

**2 - ANÁLISIS MORFOLÓGICO, SINTÁCTICO, SEMÁNTICO
Y/O PRAGMÁTICO**

SSG: UNA GRAMÁTICA DEL ESPAÑOL BASADA EN EL FORMALISMO SLOT GRAMMAR

Gerardo Arrarte

Instituto de Tecnología Multimedia IBM
Santa Hortensia, 26-28
28002 - Madrid
gerardo@emdcci11.bitnet

Introducción

En el presente documento se describe el prototipo de gramática SSG (*Spanish Slot Grammar*) desarrollado por el autor en el marco del proyecto de Traducción Automática LMT (*Logic-programming-based Machine Translation*) [McCord 89a, 89b, 89c]. El propósito de esta primera versión de SSG ha sido el de servir como módulo de análisis en el prototipo de sistema de traducción para el par de lenguas español-inglés. No obstante, es susceptible de ser utilizada con otros fines, ya que, como veremos, el formalismo proporciona unas estructuras de análisis del texto con información lo suficientemente exhaustiva para adaptarse a distintos cometidos. El presente documento contiene en primer lugar una breve reseña del formalismo gramatical en que SSG está basada y de los tipos de reglas que constituyen la gramática, para centrarse a continuación en algunos fenómenos que, por ser específicos del español, han requerido un tratamiento original con respecto a versiones de *Slot Grammar* desarrolladas previamente para otras lenguas. Finalmente, se realiza una somera evaluación de los resultados obtenidos en este primer prototipo.

El formalismo gramatical

¿Qué es «Slot Grammar»?

El formalismo gramatical *Slot Grammar* [McCord 92, McCord et al.], desarrollado por M. C. McCord en el Centro de Investigación Thomas J. Watson de IBM, está basado en la expresión sistemática de las reglas gramaticales y de la información lingüística en términos de *slots* y *slot frames*.

Los *slots* (ranuras) representan relaciones gramaticales y son designados por nombres como *subj* (sujeto), *obj* (objeto directo), *nadj* (adjetivo modificador del nombre), etc. Los hay de dos tipos: *complementos* y *adjuntos*. Cada acepción de una palabra lleva asociada una lista de *slots* complementos en el léxico. Esta lista constituye el *slot frame* de la palabra (para ser exactos, de la acepción de la palabra). Además, para cada categoría léxica o parte de la oración se especifica en la gramática un conjunto de *slots* adjuntos. La información contenida en el léxico determina, por lo tanto, todos los posibles *slots* para una acepción, de forma explícita para los complementos y de forma implícita (a través de la categoría léxica) para los adjuntos. El slot *obj*, por ejemplo, es de tipo complemento ya que es un rasgo propio de determinados verbos (los verbos transitivos) y debe ser incluido en la información léxica de cada uno de ellos. En cambio, el slot *nadj* es de tipo adjunto: está asociado a todas las palabras clasificadas como nombres en el léxico.

Dados una frase o sintagma, una palabra P del mismo (con un análisis léxico determinado) y un *slot* S asociado a P, la gramática establece cuándo se puede decir que, para otra palabra Q (con un análisis léxico determinado),

Q llena el slot S de P.

Por ejemplo, según nuestra gramática en

El saber no ocupa lugar.

saber llena el *slot* subj de *ocupa*. El proceso de análisis gramatical de una frase consiste básicamente en determinar qué relaciones del tipo 'llenar el slot S de' pueden establecerse entre los posibles análisis léxicos de sus palabras.

El análisis resultante consta de una combinación de estructura superficial y forma lógica en una única estructura. La estructura superficial está expresada en forma de árboles de dependencia, en los que cada nodo lleva asociada una palabra núcleo P y cada uno de sus nodos hijos tiene por núcleo una palabra que llena un *slot* de P. Además, cada nodo va acompañado por el *slot frame* de su núcleo, en el cual se especifican sus relaciones gramaticales, ya sean éstas superficiales o profundas, inmediatas o remotas. Esta información referente a las relaciones establecidas entre las distintas palabras de la frase, puede ser usada para construir la forma lógica correspondiente.

Entre las características del sistema, cabe destacar las siguientes:

- La totalidad del sistema está implementado en Prolog.
- Se trata de un sistema *lexicalista*, ya que gran parte de la información necesaria para el análisis se encuentra en los *slot frames* especificados en el diccionario. Esta distribución de la información facilita la ampliación y mantenimiento tanto de la gramática como del diccionario.
- Las reglas se dividen de manera modular en diferentes tipos que tratan diversos fenómenos relacionados con los *slots*. Así, por ejemplo, las reglas básicas que llenan los *slots* (*slot filler rules*), no especifican normalmente las posibles restricciones en el orden de precedencia de los sintagmas que llenan los distintos *slots* o entre cada uno de estos y la palabra núcleo. Esta clase de restricciones se fijan a través de otro tipo de reglas: las reglas de orden sintáctico (*slot ordering rules*). Por otra parte, no se hace uso en ningún momento de reglas de estructura de frase.
- El sistema incluye un *shell* o serie de módulos independientes de cada lengua, que incluye, además del analizador de base, el tratamiento general de fenómenos gramaticales de carácter universal, como la coordinación, extraposición, puntuación y parentización. No obstante, parte de la información referente a estos fenómenos puede ser especificada en la gramática de cada lengua.
- Se hace uso de un *sistema de evaluación numérica* de los análisis basado en una estrategia heurística que compara los análisis obtenidos estableciendo un orden de preferencias. Este sistema de evaluación es utilizado no sólo para ordenar los análisis finales, sino también para ordenar los análisis parciales obtenidos a lo largo del proceso y, de manera opcional, descartar los de menor preferencia. Esto redundará en una mayor eficiencia de análisis y en la posibilidad de establecer restricciones gramaticales que evitan la proliferación de análisis alternativos, reduciendo de forma considerable el conjunto de análisis finales. Buena parte del sistema de evaluación se encuentra en el *shell* del sistema, aunque también en la gramática y en el léxico específicos de una lengua pueden definirse la información o reglas de evaluación convenientes.

Además de la descrita en este documento para el español, existen versiones de *Slot Grammar* en distintas etapas de desarrollo para inglés (M. C. McCord), alemán (U. Schwall), danés (A. Bernth), noruego (M. C. McCord y A. Bernth) y hebreo (S. Wintner).

El análisis léxico

Slot Grammar se basa en un analizador ascendente de tipo *chart parser*. El análisis de una secuencia de palabras se construye de manera incremental a partir de los análisis léxicos de cada una

de ellas, mediante un proceso que consiste en crear estructuras de análisis de mayor extensión añadiendo sucesivamente sintagmas modificadores a la izquierda o derecha de una palabra núcleo.

Debemos tener presente que una palabra puede tener múltiples análisis léxicos. Esto es debido a que:

- una palabra puede tener distintas acepciones en el diccionario,
- una palabra flexionada puede corresponder a distintas formas base (v.g., *pesas* como plural del nombre *pesa* y como forma de presente de *pesar*), o a distintas formas flexivas de una misma forma base (*llegamos* como forma de presente y de pasado de *llegar*).

Cada análisis léxico (o *word frame*) de una palabra consiste en una cláusula de Prolog que contiene la siguiente información:

- La posición secuencial de la palabra en la frase.
- La forma base (es decir, no flexionada) de la palabra y la acepción o subentrada léxica a la cual corresponde el análisis.
- La categoría léxica o parte de la oración y los rasgos morfológicos (género, número, persona, tiempo, etc.) correspondientes a la forma analizada.
- El *slot frame* o lista de los *slots* de tipo complemento de la acepción en cuestión.

Veamos con más detenimiento en qué consiste la información referente al *slot frame* que aparece en los análisis léxicos. Esta lista de *slots* de tipo complemento está especificada en forma abreviada en la subentrada correspondiente del diccionario. Por ejemplo, una determinada acepción de *saber* puede tener *slots* de: a) sujeto, b) objeto directo, y c) complemento preposicional introducido por *de*, *sobre* o *acerca de*.

Para cada uno de los *slots* del *slot frame*, el análisis léxico contiene la siguiente información:

- El *nombre del slot*. Consiste en un término de Prolog, que es usado en la gramática para referirse al *slot* y también aparece en la representación gráfica de los árboles de análisis. Por ejemplo, para el verbo *saber*, algunos de los *slots* que se acaban de mencionar pueden tener nombres como *obj(n|fin|inf)* (un objeto directo que debe ser llenado ya sea por un sintagma nominal, por una cláusula completiva o por una construcción de infinitivo), o *comp(p(de|sobre|acerca.de))* (un complemento que debe ser llenado por un sintagma preposicional con una de las preposiciones indicadas).

