

Índice

CAPÍTULO 1: INTRODUCCIÓN.....	5
1.1 ANTECEDENTES	7
1.1.1 <i>Lingüística computacional</i>	7
1.1.2 <i>Sintaxis</i>	10
1.1.3 <i>Gramática formal</i>	11
1.2 MOTIVACIÓN	12
1.2.1 <i>Verbos auxiliares del español</i>	12
1.2.2 <i>El proyecto DIME</i>	14
1.3 OBJETIVO Y METODOLOGÍA	15
1.4 ORGANIZACIÓN DE LA TESIS	16
CAPÍTULO 2: CONCEPTOS GENERALES	17
2.1 FENÓMENOS LINGÜÍSTICOS	20
2.1.1 <i>Ambigüedad</i>	20
2.1.2 <i>Cabeza de frase</i>	21
2.1.3 <i>Complementariedad</i>	22
2.1.4 <i>Especificación</i>	23
2.1.5 <i>Modificación</i>	24
2.1.6 <i>Árbol sintáctico</i>	25
2.2 MECANISMOS COMPUTACIONALES.....	27
2.2.1 <i>Estructuras atributo valor</i>	28
2.2.2 <i>Jerarquía de tipos</i>	30
2.2.3 <i>Unificación</i>	33
CAPÍTULO 3: HPSG.....	36
3.1 COMPONENTES DE HPSG.....	39
3.1.1 <i>Atributo HEAD</i>	39
3.1.2 <i>Listas de estructuras atributo valor</i>	40
3.1.3 <i>Reglas gramaticales</i>	42
3.1.4 <i>Principios</i>	45
3.1.5 <i>Semántica en HPSG</i>	46
3.1.6 <i>Satisfacción léxica</i>	49
3.1.7 <i>Reglas léxicas</i>	51
3.2 EJEMPLO: EL NIÑO ALTO ENTREGÓ EL JUGUETE A JUAN	52

CAPÍTULO 4: VERBOS AUXILIARES	57
4.1 CONSTRUCCIONES PERIFRÁSTICAS.....	60
4.1.1 <i>El verbo</i>	60
4.2 CONJUGACIÓN PERIFRÁSTICA	61
4.3 COMPORTAMIENTO DE LOS VERBOS AUXILIARES.....	62
4.3.1 <i>Conjugación perifrástica de verbos auxiliares</i>	63
4.3.2 <i>Propiedades sintácticas de los verbos auxiliares</i>	65
4.3.3 <i>Construcciones anidadas</i>	66
4.4 ANÁLISIS DE LOS VERBOS AUXILIARES	66
4.4.1 <i>Verbos auxiliares + infinitivo</i>	67
4.4.2 <i>Echar a</i>	71
4.4.3 <i>Haber que</i>	74
4.4.4 <i>Verbos auxiliares + gerundio</i>	76
4.4.5 <i>Verbo auxiliar + participio</i>	80
CAPÍTULO 5: MODELO EN HPSG	84
5.1 LA JERARQUÍA DE VERBOS EN HPSG	87
5.1.1 <i>El tipo verb-lxm</i>	87
5.1.2 <i>Verbos transitivos</i>	88
5.1.3 <i>Verbos intransitivos</i>	90
5.1.4 <i>Semántica de verbos transitivos e intransitivos</i>	91
5.1.5 <i>Complementos no personales</i>	92
5.2 AUXILIARES DEL ESPAÑOL EN HPSG.....	94
5.2.1 <i>Verbo auxiliar + infinitivo</i>	95
5.2.2 <i>Verbo auxiliar + gerundio</i>	112
5.2.3 <i>Verbo auxiliar + participio</i>	113
OTROS USOS DEL TIPO <i>SRV-LXM</i>	114
5.3.1 <i>Voz pasiva</i>	114
5.3.2 <i>Oraciones atributivas</i>	117
CAPÍTULO 6: MODELO EN LKB.....	119
6.1 SISTEMA DE ARCHIVOS DE LKB	121
6.1.1 <i>Archivos del sistema de tipos</i>	121
6.1.2 <i>Archivo de entradas léxicas</i>	123
6.1.3 <i>Archivos de reglas léxicas y morfología</i>	123
6.1.4 <i>Archivo de irregularidades</i>	125
6.1.5 <i>Archivo de reglas gramaticales</i>	125
6.2 IMPLEMENTACIÓN DEL SISTEMA DE VERBOS AUXILIARES	127

6.2.1	<i>Definición de tipos</i>	127
6.2.2	<i>El verbo poder</i>	130
6.2.3	<i>Otros verbos auxiliares</i>	136
6.3	OTROS FENÓMENOS LINGÜÍSTICOS	142
6.3.1	<i>Verbos unipersonales</i>	143
6.3.2	<i>Anidamiento de perífrasis</i>	144
6.3.3	<i>Subject control preposition</i>	147
6.3.4	<i>Voz pasiva</i>	148
6.3.5	<i>Verbos atributivos</i>	150
CAPÍTULO 7: CONCLUSIONES		154
7.1	ANÁLISIS SINTÁCTICO	156
7.1.1	<i>Pruebas sintácticas</i>	156
7.1.2	<i>Propiedades sintácticas</i>	158
7.1.3	<i>El tipo <i>srv-lxm</i></i>	159
7.2	IMPLEMENTACIÓN	159
7.2.1	<i>Sistemas de verbos auxiliares</i>	159
7.2.2	<i>Otros fenómenos</i>	161
APÉNDICE A: REGLAS Y PRINCIPIOS DE HPSG		162
A.1.	REGLAS GRAMATICALES	162
A.2.	PRINCIPIOS	162
APÉNDICE B: JERARQUÍA DE TIPOS EN HPSG		164
APÉNDICE C: GRAMÁTICA DE LKB		167
APÉNDICE D: TIPOS DEL SISTEMA		168
D.1.	TIPOS PARA VERBOS	168
D.2.	VERBOS INTRANSITIVOS	168
D.3.	VERBOS TRANSITIVOS	168
D.4.	VERBOS UNIPERSONALES	168
D.5.	VERBOS AUXILIARES	169
D.6.	VERBOS DE CONTROL	170
D.7.	SUBJECT RAISING PREPOSITION	171
D.8.	VERBOS ATRIBUTIVOS	171
D.9.	REGLAS LÉXICAS	171
APÉNDICE E: ENTRADAS LÉXICAS		172

APÉNDICE F: LISTA DE ORACIONES PARSEADAS.....	184
F.1. ORACIONES CON VERBOS AUXILIARES	184
<i>F.1.1. Con un verbo intransitivos (comer).....</i>	<i>184</i>
<i>F.1.2. Con un verbo intransitivo preposicional (vivir en).....</i>	<i>184</i>
<i>F.1.3. Con un verbo transitivo (escribir).....</i>	<i>184</i>
<i>F.1.4. Con un verbo ditransitivo (entregar).....</i>	<i>184</i>
<i>F.1.5. Con un verbo unipersonal (llover).....</i>	<i>184</i>
<i>F.1.6. Oraciones anidadas.....</i>	<i>185</i>
F.2. SUBJECT RAISING PREPOSITION	185
F.3. VOZ PASIVA (ESCRIBIR TRANSITIVO Y ENTREGAR DITRANSITIVO)	185
F.4. VERBOS ATRIBUTIVOS	185
BIBLIOGRAFÍA	186

Capítulo 1: Introducción

En esta tesis se presenta un modelo del sistema de verbos auxiliares del español, así como su implementación y validación computacional. Intuitivamente, los verbos auxiliares son verbos que aparecen en oraciones seguidos de otro verbo. El primero, el verbo auxiliar, se encuentra conjugado, y el segundo en una de las formas no personales (infinitivo, gerundio y participio). Por ejemplo, en la oración *he comido*, el verbo *haber* se presenta como verbo auxiliar de la conjugación compuesta. La expresividad que aportan a una lengua y su uso tan común hacen de los verbos auxiliares un fenómeno central en el estudio del lenguaje y en particular en la implementación de sistemas de procesamiento de lenguaje natural. A continuación se introduce de manera breve los elementos involucrados en la interpretación computacional del lenguaje.

La computación lingüística es una área de la computación que tiene como propósito proponer modelos formales para la representación e interpretación computacional del lenguaje. Estos modelos, conocidos como gramáticas formales, representan por medio de reglas a los fenómenos del lenguaje. Las gramáticas formales a través del proceso de *parseo* reproducen un análisis gramatical de las oraciones. Este proceso, cuya implementación se denomina *parser*, consiste en tomar una oración como entrada y asignarle una o más representaciones de las estructuras sintácticas y semánticas que la componen. Una gramática formal robusta del español debe considerar a los fenómenos gramaticales que integran al español; en especial, el sistema de los verbos auxiliares, cuya importancia de contar con un modelo que lo represente es resaltada por Noam Chomsky [Chomsky, 1956]. Para poder incluir el fenómeno de los verbos auxiliares en una gramática formal se requiere contar con una definición puntual del fenómeno. En seguida, se presenta la problemática existente entorno a la definición de los verbos auxiliares.

El estudio del fenómeno de los verbos auxiliares del español ha sido muy importante en las gramáticas tradicionales. Sin embargo, dentro de la tradición gramatical del español no existe un consenso en la definición de verbo auxiliar. Las gramáticas tradicionales del español definen a los verbos auxiliares en función de las llamadas construcciones perifrásticas, las cuales consisten en el uso de un verbo auxiliar seguido de un verbo en forma no personal. Esta definición no es suficiente debido a que el contexto de las perífrasis

se define en término de los auxiliares y los auxiliares en término de dicha construcción. Por la circularidad en la definición, las gramáticas tradicionales incorporan otras consideraciones adicionales. Algunas gramáticas recurren a un enfoque semántico, poniendo atención en el cambio de significado que presentan los verbos auxiliares [Gili, 1991]. Por ejemplo, en la oración *el niño alto volvió a entregar el juguete a Juan*, el verbo *volver* ha perdido su significado de regresar a un lugar para marcar que la acción de *entregar el juguete a Juan* ya había sido realizada con anterioridad. Sin embargo, las consideraciones que se proponen para identificar cuando un verbo cambia su significado son subjetivas, son aún tema de discusión y están lejos de ser formales. Los verbos auxiliares no son los únicos que aparecen en construcciones que involucran a dos verbos; verbos como *querer*, que no se comporta como auxiliar, también ocurren en construcciones perifrásticas. Sin embargo, como el concepto de perífrasis sólo se aplica a los verbos auxiliares, algunas gramáticas los incluyen en el conjunto de verbos auxiliares [Lenz, 1935]. Otras gramáticas proponen una escala de auxiliares donde solamente los verbos *ser*, *estar* y *haber* son considerados como auxiliares, mientras que los verbos con un comportamiento similar son clasificados como semiauxiliares [Alonso y Henríquez, 1967]. Factores como los anteriores hacen que la definición computacional de los verbos auxiliares del español no sea trivial.

Este trabajo propone definir a los verbos auxiliares bajo un enfoque sintáctico, es decir, poniendo atención a las relaciones que se establecen entre el verbo auxiliar y los elementos que lo rodean. Como se discutirá, desde este punto de vista no existe duda sobre si un verbo se comporta como auxiliar o no; además, se abre la posibilidad de analizar el comportamiento de aquellos verbos que no son auxiliares bajo un enfoque similar. Adicionalmente, este punto de vista permitirá reinterpretar trabajos más recientes sobre el tema, que enumeran un conjunto de construcciones posibles en las que participan los verbos auxiliares [Luna, 1991].

El modelo de los verbos auxiliares se desarrolla como parte del proyecto DIME, el cual consiste en la creación de un asistente computacional para el diseño de cocinas que interactúe por medio del lenguaje hablado y de una interfaz gráfica con el usuario [Pineda, et. al., 2001]. De manera específica, este trabajo se desarrolla como parte de la integración del módulo de análisis gramatical para el español del proyecto DIME, el cual se encarga de

asignar uno o más posibles significados a una oración. Dentro del proyecto DIME se ha escogido al formalismo HPSG (Head-driven Phrase Structure Grammar)[Sag y Wasow, 1999] para definir la gramática y su ambiente de desarrollo LKB (Linguistic Knowledge Building)[Copestake, 2000] para implementarla. Los resultados de esta investigación han sido validados en el proyecto DIME con la integración del sistema de verbos auxiliares del español en el módulo de análisis gramatical.

En las siguientes secciones de este capítulo se exponen con mayor profundidad los puntos antes expuestos. En la sección 1.1 se presenta el área en la cual se desarrolla este trabajo, la computación lingüística, y los conceptos de sintaxis y gramática bajo un enfoque computacional, los cuales permitirán ubicar el tipo de modelo que se propone en esta tesis. En la primera parte de la sección 1.2. se expone con mayor extensión la problemática que gira entorno a la definición de los verbos auxiliares del español y en la segunda parte se presenta el contexto del proyecto DIME, bajo el cual se realiza la presente tesis. En la sección 1.3 se especifican los objetivos que persigue la presente tesis y la metodología a seguir para alcanzarlos.

1.1 Antecedentes

La computación lingüística es un área de la computación que estudia al lenguaje desde el punto de vista de la información necesaria para la representación e interpretación computacional del lenguaje. Un enfoque común en la lingüística computacional es el de centrar la atención en las relaciones que se establecen entre los elementos de la oración, al cual se le denomina análisis sintáctico; en particular, en este trabajo se realiza un análisis sintáctico de los verbos auxiliares del español. El resultado de este análisis se modela a través de un conjunto de reglas que representan al fenómeno de los verbos auxiliares en una gramática formal. En seguida, se presenta un panorama más general del estudio del lenguaje, en especial de la computación lingüística, la sintaxis y gramáticas formales.

1.1.1 Lingüística computacional

La lingüística computacional es un área de la computación que tiene el propósito de crear modelos computacionales de los fenómenos que integran al lenguaje hablado y escrito. El estudio del lenguaje es tan amplio que existen varias ciencias que lo estudian, como la lingüística, la psicología, las neurociencias y las matemáticas; incluso, dentro de la

computación encontramos otras dos áreas dedicadas al estudio del lenguaje: el procesamiento del lenguaje natural (NLP) y el reconocimiento y generación del habla. Adicionalmente, la complejidad del lenguaje hace que su estudio se separe en los niveles fonético y fonológico, morfológico, sintáctico, semántico, pragmático y del discurso, con el fin de simplificar el objeto de estudio. Por otro lado, dentro de la lingüística computacional existen dos enfoques sobre el tipo de modelo que representa al lenguaje: el simbólico y probabilístico.

