

Gramática contrastiva computacional: MTT vs. HPSG

Oscar García Marchena

Universidad Paris 7 - Denis Diderot, Laboratorio de Lingüística formal (LLF)
2, place Jussieu - 75251 Paris
osqvar@gmail.com

Resumen

Este trabajo analiza los métodos empleados por dos gramáticas formales diferentes (una gramática sintagmática, HPSG y una gramática de dependencias, MTT) para implementar algunos elementos idiosincrásicos de la gramática del español. Se contrastan los mecanismos empleados por cada una, y los resultados obtenidos en cuanto a su capacidad de formalizarlos dentro del sistema, para determinar cuál es más apto para integrar los datos de la gramática española.

Palabras clave: gramática formal, HPSG, MTT,

Abstract

This paper analyses the methods employed by two different formal grammars (a phrase structure grammar, HPSG, and a dependency grammar, MTT) for implementing certain idiosyncrasies of Spanish grammar. We contrast the efficiency of the mechanisms they use for accounting for them, and their ability to formalise them, which will allow us to determine the fittest for Spanish grammar.

Keywords: formal grammar, HPSG, MTT

Resumé

L'analyse présentée ici présente les méthodes employées par deux grammaires formelles différentes (une grammaire syntagmatique, HPSG, et une grammaire de dépendances) afin d'implémenter certains aspects idiosyncratiques de la grammaire de l'espagnol. Les mécanismes utilisés seront comparés, et leur flexibilité pour les formaliser, ce que nous permettra de déterminer quel est le model le plus apte pour rendre compte de la grammaire de l'espagnol.

Palabras clave: grammaire formelle, HPSG, TST

Tabla de contenidos

1. Introducción
2. HPSG
3. MTT
4. Semejanzas y diferencias entre los mecanismos de ambos modelos
5. Algunos ejemplos
6. Conclusión
7. Referencias bibliográficas

1. Introducción

Varios son los formalismos gramaticales que proponen una descripción formal detallada del funcionamiento de los lenguajes naturales y de la interacción de sus diferentes módulos: léxico, fonología, morfología, sintaxis, semántica y pragmática. Estos formalismos han sido en su mayoría implementados para realizar diversas aplicaciones informáticas.

De estos formalismos, un tipo específico que ha recibido una atención especial tanto por informáticos como por lingüistas son las llamadas *gramáticas de unificación*. Cabe destacar los modelos de la gramática léxica funcional (*Lexical Functional Grammar* o *LFG*), la gramática sintagmática nuclear (*Head-driven Phrase Structure Grammar* o *HPSG*), la gramática de árboles adjuntos (*Tree Adjoining Grammar* o *TAG*) y la teoría sentido-texto (*Meaning-Text Theory* o *MTT*). Entre estos, los modelos HPSG y MTT han sido implementados con éxito para la gramática del español, por los laboratorios de la *Universidad Nacional Autónoma de México* (REF) y de la empresa de Inteligencia Artificial *VirtuOz* en Francia (REF) respectivamente.

Estos formalismos, que sirven tanto de teorías como de instrumentos para el tratamiento del lenguaje, difieren en el modo de conceptualizar una misma descripción sintáctica del funcionamiento de la lengua. Estos formalismos diferentes, con métodos diferentes, son adaptables a cualquier lengua, pero al estar en principio diseñados para tratar problemas de una lengua concreta, ¿son adaptables con facilidad a otras lenguas diferentes? ¿Hay formalismos que se adaptan mejor a unas lenguas que otras? Los modelos chomskianos (G&B, PP, teoría Minimalista, etc.) fueron desarrollados para explicar fenómenos del inglés. MTT fue creado en Rusia durante la guerra fría, para el ruso.

Su aplicabilidad al español requiere un trabajo de análisis de sus particularidades sintácticas y de sus posibilidades de integración en el formalismo. Por ejemplo, el orden de palabras: Para explicar la asimetría entre representaciones sintagmáticas y la movilidad de algunos sintagmas, se distinguen las reglas de precedencia lineal (LP) de reglas de dominación inmediata (ID).

HPSG analiza el orden de palabras, pero es una teoría construida para explicar fenómenos lingüísticos del inglés, con orden de palabras rígido. Su aplicación a otras lenguas, como el alemán, con distintas configuraciones de orden de palabras, precisa módulos suplementarios para gestionar el orden de palabras.

