases generalizada de la oración española: el orden de los constituyentes mayores." Boletín de SEPLN, 5:106-110.
- Maling, J. & A. Zaenen (1982) : "A phrase structure account of Scandinavian extraction phenomena." (En Jacobson & Pullum, eds. The Nature of Syntactic Representation. Dordrecht: Reidel)
- Piera, C. (1985) : "Gaps in gaps in GPSG". Linguistic Inquiry, 16(4): 681-683.
- Ross, J.R. [1967] (1986) : Infinite Syntax! (Constraints on Variables in Syntax). Norwood, N.J. : Ablex Publ.

#####

Una gramática de Afijo Extendida para el español

Jos Hallebeek
Katolieke Universiteit

Resumen

El formalismo de la gramática de Afijo Extendida (*Extended Affix Grammar*) fue desarrollado en los años setenta para la definición de lenguajes artificiales, de programación principalmente. Sin embargo, este tipo de gramática ha resultado ser adecuado también para la descripción de las lenguas naturales. Actualmente existen aplicaciones al inglés, al árabe estándar moderno y al español. La base de la gramática de Afijo extendida la constituye una gramática libre de contexto. Los símbolos no-terminales de sus reglas se complementan con una especie de parámetros llamados 'afijos', que a su vez se definen también mediante otra gramática libre de contexto. Así se forma en realidad un conjunto a dos niveles: la gramática de la base y la de los 'afijos'. Como las dos gramáticas libres de contexto se juntan y entrelazan, el nivel de los 'afijos' le concede al conjunto el poder de una gramática sensible al contexto.

A diferencia de otros formalismos empleados en análisis automatizados no se describen las reglas de la lengua directamente en un lenguaje de programación. Tienen la forma de reglas rescriturales que tendrán que ser convertidas automáticamente mediante un generador de analizador (*parse generator*) en analizador automático (*parser*). Esto significa que no se le exigen conocimientos de programación al lingüista que utiliza el formalismo de la Gramática de Afijo Extendida.

En la Universidad de Nimega un grupo de investigadores de los departamentos de Inglés, Árabe, Español e Informática se dedica entre otras cosas a componer Gramáticas de Afijo Extendida para cada una de las tres lenguas y a perfeccionar el *parser generator*. La gramática del español, que va a servir para el análisis sintáctico automatizado de un corpus representativo de textos españoles actuales de 500.000 palabras, está llegando a su conclusión. El correcto funcionamiento de cualquier fase intermedia de esta gramática es probada mediante la conversión en *parser* y el subsiguiente análisis automático de las series limitadas de enunciados. En la ponencia el responsable de la composición de la gramática se propone ofrecer una idea más pormenorizada del formalismo gramatical empleado, así como del entorno en que se desarrollan las actividades de nuestro grupo de investigación.

UNA GRAMÁTICA DE AFIJO EXTENDIDA PARA EL ESPAÑOL

1. La Gramática de Afijo Extendida (*Extended Affix Grammar*)

En la Informática, dos investigadores desarrollaron, independientemente el uno del otro, dos tipos de gramática fundamentalmente equivalentes llamados respectivamente *Attribute Grammar* (Gramática de Atributo), elaborada por Knuth (1968), y *Affix Grammar* (Gramática de Afijo), elaborada por Koster (1970). Las dos gramáticas son destinadas a la definición de lenguajes artificiales y empleadas como formalismo para escribir compiladores (*compilers*). Tienen como base una serie de reglas rescriturales que forman una gramática libre de contexto (GLC). A los símbolos no-terminales de estas reglas se añade cierto número de 'atributos' o 'afijos'. Cada uno de esos símbolos tiene un número fijo de ellos, es decir que cualquier ocurrencia del mismo símbolo ha de poseer idéntico número y tipos de 'afijos' o 'atributos'. Y también su dominio es fijo. Sirven para trasladar información de una parte del árbol sintáctico a otra. A diferencia de la Gramática de Atributo (que trabaja con listas de valores) la de Afijo formula otra GLC para generar los valores permitidos de los 'afijos' o 'atributos'.