Los *slots* *obj(_)* y *comp(_)* son ejemplos de *or-slots*, un tipo especial que se caracteriza por permitir especificar una disyunción de opciones respecto al tipo de sintagmas que pueden llenar el *slot*. Esto permite clasificar debidamente en el diccionario las distintas construcciones a que puede dar lugar una palabra¹, y a la vez mantener agrupados distintos tipos de modificadores que son mutuamente excluyentes y responden a conceptos sintácticos o semánticos equiparables.

- Una indicación sobre si el *slot* es obligatorio u opcional. Más adelante, hablaremos en más detalle acerca de la obligatoriedad de un *slot*.
- En caso de que se especifiquen restricciones de este tipo en el diccionario, una combinación booleana de rasgos sintácticos o semánticos que deberá reunir el sintagma que llene el *slot*.

Un caso especial de análisis léxico es el de los *word frames* generados por la gramática para tratar las construcciones pasivas. Este fenómeno es descrito más detalladamente en el apartado «Pasiva

¹ Por ejemplo, no todos los verbos transitivos (y no todas las acepciones transitivas de un mismo verbo) pueden necesariamente ser modificados por objetos directos de los tres tipos mencionados.

refleja, verbos pronominales y otros usos del pronombre reflexivo». Por ahora, baste decir que, para las formas impersonales y para las de tercera persona de los verbos transitivos, la gramática efectúa una serie de transformaciones sobre los *word frames* ya existentes, generando otros nuevos que vienen a añadirse al conjunto de análisis léxicos de dichas formas verbales. Para el participio, la transformación consiste en sustituir los *slots* correspondientes a sujeto y objeto directo, por uno para el agente y otro para el sujeto paciente. De modo similar, para las formas de tercera persona y para las impersonales, se sustituyen dichos *slots* por uno para el indicador de pasiva refleja *se* y otro para el sujeto paciente. (Al igual que los restantes análisis léxicos, estos *word frames* especiales para el tratamiento de construcciones pasivas son generados previamente al análisis sintáctico, de manera que serán las reglas de la gramática las que posteriormente determinarán si los mismos son susceptibles de ser empleados en el análisis de la frase.)

Formato de las entradas de diccionario

Las entradas de los diccionarios utilizados para el análisis léxico están escritas en un formato especial, llamado *ELF* (*External Lexical Form*). Las entradas corresponden principalmente a las formas base (no flexionadas) de las palabras. Las formas flexivas correspondientes a cada forma base son tratadas por un analizador morfológico².

Las entradas de diccionario en formato *ELF* son cláusulas de la siguiente forma:

Palabra < A1 < A2 < ... < An.

siendo cada *Ai* un *elemento* de la entrada.

He aquí algunos ejemplos³:

```
alto < adj(501)
      < adv
      < n(909).
traducir < v(obj.p(de).p(a),28).
tirar < v(obj.iobj,1)
      < v(pron1.goal,1).
```

El primer ejemplo tiene tres elementos o subentradas para la palabra *alto*, que corresponden a sus tres acepciones como adjetivo, adverbio y nombre⁴. Vemos que en esta entrada no se ha especificado, aparte de este código, más información que la parte de la oración. No hay especificación de *slots* de tipo complemento, por lo cual la palabra sólo podrá ser modificada por los slots de tipo adjunto que correspondan a la parte de la oración de cada acepción.

La entrada correspondiente a *traducir* presenta una sola acepción para este verbo, que tiene un *slot* para objeto directo (*obj*), y dos para complementos preposicionales introducidos por *de* y por *a* (*p(de)* y *p(a)*, respectivamente). No es necesario incluir en la lista el *slot* correspondiente al sujeto,

² El análisis morfológico del español se realiza mediante dos módulos llamados *SMOR* (*Spanish Morphology*) y *SAMOR* (*Spanish Analysis Morphology*), desarrollados por G. Arrarte e I. Zapata [Arrarte et al. 91a]. Estos módulos se basan en la clasificación en paradigmas flexivos de las palabras (unas 34.000 entre verbos, nombres, adjetivos y determinantes). Dicha clasificación procede de la labor realizada por C. Rodríguez, L. Sopeña y C. Villar [Rodríguez et al.], del Centro Científico de IBM en Madrid, en colaboración con *IBM Application Systems Division* de Bethesda, en el marco del proyecto *Spanish Dictionary Development* (LEXIS).

³ El contenido de algunas entradas ha sido simplificado para mayor claridad.

⁴ Los enteros que aparecen como último argumento en los elementos correspondientes a partes flexivas de la oración son códigos que indican el paradigma flexivo de la palabra. Estos códigos son utilizados para el análisis morfológico y resultan irrelevantes para la gramática una vez que se ha obtenido el análisis léxico.

ya que este queda sobrentendido para todos los verbos⁵. No obstante, puede incluirse cuando se quiere especificar alguna característica especial para el mismo. Además, existen formas abreviadas para los nombres de algunos *slots*. Por ejemplo, *obj*, *p(de)* y *p(a)* son formas abreviadas que se convierten respectivamente en *obj(n)*, *comp(p(de))* y *comp(p(a))* en el formato interno del *word frame*.

El tercer ejemplo muestra dos acepciones del verbo *tirar*. La primera tiene *slots* de objeto directo e indirecto. La forma abreviada *iobj* se traduce en forma interna por *iobj(n|a)*, cuyas opciones corresponden a los dos tipos de sintagmas que pueden llenarlo: un sintagma nominal o un sintagma preposicional con *a*. La segunda acepción tiene un *slot* *pron*, que indica el uso pronominal⁶ del verbo y se llena con un pronombre reflexivo, y otro *slot* *goal* correspondiente a un complemento circunstancial con valor de dirección o lugar de destino.

Recordemos que en formato interno cada *slot* tiene un indicador de obligatoriedad: *spot* *id=obedic..* El carácter obligatorio de un *slot* se determina de la siguiente manera. En primer lugar, un *slot* puede ser siempre obligatorio (por ejemplo, el objeto de una preposición), en cuyo caso esto se indica en la gramática según veremos en el apartado «Reglas de declaración de slots obligatorios». Y en segundo lugar, puede serlo para algunas palabras o acepciones. En este caso, la obligatoriedad del *slot* es indicada en el diccionario mediante el carácter 1 añadido como sufijo al nombre del *slot*. En los restantes casos, el *slot* recibe el rasgo de opcional.

En los ejemplos anteriores, el único *slot* obligatorio es *pron*. Esta indicación de obligatoriedad actúa de hecho como una restricción. Su efecto es que la frase en la que aparezca el verbo *tirar* sólo podrá producir un análisis en que se emplee esta acepción en los casos en que el *slot* *pron* haya sido llenado, es decir, cuando el verbo vaya acompañado de un pronombre reflexivo que concuerde con él en persona y número.

Además del *slot frame*, el formalismo léxico permite asociar tipos semánticos o sintácticos a una determinada acepción, así como las áreas temáticas relacionadas con la misma. He aquí algunos ejemplos de entradas en que se utilizan tipos semánticos:

```
secretario < n(h,nil,952).
oveja < n(a,nil,910).
espacio < n(loc,nil,909).
miércoles < n(tm,nil,903).
debajo < adv(loc,*).
acá < adv(goal,*).
abajo < adv(loc&goal,*).
```

En estos ejemplos, el tipo semántico corresponde al primer argumento de cada subentrada. La especificación *nil* corresponde al *slot frame*, que en estos casos es la lista vacía ya que estas palabras no admiten *slots* de tipo complemento. Los tipos semánticos mostrados para los cuatro ejemplos correspondientes a sustantivos son: humano (*h*), animado (*a*), lugar (*loc*) y tiempo (*tm*). En los ejemplos correspondientes a adverbios, los tipos mostrados son: *lugar donde* (*loc*) y *lugar a donde* (*goal*). Obsérvese que en la entrada de *secretario* no es necesario especificar el tipo *animado*, implícito en el tipo *humano*, ya que este está declarado como subtipo de aquél en la gramática, de modo que el sistema le asigna automáticamente el tipo *animado* a cualquier acepción con tipo *humano*.

Durante el análisis léxico, se crea para cada acepción con especificaciones de tipos semánticos o sintácticos una o más cláusulas de Prolog que indican la pertenencia de la acepción a cada uno de dichos tipos. Esto permite establecer en las reglas de la gramática las restricciones necesarias en cuanto

⁵ Para los verbos españoles que no admiten sujeto, se incluye en la entrada una especificación de tipo sintáctico que permite establecer una restricción en las reglas de la gramática correspondientes a este *slot*.