Algunas de las ciencias que estudian al lenguaje son la siguientes:

- Lingüística: Estudia a la lengua¹.
- Neurociencias: Estudia a las estructuras cerebrales involucradas en el habla.
- Psicología : Estudia los procesos que intervienen en el habla.
- Matemáticas: Estudia los mecanismos para la representación del lenguaje, principalmente se enfoca en lógica formal, la probabilidad y la estadística.

La relación entre la lingüística y la lingüística computacional es muy estrecha, de hecho la lingüística computacional surge en los departamentos de investigación en lingüística cuando Noam Chomsky propone una caracterización formal de los fenómenos del lenguaje[Chomsky, 1956]. Uno de los objetivos de la lingüística computacional es el reproducir el lenguaje de una forma semejante a como se utiliza por el ser humano para lo cual los resultados obtenidos por las neurociencias y la psicología contribuyen a definir los parámetros a seguir. Por otro lado, las matemáticas se encargan de caracterizar a los mecanismos que permitirán implementar los modelos del lenguaje que proponga la lingüística computacional. Como podemos ver, existe una interacción entre la lingüística computacional y las otras ciencias.

Dentro de la computación han surgido tres áreas que se han enfocado al estudio del lenguaje, la lingüística computacional, el NLP y el reconocimiento y generación del habla. La principal diferencia entre estas tres áreas es el departamento de investigación del que surgen; la lingüística computacional de los departamentos de lingüística, el NLP de los departamentos de ciencias de la computación y el reconocimiento y generación del habla de los departamentos de ingeniería eléctrica. El reconocimiento y generación del habla busca

¹ La lengua se compone del lenguaje escrito y hablado

obtener una representación del lenguaje partiendo de una señal acústica. El NLP busca definir los mecanismos que permitan procesar al lenguaje; en especial, NLP está muy ligado a la inteligencia artificial (AI) en el campo de razonamiento y lógica computacional. Sin embargo, estas barreras se han roto haciendo que los enfoques y técnicas se empezaran a compartir entre las áreas creándose el campo de estudio que se conoce como procesamiento del habla y lenguaje (Speech and Language Processing, SLP) [Jurafsky, 2000].

Los tipos de modelos formales propuestos en la computación lingüística se puede enmarcar en dos tendencias: la simbólica y la probabilística. En un principio, los modelos de la lingüística computacional sólo se basaban en la teoría de los lenguajes formales propuesta por Noam Chomsky [Chomsky, 1956], pero con el tiempo modelos probabilísticos migraron del área de reconocimiento de voz para ser utilizados en la caracterización del lenguaje. En la tendencia simbólica se encuentran involucradas conceptos de la teoría de los lenguajes formales y lógica, como autómatas y gramáticas libres de contexto; mientras que la tendencia probabilística se relaciona con conceptos probabilísticos y estadísticos, como N-gramas y modelos ocultos de Markov. Hoy en día, ambas tendencias son muy sólidas y han encontrado campos específicos para su aplicación. En tareas que involucran un proceso de razonamiento sobre las unidades lingüísticas prevalecen los modelos simbólicos, y en las tareas en las que no, los probabilísticos. En particular, el modelo que se propone en este trabajo se trata de una gramática formal que corresponde a un modelo simbólico.

El lenguaje es un objeto de estudio muy complejo; está compuesto por varios tipos de unidades lingüísticas como lo son sonidos, las palabras, las oraciones o diálogos. Para acotar el alcance de un análisis generalmente el estudio del lenguaje se clasifica en uno de los siguientes campos correspondientes al tipo de unidad lingüística involucrada:

- Fonético y fonológico: El estudio de los sonidos que integran al lenguaje.
- Morfológico: El estudio de los componentes con significado de las palabras.
- Sintáctico: El estudio de las relaciones estructurales entre palabras.
- Semántico: El estudio del significado.
- Pragmático: El estudio del lenguaje con el fin de cumplir metas.
- Del discurso: El estudio de las unidades mayores a una oración.

En particular, este trabajo se realiza un análisis sintáctico donde se efectúa un estudio de las relaciones entre los verbos auxiliares y los elementos que integran a la oración. En la siguiente sección se hablará sobre la sintaxis que corresponde al área en la que se realizan análisis sintácticos.

En resumen, este trabajo, que se realiza dentro del área de la lingüística computacional y consiste en un análisis sintáctico de los verbos auxiliares del español cuyos resultados servirán para proponer un modelo simbólico, una gramática formal, del fenómeno.

1.1.2 Sintaxis

La sintaxis se encarga de estudiar las relaciones presentes entre los elementos que forman una oración gramaticalmente correcta. Las palabras de un lenguaje presentan una estructura dentro de la oración que hace que las oraciones sean correctas gramaticalmente. Identificar a estas estructuras, los elementos que intervienen y las restricciones para su formación es parte del trabajo desarrollado por la sintaxis.

Identificar las condiciones que hacen que las palabras aparezcan en oraciones correctas gramaticalmente es parte esencial del análisis sintáctico. Por ejemplo en (1), ambas oraciones son gramaticalmente incorrectas², es decir no pertenecen al lenguaje del español.

- (1) *El niño es
*El niño es alta

En la primera oración, se observa que falta un elemento después del verbo *ser*. En la segunda, aparece un elemento después del verbo *ser*, pero no es compatible con los otros elementos de la oración, ya que la concordancia entre *el niño* y *alta* no se da. A estas condiciones de que necesita una palabra para aparecer en una oración es lo que denominamos en este trabajo como comportamiento sintáctico. En este caso, para (1) el comportamiento de *ser* indica que necesita un elemento después de él, el cual debe concordar con el elemento que se encuentra antes del verbo. En esta tesis, el análisis sintáctico que se presenta discute el comportamiento de los verbos auxiliares del español.

² El asterisco señala que la oración no es gramaticalmente correcta la oración

En el lenguaje, conjuntos de palabras comparten un mismo comportamiento; una de las tareas de la sintaxis es identificar estos comportamientos. A los tipos de comportamientos más generales se les conoce como categorías gramaticales. Las categorías gramaticales más comunes son las de verbo, sustantivo, preposición y adjetivo; ejemplos de palabras de estas categorías son *entregar*, *Juan*, *a* y *alto* respectivamente. Cada una de estas categorías esta formada por tipos de comportamientos más específicos. Por ejemplo, en la categoría de verbo existen los verbos transitivos como *entregar* que necesitan de un objeto directo e indirecto, e intransitivos como *correr* que no necesitan de un objeto directo. Estos tipos representan un comportamiento más específico que el que representa la categoría de verbo; pero aún existen otros comportamientos más específicos. Por ejemplo, en los verbos transitivos encontramos aquellos que sólo necesitan de un complemento, como *comer* en *comió una manzana*, y los que adicionalmente necesitan de un objeto indirecto, como *entregar* en *entregó un juguete a Juan*. En la modelación formal del lenguaje los comportamientos generales no ayudan a definir un modelo con la precisión requerida por la computadora, por lo que es necesario identificar los comportamientos más específicos de las palabras. En particular, en esta tesis se explora el comportamiento sintáctico de un tipo de verbo: el conjunto de verbos auxiliares del español.

1.1.3 Gramática formal

Una gramática formal es un conjunto finito de elementos que nos permite generar o identificar una secuencia de símbolos pertenecientes a un conjunto definido de secuencias, a este último conjunto se le denomina lenguaje formal. Una gramática formal permite capturar los fenómenos sintácticos del lenguaje; en el contexto del lenguaje y la sintaxis los símbolos son palabras, las secuencias de símbolos oraciones y el lenguaje formal el conjunto de oraciones gramaticalmente correctas. En consecuencia, en este trabajo una gramática formal es un conjunto finito de elementos que nos permite generar o identificar oraciones gramaticalmente correctas; en particular, en este trabajo se busca la generación o identificación de oraciones en las que intervienen verbos auxiliares del español.

Existen diferentes tipos de gramáticas formales: autómatas, gramáticas libres de contexto (*Context Free Grammar*, CFG), gramáticas dependientes del contexto y la máquina de Turing. Dependerá del tipo de gramática las secuencias de símbolos que pueda generar o identificar la gramática. A principio de la década de los 1960s se creía que las

gramáticas dependientes del contexto eran las más adecuadas para representar al lenguaje natural; sin embargo, las CFG fueron usadas en los primeros sistemas computacionales durante las siguientes dos décadas, debido a que son computacionalmente menos complejas que las dependientes del contexto. En 1979 Pullum y Gazdar ponen en duda los argumentos que defendían la idea que las CFG no tenían la capacidad de representar a los fenómenos del lenguaje natural [Pullum & Gazdar, 1982]. De la mano de esta controversia, Gazdar explora la idea de utilizar gramáticas de estructura de frase libres de contexto para representar al lenguaje natural, el resultado son las gramáticas de estructura de frase generalizada (GPSG - *Generalized Phrase Structure Grammar*) [Gazdar, Klein, Pullum & Sag 1985]. Las gramáticas GPSG evolucionaron rápidamente a las gramáticas de estructura de frase guiada por la cabeza (*Head-driven Phrase Structure Grammar*, HPSG) [Sag & Pollard, 1987]. El capítulo dos y tres están dedicados a exponer los elementos y características que distinguen a las gramáticas basadas en el formalismo de HPSG en el cual está basado el modelo de verbos auxiliares del español presentado en esta tesis. En las siguientes capítulos y secciones nos referiremos a las gramáticas formales solamente como gramáticas.

1.2 Motivación

El tema esta tesis surge del trabajo que se vienen haciendo en la gramática del español del proyecto DIME. Esta gramática necesita considerar a los fenómenos gramaticales del español, uno de estos fenómenos es el de verbos auxiliares. Debido a la problemática que gira entorno a la definición de verbo auxiliar era necesario realizar un análisis su comportamiento sintáctico con el fin de obtener un modelo del fenómeno e así implementarlo al *parser* del proyecto DIME.

1.2.1 Verbos auxiliares del español

Los verbos auxiliares del español son verbos que van seguidos de otro verbo en un forma no personal (infinitivo, gerundio y participio). El verbo auxiliar agrega un matiz diferente a la forma en que se realiza la acción representada por el verbo en forma no personal; observemos:

- (2) a. Tener que escribir
- b. Estar escribiendo

c. Haber escrito

Los tres ejemplos nos señalan la ejecución de una acción, *escribir*, en (2.a) notamos que al concepto de *escribir* le añade *tener que* la obligación de ejecutar la acción; mientras que en (2.b) significa la duración o continuidad del hecho; y con (2.c) que la acción ya ha finalizado. En estos ejemplos notamos la acción central es la representada por *escribir*, pero un segundo verbo da un pequeño giro a la forma en como se realiza esta acción. Este tipo de construcción es llamada perífrasis, la cual se define como el empleo de un verbo auxiliar conjugado seguido del infinitivo, gerundio o participio de otro verbo.

La definición de perífrasis supone la definición del contexto sintáctico en el que se aparecen los verbos auxiliares; sin embargo, los autores de diferentes gramáticas del español no logran coincidir en la definición de verbo auxiliar. En su libro “Curso de superior de sintaxis española” Gili Gaya asume la posición más estricta sobre la naturaleza de los verbos auxiliares [Gili, 1991]³:

“Para distinguir si un verbo está empleado como auxiliar basta fijarse en si ha perdido su significado propio. Cuando decimos *voy a contestar a su carta*, el verbo *ir* es auxiliar puesto que no conserva su acepción de movimiento de un lugar a otro, como no la conservan tampoco los verbos *andar* y *venir* en expresiones como *andaba mirando las láminas de un libro*, *venía sospechando de este hombre*. *Deber* se ha vaciado de su sentido obligatorio, para cumplir sólo el papel de auxiliar, en la expresión *deben de ser las siete*; y el verbo *tener* se halla despojado de la significación posesiva en *se lo tengo rogado*. Como todos estos verbos (con excepción de haber) conservan en la lengua moderna su acepción propia, el sentido habrá de decidir, en cada oración en que aparezcan tales perífrasis, si su significado se ha perdido o se ha oscurecido en grado suficiente para estimarlos como verbos auxiliares”.

Esta definición de Gili Gaya niega el carácter perifrástico a construcciones que otros gramáticos no *creen* discutibles, tales como *empezar* (*comenzar*, *principiar*) *a + infinitivo*, *seguir + gerundio*. Rodolfo Lenz asume la posición contraria al aceptar como perífrasis a las oraciones factitivas, argumentando que la pérdida de significado sólo es una de las

³ p. 127

propiedades de los auxiliares [Lenz, 1937]. A lo anterior Gili Gaya hace la siguiente observación:

“Por estos motivos creemos que Lenz, extrema la cuestión acerca del número de estas construcciones posibles, y confunde a menudo los medios gramaticales de expresión con los puramente lexicológicos”.

La mayoría de los gramáticos toman una posición mediativa entre ambos autores, proponiendo una escala para calificar a un verbo de auxiliar. Califican de auxiliar a los verbos *ser*, *estar* y *haber* y los verbos modales como semiauxiliares. [Alonso y Henríquez, 1967]. La indecisión en la definición de verbo auxiliar es la causa que el conjunto de verbos auxiliares sea variable dependiendo de la literatura consultada. Estudios más recientes presentan una clasificación más precisa de los verbos auxiliares del español basada en el tipo de construcciones en las que participan; sin embargo, estas no se han consolidado en una definición sintáctica que permita construir un modelo formal de los verbos auxiliares [Luna, 1991].

Es indiscutible el cambio de significado que presentan algunos de los verbos auxiliares del español; observemos el siguiente ejemplo:

- (3) a. debo cien pesos
 b. debe entregar el juguete a Juan

En (3.a) el verbo *deber* en su significación original refiere a la aceptación de una deuda; mientras que en (3.b) refiere a la obligación de realizar una acción. Las oraciones en (3) nos muestran que el significado del verbo *deber* no es el mismo; pero además notamos que el comportamiento sintáctico no es el mismo en las dos oraciones. El verbo *deber* en (3.b) acepta como complemento una frase verbal y no un objeto directo como lo hace en (3.a). Explorar el comportamiento sintáctico de los verbos auxiliares para proponer una definición puntual del fenómeno, en la cual no haya duda de cuando un verbo auxiliar se comporta como tal, es una de las motivaciones que originó el presente trabajo.