Así el formalismo de la Gramática de Afijo funciona a dos niveles: un nivel en que se describe la GLC de la base y otro nivel en que se enumeran los valores legales de los 'afijos'. Estos no sólo se usan para pasar información de un lugar a otro sino también para especificar restricciones sensibles de contexto añadiendo reglas de evaluación y de restricción a las rescriturales de la gramática libre de contexto. Dado que es necesario explicitar las reglas de restricción y evaluación, escribir una Gramática de Afijo parece ser más bien tarea del informático o del programador que del lingüista. Este inconveniente fue solucionado por Watt (1974) al lanzar su *Extended Affix Grammar* (Gramática de Afijo Extendida). La extensión señalada en el nombre de la gramática consiste en la incorporación al formalismo de la condición inherente de que los 'afijos' del mismo nombre presentes en diferentes posiciones dentro de la misma regla tengan el mismo valor. De modo que si por ejemplo aparece usado el 'afijo' género más de una vez en la misma regla, necesariamente tendrá todas las veces el mismo valor: o masculino o femenino. Así se suprimen las reglas explícitas de evaluación y de restricción y la gramática obtiene el valor de una gramática sensible de contexto (GSC). Esto significa en comparación con una gramática libre de contexto una importante reducción del número de reglas. Además, Postal (1964) ya demostró que la GLC no sirve como instrumento para dar una definición completa de la sintaxis de una lengua natural.

2. Una Gramática de Afijo Extendida para la oración española

A fin de dar una idea de cómo son y funcionan las reglas de la Gramática de Afijo Extendida presentamos aquí un ejemplo sencillo de posibles reglas para la oración española. Tomamos como punto de

partida el enunciado: *Los tíos viven contentos*. Las reglas de la gramática libre de contexto que corresponden a esta oración son:

- (1) ORACIÓN : SINTAGMA NOMINAL, SINTAGMA VERBAL.
- (2) SINTAGMA NOMINAL : DETERMINANTE, NÚCLEO.
- (3) DETERMINANTE : ARTÍCULO DEFINIDO.
- (4) ARTÍCULO DEFINIDO : " los".
- (5) NÚCLEO : NOMBRE.
- (6) NOMBRE : RADICAL, SUFIJO1, SUFIJO2.
- (7) RADICAL : " tí".
- (8) SUFIJO1 : "o".
- (9) SUFIJO2 : "s".
- (10) SINTAGMA VERBAL : NÚCLEO, ATRIBUTO.
- (11) NÚCLEO : VERBO.
- (12) VERBO : RADICAL, SUFIJO1, SUFIJO2.
- (13) RADICAL : " viv".
- (14) SUFIJO1 : "e".
- (15) SUFIJO2 : "n".
- (16) ATRIBUTO : ADJETIVO.
- (17) ADJETIVO : RADICAL, SUFIJO1, SUFIJO2.
- (18) RADICAL : "content".
- (19) SUFIJO1 : "o".
- (20) SUFIJO2 : "s".

Las convenciones notacionales empleadas son las siguientes:

- : (punto doble) equivale a *rescribase como*;
- , (coma) sirve para unir los miembros de la misma regla;
- . (punto final) marca el final de la regla.

Los símbolos no-terminales se escriben con mayúsculas y los terminales lexicales van en minúsculas y entre comillas.