⁶ El tratamiento de los verbos pronominales para el español es descrito con más detalle en el apartado «Pasiva refleja, verbos pronominales y otros usos del pronombre reflexivo».

a tipos semánticos o sintácticos. Además, estas restricciones también pueden establecerse en el propio diccionario. Por ejemplo, si un verbo sólo admite objetos directos de tipo humano, basta especificar en la entrada correspondiente `obj:h`. No obstante, en SSG, se hace un uso limitado de este tipo de especificaciones en los diccionarios, debido a la inconveniencia en la mayoría de los casos de establecer restricciones absolutas de este tipo.

Como se ha dicho antes, además de los tipos semánticos o sintácticos, también se puede especificar, para una determinada acepción, a cuáles áreas temáticas corresponde. Por ejemplo, si queremos establecer una acepción para la palabra *activo* que se aplique sólo para el área temática *comercio*, podemos especificarlo de la siguiente manera en el diccionario:

```
activo < adj(501)
      < n(sa(comm),nil,909).
```

De este modo, *activo* será analizado en general como adjetivo, pero también como nombre cuando el texto corresponda al área temática (*subject area*) `comm`.

El formalismo admite también incluir subentradas para *multipalabras*. Veamos algunos ejemplos correspondientes a distintas categorías léxicas o partes de la oración:

```
ruido < n(909)
      < mn(«= de fondo»,909).
corto < adj(501)
      < madj(«= de vista»,501).
encoger < v(obj,57)
        < v(pron1,57)
        < mv(«= de hombros»,pron1,57).
debajo < adv(loc,*)
        < mprep(«= de»).
nado < madv(«a =>).
```

Para las partes flexivas de la oración, el símbolo (=) sustituye tanto a la forma base de la palabra como a sus distintas formas flexivas. Por lo tanto, a partir de las entradas anteriores, pueden analizarse expresiones como: ruidos *de fondo*, corta *de vista*, encogiéndose *de hombros*, etc.

Estructura interna y representación gráfica del análisis sintáctico

Durante el proceso de análisis, la información referente a cada sintagma analizado se almacena en una estructura de datos, llamada *phrase*. Esta estructura *phrase* contiene la siguiente información:

- La extensión del sintagma (posición secuencial de la primera y de la última palabras del mismo) y la posición de la palabra núcleo del sintagma.
- La forma base de la palabra núcleo y la acepción a la cual corresponde el análisis (tomadas ambas del *word frame* de dicha palabra).
- Una estructura de datos que llamamos *features*, la cual contiene:
 - la información morfológica de la palabra núcleo (parte de la oración y rasgos de género, número, persona, tiempo, etc.), tomada asimismo del *word frame* de dicha palabra⁷;

⁷ La información contenida en *Features* para una palabra dada puede ser ampliada durante el análisis mediante un mecanismo de unificación al aplicarse las reglas de la gramática. Por ejemplo, para la palabra *gran*, la forma inicial de *Features* en el análisis léxico sería:

```
adj(X.sg.pre,Y)
```

donde la variable *x* corresponde al rasgo de género, que no está instanciado en este caso por tratarse de una forma que carece de distinción de género. En el momento de aplicarse una de las reglas de la gramática (la que se da como

- y, en algunos casos, información referente al sintagma en su conjunto.
- El *slot frame* de la palabra núcleo. Se corresponde en términos generales con el *slot frame* que aparece en el *word frame* de dicha palabra, indicándose además la siguiente información para aquellos *slots* que han sido llenados:
 - la posición secuencial de la palabra núcleo del sintagma que ha llenado el *slot*;
 - en caso de tratarse de un *or-slot*, indicación de la opción que ha sido empleada (por ejemplo, para el *slot* *obj(n|fin)*, la opción empleada será *obj(fin)*, cuando el objeto directo sea llenado por una cláusula completiva);
 - para construcciones pasivas, indicación de la función que tiene, en la estructura profunda de la frase, el sintagma que ha llenado el *slot*.
- Una lista de *slots* *extrapuestos*, de los cuales hablaremos en el apartado «Reglas de extraposición».
- Sendas listas de *modificadores por la izquierda* y *modificadores por la derecha* de la palabra núcleo. Cada elemento de estas listas de modificadores tiene la forma *slot:Phrase*, donde *slot* es uno de los *slots* (de tipo complemento o de tipo adjunto) de la palabra núcleo, y *Phrase* es la estructura de datos de tipo *phrase* correspondiente al sintagma que ha llenado dicho *slot*.

La estructura de análisis que acabamos de describir puede resultar considerablemente compleja para una frase de mediana longitud. Por ello, el sistema dispone de la posibilidad de visualizarla gráficamente en forma de árbol de análisis.

En el siguiente gráfico, podemos ver el análisis resultante de la frase *Mi amigo tradujo el libro del inglés al español*⁸.

(1)	Mi	amigo	tradujo	el	libro	de	el	inglés	a	el	español.
	ndet	mil(1)						det(m.sg,poss)			
	subj(n)	amigo1(2)						noun(cn,m.sg,X1)			
	top	traducir1(3,2,5,8,11)						verb(fin(pers3.sg,past,ind:dcl:nwh))			
	ndet	el2(4)						det(m.sg,def)			
	obj(n)	libro1(5)						noun(cn,m.sg,X4)			
	comp(p(de))	del(6,8)						prep(de,X3,8)			
	ndet	el2(7)						det(m.sg,def)			
	objprep(sg)	inglés1(8)						noun(cn,m.sg,X3)			
	comp(p(a))	al(9,11)						prep(a,X2,11)			
	ndet	el2(10)						det(m.sg,def)			
	objprep(sg)	español2(11)						noun(cn,m.sg,X2)			

ejemplo en la página) para llenar el *slot* *nadj* del nombre con este adjetivo, la variable quedará unificada al rasgo de género del nombre. Por ejemplo, en *gran ciudad*, la información asociada al nodo correspondiente al adjetivo en el árbol de análisis será:

adj(f.sg.pre,Y)

(El rasgo *pre* indica que se trata de una forma de uso exclusivamente proclítico.)

⁸ Hemos utilizado la entrada de *traducir* tal como aparece en el ejemplo de la página, de modo que su *slot frame* consta (por este orden) de los siguientes *slots*: *subj(n)*, *obj(n)*, *comp(p(de))*, *comp(p(a))*.

Obsérvese además que las contracciones *del* y *al* han sido separadas en sus componentes durante el análisis léxico. Lo mismo ocurre cuando aparecen pronombres enclíticos concatenados a formas verbales.

Vemos que la representación gráfica del análisis consta de una línea por nodo, y que la información que aparece en cada línea se centra en torno a la palabra núcleo del nodo en cuestión. Dicha información consiste en: (1) las líneas de conexión del árbol, (2) el nombre del *slot* que ha sido llenado por el nodo, (3) la *predicación de acepción*, y (4) la información morfológica (y sintáctica) del nodo (*features*). La predicción de acepción tiene el aspecto de una predicción de lógica de predicados en la que el predicado es la acepción de la palabra (formada por la concatenación de la forma base y el número de acepción) con los siguientes argumentos. El primer argumento es el identificador del nodo, es decir, la posición secuencial de la palabra núcleo. Los siguientes, si los hay, son los identificadores de los nodos que llenan los *slots* del *slot frame* de la palabra núcleo. El orden de estos *slots* es aquel en que aparecen en la correspondiente subentrada del diccionario. (Por lo tanto, aunque este orden no es relevante para la aplicación de las reglas de la gramática, sí lo es a efectos de la representación gráfica.)

Veamos ahora el árbol de análisis obtenido con una frase pasiva:

(2) El libro fue traducido por mi amigo de el inglés a el español.		
ndet	e12(1)	det(m.sg,def)
subj(n)	libro1(2)	noun(cn,m.sg,X1)
top	ser1(3,2,4)	verb(fin(pers3.sg,past,ind:cl:nwh))
pred	traducir1(4,7,2,10,13)	verb(pastpart(m.sg))
agent	por1(5,7)	prep(por,X5,7)
ndet	mi1(6)	det(m.sg,poss)
objprep(sg)	amigo1(7)	noun(cn,m.sg,X5)
comp(p(de))	de1(8,10)	prep(de,X4,10)
ndet	e12(9)	det(m.sg,def)
objprep(sg)	inglés1(10)	noun(cn,m.sg,X4)
comp(p(a))	a1(11,13)	prep(a,X3,13)
ndet	e12(12)	det(m.sg,def)
objprep(sg)	español2(13)	noun(cn,m.sg,X3)

Nótese cómo los argumentos de la predicción de acepción de *traducido* revelan las funciones profundas de sus modificadores: el segundo argumento, correspondiente al sujeto, señala al nodo 7 (*mi amigo*), que llena el *slot* de complemento agente; el tercero, correspondiente al objeto directo, se unifica con el identificador del nodo 2 (*el libro*), que llena el *slot* de sujeto en la estructura superficial.