MTT, al estar construida para el una lengua en la que el orden de palabras es más flexible, está construida de forma diferente. El análisis sintáctico separa dependencias de orden de palabras de forma diferente. El orden de palabras (LI) conforma la sintaxis superficial, y las dependencias (DO) la sintaxis profunda. La separación es más acentuada, porque están dispuestos en dos módulos diferentes: dependencias y topología, con poca interacción entre ellos. Órdenes de palabras no canónicos como el *fronting* se convierten en este modelo en simples etiquetas para indicar emplazamientos “no canónicos” de algunos complementos.

Entonces, ¿es mejor, para el español, un formalismo que separe el orden de palabras, o uno que explique las interacciones de dependencias y orden? Responder a esta pregunta requiere responder primero a la pregunta: ¿Los formalismos son dependientes de la lengua? Las modelos formales ¿son suficientemente generales para explicar los fenómenos de cualquier lengua?, ¿o hay modelos más adaptados que otros a los datos de una lengua dada?

Para responder a estas preguntas vamos a contrastar dos formalismos, dos gramáticas formales que son a al vez teorías sobre el funcionamiento de la lengua, y

base de implementaciones informáticas para la creación de aplicaciones industriales de procesamiento del lenguaje natural. Tomamos algunos elementos idiosincrásicos de la sintaxis del español, y analizamos su tratamiento en HPSG y en MTT, para ver cómo puede modelizarlos cada uno, ver si un método es más fácilmente aplicable que el otro para explicar estos fenómenos del español.

Estos dos formalismos han sido desarrollados ampliamente estas dos lenguas. Podemos formalizar sin problema los fenómenos lingüísticos del español cuyo comportamiento sintáctico sea similar. Por eso, nos centramos en los elementos distintos a las lenguas que han sido formalizadas, puesto q el análisis de los fenómenos comunes ha sido ya hecho para esas lenguas

2. HPSG

HPSG y MTT son dos modelos fuertemente lexicalizados de *gramáticas de construcciones*. En ellos cada unidad de la lengua es una instancia de una categoría, y cada categoría está descrita en función de su relación con otras categorías: qué subcategoriza o por quién puede ser subcategorizada, qué relación (función sintáctica y rol semántico) se establece entre ellas, etc.

El análisis gramatical equivale así a resolver un puzzle uniendo cada unidad según el molde dado por la categoría a la que pertenece. Por ejemplo, en MTT, el análisis de una frase como “María lee un libro” es el resultado de unir cada palabra: *María* y *libro* no seleccionan nada, *un* selecciona un nombre común, *libro*; por otro lado *lee* selecciona un grupo nominal, con quien mantiene la relación de sujeto (*María*), y otro más, su objeto (*un libro*). La identificación de cuál es el sujeto y cuál el objeto dependería de unas reglas de ordenación según el orden lineal.

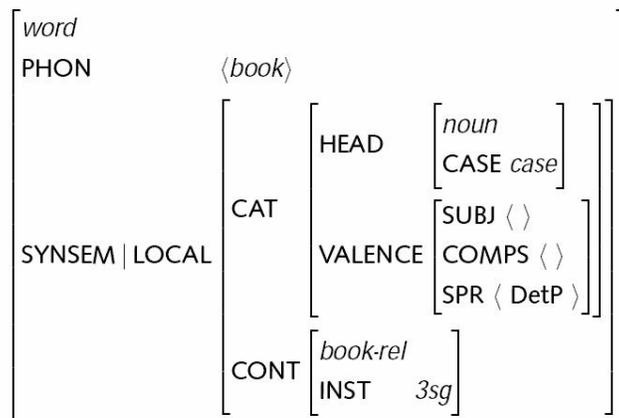
El formalismo utilizado para la confección de estas gramáticas es HPSG (Head-driven Phrase Structure Grammar o gramática sintagmática nuclear), y está implementada en la plataforma de desarrollo LKB (Linguistic Knowledge Building, Copenstake, 2002). Esta plataforma ha sido empleada ara diversos proyectos de innovación tecnológica, como *Verbmovil*, *DeepThought*, o *Aquilex* (Copenstake 2005).