1.2.2 El proyecto DIME

La presente tesis se desarrolla como parte de la integración de una gramática del español para el proyecto DIME, el cual consiste en la creación de un asistente computacional para el diseño de cocinas que interactúe por medio del lenguaje hablado y de una interfaz gráfica con el usuario. El proyecto cuenta con las siguientes áreas de estudio: reconocimiento del

habla, gramática del español y análisis del discurso. La interfaz puede recibir como entrada una señal de audio, correspondiente a una oración. Esta señal es transformada a una secuencia de palabras en texto, la cual es analizada por un *parser* para obtener una representación del significado de la oración. La representación del significado permite al sistema decidir las acciones que necesita realizar para apoyar al usuario en la tarea. Con el fin de definir el tipo de diálogos que debe sostener la interfaz del proyecto DIME, se recurrió a recabar grabaciones simulando una interacción entre el usuario final y el sistema. Estas grabaciones constituyen el Corpus DIME, el cual es la base empírica para realizar el análisis del lenguaje. Es precisamente del Corpus DIME de donde surge la necesidad de estudiar al sistema de verbos auxiliares, ya que aparecen oraciones como la siguiente:

(4) ¿me puedes mostrar el catálogo de muebles?

En (4) observamos la aparición de una construcción perifrástica, el verbo *poder* seguido por *mostrar*. El verbo *poder* no es considerado como auxiliar en las gramáticas tradicionales; sin embargo, trabajos más recientes lo muestran como tal. Esclarecer esta situación entorno a la auxiliaridad del verbo *poder* fue uno de los factores que motivó el análisis de los verbos auxiliares que se presenta en este trabajo.

1.3 Objetivo y metodología

Para cumplir con el objetivo de modelar de manera computacional al sistema de verbos auxiliares del español es necesario contar con una definición puntual de verbo auxiliar. Sin embargo, como se señaló en la sección anterior, no se cuenta con una definición que cubra este requisito, por lo que un objetivo primordial de este trabajo consiste en definir de manera formal a los verbos auxiliares del español.

Para lograr los objetivos aquí expuestos la tesis se desarrolló de la siguiente forma:

- A través de la literatura especializada se definió el comportamiento sintáctico de los verbos auxiliares del español
- Se verificó este comportamiento con el conjunto de verbos auxiliares propuesto por Gili Gaya
- Se identificó, partiendo del comportamiento sintáctico, un conjunto de propiedades de los verbos auxiliares del español.

- Se identificó en HPSG un tipo que permite capturar las propiedades de los verbos auxiliares, lo que permitió definirlos en función de HPSG.
- Se creó el modelo del sistema de verbos auxiliares del español en HPSG, en particular para los verbos que integran el conjunto propuesto por Gili Gaya.
- Se implementó y verificó este modelo en la gramática del proyecto DIME.

En el transcurso de los capítulos cuatro, cinco y seis de esta tesis se expone con mayor amplitud las consideraciones seguidas para cumplir con los objetivos planteados en este trabajo.

1.4 Organización de la tesis

Hasta ahora se ha planteado el tipo de trabajo que se realiza en esta tesis: un análisis sintáctico de los verbos auxiliares del español dentro de la lingüística computacional que se reflejará en un modelo formal del fenómeno, que surgió dentro del proyecto DIME y ante la falta de un modelo formal de los verbos auxiliares. Los siguientes capítulos tienen el objetivo de presentar el trabajo realizado en el desarrollo de esta tesis. El capítulo dos y tres comprenden los antecedentes necesarios para comprender el modelo que se expone en este trabajo; en particular, en el capítulo dos se presenta una serie de fenómenos lingüísticos y mecanismos necesarios para crear el modelo de los verbos auxiliares y en el capítulo tres las características que componen a HPSG. En el capítulo cuatro se determinan las propiedades sintácticas que poseen los verbos auxiliares del español. A partir de las propiedades identificadas el capítulo cuatro, en el capítulo cinco se presenta el modelo del sistema de verbos auxiliares del español en HPSG y en el capítulo seis se presenta la implementación del modelo en el ambiente de desarrollo LKB. Por último, en el capítulo siete se presentan las conclusiones a las que se arribaron con la elaboración de esta tesis.

Capítulo 2: Conceptos generales

En este capítulo se presentan a los fenómenos lingüísticos y los mecanismos computacionales necesarios para modelar el sistema de verbos auxiliares del español. Aunque la tesis se enfoca a los verbos auxiliares, en este capítulo se introduce a los fenómenos lingüísticos de ambigüedad, cabeza de frase, complementariedad, especificación y modificación. La comprensión de estos fenómenos permitirá establecer una visión integral del modelo sintáctico que se utiliza en esta tesis. Por otro lado, un modelo computacional del lenguaje necesita de mecanismos que le permitan abstraer y representar la información presente en el lenguaje; las estructuras de atributo valor, la jerarquía de tipos y la unificación son los mecanismos computacionales utilizados por el formalismo de Head-driven Phrase Structure Grammar (HPSG) [Sag y Wasow, 1999]. Enseguida se hace una breve introducción a los fenómenos lingüísticos arriba mencionados.

La ambigüedad se encuentra en todos los niveles del lenguaje, se presenta cuando existe un símbolo el cual tiene más de una interpretación. En especial, esta tesis se centra en la ambigüedad sintáctica, la cual consiste en que en una oración las palabras se pueden relacionar de distintas formas produciendo diferentes interpretaciones. La oración *el señor observa al niño con el telescopio* es ambigua sintácticamente, porque la relación que se establece entre la frase *con el telescopio* y los demás elementos de la oración determina si el señor o es el niño es quien tiene el telescopio. Las construcciones perifrásticas son ambiguas. Por ejemplo, la oración *el niño alto puede entregar el juguete a Juan* tiene dos significados: uno donde se expresa la capacidad del niño para entregar el juguete y uno que hace referencia a la posibilidad de que el niño entregue el juguete; el último significado corresponde a un comportamiento de verbo auxiliar, mientras que el primero, no lo es, como se explica más adelante en el capítulo cinco.

Dentro de una oración, las palabras se agrupan en frases o cláusulas, las cuales consisten en que dos o más palabras se relacionen de forma sintáctica. En la oración *el niño alto entregó el juguete a Juan*, algunas de las frases presentes son *el niño alto*, *entregó el juguete a Juan*, *el juguete* y *a Juan*. Una propiedad de las frases es que con una de las palabras que la integran comparte un conjunto de propiedades; a esta palabra se le denomina en este trabajo *cabeza de frase*. Estas propiedades compartidas dan unidad

interna a la frase y le permiten relacionarse con otros elementos de la oración. Por ejemplo, en la frase *el niño*, la cabeza de frase es la palabra *niño* y una de las propiedades compartidas es la concordancia, en este caso género masculino y tercera persona del singular. Esta concordancia caracteriza a toda la frase y le permite relacionarse con un adjetivo de género masculino, *alto*, y con una frase verbal en tercera persona del singular, *entregó el juguete a Juan*.

Los fenómenos de complementariedad y especificación son fenómenos donde una palabra necesita de otras palabras o frases para lograr su completa significación. En estos fenómenos una vez que una palabra se relaciona con sus argumentos, los elementos necesitados, se conforma una frase en donde esta palabra es la cabeza de frase. El fenómeno de complementariedad se presenta con los verbos, los cuales necesitan de elementos adicionales que definan el alcance de la acción sobre los objetos. En la oración *el niño alto entregó el juguete a Juan*, el verbo *entregar* necesita de un objeto directo que cumpla con el papel de objeto entregado, *el juguete*, y de un beneficiario de la ejecución de la acción, el objeto indirecto *Juan*. Un ejemplo de especificación se presenta en los sustantivos que requieren de un artículo para limitar la extensión de su significado. En la frase *el niño alto*, el artículo *el* marca que el niño al que hacemos referencia está plenamente identificado. Ambos fenómenos son muy importantes para la definición de los tipos de argumentos que necesitan los verbos auxiliares del español.

El fenómeno en el que ciertas palabras aparecen opcionalmente en las oraciones con el fin de definir con mayor precisión algún elemento de la oración se conoce como modificación. La palabra *alto* en la frase *el niño alto* no es una palabra indispensable para que la frase logre su significación y solamente agrega información que nos permite determinar con mayor precisión a la persona de la que se habla. Es importante notar que en el fenómeno de la modificación la palabra o frase que es modificada es la cabeza de frase. La modificación no interviene directamente con los verbos auxiliares, pero es importante tenerlo en cuenta para que el sistema de verbos auxiliares no interfiera con el fenómeno.

A continuación se introducen los conceptos computacionales que permiten modelar fenómenos sintácticos del lenguaje. Las estructuras atributo valor es el mecanismo que nos permite representar la información lingüística de los elementos que integran la oración. En términos computacionales se trata de una estructura de datos que almacena información

lingüística. Estas estructuras están compuestas de atributos que representan una característica general de cierto objeto y de valores que definen a la propiedad específica que el objeto posee. Por ejemplo, en el contexto de las palabras, la información de concordancia, que corresponde a número, género y persona, se representa en las palabras por los atributos NUM, GEN y PER, los cuales, para la palabra *niño* en particular, tienen los valores de singular, masculino y tercera respectivamente. En HPSG las estructuras atributo valor son ampliamente utilizadas.

La jerarquía de tipos consiste en la organización de las estructuras atributo valor bajo una estructura jerárquica. Una estructura jerárquica permite establecer relaciones entre elementos comunes y diferenciar aquellos que no lo son. El tipo de relación que se establece en una jerarquía es de padre e hijo. Las estructuras hijo comparten, a través de un mecanismo de herencia, ciertas características con su o sus padres. La estructura jerárquica que se utiliza en HPSG es de una *lattice*. El conjunto de características y reglas que definen las relaciones en una *lattice* es presentado en este capítulo.

La unificación es la operación que se encarga de validar las estructuras que intervienen en el análisis sintáctico. La unificación verifica si dos estructuras atributo valor son compatibles entre sí; en caso de serlo, la unificación resulta en una estructura mínima que involucre a las dos estructuras atributo valor. La compatibilidad de estructuras involucra también las relaciones que se definen en la jerarquía de tipos. La unificación es una operación que se utiliza ampliamente en la definición tanto de HPSG como en gramáticas basadas en este formalismo.

En las secciones que integran este capítulo se abordarán los conceptos aquí introducidos con mayor amplitud. La sección 2.1 contiene a los fenómenos lingüísticos aquí presentados, ambigüedad, complementariedad, especificación y modificación. En la sección 2.2 se explican con mayor precisión los conceptos computacionales de estructuras atributo valor, jerarquía de tipos y unificación. Adicionalmente, en la primera sección se presenta el concepto de árbol sintáctico, el cual es una representación de las estructuras sintácticas que integran a la oración.

Los fenómenos y herramientas que en este capítulo se presentan son parte esencial de la maquinaria conceptual y computacional de HPSG. Estos fenómenos repercuten directamente en la definición misma de HPSG como se aprecia en el capítulo tres. Además,

estos fenómenos juegan un papel fundamental en la discusión del comportamiento sintáctico de los verbos auxiliares del español que se lleva a cabo en los capítulos cuatro y cinco. Los mecanismos computacionales, que aquí se presentan, son parte fundamental de la definición de los elementos que integran a HPSG como se ve en el siguiente capítulo.

2.1 Fenómenos lingüísticos

2.1.1 Ambigüedad

La ambigüedad es un fenómeno del lenguaje que se presenta cuando un símbolo lingüístico tiene varias interpretaciones. La ambigüedad se presenta en los niveles fonético y fonológico, léxico, sintáctico y pragmático del lenguaje. La diversidad que presenta hace de la ambigüedad un problema central en la interpretación computacional del lenguaje, en donde es importante identificar a las posibles interpretaciones para escoger aquella que sea la correcta; pero como es de suponerse ambas tareas no son nada triviales. A continuación se presentan los tipos de ambigüedades correspondientes a los niveles de estudio del lenguaje.

La ambigüedad fonética y fonológica se refleja en la existencia de distintas pronunciaciones para una palabra. La diversidad en la pronunciación de las palabras se debe a distintos factores de muy variada naturaleza. En primer lugar, encontramos a las características físicas de la voz de cada individuo, la voz varía de persona a persona lo que hace que la señal de audio sea diferente entre las personas. Otro factor que influye corresponde a las variaciones alofónicas de los sonidos, las cuales son pronunciaciones alternativas para una misma unidad fonética. En la pronunciación las palabras también influyen las palabras que están antes o después de una palabra, esto se puede observar en la frase *un peso* donde la pronunciación de la *n* se acerca a una *m*. Factores como los anteriores hacen del reconocimiento y generación del habla un área de estudio muy activa en la actualidad.

La ambigüedad léxica se presenta cuando una misma palabra representa a más de un significado. Por ejemplo, la palabra *corredor* es léxicamente ambigua ya que hace referencia a un pasillo en la oración *el corredor está lleno de sillas*, pero adicionalmente se puede referir a alguien que corre como en la oración *el corredor mexicano ganó*. Se le denomina ambigüedad léxica porque para estas palabras debe haber una entrada léxica que

consiste en una descripción de sus propiedades tantas veces como significados tenga la palabra.

La ambigüedad sintáctica se presenta cuando en una oración las relaciones entre las palabras se pueden establecer de distintas formas. Por ejemplo, en la oración *el señor observa al niño con el telescopio* la frase *con el telescopio* puede estar asociada con la frase *al niño* o con la de *observa al niño*, que dependerá de la relación el significado final que tome la oración; en el primer caso se hace referencia a que el niño tiene el telescopio, mientras que en el segundo a que la acción de *observar* se lleva a cabo con el telescopio. Este tipo de ambigüedad se volverá a retomar en la siguiente sección en donde se introduce el concepto de árbol sintáctico en donde se representan las relaciones sintácticas y es posible observar la ambigüedad sintáctica.

La ambigüedad pragmática se presenta en aquellas oraciones que por su uso toman un nuevo significado al textual. Un ejemplo de este tipo ambigüedad son las oraciones que siguen la construcción *echar un ojo a algo* que hacen referencia a cuidar de algo y no lanzar físicamente un ojo a algo.

La ambigüedad que presentan los verbos auxiliares del español es del tipo léxica, es decir, existen varias entradas léxicas para un verbo auxiliar. Por ejemplo, para las oraciones *debo cien pesos* y *debo entregar el juguete a Juan* los verbos *deber* son diferentes. En el primero caso se refiere a la aceptación de una deuda y el segundo, que se comporta como auxiliar, a la obligación de realizar una acción. Pero, como se discute en los capítulos cuatro y cinco, conjuntamente con el cambio de significado, los verbos auxiliares adicionalmente presentan un cambio en su comportamiento sintáctico que hace que se relacionen de manera distinta con los elementos que integran a la oración. En consecuencia, la ambigüedad léxica que presentan los verbos auxiliares se refleja en una ambigüedad sintáctica de las oraciones en las que participan.

2.1.2 Cabeza de frase

El fenómeno de cabeza de frase es una propiedad que presentan las frases. Las frases o cláusulas son subelementos de la oración, los cuales consisten en una palabra o grupo de palabras que se encuentran relacionadas sintácticamente de tal forma que no necesitan de otro elemento para lograr su significación. En (1) podemos observar a las frases que integran a la oración *el niño alto entregó el juguete a Juan*.