Tipos de reglas de la gramática

En *Slot Grammar* existen nueve tipos básicos de reglas relacionados con: (1) declaración de *slots* de tipo adjunto, (2) llenado de *slots*, (3) declaración de *or-slots*, (4) orden sintáctico, (5) declaración de *slots* obligatorios, (6) extraposición, (7) coordinación, (8) puntuación, y (9) evaluación de análisis. Cada tipo de reglas tiene un formato especial, aunque aquí no nos detendremos en la descripción de cada uno de ellos. Un compilador de reglas existente en el *shell* se encarga de convertirlas en cláusulas de Prolog.

• Reglas de declaración de slots de tipo adjunto

Como queda dicho, los *slots* de tipo complemento que pueden modificar a una palabra son propios de la misma y, por ello, aparecen listados en la correspondiente entrada de diccionario. En cambio, los *slots* de tipo adjunto que pueden modificarla dependen exclusivamente de su categoría léxica o parte de la oración. Para cada categoría léxica que lleva asociados *slots* de este tipo, existe en la gramática una regla en la cual se declara la lista de los mismos.

Los *slots* de tipo adjunto, a diferencia de los de tipo complemento, son siempre opcionales y pueden ser llenados cualquier número de veces. Por ejemplo, un nombre puede tener como modificadores un número indefinido de adjetivos, cada uno de los cuales llenará el *slot* *nadj* (adjunto). En cambio, un verbo transitivo admite como máximo un modificador que llene el *slot* *obj* (complemento).

• Reglas de llenado de slots

Las reglas de llenado de *slots* (*slot filler rules*) constituyen el núcleo de la gramática. Para cada *slot*, ya sea de tipo complemento o de tipo adjunto, existe una o más reglas de este tipo. El proceso de análisis consiste básicamente en una serie de pasos, en cada uno de los cuales se toma uno de los *slots* asociados a la palabra núcleo de un sintagma y se comprueba si existe otro sintagma adyacente al mismo por la izquierda o por la derecha que satisfaga las condiciones de una de las reglas de llenado para dicho *slot*. Si esto es así, dicho sintagma llena el *slot* y se añade a los modificadores de la palabra núcleo. El resultado es un sintagma de extensión mayor, a partir del cual se continúa este proceso.

Las reglas de llenado de *slots* pueden ser de dos tipos: normales y de *slots* *extrapuestos*. Nos ocuparemos por el momento de las normales. El formato de este tipo de reglas es

Slot ==> Cuerpo.

donde Slot es el nombre del *slot* y Cuerpo tiene un formato similar al cuerpo de una cláusula de Prolog, es decir, una serie de objetivos combinados mediante operadores lógicos *and* (&) y *or* (|). Existe una serie de *objetivos especiales* que son reconocidos por el compilador de reglas y que permiten acceder a datos concretos ya sea del sintagma con que se intenta llenar el *slot* o bien del nodo superior.

A modo de ejemplo, veamos una de las reglas correspondiente al *slot* *nadj*⁹:

nadj ==> f(adj(G.N.*,*)) &
 hf(noun(cn,G.N.*)).

Los objetivos especiales *f*(Features) y *hf*(HFeatures) se satisfacen cuando Features y HFeatures se unifican con los rasgos morfológicos del núcleo del sintagma que llena el *slot* y del nodo superior, respectivamente. Por lo tanto, esta regla establece que un adjetivo puede llenar el *slot* *nadj* de un nombre de tipo *cn* (común), siempre que ambos concuerden en género y número.

• Reglas de declaración de or-slots

Un *slot* de tipo complemento puede ser declarado *or-slot* (ver explicación en la página), mediante este tipo de reglas. Naturalmente, un *slot* del tipo *or-slot* requiere una o más reglas de llenado para cada opción específica.

Este tipo de reglas permiten además establecer condiciones que deben cumplirse para el llenado del *slot*, similares a las de las reglas de llenado de *slots* pero que son de carácter general para todas las opciones. Esto evita el tener que repetirlas en cada regla.

• Reglas de orden sintáctico

Cada vez que se aplica una regla de llenado de un *slot*, el analizador comprueba que la posición del nuevo modificador es válida de acuerdo con lo establecido en las reglas de orden sintáctico. Estas son de dos tipos:

⁹ La regla dada a modo de ejemplo es una versión algo simplificada pero válida desde el punto de vista del formalismo.

Reglas de precedencia núcleo/slot: Indican cuándo un sintagma que llena un determinado *slot* puede modificar a otro por la izquierda y cuándo por la derecha. Por ejemplo, para el *slot* *nadj* podemos establecer restricciones generales de este tipo:

- el sintagma (adjetivo) que lo llene puede ser un modificador por la derecha del nodo superior (nombre), siempre que no se trate de un adjetivo marcado con el rasgo *pre*, correspondiente a formas proclíticas apocopadas;
- por otra parte, puede ser un modificador por la izquierda, siempre que no tenga a su vez modificadores por la derecha¹⁰.

Reglas de precedencia entre slots: Regulan el orden de precedencia entre dos *slots* que modifican, ambos por la izquierda o ambos por la derecha, a un mismo núcleo.

Por ejemplo, una de estas reglas establece que el objeto indirecto debe preceder al objeto directo siempre que ambos sean llenados por pronombres personales.

• Reglas de declaración de slots obligatorios

Como hemos visto, el hecho de que un *slot* sea o no obligatorio puede afectar el proceso de análisis. El sintagma correspondiente a un análisis parcial no puede modificar a otro nodo a menos que todos los *slots* obligatorios de dicho sintagma hayan sido llenados previamente¹¹.

Hay dos maneras de hacer que un *slot* sea obligatorio. La primera, ya descrita en el apartado «Formato de las entradas de diccionario», consiste en declararlo en el diccionario añadiendo el carácter 1 al nombre del *slot*. La segunda consiste en declarar de manera general un *slot* como obligatorio en la gramática mediante reglas especiales.

Por ejemplo, el *slot* *objprep*(_) , correspondiente al objeto de una preposición¹², está declarado en la gramática como *slot* obligatorio. Por consiguiente, una preposición no puede modificar otra palabra mientras su *slot* *objprep*(_) permanezca vacío.

• Reglas de extraposición

Normalmente, el formalismo sólo permite que un sintagma llene un *slot* perteneciente al núcleo de otro sintagma adyacente al mismo. No obstante, hay casos en que la relación de modificación se establece entre palabras que no cumplen este requisito. Existe un mecanismo de extraposición de *slots* que permite tratar este problema.

¹⁰ El adjetivo o participio *adecuado*, por ejemplo, admite un modificador por la derecha consistente en un sintagma preposicional con *a* o *para*:

lecturas adecuadas para niños

También puede preceder al nombre, pero en ese caso no admite modificadores por la derecha:

adecuadas lecturas

**adecuadas para niños lecturas*

¹¹ Existen excepciones a este requisito relacionadas con el fenómeno de extraposición.

¹² Ver ejemplo de árbol de análisis en la página .

En el siguiente ejemplo de análisis para la frase *Te quiero hablar de cierto asunto*, el *slot* de objeto indirecto del verbo *hablar* ha sido llenado en el nodo superior, correspondiente al verbo *querer*¹³:

(4)	Te quiero hablar de cierto asunto.	
iobj(n)	tú2(1)	noun(pron(pers), pers2.X3.sg.dat.X4,X5)
top	querer1(2,u,3)	verb(fin(pers1.sg,pres,ind:cl:nwh))
obj(inf)	hablar3(3,u,1,6)	verb(inf(X6))
comp(p(de))	del(4,6)	prep(de,X7,6)
nadj	cierto1(5)	adj(m.sg.X8,X9)
objprep(sg)	asunto1(6)	noun(cn,m.sg,X7)

Este cambio de nivel del *slot* ha sido posible de la siguiente forma:

- Cuando el sintagma *hablar de cierto asunto* llena el *slot* obj(inf) de *querer*, el *slot* iobj(n|a) de *hablar* aún está vacío.
- iobj(n|a) es un *slot* susceptible de ser extrapuesto a un nivel superior. Por otra parte, obj(inf) es un *slot* a través del cual está permitida la extraposición.
- Por lo tanto, el *slot* iobj(n|a) es extrapuesto al nodo superior, siendo incluido en la lista de *slots* extrapuestos de dicho nodo.
- Posteriormente, el sintagma *te* llena dicho *slot* mediante la aplicación de una regla de llenado de *slots* extrapuestos.
- Finalmente, se deja constancia en la estructura de análisis del vínculo establecido entre el sintagma *te* y la palabra *hablar*, incluyendo el identificador de nodo 1, correspondiente a *te*, en la información referente al *slot* iobj(n) en el *slot frame* de *hablar*.