HPSG (Ginzburg, J. & Sag, I, 2000) es, como indica su nombre, una gramática de estructuras sintagmáticas. Es un sistema formal compuesto por un léxico, al cual se le asigna la (sub)categoría de la que es un ejemplar. Cada categoría puede estar definida por sus posibilidades combinatorias, su semántica y sus propiedades morfológicas. Para ello, a cada categoría se le asigna una construcción especificando el conjunto de sus propiedades en forma de rasgos gramaticales. Estas categorías están organizadas jerárquicamente, de más general a más específica, y cada subtipo va heredando las restricciones de su supertipo y añadiendo otras. Cada palabra de este modo un ejemplar de signo (sintáctico-semántico), definido como un conjunto de características (rasgos): género, número, persona... que serán distintos según el tipo de categoría.

Los componentes principales de la gramática son así un léxico (por lo que esta gramática pertenece al conjunto de las gramáticas lexicalizadas) y un conjunto ordenado de construcciones (por lo que se le dice de construcciones) donde se indican las condiciones en las que puede aparecer (por lo que es de restricciones). Finalmente, estas construcciones se ensamblan entre sí gracias al mecanismo de unificación, por lo que también es una gramática de unificación. Otros componentes son los principios de buena formación, que explicitan los mecanismos de subcategorización, y la transmisión de rasgos gramaticales, y la ontología, que organiza también jerárquicamente los significados de cada unidad.

El signo lingüístico en HPSG tiene la siguiente estructura:

(1)



El signo queda descrito¹ entonces como una matriz de rasgos, en la que cada rasgo es un par de atributo-valor, y el valor puede ser a su vez una matriz de pares. Los valores Phon, para el valor fonológico, Synsem, para describir el conjunto de propiedades sintáctico-semánticas (Cat para la información sintáctica de la categoría, y Cont para su información semántica). *Local* se opone a *no local*, que contiene los rasgos necesarios para codificar las dependencias a distancia o extracciones. Limitamos aquí la descripción del signo a los rasgos locales. El núcleo tendrá como valor el tipo de categoría, y las valencias contendrán la información relativa a los potenciales sujeto, especificador (SPR), o complementos. Por último, en cuanto a la semántica, se especifican los índices (INST) de persona, número y género, y su significado, como un conjunto de relaciones semánticas (*X-rel*).

3. MTT

Este modelo parte del concepto de paráfrasis sinonímica: la asociación de un contenido o sentido a diversos *textos* o formas de expresión de una misma idea. La lengua se percibe así como un mecanismo de asociaciones entre un sentido y varios *textos*. La riqueza connotativa es sacrificada por la aplicabilidad computacional, por lo que este modelo ha tenido un éxito particular en su aplicación a los *lenguajes controlados* de la documentación técnica (Nasr: 1996). Esta asociación se lleva a cabo gracias a un modelo formal de transformación (de reducción de) una expresión lingüística a su significado, un metalenguaje lo suficientemente abstracto para formalizar la expresión de un mensaje en cualquier lengua.

MTT se inscribe así en el marco de las gramáticas formales, que integran datos sobre la semántica, sintaxis de dependencias, morfología y fonética, que interactúan según la estructura siguiente:

(2) Semántica $\langle == \rangle$ sintaxis profunda $\langle == \rangle$ sintaxis superficial
 Morfología profunda $\langle == \rangle$ morfología superficial $\langle == \rangle$ fonología

¹ De (Tseng 2004)

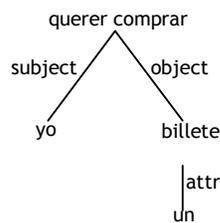
Es decir, que una producción lingüística pasa por varios estados de análisis. Primero, del análisis fonológico se localizan los fonemas que constituyen morfemas, y posteriormente esos morfemas se reducen a su forma « canónica », o flexión menos marcada (lo que equivale a hacer una tripartición del binomio lema – token, correspondiendo el análisis lemativo a la morfología profunda). Igualmente, las relaciones entre palabras se analizan en tres etapas: sintaxis superficial, en la que se perciben las relaciones sintagmáticas de las palabras en el orden en que aparecen, sintaxis profunda, donde las palabras están representadas según sus relaciones de dependencias, siguiendo el modelo de Tesnière (Tesnière: 1959), y finalmente la semántica, donde se incluyen el significado de cada palabra, y el rol semántico de cada relación entre palabras, para así generar una representación completa del significado de una frase.

(3) Niveles de análisis o reescritura en MTT: ejemplos.