En la oración *Los tíos viven contentos* se presentan varios casos de concordancia: entre *los* y *tíos*, entre *los tíos* y *viven*, y entre *los tíos* y *contentos*. Están implicadas las categorías gramaticales de género, número y persona. Es posible explicar los casos de concordancia señalados aquí por las reglas de una GLC incluyendo símbolos no-terminales particulares para las diferentes personas, géneros y números gramaticales y enumerando las combinaciones permitidas, una por una en una larga serie de reglas rescriturales. Está claro que semejante descripción no se consigue en dos o tres reglas sino que exige un número mucho más elevado. Otra solución para describir el mismo fenómeno gramatical se ofrece si complementamos los símbolos no-terminales con una especie de parámetros que indican las categorías gramaticales (es decir, género, número, persona) y no los valores exactos (masculino, femenino, singular, plural, etc.). Los últimos son enumerados entonces en otra GLC. Los parámetros a que nos referimos son los 'afijos' de la gramática de que hemos hablado antes. Los introducimos en algunas de las reglas rescriturales formuladas mediante las abreviaciones g (por género), n (por número) y p (por persona):

- (1a) ORACIÓN : SINTAGMA NOMINAL(g, n, p),
 SINTAGMA VERBAL(g, n, p).
 (2a) SINTAGMA NOMINAL(g, n, p) : DETERMINANTE(g, n),
 NúCLEO(g, n, p).
 (10a) SINTAGMA VERBAL(g, n, p) : NúCLEO(n, p),
 ATRIBUTO(g, n).

A los 'afijos' g, n, p corresponden unas reglas especiales que llamamos metarreglas:

- (21) g :: MASCULINO;
 FEMENINO.
 (22) MASCULINO :: "MASCULINO".
 (23) FEMENINO :: "FEMENINO".
 (24) n :: SINGULAR;
 PLURAL.
 (25) SINGULAR :: "SINGULAR".
 (26) PLURAL :: "PLURAL".
 (27) p :: 1A;
 2A;
 3A.
 (28) 1A :: "1A".
 (29) 2A :: "2A".
 (30) 3A :: "3A".

Las nuevas convenciones notacionales introducidas en (21)-(30) son:

:: (dos puntos dobles) se usan en las metarreglas con el valor de *rescribase como*;

; (punto y coma) separa las definiciones alternativas de una regla.

Los valores no-terminales de los 'afijos' se escriben en mayúsculas y los terminales en mayúsculas y entre comillas. Los nombres de las categorías (g, n, p) van escritos en minúsculas.

En la Gramática de Afijo Extendida están implícitas las reglas de evaluación y de restricción. Esto quiere decir que unos nombres iguales de 'afijos' en la misma regla exigen valores terminales idénticos. De modo que en (1a) g, n y p del Sintagma Nominal y del Sintagma Verbal representan los mismos valores para ambos: o MASCULINO o FEMENINO, o SINGULAR o PLURAL, etc. Y en (2a) y (10a) ocurre lo mismo.

Cuando queremos dar una validez más general a la regla (1a), tendremos que tener en cuenta la opcionalidad de la función del ATRIBUTO. El SINTAGMA VERBAL tiene el 'afijo' g cuyo valor según la regla (10a) es idéntico al del ATRIBUTO. Y en regla (1a) el mismo 'afijo' tiene identidad de valor con el del SINTAGMA NOMINAL. Sin embargo, el ATRIBUTO no está siempre presente, porque además de *Los tíos viven contentos* también es correcta la secuencia *Los tíos viven*. Esto significa que no se puede exigir identidad constante entre el 'afijo' g del SINTAGMA NOMINAL y del SINTAGMA VERBAL. Y dentro del SINTAGMA VERBAL (10a) tampoco entre el 'afijo' n del ATRIBUTO y del NúCLEO. Adaptamos

la regla (1a) de tal forma que se deshace la identidad de g del SINTAGMA NOMINAL y del SINTAGMA VERBAL:

(1b) ORACIóN : SINTAGMA NOMINAL(g, n, p),
SINTAGMA VERBAL(g1, n, p),
coord género(g, g1).