En el apartado anterior, se ha dicho que un sintagma no puede modificar a otra palabra de la frase hasta que todos sus *slots* obligatorios hayan sido llenados. Ahora bien, el mecanismo de extraposición permite que esto sí ocurra cuando un *slot* obligatorio aún vacío puede ser extrapuesto a un nivel superior. En ese caso, el llenado del *slot* queda pendiente de ser efectuado en dicho nivel (o en niveles superiores a este en caso de que el *slot* pueda ser extrapuesto nuevamente). No obstante, el requisito de llenar un *slot* obligatorio debe satisfacerse en algún nivel. Si tras sucesivas extraposiciones del mismo no se encuentra un sintagma que lo llene y, por otra parte, el *slot* ya no puede ser nuevamente extrapuesto, el análisis parcial resultante no podrá actuar como modificador de otra palabra (o, en el caso de que corresponda a la totalidad de las palabras de la frase, ser considerado un análisis válido de la misma).

El *shell* del sistema se ocupa de algunos de los problemas relacionados con la extraposición, pero hay tres tipos de reglas para este fenómeno que son específicas de la gramática de cada lengua: (1) declaración de *slots* que pueden ser extrapuestos, (2) declaración de *slots* a través de los que se puede extraponer otro *slot*, y (3) reglas de llenado de *slots* extrapuestos.

¹³ La letra u (*unfilled*) que aparece en el argumento correspondiente al sujeto en la predicación de acepción de los verbos *querer* y *hablar* indica que dicho *slot* no ha sido llenado por ningún sintagma.

El tercer tipo constituye en realidad un caso especial de las reglas ya descritas en el apartado «Reglas de llenado de slots». No obstante, su aplicación difiere de las mismas en que: (1) no se aplican las restricciones de orden sintáctico que hemos visto en el apartado «Reglas de precedencia entre slots», y (2) no admiten la extraposición al nodo superior de los *slots* que estén vacíos en el sintagma que llena el *slot* extrapuesto.

• Reglas de coordinación

No entraremos aquí en detalles sobre el mecanismo mediante el cual distintos sintagmas se unen en una estructura de análisis coordinada, ya que esto forma parte del *shell* del sistema. No obstante, se da a continuación una breve descripción de la información relativa a este fenómeno que, por ser específica de cada lengua, forma parte de la gramática.

Las reglas de la gramática referentes a este fenómeno sirven para establecer cuáles son los rasgos (*features*) que deben asignarse a una estructura coordinada en función de los rasgos de sus elementos y de la conjunción que los une.

Veamos, mediante algunos ejemplos sencillos, cómo actúan estas reglas en el caso de la coordinación de sintagmas nominales y el efecto que esto tiene sobre el análisis. En el siguiente ejemplo, un sintagma formado mediante la coordinación con la conjunción y de dos estructuras con rasgos similares, hereda dichos rasgos:

```
(5) Acciones y palabras.
├ lconj acción1(1) noun(cn,f.pl,nwh)
├ top y(2,1,3) noun(cn,f.pl,nwh)
└ rconj palabra1(3) noun(cn,f.pl,nwh)
```

En el ejemplo anterior los rasgos morfológicos de ambos nombres son totalmente equivalentes, ya que ambos son nombres comunes femeninos en plural. Pero cuando esto no es así, las reglas de coordinación los combinan para producir una estructura de rasgos nueva. En el siguiente ejemplo, vemos cómo a partir de dos sintagmas nominales de distinto género, se construye otro de género masculino:

```
(6) Hombres y mujeres.
├ lconj hombre1(1) noun(cn,m.pl,nwh)
├ top y(2,1,3) noun(cn,m.pl,nwh)
└ rconj mujer1(3) noun(cn,f.pl,nwh)
```

Esta información es esencial para las ulteriores etapas del análisis de la frase, en las que se apliquen reglas de concordancia del sintagma coordinado con otros elementos. Según el análisis que acabamos de ver, *hombres* y *mujeres* concordará en género con *adultos*, pero no con *adultas*:

```
(7) Hombres y mujeres adultos.
├ lconj hombre1(1) noun(cn,m.pl,nwh)
├ top y(2,1,3) noun(cn,m.pl,nwh)
├ rconj mujer1(3) noun(cn,f.pl,nwh)
└ nadj adulto2(4) adj(m.pl,X1,X2)

(8) Hombres y mujeres adultas.
├ lconj hombre1(1) noun(cn,m.pl,nwh)
├ top y(2,1,3) noun(cn,m.pl,nwh)
├ rconj mujer1(3) noun(cn,f.pl,nwh)
└ nadj adulto2(4) adj(f.pl,X1,X2)
```

En (7) el adjetivo modifica al conjunto de la estructura coordinada, mientras que en (8) modifica sólo al segundo elemento.

En el siguiente ejemplo, los dos sintagmas coordinados concuerdan en género, pero el rasgo de número ha quedado sin instanciar:

(9)	Él o su hermano.	
└─	lconj él1(1)	noun(pron(pers), pers3.m.sg.nom.X3, nwh)
└─	top o(2,1,4)	noun(pron(pers), pers3.m.X1.nom.X2, nwh)
└─	ndet su1(3)	det(m.sg, poss)
└─	rconj hermano1(4)	noun(cn, m.sg, nwh)

Esto se debe a que una estructura coordinada de estas características puede concordar en español tanto en singular como en plural:

Vendrá él o su hermano.

Vendrán él o su hermano.

En estas frases, el rasgo de número de la estructura coordinada *él o su hermano* quedará instanciado por unificación a *sg* o a *pl* al aplicarse las condiciones de la regla de llenado del *slot* de sujeto que requieren que este concuerde en número con el verbo.

• Reglas de puntuación

La puntuación, al igual que la coordinación (con la que a menudo se solapa ya que algunos signos de puntuación, especialmente la coma, pueden actuar como elementos coordinantes), es tratada sobre todo en el *shell* del sistema, de modo que tampoco nos detendremos en pormenores sobre este tema. No obstante, existe en la gramática un tipo de reglas que permite especificar posibles peculiaridades de una lengua en el uso de los signos de puntuación. Se trata de las *reglas de separadores*.

Los *separadores* son determinados signos de puntuación que separan a un modificador del sintagma al cual modifica. No cualquier tipo de modificador admite un separador. Por ejemplo, en español normalmente no se intercala una coma entre el verbo y el sujeto o entre el verbo y el objeto directo:

*Ha venido, tu hermano.

*Vi, a tu hermano.

Cuando el analizador intenta llenar un *slot* con un sintagma, comprueba la existencia de separadores y si estos son o no compatibles con lo establecido por las reglas de separadores para dicho *slot*. Sin embargo, no es necesario declarar reglas de este tipo para cada *slot*, ya que en ausencia de las mismas, el analizador aplica unos valores por defecto. Si en la gramática no se especifica de otra manera, los *slots* de tipo complemento no admiten la presencia de separadores. Para los de tipo adjunto, se permite opcionalmente una coma. Estos valores por defecto coinciden generalmente con las normas que rigen el uso de separadores para la lengua española, por lo cual, en la gramática del español, prácticamente no se usa este tipo de reglas.

• Reglas de evaluación de análisis

Como se ha dicho en la página, el analizador hace uso de un sistema de evaluación numérica de análisis que permite no sólo establecer un orden de preferencia entre los análisis finales obtenidos (en caso de haber más de uno), sino también, aunque de manera opcional, descartar aquellos análisis parciales de menor preferencia de entre los obtenidos para un mismo segmento de la frase.

Este sistema de evaluación forma parte del *shell* del sistema, pero buena parte de la información que utiliza está contenida en la gramática, en forma de reglas de evaluación de análisis¹⁴. Estas reglas permiten *penalizar* o *premiar* de manera ponderada un análisis cuando se cumplen determinadas condiciones. Hay dos clases de reglas de evaluación de análisis, que veremos a continuación.