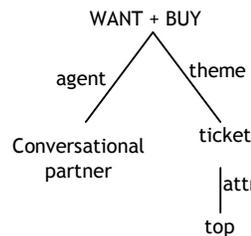
<i>ejemplo</i>	<i>nivel de análisis</i>
quiero comprar un billete	Fonología
quiero comprar un billete	Morfología superficial
querer[V: 1ªp, sg, pres, ind] comprar[V: inf] un[Det: num=1] billete[Nc: Gen=com, num=sg]	Morfología profunda
orden lineal (6)	Sintaxis superficial
árbol de dependencias (4)	Sintaxis profunda
grafo semántico (5)	Semántica

El paso entre cada fase de análisis queda asegurado por un mecanismo formal preciso, consistente en sucesivas “reescrituras” en el sentido generativo. Una cadena de caracteres (*morfología superficial*) se reescribe normalizada (con su *morfología profunda*), y posteriormente su *sintaxis superficial* se reescribe en un árbol de dependencias, y éste en un grafo semántico. MTT se inscribe así en la corriente de gramáticas generativas no transformacionales, gramáticas formales, de restricciones o de reescritura, como lo son HPSG (*Head-driven Phrase Structure Grammar*), LFG (*Lexical Functional Grammar*), o TAG (*Tree Adjoint Grammar*). Como ellas, MTT recurre al mecanismo de unificación para asegurar las asociaciones entre las diferentes fases en su implementación. La gramática resultante recibe el nombre de GUST, o *Gramática de Unificación polarizada sentido-texto* (Kahane: 2004).

(4) Sintaxis profunda (árbol de dependencias)



(5) Grafo semántico



Tanto el árbol de dependencias como el árbol semántico se representan mediante grafos, lo que permite poner en común la sintaxis de dependencias con los modos operativos de la teoría de grafos. Cada palabra está representada por un nodo, y las relaciones entre los nodos por arcos, etiquetados según su función sintáctica o semántica.

El paso de la sintaxis superficial a las sintaxis profunda o de dependencias, la llamada *cartografía*, es el resultado de la aplicación de unas reglas de orden lineal, que serán propios a cada lengua, según el orden de palabras. Los parámetros que se tienen en cuenta son la relación entre dos categorías (*Cat 1* y *Cat 2*), y la función sintáctica y semántica que se establece entre ellos. Dados unos requerimientos respecto a la naturaleza de la categoría dada, se establece un índice de linealidad que marca la posición de un elemento respecto al otro. El orden de palabras se conceptualiza como una línea horizontal en la que la primera categoría (*Cat 1*) ocupa la posición central, y el índice indica la posición de la segunda categoría (*Cat 2*) respecto a la primera. Así, un índice negativo señala que la *Cat 2* se sitúa a la izquierda de *Cat 1* y viceversa.

(6) Orden lineal para « *quiero comprar un billete* »

Función	Cat 1	Cat 2	Lin. Index	Lin 1 Req	Lin 2 Req
sujeto	V	Pron	-3	modetemps = flex	case = nom
OD	V	Nc	+5	---	---
Attr	Nc	Det	-1	---	---

En (6) se indican las reglas de linealidad que rigen el paso de la sintaxis superficial a la sintaxis profunda. Un pronombre que tiene con un verbo una relación de sujeto se sitúa a la izquierda de éste, siempre que el pronombre vaya en caso nominativo y el verbo tenga marcas de flexión (es decir, que no sea un infinitivo). La movilidad de complementos puede indicarse con la presencia de reglas para diferentes posiciones en la frase. Así, para permitir la posibilidad de un sujeto postverbal, basta con añadir una regla en la que el índice de linealidad sea positivo, con una distancia que tenga en cuenta la posible inserción de complementos entre el verbo y el sujeto. Esta computación del orden de palabras, si bien resulta poco visible al lingüista humano, permite describir con precisión un modelo intuitivo en el que el orden de palabras está estrechamente ligado a la estructura de dependencias: en una lengua donde la posición de los complementos goza de cierta movilidad, como es el caso en español, es fácil distinguir cómo los complementos de un núcleo *orbitan* alrededor de él en jerarquías concéntricas. De este modo, resulta sencillo marcar las posiciones posibles de un complemento, y la información añadida que aporta esta posición. Por ejemplo, este sistema es capaz de describir las variaciones de (7a-e), añadiendo a la categoría en una posición marcada un rasgo (p.e.: [*focus = +*]) que indique el aporte potencial de esa posición al significado global de la frase. Sin embargo, en el estado actual de aplicabilidad del modelo, esta información se tiene en cuenta en pocos casos, como la indicación de la alteración de orden en las oraciones interrogativas.