- (1) a. *el niño*
- b. *el niño* alto
- c. *el juguete*
- d. a *Juan*
- e. *entregó* el juguete a Juan

Todas las frases listadas en (1) presentan la propiedad de compartir un conjunto de características con una de las palabras que la integran. A la palabra con la que comparte las características se denomina *cabeza de frase*; en (1) este tipo de palabras están marcadas con itálicas. El tipo de la frase es definido por la categoría gramatical de la cabeza frase; así donde la cabeza de frase sea sustantivo la frase es sustantiva, donde sea un verbo es verbal y donde sea una preposición es preposicional. Al conjunto de propiedades que son compartidas entre frase y cabeza frase se le denomina “propiedades cabeza”. Por ejemplo, la frase sustantiva (1.b) tiene como cabeza de frase al sustantivo *niño*, esta palabra comparte su propiedad cabeza de concordancia de tercera persona masculina del singular con la frase, dando como resultado una frase sustantiva en tercera persona masculina del singular. Por otro lado, la frase verbal (1.e) tiene como cabeza de frase al verbo *entregó*, con la cual comparte su propiedad cabeza de concordancia de tercera persona del singular, de manera similar como ocurrió con (1.b). Pero adicionalmente, frase y verbo comparten la necesidad de un sujeto, lo que permitirá relacionarse a esta frase con la de (1.b) para así agrupar toda la oración.

La condición de cabeza de frase es definida por el tipo de relación sintáctica que se establece entre las palabras que conforman una frase. En las tres siguientes secciones se describirán tres fenómenos gramaticales que definen tres posibles relaciones sintácticas; adicionalmente se discutirá con mayor amplitud las condiciones necesarias para que una palabra sea cabeza de frase dentro de una frase.

2.1.3 Complementariedad

La complementariedad es un fenómeno lingüístico en donde una palabra requiere de otras para lograr su significación. El fenómeno de complementariedad no es el único que captura la necesidad de una palabra por otra, ya que el fenómeno de especificación también lo hace. En general, las palabras requeridas por otra palabra se les denomina argumentos sin importar si es por el fenómeno de complementariedad o especificación. Los argumentos de

complementariedad son llamados complementos. En el español los complementos generalmente se encuentran después de la palabra que lo solicita; por ejemplo, en (1.d) y (1.e) las preposición *a* y el verbo *entregó* son seguidas por sus complementos. Una vez que la palabra se encuentra seguida por sus complementos se agrupa en una frase que la tiene como cabeza de frase. En el español las preposiciones y los verbos son ejemplos de palabras que requieren de complementos; en las frases (1.d) y (1.e) se observa *a* y *entregar* como cabezas de una frase preposicional y una verba respectivamente una vez que se han agrupado con sus complementos.

La naturaleza de la palabra define el tipo de argumentos que necesita. Por ejemplo, el verbo *entregar* es de los llamados verbos transitivos; este tipo de verbos necesitan de dos complementos para lograr su significación. El primero hace referencia al objeto sobre el cual recae la acción y el segundo a quien se beneficia con la acción. En la oración (2.e) la frase *el juguete* hace referencia a lo entregado y la frase *a Juan* a quien se le entrega. La naturaleza de *entregar* indica que el primer complemento debe ser una frase sustantiva y el segundo una preposicional. Existen otros fenómenos que interactúan con los complementos de las palabras; por ejemplo, aquellos fenómenos que involucran a los complementos de verbos transitivos: la omisión de algún complemento, como en *el niño alto entregó el juguete*; el orden en que aparecen los complementos, como en *el niño alto entregó a Juan el juguete* o la utilización de clíticos *el niño alto se lo entregó*; sin embargo, explorar estos fenómenos queda fuera de alcance del presente trabajo por lo que no serán considerados durante el desarrollo de este trabajo.

En particular, para los verbos auxiliares del español es muy importante definir el tipo de complemento que posee, es decir, aquellos elementos que lo siguen y con los cuales forma una frase. El capítulo cuatro está dedicado a definir las características sintácticas que poseen los verbos auxiliares y una parte fundamental de este análisis es definir el tipo complemento que tiene un verbo auxiliar.

2.1.4 Especificación

La especificación es el segundo fenómeno que se relaciona con la necesidad de una palabra por otras. A diferencia de la complementariedad, el argumento en la especificación del español se encuentra generalmente a la izquierda de la palabra, se le denomina especificador y tiene el propósito de delimitar el alcance en el significado de la palabra que

lo requiere. Un caso de especificación es el de los sustantivos comunes los cuales generalmente aparecen precedidos por un artículo, los cuales son palabras que definen el significado de las palabras. Por ejemplo, en las frases (2.a) y (2.c) el artículo *el* determina que el niño y juguete al cual hace referencia la frase ha sido identificado con anterioridad. Otro caso de especificación importante es el sujeto de los verbos. La mayoría de los verbos necesitan de un especificador que identifique a la entidad quien realiza la acción verbal; en particular, a este especificador se le conoce como sujeto de la oración y esta compuesto por una frase sustantiva.

En forma similar como ocurrió con los complementos, cuando una palabra se encuentra precedido por su especificador se forma una frase en donde la palabra es la cabeza de frase. Esto lo podemos apreciar en (2.a) y (2.c) en donde las cabezas de frase son *niño* y *juguete* respectivamente. En el caso de la frase verbal (2.e) cuando su especificador, el sujeto, le precede, la nueva frase que agrupa a sujeto y frase verbal representa la oración completa, en este caso: *el niño alto entregó el juguete a Juan*.

En particular, el especificador de los verbos auxiliares juega un papel muy importante en las características que poseen este tipo de verbos. En el capítulo cuatro se establecerá la naturaleza del especificador de los verbos auxiliares.

2.1.5 Modificación

La modificación es un fenómeno gramatical en donde la aparición de una palabra en un frase es opcional. A diferencia de los fenómenos de complementariedad y especificación, donde la presencia de una palabra es requerida para lograr una significación completa, en el fenómeno de modificación la presencia de la palabra sirve para agregar información sobre la situación a que hace referencia la frase. En el español los adjetivos, adverbios y frases preposicionales cumplen con el fenómeno de modificación; en (2) se presentan algunos ejemplos de oraciones donde se presenta el fenómeno de la modificación

- (2) a. el niño alto entregó el juguete a Juan
- b. el niño entregó rápidamente el juguete a Juan
- c. el niño entregó el juguete a Juan por la mañana

Para (2) la oración sin modificación es: *el niño entregó el juguete a Juan*; en todos los casos la información proporcionada por el fenómeno de la modificación no es indispensable para que la oración esté completa; sin embargo, agrega información sobre la situación que

describe la oración. En (2.a) la palabra *alto* es opcional a la frase *el niño alto*; esta palabra tiene el objetivo de determinar con mayor precisión al niño que entregó el juguete. En (2.b) la palabra *rápidamente* tiene el objetivo de describir la forma en como se realiza la acción de entregar. Finalmente, en (2.c) la frase *por la mañana* tienen el objetivo de ubicar el tiempo en el que se realizó la acción de *entregar*.

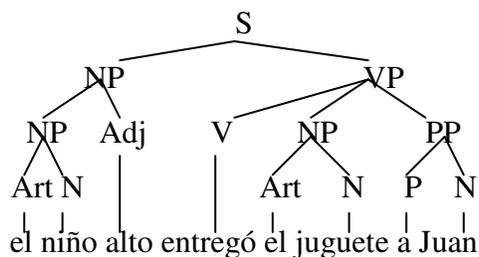
Otra diferencia entre modificación y los fenómenos de complementariedad y especificación se relaciona con quien desempeña el papel de cabeza de frase. Mientras que en los fenómenos de complementariedad y especificación la cabeza de frase es la palabra que se agrupa con sus argumentos, en la modificación es la palabra o frase modificada quien desempeña este papel y no así la palabra que modifica. Esto se aprecia en (2.b) en donde la frase *el niño* está marcada como cabeza de frase, esto se comprueba con el hecho de que la frase (2.b) no es una frase adjetiva sino sustantiva.

En particular, el fenómeno de los verbos auxiliares del español no está relacionado con el fenómeno de la modificación. Pero será muy importante que el fenómeno pueda interactuar con los verbos auxiliares ya que es muy común en el español.

2.1.6 Árbol sintáctico

Hasta ahora hemos visto un total de cinco fenómenos sintácticos: ambigüedad, cabeza de frase, complementariedad, especificación y modificación. Los tres últimos fenómenos son comunes entre sí porque permiten agrupar conjunto de palabras en frases, es decir, proporcionan una estructura a los elementos de una oración. Un árbol sintáctico nos permite representar la estructura que tienen las palabras y las frases en una oración. Por ejemplo, (3) se observa el árbol sintáctico de la oración *el niño alto entregó el juguete a Juan*.

(3)



Un árbol consta de nodos y ramas; los nodos están representados por letras y palabras y las ramas por líneas. El nodo que se encuentra en un nivel arriba de otro se dice lo domina; por ejemplo, el nodo marcado con la letra *P* domina al nodo *a*, pero es dominado por el nodo

PP. Los nodos al final del árbol que no dominan a ningún otro se les llama nodos terminales o léxicos⁴. En un árbol sintáctico estos coinciden con las palabras que forman a la oración. Al nodo inicial que no es dominado por ningún otro nodo se le llama la raíz y representa a toda la oración. La relación que se establece entre nodos inmediatos, es decir aquellos que los separa una rama, es de padre a hijo.

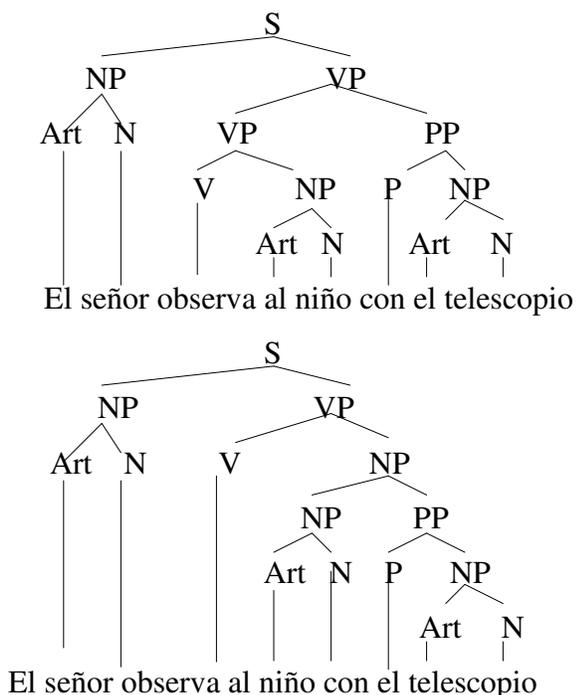
El objetivo principal de un árbol es representar la estructura interna de la oración. En (3) se observan los elementos que integra a la oración. En primer lugar tenemos las palabras que conforman la oración, los nodos terminales. El siguiente nivel del árbol corresponde a la representación de la categoría gramatical que posee cada palabra, esto es con el fin de caracterizar la función que desempeña cada palabra en la oración a este nivel se le llama categorización léxica; se ha adoptado la siguiente notación: sustantivo *N* (nombre), artículos *Art* (artículo), adjetivos *Adj* (adjetivo), verbos *V* (verbo) y preposiciones *P* (preposición). Los siguientes niveles en el árbol sintáctico corresponden a las frases que se forman por los fenómenos de complementariedad, especificación y modificación. Por ejemplo, las dos *NP* del tercer nivel relacionan a un sustantivo con su especificador, por lo que la frase, marcada con la *P* (del inglés *phrase*), corresponde a una frase sustantiva, marcada por la *N*; esta notación captura parte del fenómeno de cabeza de frase. Por otro lado la *VP*, frase verbal, que aparece en el cuarto nivel relaciona al verbo con sus complementos, otra vez se puede observar que la cabeza de frase define el tipo de frase. En este mismo nivel observamos a otra frase sustantiva, *NP*, esta representa la relación de modificación que se establece entre la frase *el niño* y el adjetivo *alto*; como se observa la cabeza de frase no corresponde al adjetivo tal y como se estableció dentro de las características de la modificación. El último nivel establece la relación entre la frase verbal y su sujeto por medio de la especificación. En el siguiente capítulo se expondrá las características de HPSG, el cual cuenta con un conjunto de reglas que son capaces de reproducir la estructura sintáctica de una oración de manera similar a como se observa en este árbol.

Por último, en la sección 2.1.1 correspondiente a la ambigüedad se había establecido que una oración puede ser ambigua sintácticamente si existe más de una forma en que sus

⁴ También conocidos como hojas

elementos se pueden relacionar sintácticamente. La existencia de ambigüedad sintáctica repercute en la existencia de dos o más árboles sintácticos para una misma oración. A continuación se presenta los árboles correspondientes a la oración *el señor observa al niño con el telescopio*, la cual fue un ejemplo de oración sintácticamente ambigua.

(4)



La ambigüedad surge en la forma como se puede relacionar la frase preposicional con los elementos de la oración. En el primer árbol vemos que la *PP* modifica a la frase verbal, mientras que en el segundo modifica a una *NP*. El primer árbol correspondería a la interpretación de que el señor utiliza un telescopio para observar al niño y el segundo a la de que el niño observado tiene un telescopio. Como se había señalado los verbos auxiliares del español son ambiguos; como se mostrará en el capítulo seis, que corresponde a la implementación del sistema, en donde el análisis sintáctico computacional de algunas oración que incluyan verbos auxiliares dará como resultado más de un árbol sintáctico.

2.2 Mecanismos computacionales

En esta sección se describen los mecanismos computacionales que son la base del formalismo HPSG: estructuras atributo valor, jerarquía de tipos y unificación. Los ejemplos que aquí se presentan están orientados a comprender el funcionamiento de estos

mecanismos. Estos ejemplos son independientes de la información sintáctica necesaria para modelar la sintaxis de un lenguaje. La adaptación de estos mecanismos para capturar información sintáctica se presenta en el siguiente capítulo.

2.2.1 Estructuras atributo valor

Las estructuras atributo valor nos permiten representar las propiedades de un objeto. Computacionalmente, se trata de una estructura de datos compuesta por un grafo dirigido sin ciclos (DAG, *direct acyclic graph*); las propiedades de este tipo de grafos se discutirá más adelante. La información en una estructura atributo valor se representa en pares de atributo valor. Las características que son generales a objeto son represadas por los atributos, mientras que las propiedades que son específicas corresponden a los valores. Por ejemplo, para caracterizar los libros de nuestra biblioteca algunos de los atributos son el título, el autor y la editorial; pero los valores correspondientes para referirnos a un libro en particular de la biblioteca son Balún Canán, Rosario Castellanos y Fondo de Cultura Económica respectivamente. En (5) se ilustran los pares atributo valor correspondientes al libro Balún Canán; en busca de claridad se acostumbra que los atributos vayan en mayúsculas y los valores en minúsculas.