Reglas de evaluación de slots: Durante el proceso de análisis, el analizador asigna a cada sintagma o segmento de frase analizado una evaluación numérica que permite compararlo con otros análisis obtenidos para el mismo segmento. En el cálculo de dicha evaluación intervienen básicamente tres factores, relacionados con: (1) proximidad de los modificadores con la palabra que modifican (*close attachment*), (2) paralelismo de los elementos coordinados, y (3) evaluación de *slots*. Dado que de los dos primeros factores se ocupa el *shell* del sistema, aquí sólo nos ocuparemos del tercero.

La evaluación de *slots* de un análisis es la suma de una serie de valores, asociado cada uno de ellos a uno de los *slots* que han sido llenados en los distintos nodos del árbol de análisis. El valor de evaluación asociado a cada *slot* es el establecido por la regla de evaluación de *slots* aplicable al mismo. En caso de no existir reglas de evaluación para un *slot* determinado o de no cumplirse las condiciones establecidas en las mismas, se toma un valor por defecto, que es 0 para los *slots* de tipo complemento y 1 para los de tipo adjunto.

Veamos algunos ejemplos sencillos tomados de SSG:

- Para el *slot* *ndet* (determinante que modifica a un nombre), se establece una evaluación de 0 cuando el nodo superior corresponda a un nombre común. Esto equivale a *despenalizar* en ese caso el empleo de este *slot*, ya que, al tratarse de un *slot* de tipo adjunto, el valor por defecto sería 1.
- Por el contrario, el *slot* *objprep*(_) (objeto de la preposición), cuya evaluación por defecto sería 0, es penalizado con la evaluación 1 cuando lo llena un sintagma adjetivo, ya que aunque está permitido que este *slot* lo ocupe un adjetivo, es más frecuente que lo ocupen otros tipos de sintagmas (sintagmas nominales o adverbiales, construcciones de infinitivo, oraciones completivas).
- También está penalizado (con la evaluación 3) el *slot* *vprep* (sintagma preposicional que modifica a un verbo), cuando la modificación se haya realizado por la izquierda y no haya separador.

Reglas de evaluación de sintagmas: Estas reglas permiten

- alterar la evaluación de un análisis, o bien
- descartar un análisis como no válido,

en el momento en que el sintagma analizado va a modificar a otra palabra de la frase (o una vez obtenido un análisis para la totalidad de la frase), si se cumplen para dicho sintagma determinadas condiciones establecidas en las mismas.

¹⁴ También puede incluirse información que afecte a la evaluación de análisis en las entradas de diccionario.

Las reglas de evaluación de sintagmas se usan en mucho mayor medida en la gramática del español que en otras versiones de *Slot Grammar* desarrolladas hasta ahora para otras lenguas (principalmente, lenguas germánicas). La mayor flexibilidad de la estructura sintáctica del español en comparación con esas lenguas hace que no se puedan establecer restricciones demasiado rígidas en las reglas de llenado de *slots* o en las de orden sintáctico. De modo que, en muchos casos, se permite una cierta flexibilidad en cuanto al llenado de determinados *slots*, pero en un estadio posterior del proceso se comprueba si el análisis obtenido reúne determinados requisitos.

Veamos cómo se aplica una de estas reglas en el análisis de la frase

Cuando llueva, lea un libro.

La palabra *lea* tiene cuatro posibles análisis léxicos para una determinada acepción del verbo *leer*: uno para el imperativo, y tres para las formas de presente de subjuntivo correspondientes a las personas primera, segunda (tratamiento de cortesía) y tercera. Evidentemente, en el contexto de esta frase, sólo puede ser interpretada como una forma de imperativo. Pero en un analizador ascendente no se pueden descartar las otras tres posibilidades hasta completar el análisis de la frase, ya que durante el proceso de análisis pueden encontrarse elementos de la frase que permitan que este verbo sea analizado como subjuntivo. Por ejemplo, la aparición de un determinado adverbio:

Quizá, cuando llueva, lea un libro.

Una regla de la gramática descarta los análisis correspondientes a sintagmas verbales con rasgo de subjuntivo cuando:

- existe un análisis alternativo con el rasgo de imperativo para la palabra núcleo del sintagma,
- no existe un modificador consistente en un determinado tipo de adverbios (como *quizá*),
- el sintagma no tiene el rasgo *dep* que señala una oración dependiente o subordinada.

Veamos otro ejemplo. Las palabras *el día nublado* pueden ser analizadas como sintagma nominal. Sin embargo, en la frase

Ayer amaneció el día nublado.

no cabe interpretar *el día nublado* como sujeto, ya que aquí *nublado* hace la función de atributo del sujeto, por lo cual no debe modificar al nombre, sino al verbo. Para ello, debe llenar el *slot* *comp(a)* reservado a este tipo de atributos. Nuestra gramática consigue dar preferencia a esta lectura correcta de la frase, gracias a una regla que incrementa en 1 la evaluación de un sintagma nominal cuando:

- contiene un modificador que llena un *slot* *nadj*, y
- el núcleo del nodo superior (es decir, la palabra que va a ser modificada por el sintagma) es un verbo cuyo *slot frame* contiene un *slot* *comp(a)* que aún no ha sido llenado.

Algunos fenómenos propios del español

A continuación nos detendremos en algunos de los fenómenos que, por ser característicos (si bien no exclusivos) del español, han requerido un tratamiento gramatical diferente del propuesto en las versiones de *Slot Grammar* desarrolladas con anterioridad para otras lenguas.

• La ambigüedad morfosintáctica

Quizás sea este el problema de mayor envergadura al que se ha tenido que hacer frente durante el desarrollo de la gramática. La mayor ambigüedad morfosintáctica inherente al español en comparación con otras lenguas¹⁵, se debe sobre todo a tres factores:

- mayor flexibilidad en el orden sintáctico superficial,
- coincidencia de formas verbales y pronominales correspondientes a distintas personas gramaticales, y
- sujetos elípticos.

Estos dos últimos factores se combinan a menudo de tal manera que una frase aislada del contexto puede tener múltiples lecturas alternativas. Veamos, por ejemplo, esta oración final:

para que se la describa

La forma verbal *describa* tiene cuatro posibles análisis léxicos para una determinada acepción del verbo *describir*:

- primera persona singular, presente de subjuntivo
- tercera persona singular, presente de subjuntivo
- segunda persona singular (tratamiento de cortesía), presente de subjuntivo
- segunda persona singular (tratamiento de cortesía), imperativo

Sabemos que no cabe analizar en este contexto *describa* como forma de imperativo, ya que: (1) pertenece a una oración subordinada, y (2) los pronombres aparecen en posición proclítica. Aparte de esto y dada la omisión del sujeto, la asignación de esta forma a una de las tres personas del presente de subjuntivo no puede hacerse a partir de criterios morfosintácticos, sino exclusivamente según el contexto o la situación. Por lo tanto, la gramática generará análisis alternativos de la frase para cada una de ellas¹⁶.

Por otra parte, los pronombres *se* y *la* tienen también distintos análisis léxicos posibles. Para el primero tenemos al menos ocho análisis correspondientes a las distintas combinaciones de los rasgos de persona, número y caso:

- tercera persona / segunda persona (tratamiento de cortesía)
- singular / plural
- dativo / reflexivo

¹⁵ Téngase en cuenta que tomamos como punto de referencia aquellas lenguas para las que se han desarrollado prototipos de *Slot Grammar*. Sin embargo, algunos de los fenómenos de los que tratamos tienen una correlación bastante estrecha en otras lenguas románicas.

¹⁶ La información referente a persona gramatical puede ser imprescindible, por ejemplo, a la hora de traducir la frase a otro idioma en el que esta distinción deba hacerse explícita.

El pronombre *la* sólo puede corresponder al acusativo de la tercera persona o de la segunda persona (tratamiento de cortesía). Además, esta palabra puede ser analizada como forma femenina singular del artículo *y*, también, como nombre común (nota musical).

De modo que para las tres palabras *se*, *la* y *describa* obtendremos un total de 128 permutaciones de posibles análisis léxicos. Las restricciones establecidas en las reglas de la gramática evitan que obtengamos análisis para todas ellas. Así, por ejemplo, un pronombre reflexivo sólo puede modificar a un verbo con el que concuerde en persona y número; el objeto directo y el indirecto no pueden corresponder a un mismo referente, por lo cual se puede descartar el análisis simultáneo de ambos pronombres como formas de segunda persona; etc. No obstante, aún teniendo en cuenta las distintas restricciones de carácter morfosintáctico, se tiene un conjunto considerable de posibles análisis, según los rasgos de persona y número que se atribuyan al sujeto, al objeto directo y al objeto indirecto. La siguiente tabla muestra las 17 posibles lecturas correctas de la frase. La elección del análisis correcto requiere la desambiguación de la oración en función del contexto mediante la resolución de las anáforas y de la elipsis del sujeto.