- (7a) *He visto a la chica*
- (7b) *A la chica la he visto*
- (7c) *A la chica la he visto yo*
- (7d) *He visto yo a la chica*
- (7e) *A la chica yo la he visto*

Todo este proceso de reescrituras sucesivas, ejemplificado en (4), requiere ciertos componentes formales para su realización: un diccionario de lemas con su naturaleza categorial y sus flexiones posibles que permita identificar la morfología superficial y pasar a la profunda. Igualmente es necesaria una gramática de construcciones que indique las posibilidades de combinación de las categorías, y sus relaciones de dependencia. Estas descripciones deben ser a la vez sintácticas y

semánticas, para permitir el paso de la sintaxis profunda a la semántica. Finalmente, la correspondencia entre la sintaxis profunda y la superficial requiere unas reglas que describan la ordenación lineal de los sintagmas respecto a sus dependencias, es decir, la cartografía proyectada físicamente por el nivel abstracto de dependencias.

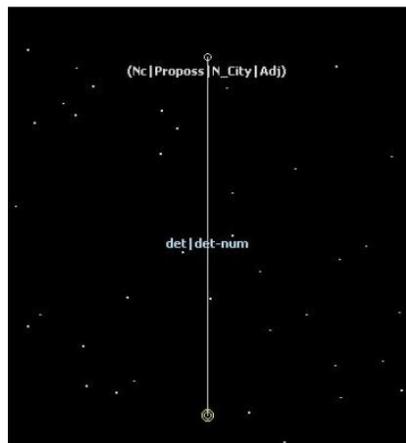
4. La Gramática de Unificación Polarizada (GUP)

Este modelo representacional ha sido formalizado como una de las llamadas “nuevas gramáticas” basadas en la unificación. Estas gramáticas generativas no transformacionales de descripciones sintagmáticas, altamente lexicalizadas a veces (como HPSG), que contienen restricciones a las combinaciones de unidades léxicas y sintagmáticas en forma de rasgos codificados en tipos léxicos, proporcionan a MTT la mecánica necesaria para su implementación. Algunas de estas gramáticas son HPSG, o *gramática sintagmática nuclear*, LFG o *gramática léxica funcional* y TAG, o *gramática de árboles adjuntos*. Siguiendo estos modelos, la teoría Sentido-Texto ha sido implementada primero bajo el nombre de *gramática de unificación de la teoría sentido-texto* o GUST (Kahane: 2002); y más tarde como *gramática de unificación polarizada* o GUP (Kaen 2004).

Partiendo del hecho de que contamos con un léxico categorizado, y con una descripción de la estructura sintagmática en la que aparece cada categoría, GUP permite conectar todas las construcciones en una sola, generando así un análisis completo de la estructura de la frase. Las construcciones están codificadas en forma de grafos, compuestos por nodos y arcos donde los nodos representan las categorías subcategorizadas por cada construcción, y los arcos las relaciones sintácticas que se establecen entre ellos.

El mecanismo de unificación es simple: consiste en un marcado sobre los nodos que está presente como información lingüística, la saturación. Los nodos no estarán saturados si requieren la presencia de esta categoría para completar la construcción, y un nodo no saturado podrá conectarse solamente con un nodo saturado con la categoría y restricciones morfológicas indicadas por la construcción.

(8) Estructura de la categoría *Determinante*



La estructura de un nombre consistirá así en un nodo simple saturado, mientras que la de un determinante estará compuesta por dos nodos unidos por un arco donde el nodo inferior será el propio determinante, el nodo superior un nombre no

saturado, y el arco la relación de determinación. Así, el nodo no saturado de la construcción del determinante se conectará al nodo solitario saturado de la construcción del nombre.

Se precisa un léxico, que contenga la información aportada por su morfología, así como el lema que le corresponde, pues cada entrada del léxico es un token. Además es necesario identificar su categoría gramatical. En tercer lugar, una ontología permitirá asignar a cada lema un significado normalizado, que será el mismo para los lemas sinónimos, así como indicar las relaciones de hiponimia e hiperonimia.

También es necesario un conjunto de construcciones gramaticales, en forma de grafos con información sobre las funciones sintácticas entre los elementos de un sintagma, así como los roles semánticos que subcategoriza un núcleo para sus complementos. Esto no se limita a los núcleos verbales, sino a la totalidad de las relaciones entre palabras, si bien el número de roles es muy reducido. Finalmente, unas reglas de linealidad identifican qué categorías se combinan entre ellas según su posición en la frase. Estas reglas también nos sirven para preferir una combinación a otra; por ejemplo, en una frase española donde la marca semántica “a” de objeto directo está ausente, el sujeto será preferentemente el SN situado a la izquierda, y el OD el de su derecha.