- (5) TÍTULO = balún canán
AUTOR = rosario castellanos
EDITORIAL = fce

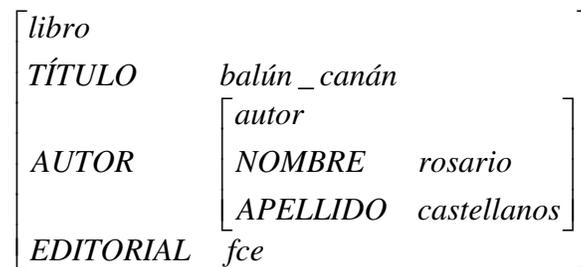
Una estructura de atributos valor es un conjunto de pares atributo valor asociados a un *identificador*. Una forma de representar una estructura atributo valor es a través de las matrices de atributo valor (AVM). El AVM correspondiente al ejemplo en (6) se muestra a continuación:

- (6)
- | | | | |
|----------|------------------|------------------------------|-----------|
| $\left[$ | <i>libro</i> | \quad | $\right]$ |
| | <i>TÍTULO</i> | <i>balún _ canán</i> | |
| | <i>AUTOR</i> | <i>rosario _ castellanos</i> | |
| | <i>EDITORIAL</i> | <i>fce</i> | |

La palabra *libro* indica el nombre del identificador del conjunto de pares atributo valor; en este caso, como esta estructura atributo valor pretende englobar a las características que presentan los libros este conjunto, se llama *libro*. Los corchetes que encierran la estructura

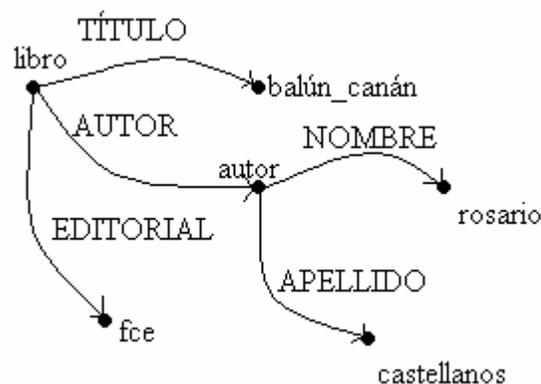
atributo valor definen el alcance sobre los pares atributo valor. Un aspecto importante de las estructuras atributo valor es que un atributo puede tomar como valor a otra estructura atributo valor. Por ejemplo, supongamos que deseamos separar nombre y apellido del autor; una posibilidad es sustituir el atributo de *AUTOR* por el de *NOMBRE* y *APELLIDO*, pero estos dos atributos siguen refiriendo a características del autor por lo que sería conveniente agruparlos en una estructura atributo valor y asignársela al atributo *AUTOR*. El AVM resultante de esta modificación se muestra en (7). Es importante notar que el atributo *AUTOR* en realidad esta tomando como valor una estructura que representa a otro objeto caracterizado, en este caso aquel que corresponde a un autor de libros.

(7)



Una representación adicional de las estructuras atributo valor es la correspondiente a los DAG, en donde los identificadores y valores son nodos y los atributos arcos estos últimos se encargan de unir a un identificador con un valor o con otro identificador. En (8) se muestra el DAG correspondiente a nuestro ejemplo (7).

(8)



Los DAG ejemplifican de manera más clara las propiedades que poseen las estructuras atributo valor:

- Todos los nodos están unidos por arcos
- Un arco sale de un identificador y se dirigen a un valor o a un identificador

- Dos arcos no podrán llegar al mismo nodo

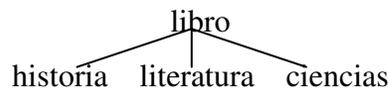
En el siguiente capítulo se presentarán un conjunto estructuras atributo valor que permiten representar información lingüística en HPSG.

2.2.2 Jerarquía de tipos

La jerarquía de tipos es un mecanismo que permite clasificar objetos. En particular, la jerarquía que se presenta en esta sección corresponde a una *lattice*, la cual es la utilizada por HPSG. Una *lattice* permite clasificar a los objetos en función de las clases que de este objeto existen. Por ejemplo, para organizar nuestra biblioteca el objeto *libro* es nuestro objeto principal; pero sabemos que el conjunto de libros que integra nuestra biblioteca es diverso, ya que tenemos los que tratan de historia, literatura o ciencias.

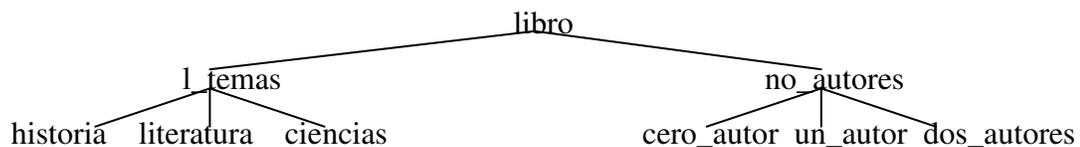
Una *lattice* esta conformada por nodos y arcos; un nodo representa a una clase de objeto y un arco establece una relación de padre a hijo entre los tipos de objetos. Por ejemplo, el objeto *libro*, para el cual existe un nodo, es el padre de libros que tratan de historia, literatura o ciencias, los cuales están representados por otro nodo cada uno; un arco se encarga de establecer esta jerarquía como se ve en (9). Un tipo hijo podrá tener también sus propios hijos y así sucesivamente; a los serie de nodos que parten de un tipo en particular se les conoce como descendientes. Una restricción muy importante es que un nodo hijo no podrá ser padre de su propio padre o de algún nodo anterior.

(9)



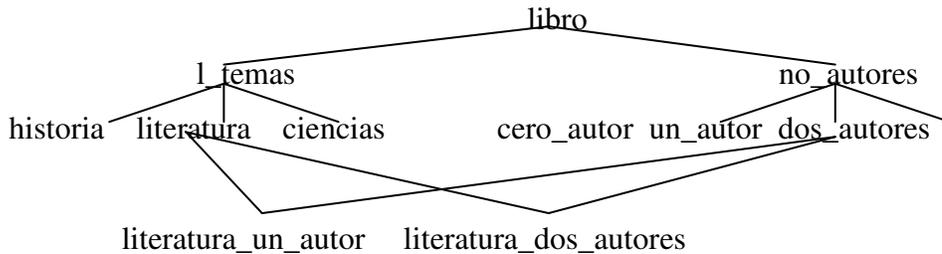
La jerarquía que hasta ahora se ha expuesto es semajante a la jerarquía de árbol; pero a diferencia de este tipo de jerarquía, una *lattice* permite que dos nodos tengan como hijo a un mismo nodo. Por ejemplo, supongamos que adicionalmente deseamos clasificar nuestros libros por el número de autor; la *lattice* correspondiente es la siguiente:

(10)



Esta jerarquía, tal y como aparece, no nos ayuda en mucho, porque el libro de Balún Canán tendría que ser de dos tipos de *literatura* y de *un_autor*; situación que hace redundante nuestra información. Pero, con la propiedad extra que poseen una *lattice* esto se puede solucionar creando un tipo que combine ambas clasificaciones como se ve en (12).

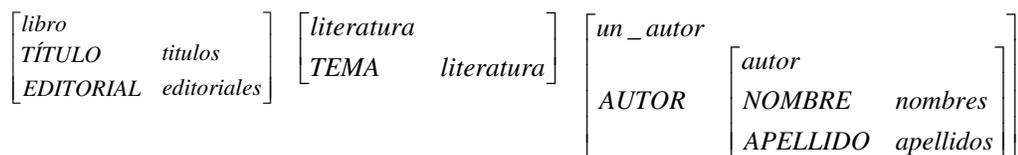
(11)



Con esta modificación ahora es posible representar al libro de Balún Canán bajo un solo tipo, *literatura_un_autor*. La jerarquía de tipos utilizada en HPSG agrega otras características adicionales a la *lattice*. En primer lugar, los nodos de la jerarquía representa a una estructura atributo valor. En segundo lugar, se establece un mecanismo de herencia entre padre e hijo.

Como un tipo representa a una estructura atributo valor, los tipos están compuestos por dos elementos; el primero corresponde al nombre del nodo y el segundo, que es opcional, es una estructura atributo valor, a la cual se le llama *restricción*, la cual representa las características que posee el objeto representado por este tipo. En la sección 2.2.1 de estructuras atributo valor se había hecho la distinción entre valor e identificador; en la jerarquía de tipos valores son aquellos que no poseen una restricción, mientras que los identificadores si la poseen. Por ejemplo, en (12) se puede observar las estructuras atributo valor correspondientes a los tipos *libro*, *literatura* y *un_autor*.

(12)



Como el nodo *libro* representa a todos los posibles libros, este agrega la restricción de que todo libro tiene un *TÍTULO* y una *EDITORIAL*; mientras que *literatura* tiene definido su atributo *TEMA* como de *literatura* y por último el nodo *un_autor* agrega la restricción de tener el atributo *AUTOR*. Es importante notar que estas estructuras atributo valor no están

haciendo referencia a un libro en particular por lo que los valores que se presentan son generales. Estos valores están representando a objetos diferentes a libros, por lo que estos tendrán que estar definidos a la par del tipo *libros*; pero, como no provienen de otro objeto con el que compartan ciertas características se declaran como hijo de **top**, el tipo que contiene a todos los objetos.

Como se menciono la jerarquía de tipos permite agregar un mecanismo de herencia entre padres e hijos. Este mecanismo consiste en que la restricción, estructura atributo valor, de un padre será heredada a su hijo. Lo anterior explica porque en la descripción del tipo *un_autor* no se ha incluido los atributos de *TÍTULO* y *EDITORIAL*. En (13) se muestra al tipo *un_autor* considerando la información que contiene tanto por una declaración explícita como la que recibe vía el mecanismo de herencia.

(13)

$$\left[\begin{array}{l} un_autor \\ TÍTULO \\ AUTOR \\ EDITORIAL \end{array} \begin{array}{l} \\ \\ \\ editoriales \end{array} \left[\begin{array}{l} \\ \\ \\ \\ \end{array} \right] \right]$$

A continuación se presenta la tabla con toda la jerarquía de tipos correspondiente al ejemplo aquí analizado⁵:

Nombre nodo	Restricción	Padre
top		
libro	[TÍTULO títulos, EDITORIAL editoriales]	top
autor	[NOMBRE nombres, APELLIDO apellidos]	top
títulos		top
editoriales		top
nombres		top
apellidos		top
l_temas	[TEMA temas]	libro

⁵ En esta tabla por cuestión de formato los AVM se encuentran en formato de texto.

no_autor		libro
l_historia	[TEMA historia]	l_temas
l_literatura	[TEMA literatura]	l_temas
l_ciencias	[TEMA ciencias]	l_temas
cero_autor		no_autores
un_autor	[AUTOR autor]	no_autores
dos_autores	[AUTOR_1 autor, AUTOR_2 autor]	no_autores
literatura_un_autor		l_literatura y un_autores
literatura_dos_autores		l_literatura y dos_autores
balún_canán		títulos
fce		editoriales
rosario		nombres
castellanos		apellidos

Tabla 1. Jerarquía de tipos ejemplo libros

2.2.3 Unificación

Mientras que las estructuras atributo valor nos permiten reproducir la información de manera estructurada, en nuestro caso lingüística, la unificación es el mecanismo computacional que nos permitirá procesar computacionalmente a esta representación. La unificación es una operación que valida la compatibilidad entre dos estructuras atributo valor; si estas son compatibles las amalgama en una mínima estructura atributo que contenga a las dos descripciones. Cuando las estructuras no son compatibles la operación no se puede llevar a cabo.

La operación de unificación se basa en la propiedad *subsume* de las estructuras atributo valor, la cual dice: *una estructura a subsume a b, si a contiene a toda la estructura b*. Por ejemplo, tenemos las siguientes estructuras:

(14)

$$\left[\begin{array}{l} a \\ UNO \quad uno \end{array} \right] \quad \left[\begin{array}{l} b \\ DOS \quad dos \\ TRES \quad \left[\begin{array}{l} a \\ UNO \quad uno \end{array} \right] \end{array} \right]$$

Es evidente que la estructura a no subsume a b ; sin embargo la estructura b si subsume a a ya que su atributo $TRES$ tiene como valor una estructura del tipo a .

La unificación en términos de esta propiedad se define de la siguiente manera: la unificación, representada en este trabajo por el símbolo $\&$ ⁶, entre A y B es una estructura C, la cual es la estructura atributo valor mínima que subsume tanto a A como a B. Ejemplo:

(15)

$$\begin{bmatrix} UNO & uno \\ DOS & dos \end{bmatrix} \& \begin{bmatrix} UNO & uno \\ TRES & tres \end{bmatrix} = \begin{bmatrix} UNO & uno \\ DOS & dos \\ TRES & tres \end{bmatrix}$$

$$\begin{bmatrix} UNO & uno \\ DOS & dos \end{bmatrix} \& \begin{bmatrix} UNO & dos \\ TRES & tres \end{bmatrix} = \times$$

En el primer ejemplo de (15) se observa como las dos descripciones se combinan para crear una estructura que contiene ambas descripciones, es decir, las descripciones son compatibles y la unificación fue posible. En el segundo ejemplo, se observa lo contrario, ambas descripciones hacen referencia al atributo UNO con valores que no son compatibles entre sí, evitándose la unificación de ambas descripciones. La compatibilidad entre valores se presenta de dos forma; la primera, cuando el tipo que lo representa es el mismo y, la segunda, cuando uno de lo tipos que lo representa es descendiente del otro. Por ejemplo, si existe un tipo que representa al valor *números* que es padre de los tipos *uno*, *dos* y *tres*, la unificación entre las siguientes estructuras es posible:

(16)

$$\begin{bmatrix} UNO & números \\ DOS & dos \\ TRES & tres \end{bmatrix} \& \begin{bmatrix} UNO & uno \\ TRES & números \end{bmatrix} = \begin{bmatrix} UNO & uno \\ DOS & dos \\ TRES & tres \end{bmatrix}$$

Es importante que notar que la unificación entre dos valores compatibles da como resultado al tipo de menor jerarquía; es por esto que el tipo *números* ya no aparece en la descripción.