Una de las maneras en que SSG descarta lecturas de una frase incorrectas desde el punto de vista morfosintáctico es mediante una regla de evaluación de sintagmas que comprueba que no aparezcan en un mismo contexto formas verbales o pronominales de segunda persona correspondientes una al tratamiento familiar y otra al de cortesía. Por ejemplo, de la frase

Quieres que se la describa.

sólo se obtienen 5 análisis a pesar de que la oración completiva (*que se la describa*) por sí sola tiene 17. La aparición de una forma verbal de tratamiento familiar (*quieres*) permite descartar todos los análisis de *se*, *la* y *describa* como formas de segunda persona.

Además, para algunos verbos, pueden establecerse restricciones de tipo semántico en cuanto al sintagma que debe llenar determinados *slots*. El formalismo léxico prevé la posibilidad de establecer estos tipos semánticos en las entradas de diccionario. Por ejemplo, si se quisiera especificar que el sujeto y el objeto indirecto de *escribir* deben tener el rasgo *humano*, la entrada podría tener el siguiente aspecto:

`escribir < v(subj(n):h,obj.iobj:h.loc,22)`

La letra *h*, separada por dos puntos del nombre del *slot*, indica que el sintagma que lo llene deberá tener dicho tipo semántico.

Reduplicación del objeto directo y del objeto indirecto

Es un fenómeno habitual en español la reduplicación mediante un pronombre átono de los complementos directo e indirecto en determinadas circunstancias. El hecho de que un mismo argumento del verbo esté representado por dos sintagmas independientes plantea algunos problemas en nuestro formalismo.

Tabla 1. Posibles lecturas de la oración final 'para que se la describa'		
SUJETO	O.D.	O.I.
YO	ELLA	ÉL/ELLA
YO	ELLA	ELLOS/ELLAS
YO	ELLA	USTED
YO	ELLA	USTEDES
YO	USTED	ÉL/ELLA
YO	USTED	ELLOS/ELLAS
ÉL/ELLA	ELLA	ÉL/ELLA
ÉL/ELLA	ELLA	ÉL/ELLA (refl.)
ÉL/ELLA	ELLA	ELLOS/ELLAS
ÉL/ELLA	ELLA	USTED
ÉL/ELLA	ELLA	USTEDES
ÉL/ELLA	USTED	ÉL/ELLA
ÉL/ELLA	USTED	ÉL/ELLA (refl.)
ÉL/ELLA	USTED	ELLOS/ELLAS
USTED	ELLA	ÉL/ELLA
USTED	ELLA	ELLOS/ELLAS
USTED	ELLA	USTED (refl.)

Como ya se ha dicho, los *slots* de tipo complemento (a diferencia de los de tipo adjunto) sólo pueden llenarse una vez. Para analizar los pronombres reduplicativos del objeto directo e indirecto, existen en la gramática dos *slots* de tipo adjunto del verbo llamados, respectivamente, *objdup* y *iobjdup*.

Estos *slots* se pueden llenar con un pronombre personal acusativo o dativo siempre que el *slot frame* del verbo contenga el *slot* respectivo *obj* () o *iobj* (), tanto si este ha sido llenado como si permanece vacío. Una regla de evaluación de sintagmas verificará posteriormente que en cada análisis en que un pronombre llene un *slot* de tipo adjunto *objdup* o *iobjdup*, haya otro sintagma que llene el *slot* de tipo complemento correspondiente.

Veamos un ejemplo. En la frase

A ellos les veo cada día.

el segmento *les veo* producirá distintos análisis parciales correspondientes, por una parte, a los distintos análisis léxicos del pronombre *les* (segunda y tercera persona), y por otra parte a los dos *slots* (*obj* () y *objdup*) que pueden ser llenados por el pronombre. Luego, el sintagma preposicional *a ellos* modificará a *veo*, llenando el *slot obj* (n). Esto sólo es posible si este *slot* está vacío, es decir, cuando *les* ha sido analizado como *objdup*. Además, como los sintagmas que llenan ambos *slots* deben concordar en persona y número, el análisis adecuado de *les veo* será aquel en que *les* haya sido analizado con el rasgo de tercera persona. El análisis final de la frase es:

(10) A ellos les veo cada día.

obj(n)	al(1,2)	prep(a,X2,2)
objprep(pl)	ellos2(2)	noun(pron(pers),pers3.m.pl.opr.X3,X4)
objdup	ellos2(3)	noun(pron(pers),pers3.m.pl.acc.X5,X6)
top	ver1(4,u,2)	verb(fin(pers1.sg,pres,ind:dcl:nwh))
ndet	cadal(5)	det(m.sg, indef)
vadv	díal(6)	noun(cn,m.sg,X7)

En el siguiente ejemplo sin reduplicación del objeto directo, el pronombre *les* ha llenado el *slot* obj(n). El análisis en que el pronombre llena el *slot* objdup ha sido descartado. Por otra parte, la ambigüedad en cuanto a la persona del pronombre queda ahora manifiesta en la existencia de dos análisis finales:

(11) Les veo cada día.

obj(n)	ustedes1(1)	noun(pron(pers),pers2f.m.pl.acc.X2,X3)
top	ver1(2,u,1)	verb(fin(pers1.sg,pres,ind:dcl:nwh))
ndet	cadal(3)	det(m.sg, indef)
vadv	díal(4)	noun(cn,m.sg,X4)

obj(n)	ellos2(1)	noun(pron(pers),pers3.m.pl.acc.X2,X3)
top	ver1(2,u,1)	verb(fin(pers1.sg,pres,ind:dcl:nwh))
ndet	cadal(3)	det(m.sg, indef)
vadv	díal(4)	noun(cn,m.sg,X4)

Pasiva refleja, verbos pronominales y otros usos del pronombre reflexivo

Los pronombres reflexivos, y en particular el pronombre *se*, cuentan en español con una riqueza de usos y funciones diversas que representan un auténtico desafío para el análisis gramatical. En la primera versión de SSG, no se ha intentado tratar este tema de manera exhaustiva, sino que se han agrupado estas funciones de la siguiente manera:

- Pasiva refleja
- Uso reflexivo (o recíproco) como objeto directo o indirecto
- Verbos pronominales

He aquí algunos ejemplos:

El partido se jugó ayer.
Carmen se considera inteligente.
Ella se aseguró de que todo estaba bien.

La primera frase es un caso de pasiva refleja, de significado equivalente a

El partido fue jugado ayer.

El sintagma nominal *el partido* ocupa el *slot* de sujeto en la estructura superficial de la frase, pero corresponde al objeto directo del *slot frame* de *jugar*. En estas construcciones de pasiva refleja, el pronombre *se* llena el *slot* especial *pass*, que sustituye al sujeto en el *slot frame* del verbo. Obsérvese que estas construcciones, a diferencia de la pasiva convencional, no admiten la presencia de un

complemento agente. Ya hemos visto en la página un ejemplo de análisis de una frase pasiva convencional. He aquí el análisis correspondiente a nuestro ejemplo de pasiva refleja:

(12)	El partido se jugó ayer.		
	ndet	el2(1)	det(m.sg,def)
	└─ subj(n)	partido2(2)	noun(cn,m.sg,X1)
	└─ pass	se9(3)	noun(pron(reflprn),X2.X3.X4.prefl.X5,X6)
	└─ top	jugar1(4,3,2)	verb(fin(pers3.sg,past,ind:dc1:nwh))
	└─ vadv	ayer1(5)	adv(X7,X8)

Este tipo de análisis no se construye sobre el *slot frame* del verbo que se especifica en la entrada de diccionario, sino sobre un *slot frame* especial generado según lo ya descrito en el apartado «El análisis léxico». Por ejemplo, si el *slot frame* original de la acepción transitiva de *jugar* contiene un *slot* de sujeto y otro de objeto directo, se genera, para las formas de tercera persona (y para las impersonales), un nuevo *slot frame* de pasiva refleja en el que los nuevos *slots* de *pass* y de sujeto paciente quedan vinculados a los *slots* originales.