4. Semejanzas y diferencias entre los mecanismos de ambos

A pesar de las diferencias entre estos dos modelos, sólo la separación del tratamiento de orden de palabras en un módulo separado (topología) distingue, en principio, la configuración básica de ambos formalismos. La consecuencia de esto es, en consecuencia, que MTT no estudia las frases de una lengua en cuanto a su estructura sintagmática, sino sólo en cuanto a las relaciones de selección.

Las demás diferencias son principalmente consecuencia del distinto desarrollo de ambas; en HPSG la ergonomía arquitectural es un objetivo primordial, por lo que incorpora ciertos mecanismos de optimización, tomados de la lógica formal, que podrían incorporarse a MTT sin que estos añadidos afecten a su arquitectura. Ambos sistemas parten del **lexicalismo**: la inclusión en las entradas léxicas de un máximo de información; no solo los valores correspondientes a los atributos de género, número, persona, tiempo, aspecto, etc., sino también la indicación de la subcategorización, las restricciones de coaparición con otros elementos, restricciones sobre los valores de ciertos atributos de unidades subcategorizadas, etc.

Cada tipo (categoría o construcción) está presentado como una matriz de rasgos. La diferencia principal es que en HPSG podemos encontrar valores recursivos, es decir, que los valores pueden ser a su vez rasgos (pares de atributo y valor), mientras que en MTT solo tenemos rasgos binarios (un atributo y un valor).

Igualmente, ambos formalismos, asocian las entradas léxicas a categorías, presentadas como construcciones. También los dos cuentan con ontologías semánticas; la diferencia es que en HPSG, las relaciones semánticas que se establecen entre las unidades de la lengua vienen dadas por el significado de cada una de ellas, mientras que en MTT las relaciones semánticas entre palabras o sintagmas están descritas como roles semánticos (agente, paciente, tema, etc.), y son independientes del significado de las unidades. Un análisis detallado de la interacción entre semántica léxica y roles semánticos enriquecería ampliamente el modelo MTT.

Una de las diferencias sistémicas es el tratamiento de la correferencia: en HPSG los valores correferentes, ya sea de valores de atributos o puramente léxicos, está indicado

por la coindización: un índice numérico idéntico en varios valores indica que este valor es compartido. Esto resulta particularmente útil para las descripciones de concordancia, ya que los modificadores pueden especificar que sus valores concordantes serán, sean los que sean, los mismos que los de los núcleos de los que son adjuntos. MTT podría integrar este sistema de coindización, más flexible que el que posee en su implementación GUP.

En este sistema, los dependientes pueden transmitir valores a los núcleos de los que dependen (hacerles *percolar*). Sin embargo, este reparto de valores es unidireccional. Por otro lado, para las anáforas léxicas, GUP incorpora un tipo de relaciones, *cuasidependencias* que indican la correferencia de valores léxicos, por ejemplo, entre el sujeto de un verbo de control sujeto y el de su infinitivo, etc. La consecuencia de esto es que las estructuras de GUP no tienen ya forma de árboles, sino de grafos.

Sin embargo, este sistema no es trata de manera homogénea la anáfora léxica, puesto que sólo puede codificarse en relaciones de subcategorización, quedando así sin marcar en los adjuntos (En *Te prometo que vendré* se indica la correferencia del sujeto, mientras que en *Estoy en mi casa* no se hace), ya que las cuasidependencias se codifican a nivel léxico, y la anáfora de *Estoy en mi casa* identifica los valores de sintagmas separados, pero MTT no distingue sintagmas.

La gran diferencia entre los dos modelos es la disposición de los elementos; HPSG dispone categorías y valores en una jerarquía de herencia, donde cada elemento hereda los valores de sus ancestros y añade otros más específicos. Para lograr esto se necesita un gran número de categorías auxiliares, que recogen las características comunes de varios elementos, lo que aumenta el tamaño de la arquitectura, pero la dota de solidez en cuanto a explicación teórica. MTT, por el contrario, carece de esta ergonomía arquitectural y solidez teórica, lo que la hace más simple y legible.