En este capítulo se han presentado un conjunto de fenómenos lingüísticos y mecanismos computacionales en los cuales se basa el formalismo de HPSG para crear un modelo de la sintaxis del lenguaje. Los fenómenos aquí expuestos son la base del

⁶ Generalmente la unificación se representa con el símbolo \prime ; pero por razones de tipografía en este trabajo utilizamos el $\&$, el cual es utilizado por LKB.

formalismo; pero además, son de suma importancia para identificar las propiedades que poseen los verbos auxiliares del español. Los mecanismos computacionales constituyen la maquinaria de la que se vale HPSG para crear una gramática; la definición de los elementos que componen a HPSG se discute en el siguiente capítulo.

Capítulo 3: HPSG

En este capítulo se presentan las principales características que componen al formalismo Head-driven Phrase Structure Grammar (HPSG) [Sag y Wasow, 1999]. HPSG es un teoría gramatical que permite construir un modelo sintáctico del lenguaje, una gramática. Las gramáticas en HPSG están compuestas por reglas gramaticales, principios generales, reglas léxicas, una jerarquía de tipos y un lexicón. HPSG captura los fenómenos generales de la lengua por medio de las reglas gramaticales y los principios. Los fenómenos particulares a una lengua son capturados a través de las reglas léxicas y las entradas léxicas. La jerarquía de tipos engloba a tanto a los tipos que representan fenómenos generales como particulares de la lengua. Los verbos auxiliares corresponden a un fenómeno general de la lengua para el cual HPSG cuenta con tipos especializados. Sin embargo, el fenómeno de verbos auxiliares es particular en el contexto del español; por lo cual, en la presente tesis se propone un conjunto tipos y entradas léxicas que integran el sistema de verbos auxiliares del español.

Una de las características principales de HPSG es que está orientado al lexicón; éste corresponde al listado de los símbolos básicos que componen estructuras sintácticas. En el lenguaje estos símbolos básicos son las palabras. En una gramática orientada al lexicón las palabras van acompañadas de una descripción de sus características sintácticas con el fin de definir las estructuras en las que toman parte. Adicionalmente, el hecho que HPSG esté orientado al lexicón hace el conjunto de reglas gramaticales que lo integran sea mínimo. En la presente tesis se abordan un total de cuatro reglas gramaticales: *Head Complement Rule* (Regla de complemento cabeza, HCR), *Head Specifier Rule* (Regla de especificador cabeza, HSR), *Head Modifier Rule* (Regla de modificador cabeza, HMR) y *Subject-Pro Drop Rule* (Regla de sujeto-Pro omitido, SPDR). A continuación se describirá de manera breve la función que cumple cada una de esta reglas, con excepción de SPDR por ser una regla específica del español, y los principios generales de HPSG.

Las reglas *gramaticales* se encargan de generar a las frases que integran a una oración. La regla HCR captura el fenómeno de complementaridad descrito en el capítulo anterior. En la oración ejemplo *el niño alto entregó el juguete a Juan*, el verbo *entregar* necesita de sus complementos, de un objeto directo y uno indirecto. La regla HCR es la

encargada de combinar a este verbo con sus respectivos complementos, *el juguete* y *a Juan*, para así obtener la frase verbal de *entregó el juguete a Juan*, donde *entregó* es la cabeza de frase. Por otro lado, la regla HSR captura el fenómeno de especificación. En la frase *el niño*, el sustantivo *niño*, al ser un sustantivo común, necesita de un artículo, un especificador, que sea compatible con su concordancia de género masculino y tercera persona del singular. HSR es la regla encargada de combinar a *niño* con su especificador *el*, el cual cumple con las propiedades requeridas. Por último, la regla de HMR captura el fenómeno de modificación. Por ejemplo, la palabra *alto* no es una palabra que sea necesaria para la frase *el niño*, pero adjetivos como *alto* comúnmente agregan información que ayuda en la determinación del objeto. HMR es la regla que combina a los modificadores, como *alto*, con los elementos a los que puede modifica, en este caso *el niño*. En la definición de los verbos auxiliares será muy importante distinguir las características que tienen su complemento, el verbo en forma no personal, y su especificador, el sujeto.

Los principios consisten en un conjunto de condiciones que se deben satisfacer en la aplicación de una regla. En los ejemplos de las reglas de HCR, HSR y HMR que se mencionaron antes, intervienen los principios de *Head Feature Principle* (Principio de propiedad cabeza, HFP) y de *Valence Principle* (Principio de valencia, VP). El principio de HFP garantiza que ciertas propiedades sintácticas clasificadas como *head* se compartan entre la cabeza frase y frase. Por ejemplo, HFP garantiza que la concordancia, género masculino y tercera persona del singular, en la oración *el niño alto* sea la misma que la de la cabeza palabra *niño* después de haberse aplicado las reglas HSR y HMR. Por otro lado, VP garantiza que los atributos de una palabra, complementos y especificador, se sigan solicitando después de la aplicación de una regla, siempre y cuando la regla no involucre a algunos de estos términos. Por ejemplo el verbo *entregar* se ha combinado con sus complementos en la frase *entregó el juguete a Juan*, pero a este verbo aún le falta combinarse con su sujeto en tercera persona del singular. El principio de VP garantiza que el especificador, el sujeto en este caso, sea solicitando por la frase y en consecuencia, la regla HSR puede aplicarse para así obtener la estructura sintáctica de la oración *el niño alto entregó el juguete a Juan*. En seguida se describe de manera breve la función que desempeñan las reglas léxicas y su diferencia de las reglas gramaticales.

Las reglas *léxicas* se encargan de preparar a las entradas léxicas para que puedan ser utilizadas por las reglas gramaticales. En nuestra oración ejemplo una regla léxica conjuga el verbo *entregar* a *entregó* y complementa la información sintáctica que conlleva la conjugación, concordancia y tiempo. HPSG distingue entre los elementos que son tomados por una regla gramatical, *words* y los elementos que son tomados por una regla léxica, *lexemes*. Los *lexemes* representan una abstracción de los atributos de las palabras, la cual ya se encuentra interpretada en los tipos *words*. La interpretación de la abstracción se lleva a cabo por medio del principio de *Argument Realization* (Principio de realización de argumentos, ARP). La aplicación de este principio especifica que argumentos de las palabras desempeñan un papel de especificador y cuales de complemento.

Antes de finalizar debemos resaltar el tratamiento que HPSG hace de la semántica. Para la representación del significado de una oración HPSG utiliza un enfoque de composición semántica, en donde el significado total de la oración se construye a partir de los significados de los elementos que *integran* a la oración y de las relaciones sintácticas que se establecen entre ellos. Para lograr lo anterior HPSG cuenta con los principios de *Semantic Compositionality Principle* (Principio de composicionalidad semántica, SCP) y *Semantic Inheritance Principle* (Principio de herencia semántica, SIP). El principio de SCP establece que en la aplicación de una regla la representación de la semántica en la frase resultante será la suma de las representaciones semánticas de los elementos combinados. El principio de SIP establece que la representación semántica principal de una frase es la de la cabeza frase. Para el sistema de verbos auxiliares del español, SCP y SIP reproducen la composición de semántica; solo resta definir los elementos que intervienen en las representaciones semánticas y las relaciones que se establecen entre estos y las representaciones semánticas de los elementos de la oración.

En las secciones siguientes se describen con mayor precisión los conceptos de reglas gramaticales, principios, reglas léxicas, entradas léxicas y la semántica de HPSG aquí introducidos. En la sección 3.1, se presentan y analizan los elementos que componen HPSG, utilizando los conceptos *computacionales* vistos en el capítulo anterior. Adicionalmente, en la sección 3.2 se expone el análisis sintáctico de la oración *el niño alto entregó el juguete a Juan*, donde se verá la interacción de los diferentes elementos que componen a HPSG.

En este capítulo se explica el formalismo escogido para capturar el comportamiento de los verbos auxiliares del español, lo que resulta de suma importancia para comprender el modelo de verbos auxiliares que se propone en esta tesis. Los conceptos aquí presentados *son* retomados en el capítulo cinco, en donde se expone el modelo computacional de los verbos auxiliares del español en HPSG; y en el capítulo seis donde se presenta parte de la implementación de una gramática del español y el modelo del sistema de verbos auxiliares del español para su validación computacional.

3.1 Componentes de HPSG

3.1.1 Atributo HEAD

El formalismo de HPSG utiliza a las estructuras atributo valor para representar las características lingüísticas de *palabras*, frases y oraciones. Para lograr lo anterior HPSG utiliza un conjunto de pares atributos valor para capturar la información sintáctica y semántica. Algunos de estos atributos interactúan de manera directa con las reglas y principios que se discuten más adelante.

En las estructuras de HPSG es muy importante que la información este ordenada según su naturaleza. HPSG agrupa a las características sintácticas en el atributo SYN (*SYNTAX*) y las semánticas en el atributo SEM (*SEMANTICS*). Estos dos atributos son los primeros en una estructura atributo valor de HPSG. En esta sección y las siguientes nos enfocaremos en los atributos que forman parte de SYN; los atributos de SEM se abordan en la sección 3.1.5.

En la sección 2.1.2 del capítulo anterior se expuso el fenómeno de cabeza de frase, el cual consiste que una frase comparte un conjunto de propiedades llamadas de cabeza con una de las palabra que la integra; para agrupar este conjunto de propiedades HPSG utiliza al atributo HEAD. Dos de los atributos principales que contiene HEAD son la categoría gramatical CAT y la concordancia AGR (*AGREEMENT*). El atributo CAT toma algunos de los siguientes valores *verb*, *nom*, *prep*, *det*, *adj* y *adv*, los cuales corresponden a verbo, sustantivo, preposición, determinante⁷, adjetivo y adverbio. El atributo AGR agrupa a los

⁷ Los determinantes son palabras cuya función es determinar el significado de otra palabra; los artículos son un tipo de determinantes.

atributos que representan la concordancia, estos atributos son NUM para el número, PER para la persona y GEND para el género. Un atributo HEAD luce de la siguiente forma:

(1)

$$HEAD \left[\begin{array}{c} CAT \\ AGR \end{array} \left[\begin{array}{cc} agr & cat \\ NUM & num \\ GEND & gend \\ PER & per \end{array} \right] \right]$$

Los valores indicados con *cat*, *num*, *gend* y *per* representan valores genéricos de los atributos correspondientes; en la jerarquía de tipos estos valores están representados por un tipo el cual es padre de los tipos que representan valores más específicos; por ejemplo, *num* es padre de los tipos *sng* y *plur* que representan a los valores de número singular y plural. En el apéndice B de este trabajo se presenta a la jerarquía de tipos de HPSG. El atributo HEAD agrupa a las propiedades cabeza, pero no es suficiente para reproducir el fenómeno de cabeza de frase; en la sección 3.1.4 se introduce el principio HFP que utilizando el atributo HEAD reproduce el fenómeno.

3.1.2 Listas de estructuras atributo valor

En el capítulo anterior se habló ampliamente de los fenómenos de complementariedad, especificación y modificación. En los dos primeros una palabra requiere de otras, llamadas argumentos, para lograr su significación; mientras que en el tercero una palabra es capaz de determinar con mayor precisión las características de otra. El formalismo de HPSG cuenta con tres atributos para representar a las palabras requeridas y a las palabras a modificar. Estos atributos son COMPS (*COMPléments*), SPR (*SPecifieR*) y MOD (*MODifier*).

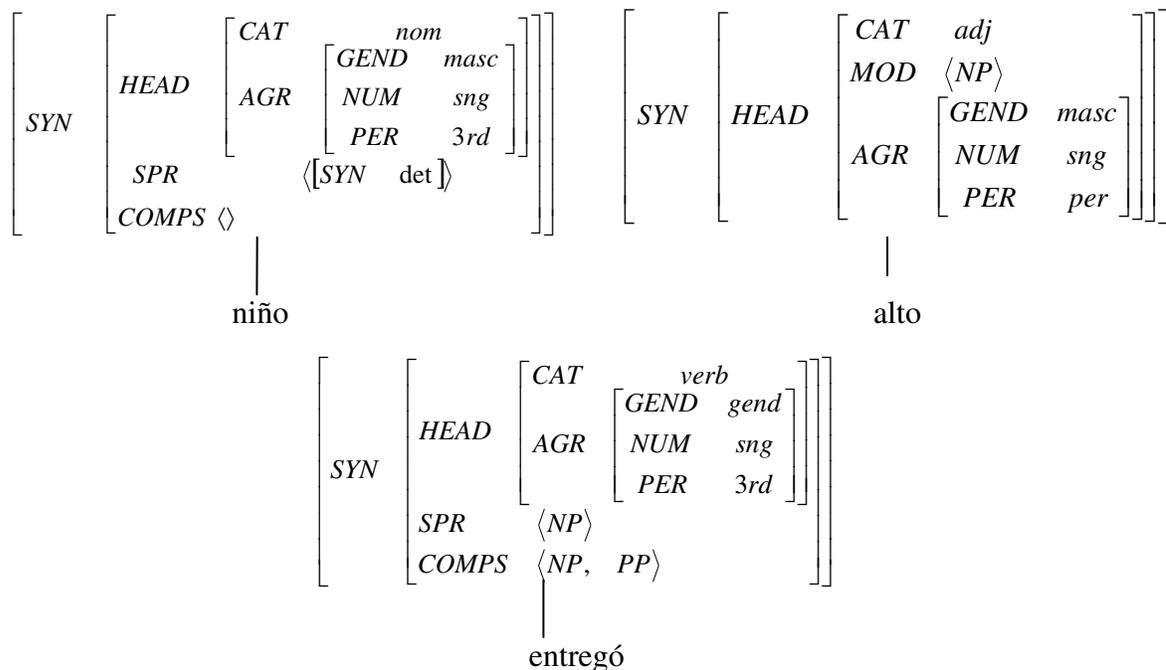
Los atributos COMPS, SPR y MOD tienen en común que son una lista de estructuras atributo valor que representan a otros elementos sintácticos, es decir, palabras o frases. El atributo COMPS especifica la secuencia de complementos requeridos por el símbolo, el de SPR a la de especificadores y el de MOD a la lista de los elementos a los que dicho símbolo puede modificar. Para representar una lista de estructura atributo valor HPSG encierra entre los símbolos “<” y “>” a los elementos de la lista y son separados por comas. Estos atributos son definidos en HPSG de la siguiente forma:

(2)

$$\left[\begin{array}{c} SYN \\ \left[\begin{array}{c} HEAD \\ SPR \\ COMPS \end{array} \left[\begin{array}{c} [MOD \langle [] \dots \rangle] \\ \langle [] \dots \rangle \\ \langle [] \dots \rangle \end{array} \right] \end{array} \right]$$

En (2) es importante observar la posición que ocupan los atributos; mientras SPR y COMPS están al mismo nivel que el atributo HEAD, el atributo MOD forma parte de este atributo. En esta estructura atributo valor el símbolo “[]” indica que debe haber una estructura atributo valor, en este caso que represente la estructura sintáctica de una palabra o frase; los símbolos “...” señalan la posibilidad de existencia de más descripciones sintácticas. Al igual que ocurre con el atributo HEAD, los atributos SPR, COMPS y MOD agrupan la información necesaria para los fenómenos de especificación, complementariedad y modificación, pero no son suficientes para capturar todo el fenómeno. En la siguiente sección se expone el conjunto de reglas que conjuntamente con los atributos hacen posible representar a estos fenómenos en HPSG. A continuación, se presenta cómo lucen los atributos antes presentados y el atributo HEAD para las palabras *niño*, *alto* y *entregó*:

(3)



Cada una de estas estructuras representa las características sintácticas de las palabras. La primera, se refiere a una palabra cuya categoría gramatical es de sustantivo (*nom*), que tiene una concordancia de tercera persona del singular y masculina (*masc*, *sng* y *3rd*), que necesita un especificador del tipo *det* (determinante, artículo) y que no requiere de complementos obligatorias. La segunda estructura, describe una un adjetivo (*adj*) cuya concordancia es singular y masculino (*sng* y *masc*), en este caso se indica que esta palabra

es capaz de modificar a una frase sustantiva (MOD); un punto importante es notar que no se hace referencia a los atributos SPR y COMPS, en las estructuras de HPSG es muy común omitir algunos de los atributos ya sea porque se encuentran descritos en algún tipo padre, porque no están o porque no son relevantes. Por último, la tercera estructura describe a un verbo (*verb*) cuya concordancia es de tercera persona del singular (3rd y *sng*), esta estructura indica que la palabra necesita como especificador una frase sustantiva (SPR) y que necesita de dos complementos, en donde el primero es una frase sustantiva y el segunda una preposicional. Como se aprecia en estas estructuras, las principales propiedades sintácticas se encuentran representadas; en las siguientes secciones se presenta a los componentes de HPSG que basándose en estos atributos reproducen los fenómenos de cabeza de frase, complementariedad, especificación y modificación.