Los nombres de los 'afijos' g y g1 ya no son idénticos de modo que sus valores tampoco lo son obligatoriamente. Además hemos añadido un miembro nuevo a la definición de la regla: *coord género(g1, g)*. Este representa un nuevo tipo de regla llamado 'predicado', que opera sobre los 'afijos' fijando los valores que puedan tomar. En este caso concreto el valor de g1 queda vacío o idéntico al del g del SINTAGMA NOMINAL. Por esto, el 'predicado' tiene estas reglas:

coord género(g, g): ;
coord género(g, VACÍO): .

Se observa que las reglas del 'predicado' tienen sólo la parte de la izquierda llenada y que la parte derecha queda vacía. La primera alternativa específica que los dos valores de g son idénticos, mientras que la segunda es para el caso que no se realiza la opción del ATRIBUTO de modo que falta un valor para g1 (VACÍO).

De manera parecida puede incorporarse a la regla para el SINTAGMA VERBAL la opcionalidad de los 'afijos' g y n del ATRIBUTO. Puesto que el valor del género del SINTAGMA VERBAL es siempre idéntico al del ATRIBUTO - el NÚCLEO no tiene 'afijo' g - basta con formular un 'predicado' para el número. En el 'predicado' entran n del NÚCLEO (el verbo) y del ATRIBUTO y se especifica entonces que n del ATRIBUTO o queda vacío o es idéntico al n del NÚCLEO (el verbo).

(10b) SINTAGMA VERBAL(g, n, p) : NÚCLEO(n, p),
ATRIBUTO(g, n1),
coord n(n, n1).
coord n(n, VACÍO): ;
coord n(n, n): .

De forma muy global acabamos de indicar los tipos de reglas que forman parte del formalismo de la Gramática de Afijo Extendida: las reglas rescriturales, libres de contexto, las metarreglas de los 'afijos', sensibles de contexto, y los predicados. Existe todavía otro recurso de esta gramática que sólo queremos mencionar aquí de paso. Nos referimos a la dirección en que la información sobre los valores exactos de los 'afijos' se transmite por el árbol sintáctico: de arriba hacia abajo o al revés. En el análisis automático una correcta selección del flujo (*flow*) puede traer consecuencias importantes para la velocidad del proceso. Los valores de los 'afijos' pueden ser inherentes o derivados. Los inherentes son propios de una categoría y van de abajo hacia arriba, los derivados son valores exigidos a una categoría por otra anterior y van por eso de arriba hacia abajo por el árbol sintáctico. Queremos explicarlo con un ejemplo concreto. Un verbo transitivo siempre va acompañado de un objeto directo. Si se

considera el objeto directo un complemento cuya presencia es exigida por el verbo, se le da uno o más 'afijos' derivados al objeto directo. Si, por el contrario, se opina que el verbo obtiene el valor de transitivo precisamente porque figura en la oración un objeto directo, todos los 'afijos' de éste son inherentes. La diferencia gramatical reside, pues, en la manera de ver la relación verbo y objeto directo.

3. Estado actual de la Gramática de Afijo Extendida para el español

Uno de los objetivos del proyecto ASATE, iniciado en la Universidad de Nimega en 1985 (ver Hallebeek 1987), es la composición de una Gramática de Afijo Extendida para el español. Esta gramática formal contiene las principales estructuras del español escrito sin tomar en cuenta los aspectos semánticos ni pragmáticos. Tenemos la mayor parte terminada y disponemos actualmente de módulos correspondientes e.o. a los siguientes constituyentes: Oración, Sintagma Nominal, Sintagma Adjetivo, Sintagma Adverbial, Sintagma Preposicional, Sintagma Verbal. Se incluyen descripciones de fenómenos gramaticales tales como el comparativo, imperativo, inversión, oraciones reflexivas y pasivas, coordinación y subordinación, formas no personales del verbo.