Si bien a priori MTT no codifica la noción de sintagma, hay ciertas construcciones que de hecho sirven para codificar el mismo fenómeno: las estructuras exocéntricas (sin núcleo). Las preposiciones que introducen adjuntos requieren una construcción para adjuntarse al núcleo verbal, lo que equivale a un sintagma adjunto. Por este motivo, no podemos decir realmente que MTT no trate la estructura sintagmática de la lengua, pues las requiere para las construcciones que no están “dirigidas por el núcleo” (o *head-driven*). Por esto, no podemos decir que la diferencia principal entre HPSG y MTT resida en el tratamiento de los sintagmas de las frases de una lengua.

(9) Comparación de mecanismos en HPSG y MTT

	HPSG	MTT
Información en el léxico	sí	Sí
Estructura de rasgos	Sí (recursiva)	Sí (binaria)
Construcciones de categorías	Sí	Sí
Construcciones de sintagmas	Sí	Sí (egocéntricos)
Ontologie semantique	Sí	Sí
Correferencia	Sí	Sí (irregular)

5. Algunos ejemplos

El tratamiento del orden de palabras es una de las principales diferencias entre HPSG y MTT. La estructura de dependencias de MTT no da ninguna información

sobre el orden de palabras; el grafo sintáctico no está ordenado. En HPSG, sin embargo, el orden de la lista de complementos indica la función sintáctica: el primero es el sujeto, segundo el objeto directo, y después los complementos oblicuos. Es entonces una lista ordenada. Sin embargo, con un atributo llamado DOM, podemos reordenar la lista de sintagmas haciendo una correspondencia entre la lista de argumentos (dependencias) y el orden de su realización.

En MTT el orden de palabras es gestionado por el modulo topológico, que indica el orden de los dependientes con respecto a su núcleo en función de la función que se establece entre ellos. Este modulo permite codificar también el orden respecto a la presencia de un determinado rasgo. De este modo podemos capturar generalidades sobre el orden de palabras.

Pero ¿qué generalidades buscar para determinar qué nombre o pronombre (con sus posibles dependientes) funciona como sujeto y cuál como objeto directo en función del orden? Frases como “*el botón no lo he tocado yo*” y “*Yo no he tocado el botón*” nos informan de qué elementos marcan el orden de las dependencias. En este caso parece ser el clítico, cuya presencia indica que el objeto se realiza en posición preverbal (fronting). Para codificar esto, podemos considerar el clítico como un afijo verbal que marca este orden. El clítico graba en el núcleo verbal una información (un rasgo) que informa de este orden (marcado). Los objetos directos podrán ser entonces interpretados como tales en posición preverbal si el núcleo contiene este rasgo. Otro elemento desambiguador parecen ser el rasgo animado del sujeto, requerido por el verbo.

Una gran diferencia de la gramática española con respecto a la inglesa es la posibilidad de expresar el sujeto sólo mediante la morfología verbal. En español podríamos codificar esta información indicando que la presencia del sujeto, como a veces el objeto, es optativo. Sin embargo hay una diferencia: el sujeto pronominal (recuperado por la morfología verbal) no tiene las mismas propiedades que el objeto, sino las del clítico: índice a un referente externo, presencia de persona, caso, género, etc.

El tratamiento del francés nos da pistas: los pronombres sujetos pueden ser analizados también como críticos, siendo así posible decir que es posible tener frases compuestas solo de un verbo. La diferencia es que en francés los pronombres personales de sujeto son átonos, mientras que los españoles son tónicos, por lo que no podemos analizarlos como prefijos. Además, su presencia puede marcar contrastes informativos con respecto a otras personas del discurso. Por esta razón podemos decir que una frase no marcada carecerá de pronombre personal sujeto, y que la presencia de éste introducirá una marca que aporta esta información. De este modo, ambos formalismos tratan el fenómeno de prodrop como la opcionalidad de emplear un pronombre personal sujeto.

La presencia de una marca “a” en el objeto directo animado ha sido implementado en HPSG por (Marimón 2007), como un fenómeno de marcado del segundo argumento bajo ciertas condiciones. En MTT esta marca recibe un tratamiento particular; no está considerada propiamente una preposición, introduce un SN como segundo argumento, pero carece de la marca preposicional. En su lugar aporta el caso acusativo al nombre, siempre que éste sea animado.