3.1.3 Reglas gramaticales

Las reglas gramaticales son los elementos de HPSG que se encargan de agrupar a las descripciones de las palabras en frases u oraciones. En la sección anterior se introdujeron tres atributos que representan la información involucrada en los fenómenos de complementariedad, especificación y modificación; estos atributos indican las estructuras que una palabra necesita para que estén agrupadas en una frase. Las reglas de *Head Complement Rule* (HCR), *Head Specifier Rule* (HSR) y *Head Modifier Rule* (HMR) agrupan a los elementos necesarios que conforman una frase en HPSG; mientras que los principios son un conjunto de condiciones que debe satisfacer la aplicación de una regla. A continuación se presenta la representación de las tres reglas utilizando estructuras atributo valor:

(4) a. Head Complement Rule

$$\left[\begin{array}{l} \textit{phrase} \\ \textit{COMPS} \langle \rangle \end{array} \right] \longrightarrow H \left[\begin{array}{l} \textit{word} \\ \textit{SYN} \textit{COMPS} \langle \#L \rangle \end{array} \right] \#2$$

b. Head Specifier Rule

$$\left[\begin{array}{l} \textit{phrase} \\ \textit{SPR} \langle \rangle \end{array} \right] \longrightarrow \#L \ H \left[\begin{array}{l} \textit{phrase} \\ \textit{SYN} \textit{SPR} \langle \#L \rangle \end{array} \right]$$

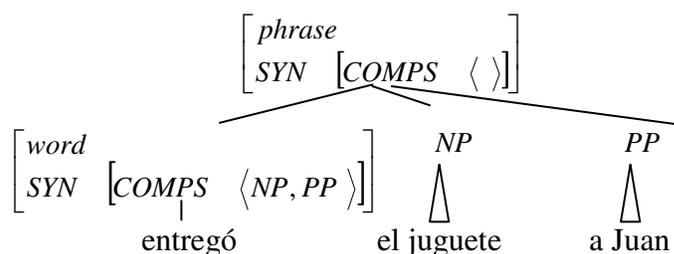
c. Head Modifier Rule

$$[\textit{phrase}] \rightarrow H \#1 [\textit{phrase}] \left[\begin{array}{l} \textit{phrase} \\ \textit{SYN} \textit{HEAD} \textit{MOD} \ \#1 \end{array} \right]$$

Antes de presentar el significado de estas expresiones debemos hacer un paréntesis referente a la notación. En primer lugar, en estas expresiones aparece el símbolo “|” el cual nos sirve para hacer referencia a una secuencia de atributos sin la necesidad de abrir y cerrar los corchetes. Por ejemplo en © el atributo MOD está contenido en el atributo HEAD, que a su vez está en SYN. También en estas expresiones aparece el símbolo “#”; este debe ser seguido por una etiqueta que puede ser un número o letra; en el caso de las dos primeras reglas es L, mientras que en la segunda es 1. Este símbolo indica que los valores en donde se encuentran estas etiquetas deben unificar, es decir, deben ser compatibles. El símbolo “→” separa en dos grupos a las estructuras que definen la regla; el de la izquierda está conformado por un elemento, al cual se le denomina madre de la regla, y el de la derecha especifica a las estructuras hijas. Una regla propone una secuencia de estructuras hijas que pueden agruparse y representarse con una estructura madre; se marca anteponiendo con una “H” aquella estructura hija que sea la cabeza de frase. A continuación se presenta la interpretación de cada una de las reglas con un ejemplo.

La regla HCR agrupa a una estructura atributo valor que representa a una palabra con sus complementos. La regla dice si una estructura tiene un atributo COMPS cuyas descripciones coinciden con los elementos que le siguen, se agrupa a los elementos en una frase cuyo atributo COMPS está vacío; en términos computacionales se dice que el atributo ha sido saturado. El árbol sintáctico (5) muestra la aplicación de la regla HCR⁸:

(5)

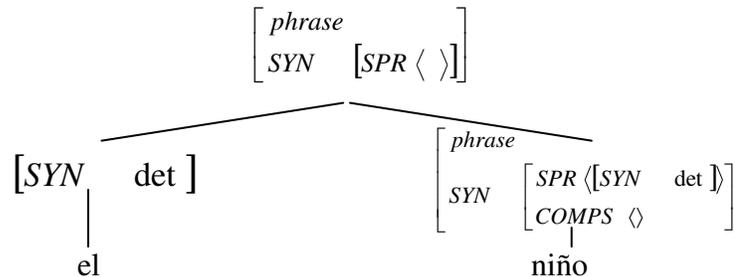


La regla HSR agrupa a una estructura atributo valor que representa a una palabra con su especificador descrito en el atributo SPR. La regla dice se agrupa a una estructura si tiene un atributo SPR cuya descripción coinciden con los elementos que le anteceden se en

⁸ En este árbol se aprecia los símbolos *NP* y *PP*, los cuales representan a una frase sustantiva y preposicional respectivamente, en la sección 3.2 se aprecia la forma en que se agrupan los elementos de la frase.

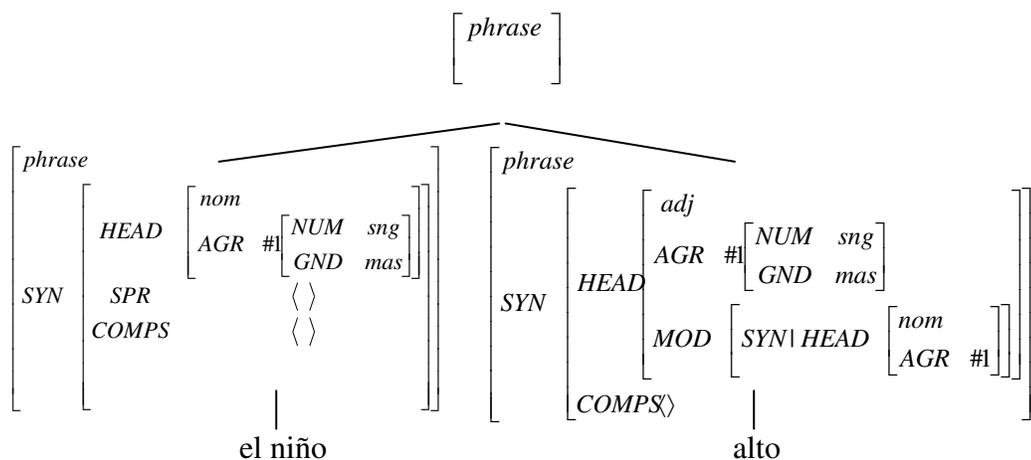
una frase cuyo atributo SPR está saturado. El siguiente árbol sintáctico muestra la aplicación de la regla HSR:

(6)



Por último, la regla HMR agrupa una estructura atributo que representa a un modificador con la palabra que puede modificar. La regla dice si una estructura atributo valor tiene un atributo MOD cuya descripción coincide con el elemento que le antecede se puede agrupar en una frase. Esta regla es diferente con respecto a las dos anteriores con respecto a la hija cabeza; mientras que en las dos anteriores la hija cabeza era la estructura que contenía al atributo SPR o COMPS, en esta última la cabeza no es la estructura con el atributo MOD. El siguiente árbol sintáctico muestra la aplicación de la regla HMR:

(7)



En este último ejemplo se introduce un factor que no había sido considerado en los ejemplos anteriores; este se refiere a la concordancia entre los elementos de una frase. La concordancia, representada en el atributo AGR, debe ser compatible, debe unificar, con la descripción de la concordancia del atributo MOD. En el ejemplo, se observa que el atributo AGR de la palabra y de MOD tienen la etiqueta #1 que indica la unificación entre ambas

estructuras. Una situación similar se presenta en el ejemplo (6) donde la concordancia de la palabra *niño* debe coincidir con la de su especificador.

3.1.4 Principios

Los principios son un conjunto de condiciones que se deben satisfacer durante la aplicación de una regla. Estas condiciones están en función de los fenómenos lingüísticos. En esta sección se presentan dos principios, uno relacionado con el fenómeno de cabeza de frase y otro con los de complementariedad y especificación. Estos principios son *Head Feature Principle* (HFP) y *Valence Principle* (VP).

El principio HFP reproduce el fenómeno de cabeza de frase descrito en el capítulo anterior, garantiza que las propiedades cabeza, que en HPSG corresponden al representadas en el atributo HEAD de la hija cabeza se compartan con la madre de la regla. HFP dice:

- (8) *El valor del atributo HEAD de la hija cabeza es el valor del atributo HEAD de la madre.*

De manera esquemática el principio se enuncia de la siguiente forma:

(9)

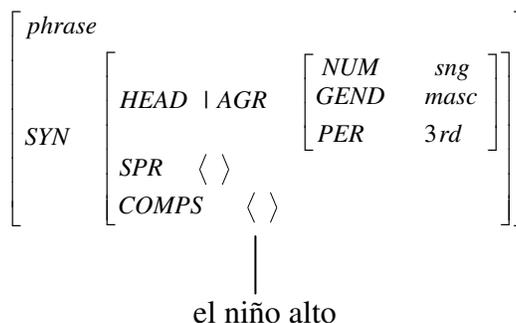
$$\left[\begin{array}{l} phrase \\ [HEAD \quad \#1] \end{array} \right] \longrightarrow \dots H [HEAD \quad \#1] ..$$

El principio de valencia garantiza que los atributos COMPS y SPR se compartan con la madre de una regla siempre y cuando la regla no los modifique. El principio se enuncia de la siguiente manera:

- (10) *El valor de SPR y COMPS de la hija cabeza se unifica con los atributos SPR y COMPS de la madre siempre y cuando no se aplique alguna regla que modifique el valor de uno de estos atributos.*

En particular, para este trabajo las únicas reglas que modifican a estos valores son HSR y HCR. La siguiente estructura, que representa a la frase *el niño alto*, se muestra los atributos teniendo en cuenta a los dos principios introducidos en esta sección.

(11)

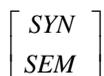


Los atributo HEAD, SPR y COMPS corresponden a los de la hija cabeza de frase, en este caso a la estructura que representa a la frase *el niño*. El hecho de que SPR y COMPS estén vacíos es porque en la frase *el niño* estos atributos estaban saturados.

3.1.5 Semántica en HPSG

Una gramática tiene el objetivo de asignar una estructura sintáctica a una oración; pero adicionalmente a través de la estructura y de los significados individuales de las palabras se puede obtener una representación del significado de la oración. El formalismo de HPSG utiliza un conjunto de atributos y principios para representar y reproducir la información semántica. Se propone que la información relacionada con semántica de una oración se agrupe en el atributo SEM el cual se encuentra al mismo nivel que el atributo SYN.

(12)



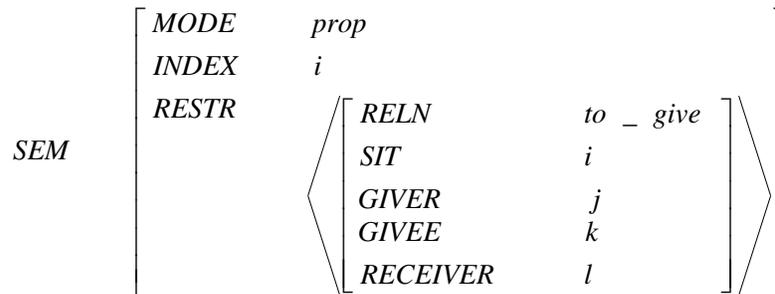
Para representar el significado de las palabra y frases se captura la siguiente información: el modo semántico, un índice semántico y las restricciones semánticas. El modo semántico nos indica el tipo de significado que trata de expresar la oración, en HPSG se utilizaran cuatro modos semánticos, que se ilustran en la siguiente tabla:

(13)

Modo semántico	Tipo de frase	Ejemplo
Proposición	Afirmativas	Juan duerme
Interrogación	Interrogativas	¿Juan duerme?
Directivas	Imperativas	¡Juan duerme!
Referencias	Frases sustantivas	La UNAM

Un índice tiene como objetivo definir de manera única a los elementos semánticos que intervienen en las restricciones semánticas. Por último, las restricciones semánticas son estructuras atributo valor que describen el significado de la palabras. Por ejemplo, en (14) se ilustra la semántica que tiene el verbo *entregar*:

(14)



El atributo MODE señala que la semántica es del tipo proposición. El atributo de INDEX es el atributo que identifica a la restricción que representa a esta estructura, este atributo es muy importante cuando intervienen varias restricciones. En este caso, este atributo corresponde a *i* que se relaciona a través del atributo SIT con la restricción que define una relación de entregar algo a alguien (*RELN to_give*) en donde existen tres argumentos, el que da (GIVER), lo dado(GIVEE) y quien recibe (RECEIVER). Esta restricción podría representarse como una proposición de la siguiente forma: *to_give(j,k,l)* la cual puede ser interpretada como la persona *j* entrega el objeto *k* a la persona *l*.