Gracias a esto, y como puede observarse en el análisis del ejemplo anterior, el argumento de la *predicación de acepción* de *jugar* correspondiente al sujeto señala al nodo 3 (*se*), mientras el correspondiente al objeto directo señala al nodo 2 (*el partido*), con lo cual se logra reflejar la estructura profunda de la frase.

También se genera, para las formas de participio, un *slot frame* correspondiente a las construcciones de la pasiva convencional (*El partido fue jugado ayer*), sustituyendo el *slot* de sujeto por uno de complemento agente, y el de objeto directo por otro de sujeto paciente.

En SSG se tratan de igual forma que la pasiva refleja otras construcciones de estructura similar pero cuyo valor semántico es diferente. Veamos estas dos frases:

Antes que nada, debe encenderse el aparato.
En caso de error, se enciende una luz amarilla.

La primera constituye un ejemplo de pasiva refleja, pero la segunda tiene más bien un valor *anticausativo* o de espontaneidad.

El pronombre *se* usado en este tipo de construcciones (de pasiva refleja y anticausativas) no varía en forma, pudiendo variar el sujeto y el verbo en número:

El partido se jugó ayer.
Los partidos se jugaron ayer.

La luz se enciende.
Las luces se encienden.

Para los otros dos tipos de construcciones en que se utilizan pronombres reflexivos, estos concuerdan en persona y número con el verbo:

Carmen se considera inteligente.
Vosotros os consideráis inteligentes.

Ella se aseguró de que todo estaba bien.
Yo me aseguré de que todo estaba bien.

De estas frases, las dos primeras corresponden a un uso propiamente reflexivo del pronombre en función de objeto directo. He aquí el análisis obtenido para la segunda:

(13)	Vosotros os consideráis inteligentes.	
└ subj(n)	vosotros1(1)	noun(pron(pers),pers2.m.pl.nom.X1,X2)
└ obj(n)	vosotros3(2)	noun(pron(pers),pers2.X3.pl.ref.X4,X5)
• top	considerar1(3,1,2,4)	verb(fin(pers2.pl,pres,ind:cl:nwh))
└ comp(a)	inteligente1(4)	adj(X6.pl.X7,X8)

Podemos considerar este tipo de construcciones como un caso especial de construcciones transitivas, en las que sujeto y objeto directo se refieren a una misma entidad y, por lo tanto, concuerdan en persona y número con el verbo. Lo dicho para el objeto directo es aplicable también para el objeto indirecto, que también puede tener valor reflexivo, como en la frase

Él se lavó las manos.

Los verbos pronominales, aunque tienen frecuentemente correlación con un verbo transitivo, se diferencian de este uso reflexivo o recíproco de los verbos transitivos en que:

- tienen normalmente otro valor semántico (no es lo mismo *acordar algo* que *acordarse de algo*, *empeñar algo* que *empeñarse en algo*, *referir algo* que *referirse a algo*), y
- suelen tener una estructura argumental propia, que incluye a menudo un *slot* de complemento preposicional.

Para estos verbos, el pronombre reflexivo llena un *slot* *pron* especificado en la correspondiente entrada de diccionario. He aquí, a modo de ejemplo, la entrada correspondiente al verbo *empeñar*, para el cual hay una acepción transitiva y otra pronominal:

empeñar < v(obj1,1)
< v(pron1.p(en),1).

Función pronominal del determinante

Otro fenómeno característico del español es la presencia del determinante con carácter pronominal en sintagmas nominales en que el nombre es omitido, como en los siguientes ejemplos:

el de todos los días (*el pan de todos los días*)
los de Madrid (*los habitantes de Madrid*)
el que vino ayer (*el chico que vino ayer*)
lo que yo te diga (*las cosas que yo te diga*)

En nuestra gramática, se considera que el núcleo de estas frases es el determinante, que es modificado ya sea por un sintagma preposicional o por una cláusula de relativo.

Para aquellos *slots* que suelen ser llenados por sintagmas nominales (sujeto, objeto directo, etc.), existen reglas que les permiten ser llenados por un determinante siempre que este esté modificado por un sintagma preposicional o por una cláusula de relativo.

Cierta similitud con las que se acaban de mencionar tienen algunas estructuras como:

el caro
los mencionados
lo dicho

No obstante, para estos casos se ha preferido considerar que el núcleo del sintagma es un adjetivo o participio nominalizado por la presencia del determinante. Al igual que en el caso anterior, existen reglas que permiten que estos adjetivos o participios nominalizados llenen *slots* normalmente reservados para sintagmas nominales.

Resultados obtenidos

Para la evaluación y desarrollo de la gramática se han utilizado una serie de corpus lingüísticos de diverso tipo recopilados a tal efecto. Las frases contenidas en estos corpus provienen de distintas fuentes:

- ejemplos procedentes de varios diccionarios y manuales de gramática española;
- textos técnicos, especialmente manuales de IBM;
- artículos de prensa diaria.

Una vez incluida la información necesaria en el diccionario, la aplicación de la gramática SSG a las frases de dichos corpus ha arrojado unos resultados considerablemente positivos si tenemos en cuenta que se trata de un primer prototipo. Los análisis correctos alcanzan porcentajes satisfactorios para los textos de carácter técnico. Las limitaciones del sistema quedan, en cambio, manifiestas en lo referente a textos pertenecientes a un dominio más general. Entre los ejemplos procedentes de diccionarios y gramáticas existen numerosas frases que presentan estructuras de considerable complejidad, a menudo propias de registros lingüísticos menos formales, o bien, cuyo análisis resulta difícil al depender en mayor medida de consideraciones semánticas, contextuales o situacionales. Los textos tomados de artículos de prensa presentan también mayor dificultad, por tratarse frecuentemente de frases de mayor longitud y complejidad.

Uno de los errores más frecuentes en los análisis es la asignación de un modificador a un nivel erróneo del árbol, especialmente en el caso de los sintagmas preposicionales (*PP-attachment*). Otros errores se deben generalmente a limitaciones en el *shell* del sistema, que no contempla, en su versión actual, determinados fenómenos que no se dan en las lenguas para las que fue originalmente diseñado.

La cobertura lingüística alcanzada abarca un gran número de fenómenos sintácticos del español, entre los que cabe señalar, además de los ya mencionados, los siguientes: perífrasis verbales, oraciones subordinadas, frases de relativo, cláusulas absolutas, aposición, construcciones de infinitivo, gerundio y participio, interrogación.

Referencias

- [ARRARTE ET AL. 91A] Arrarte, G., Zapata, I. y McCord, M. C. «Spanish Generation Morphology for an English-Spanish Machine Translation System», Research Report RC 17058, IBM Research Division, Yorktown Heights, NY 10598, 1991.
- [ARRARTE ET AL. 91B] Arrarte, G., Carranza, B. y Redondo, T. «Tratamiento de Clíticos en un Sistema de Traducción Automática», *VII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural*, Valencia, 1991.
- [MCCORD 89A] McCord, M. C. «Design of LMT: A Prolog-Based Machine Translation System», *Computational Linguistics*, vol. 15, pp. 33-52, 1989.
- [MCCORD 89B] McCord, M. C. «A New Version of the Machine Translation System LMT», *Literary and Linguistic Computing*, 4, pp. 218-229, 1989.
- [MCCORD 89C] McCord, M. C. «LMT», *Proceedings of MT Summit II*, pp. 94-99, Deutsche Gesellschaft für Dokumentation, Frankfurt, 1989.
- [MCCORD 92] McCord, M. C. «The Slot Grammar System», Research Report RC 17313, IBM Research Division, Yorktown Heights, NY 10598; próxima publicación en J. Wedekind y C. Rohrer (Eds.), *Unification in Grammar*, MIT Press, 1992.

- [MCCORD ET AL.] McCord, M. C., Bernth, A., Lappin, S. y Zadrozny, W. «Natural Language Processing within a Slot Grammar Framework», Research Report RC 17538, IBM Research Division, Yorktown Heights, NY 10598; próxima publicación en *International Journal on Artificial Intelligence Tools*, 1992.
- [MOLINER] Moliner, M. «Diccionario de uso del español», Gredos, Madrid, 1982.
- [R.A.E.] Real Academia Española «Esbozo de una nueva gramática de la lengua española», Espasa-Calpe, Madrid, 1979.
- [RODRÍGUEZ ET AL.] Rodríguez, C., Sopena, L., Valladares, C. y Villar, C. «Clasificación morfológica del léxico castellano para un analizador en ordenador», *VII Congreso de AESLA*, Sevilla, 1989.
- [SECO] Seco, M. «Diccionario de dudas y dificultades de la lengua española», Espasa-Calpe, Madrid, 1986.