6. Conclusión

Varias conclusiones nacen de estos contrastes; hemos visto cómo HPSG emplea métodos avanzados de ergonomía formal para dotar al sistema de consistencia teórica, y que estos mecanismos pueden incorporarse en MTT. El orden de palabras es, en MTT, una cuestión de reorganización de dependientes según sus rasgos; en HPSG se detalla toda la estructura sintagmática que permite cada realización. El grado de detalle de HPSG es mucho mayor, pero MTT es más flexible, al cubrir gran número de fenómenos con una mecánica relativamente simple. Esto se refleja asimismo en las posibilidades de sus aplicaciones prácticas respectivas en el dominio del procesamiento del lenguaje natural: HPSG es empleable para la generación de lenguaje, mientras que MTT se ha dirigido sobre todo para el reconocimiento del lenguaje natural.

7. Referencias bibliográficas

Abeillé, A. (1993) *Les nouvelles syntaxes: Grammaires d'unification et analyse du français*. Paris, Armand Colin Editeur.

Bender, E., Sag, I. and Wasow, T. (2003) *Syntactic Theory: A formal introduction*, Stanford: Stanford, CSLI Publications.

Bonami, O. (1999) "Constituency and word order in French subject inversion" en G. Bouma (ed), *Constraints and Resources in Natural Language Syntax and Semantics*, Stanford, CSLI Publications

Bouma, Goose (1998) *Categories, Prototypes, and Default Inheritance*. Gosse Bouma, Geert-Jan Kruijff & Richard Oehrle (eds). En *Proceedings of FHCG'98*

Copestake, A. (2005) *Deep Linguistic Processing with HPSG (DELPH-IN)* sur <http://wiki.delph-in.net/moin/FrontPage>.

C-ORAL-ROM Projet : <http://www.lablita.dit.unifi.it/cromdemo/index.html>

Garcia-Marchena, O (2005) *Une introduction à l'implémentation des relatives de l'espagnol en HPSG – LKB*. Tesina de DEA. Paris. Nouvelle Sorbonne Library.

Ginzburg, J. & Sag, I. (2000) *Interrogative investigations: the form, meaning, and use of English interrogatives*. Stanford, CSLI Publications.

Kahane, Sylvain. (2001) *The Meaning-Text Theory*. In *Dependency and Valency: An International Handbook of Contemporary Research*, Berlin: De Gruyter

Kahane Sylvain. (2002) *Grammaire d'Unification Sens-Texte – Vers un modèle mathématique articulée de la langue*, Hdr of university Paris 7, 82 p.

Kahane, Sylvain. (2003) *Les signes grammaticaux dans l'interface sémantique-syntaxe d'une grammaire d'unification*, TALN 2003

Kahane S. (2002) *Grammaire d'Unification Sens-Texte : vers un modèle mathématique articulé de la langue*, Tesis de habilitación a la investigación, Université Paris 7.

- Malouf, Robert (1997) "Mixed categories in HPSG". In Robert D. Borsley (ed), *Syntactic Categories*. Academic Press, New York.
- Mel'cuk Igor (1988) *Dependency Syntax: Theory and Practice*, State Univ. of New York Press, Albany.
- Mel'cuk Igor (1997) *Vers une Linguistique Sens-Texte, Leçon inaugurale au Collège de France*, Collège de France, Paris, 78p.
- Mitkov, R. (ed.) (2004): *The Oxford Handbook of Computational Linguistics*. Oxford, OUP.
- Nasr, A. (1996) *Un modèle de reformulation automatique fondé sur la Théorie Sens Texte: Application aux langues contrôlé*. Tesis Doctoral en informática, Universidad Paris 7.
- Pasero, R., Royauté, J. & Sabatier, P. (2004) « Sur la syntaxe et la sémantique des groupes nominaux à tête prédicative ». *Linguisticae Investigationes* 27:1 p. 83–124. Madrid. John Benjamins Publishing Company
- Pineda L. & Meza, I. (2000) *Una gramática básica del español en HPSG*. Monografía en línea; disponible en <http://leibniz.iimas.unam.mx/~luis/DIME/recursos/gramatica-esp%F1ol/gramatica-esp-hpsg.pdf>
- Pollard C. and Sag, I. (1994) *Head-Driven Phrase Structure Grammar*. Stanford, CSLI Publications.
- Tesnière, L., (1959) *Éléments de syntaxe structurale*, Klincksieck, Paris.
- Tesnière, L., (1934) "Comment construire une syntaxe" en *Bulletin de la Faculté des Lettres de Strasbourg*, 7 - 12, 219–229.
- Tseng, J. (2004) *Une grammaire HPSG / LKB pour le français*. Disponible sur www.llf.cnrs.fr/Gens/Abeille/hpsg-04.doc