Aparte de esto estamos completando la confección de un lexicon de radicales de nombres, adjetivos, verbos y palabras funcionales. Este lexicon también tiene la forma de una Gramática de Afijo Extendida. Consiste de un número limitado de nombres de categorías lexicales (Verbo, Adjetivo,...) que se definen en forma de símbolos lexicales terminales. Los símbolos terminales son realizaciones alternativas de la misma categoría. Se construye según el modelo:

RAÍZ ADJETIVAL : " alt";
 " bonit";
 " pequeñ";
 ...

La información sobre los aspectos morfosintácticos de las radicales está expresada por 'afijos' que acompañan a las radicales. Así los verbos llevan 'afijos' con indicación de la categoría sintáctica (transitivo, intransitivo,...), de su tipo de conjugación (regular, irregular,...) y de eventuales preposiciones que se combinan con ellos. El lexicon comprende actualmente más de 4000 lemas.

Como parte integrante de la gramática estamos construyendo un analizador morfológico destinado a analizar y reconocer vocablos flexionados, compuestos, derivados y simples o primitivos. La gramática morfológica contiene una lista exhaustiva de los prefijos, infijos, sufijos y morfemas flexivos. El *parser* morfológico analiza cualquier palabra de un texto escrito sugiriendo una o más radicales posibles después de separar de ella los prefijos, infijos y sufijos (con inclusión de los flexivos). Se comprueba de forma automatizada si las radicales sugeridas por el *parser* figuran en el lexicon. Las que no se encuentran son desechadas y las otras pasan al analizador sintáctico

provistas de información gramatical pertinente (la categoría del vocablo, su género, número, persona, tiempo,...). Consideremos la palabra *derecho*. Formalmente puede tratarse e.o. de un nombre o adverbio (*derecho*), o de radical de verbo, nombre o adjetivo (*derech*) + flexivo (*o*), o de un prefijo (*de*) + radical de verbo, nombre o adjetivo (*rech*) + flexivo (*o*). Tras la consulta del lexicón el *parser* toma en cuenta en este caso como radicales aceptables el adverbio *derecho* y el adjetivo *derech* produciendo como resultados finales de su análisis las categorías adverbio y adjetivo, el último provisto de los 'afijos' masculino y singular.

Habrà quedado claro de lo precedente que en nuestro proyecto no partimos del análisis automatizado de textos lematizados, con indicación de las categorías gramaticales de sus constituyentes. Nuestro propósito es desarrollar un instrumento de análisis automatizado completamente independiente de los textos a analizar. Por esto el lexicón constituye una parte esencial de nuestra gramática. Sin embargo, no se puede excluir la necesidad de complementar temporalmente el lexicón con vocabularios locales de los textos a analizar. Esto será inevitable sobre todo para el caso de los nombres propios y los neologismos que difícilmente en su totalidad podrán incorporarse al lexicón.

4. El *parser generator* y el *Linguistic Data Base*.

Nuestra gramática formal escrita en el formalismo de la Gramática de Afijo Extendida se convierte automáticamente mediante un programa de *parser generator* (generador de analizador) en analizador automático. El desarrollo de un *parser generator* exclusivamente lingüístico es tarea del Departamento de Informática de nuestra universidad. Disponemos actualmente de una versión destinada para fines informáticos pero adaptada para uso lingüístico. El analizador produce diagramas arbóreos que aparecen en la pantalla del ordenador o del terminal. Es posible pasarlos a la impresora. El árbol sintáctico tiene nudos que llevan el nombre de la función sintáctica y de la categoría sintáctica con los correspondientes 'afijos'. En las reglas de rescritura de la gramática formal se observa una sucesión sistemática y constante de función a categoría hasta llegar a los símbolos terminales lexicales. Por razones prácticas, de espacio en la pantalla, los árboles producidos crecen de izquierda a derecha, en lugar de arriba abajo. Como ejemplo reproducimos aquí el diagrama que corresponde a la oración *Los tíos viven contentos*. Es de acuerdo con las reglas formuladas en los apartados anteriores pero sin reflejar las reglas morfológicas (partición de los vocablos en radicales y sufijos).