El formalismo de HPSG utiliza un enfoque composicional de la semántica, la cual establece que la estructura de la semántica de una oración está en función de la restricciones semánticas de las palabras que la integran. El principio de *Semantic Compositionality Principle* (SCP) es el principio encargado de establecer esta función en HPSG, este principio dice:

(15) *El valor del atributo REST de la madre es la suma de los atributos REST de la hijas.*

Se entiende como suma a la concatenación de las estructuras bajo una misma lista. La interrelación de las restricciones semánticas serán definida por las propias estructuras semánticas.

El principio SCP garantiza que las restricciones llegan a la madre de la regla, pero aún nos queda decidir cual de estas restricciones es la que representa a la semántica de esta

nueva estructura, para resolver este problema HPSG cuenta con el principio de *Semantic Inheritance Principle* (SIP) el cual establece los siguiente:

(16) *El valor de los atributos MODE e INDEX de la madre en una regla son los valores de los atributos MODE e INDEX de la hija cabeza.*

A continuación se presenta la semántica de la frase *entregó el juguete a Juan*; después de la aplicación de la regla HCR como se presentó en el ejemplo (5) y tomando en cuenta los principios SCP y SIP.

(17)

$$SEM \left[\begin{array}{l} MODE \quad prop \\ INDEX \quad i \\ RESTR \quad \left\langle \begin{array}{l} RELN \quad to_give \\ SIT \quad i \\ GIVER \quad j \\ GIVEE \quad k \\ RECEIVER \quad l \end{array} \right\rangle \left[\begin{array}{l} RELN \quad name \\ SIT \quad k \\ NAME \quad toy \end{array} \right] \left[\begin{array}{l} RELN \quad name \\ SIT \quad l \\ NAME \quad juan \end{array} \right] \right\rangle \end{array} \right]$$

El principio de SCP permite que las tres restricciones sean agrupadas en esta estructura atributo valor al concatenarlas. Por otro lado, el principio de SIP asigna los mismos valores que tiene la cabeza hija, el verbo *entregó*, a los atributos MODE e INDEX de esta estructura; la semántica de la cabeza hija es la que aparece en el ejemplo (14). Los valores de los atributos GIVER y GIVEE coinciden con los atributos SIT de las otras dos restricciones debido a que la descripción de los complementos del verbo *entregar* solicita que estos compartan su índice con esta restricción. Esto se puede apreciar en la siguiente descripción parcial del verbo *entregar*.

(18)

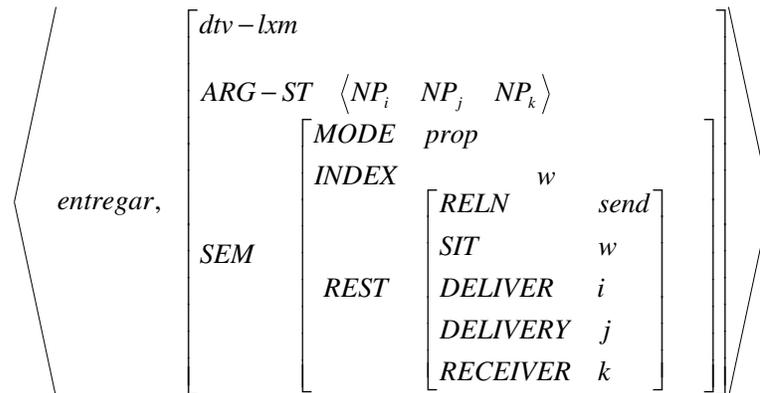
$$\left[\begin{array}{l} word \\ COMPS \quad \langle [SEM | INDEX \ #1], [SEM | INDEX \ #2] \rangle \\ SEM | RESTR \quad \left\langle \begin{array}{l} GIVEE \quad \#1 \\ RECEIVER \quad \#2 \end{array} \right\rangle \end{array} \right]$$

Esta es la forma como los atributos de las restricciones se interrelacionan entre sí. Los atributos y principios semánticos garantizan que la información semántica este representada en las estructuras atributo valor obteniendo con esto una representación de la semántica de las oraciones.

3.1.6 Satisfacción léxica

En una gramática el lexicón es el conjunto de símbolos básicos que a través de las reglas constituyen unidades más complejas. En HPSG el lexicón esta conformado una lista de entradas léxicas la cuales están formadas por la ortografía de una palabra y una estructura que define sus características; estos dos elementos comúnmente se encierran entre los símbolos “<” y ”>” y son separados por una coma. Por ejemplo, la entrada léxica para la palabra *entregar* es la siguiente:

(19)



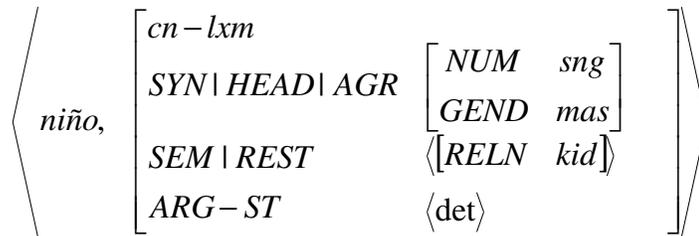
Todas las entradas léxicas de HPSG son del tipo *lxm*; este tipo permite organizar los comportamientos de las entradas léxicas en grupos de palabras a través de la jerarquía de tipos. En (19) vemos que la entrada léxica es del tipo *dtv-lxm*; en este tipo se describen las propiedades que poseen los verbos ditransitivos. Otra diferencia importante en esta descripción es la aparición del atributo ARG-ST, el cual agrupa en una a los argumentos de las palabras, es decir, a complementos y especificador. El tipo *word*, visto en la descripción de las reglas gramaticales, representa a las palabras cuando están listas para ser tomadas como un argumento de una regla gramatical; mientras que el tipo *lxm* representa a las palabras antes de que esto sea suceda. La condiciones satisfechas por una entrada léxica para convertirse del tipo *lxm* al tipo *word* es lo que se conoce como satisfacción léxica, la cual se describe de la siguiente forma: una estructura *F* del tipo *word* satisface léxicamente a la entrada léxica $\langle w,d \rangle$ si y sólo si:

1. *F* subsume a *d*
2. *F* cumple con el Argument Realization Principle, que dice:

$$\left[\begin{array}{l} SYN \\ ARG-ST \end{array} \left[\begin{array}{l} SPR \quad \#1 \\ COMPS \quad \#2 \end{array} \right] \right] \langle \#1 \oplus \#2 \rangle$$

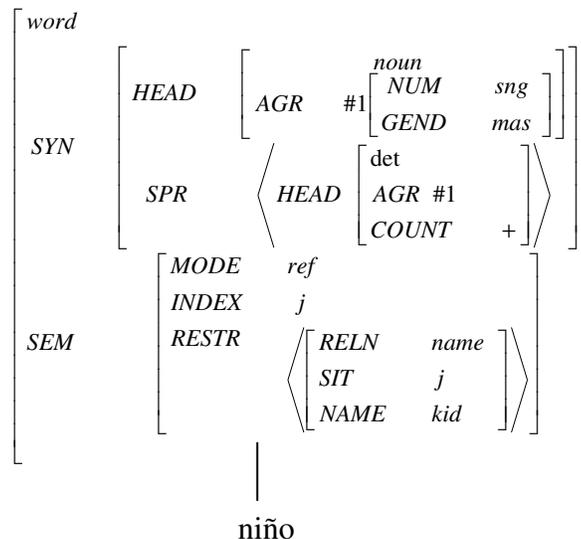
En donde w corresponde a la ortografía de la palabra, d es una descripción del tipo lxm y F es la estructura sintáctica que cumple con la satisfacción léxica. Las condiciones anteriores, establecen que para que una entrada léxica pueda ser usada como una estructura válida es necesario crear una estructura de atributos F que subsuma a la abstracción capturada en el lexicon a través de w y que la lista ARG-ST divida sus elementos en los valores de SPR y COMPS. Por ejemplo, la entrada léxica para la palabra *niño* es la siguiente:

(20)



La estructura F que satisface léxicamente a esta entrada léxica es:

(21)

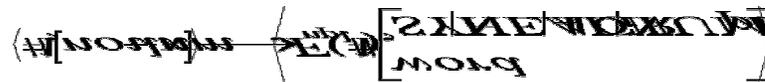


Esta descripción subsume a la descrita por la entrada léxica, la cual es la primera condición de la satisfacción léxica. La segunda condición se satisface en la separación del atributo ARG-ST de la entrada léxica en dos elementos; el primero se encuentra en SPR y el segundo en COMPS, pero debido a que este último corresponde a la lista vacía el atributo COMPS está vacío.

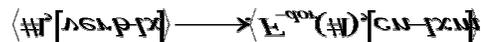
3.1.7 Reglas léxicas

El conjunto de reglas y principios que utiliza HPSG son mínimas; con estas reglas se pretende captar todos los fenómenos de sintácticos de cualquier lengua. El formalismo HPSG cuenta con un conjunto de reglas que reproducen fenómenos sintácticos más específicos de cada lengua; estas reglas son llamadas léxicas ya que solamente son aplicables a tipos *lxm*, los cuales se marcan con el sufijo de *lxm*. Se dividen en dos tipos: *reglas de inflexión* para convertir tipos *lexeme* en tipo *word* y *reglas derivacionales* para cambiar la categoría léxica de una palabra. Los elementos que conforman a la reglas léxicas son entradas léxicas; a la izquierda del símbolo “→” se encuentra la entrada léxica original y a la dercha la entrada léxica resultante de la aplicación de la regla. A continuación se pueden apreciar dos reglas léxicas, una de inflexión y una derivacional:

(22) a. Regla de inflexión: Regla léxica para los sustantivos plurales



b. Regla derivacional: Regla léxica para substantivación verbal



La primera regla tiene el objetivo de formar el plural de los sustantivos. La escritura de #1, que es la palabra, será cambiada por la función F_{npl} . La función F_{npl} es una función morfológica que es capaz de entregarnos la escritura de #1 en plural. La función solo se aplicará si la palabra está relacionada en una entrada léxica a una estructura de atributos del tipo *noun-lxm*, el cual es utilizado para los sustantivos. Aparte del cambio en la escritura también se hará un cambio en el tipo del objeto léxico, este cambio convierte de *lexeme* a tipo *word*, permitiendo con esto las reglas gramaticales la puedan utilizar como una estructura hija. Adicionalmente, en la regla se modifica un atributo, con el cambio en la ortografía el número del sustantivo ha dejado de ser singular para convertirse en plural; esto se refleja atribuyendo al valor *pl* al atributo NUM.

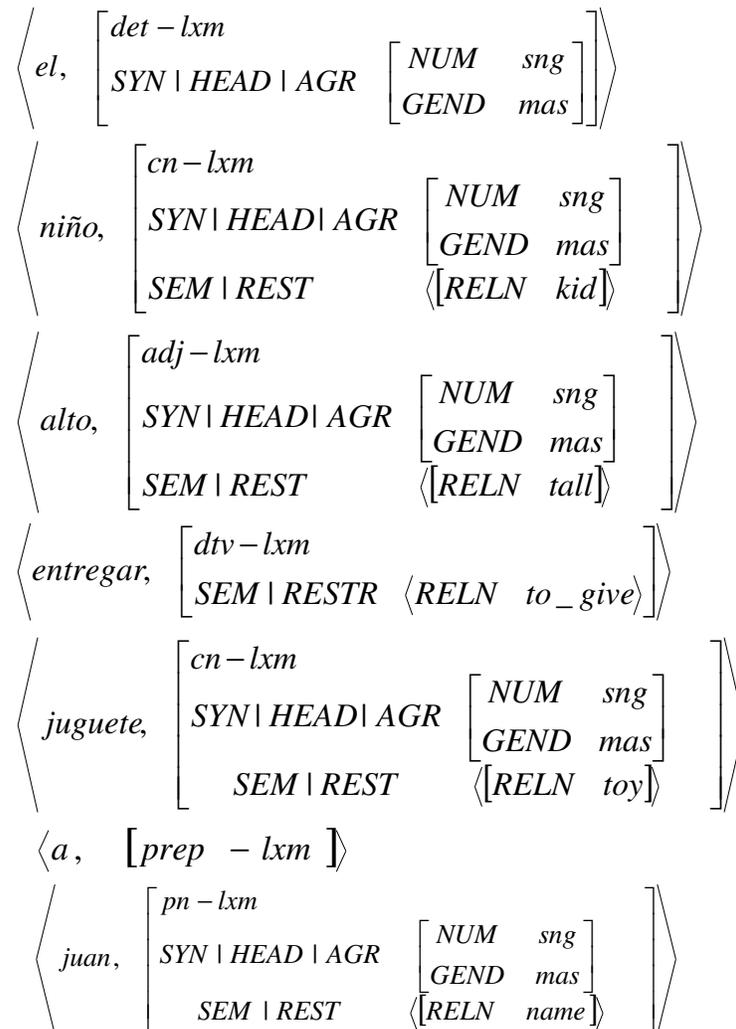
La segunda regla se encarga de substantivar a los verbos, es decir pasa un verbo a sustantivo; por ejemplo el verbo *jugar* al sustantivo *jugador*. La función morfológica F_{dor} cambia la ortografía original a la correspondiente a sustantivo. Un punto importante es que no cambia el tipo *lexeme*, por lo que la estructura resultante, podrá ser tomada por otra regla léxica pero no por una regla gramatical. Si queremos que la palabra sea utilizada por

las reglas gramaticales es necesario aplicarle otra regla, como la primera y así podemos obtener el plural *jugadores* con su estructura de atributos tipo *word*.

3.2 Ejemplo: el niño alto entregó el juguete a Juan

A continuación presentamos un pequeño ejemplo donde las reglas de HCR, HSR y HMR reconstruyen la estructura de la oración: *El niño alto entregó el juguete a Juan*. Como primer paso se exponen las entradas léxicas de cada una de las palabras:

(23)



El valor de los atributos omitidos, como SPR, COMPS o MOD, es descrito por la descripción de los tipos *lxm*; los tipos son descritos en el apéndice B de este trabajo. Para lograr obtener una estructura que defina a la palabra *entregó* es necesario cambiar la ortografía de la entrada léxica de *entregar*. Esto se logra con la siguiente regla léxica del pretérito perfecto absoluto que se enuncia de la siguiente forma:

(24)

$$\left\langle \#1, \left[\begin{array}{l} \text{verb} - \text{lxm} \\ \text{SEM} \quad [\text{REST} \quad \#A] \end{array} \right] \right\rangle \Rightarrow$$

$$\left[\begin{array}{l} \text{word} \\ \text{SYN} \mid \text{HEAD} \quad [\text{FORM} \quad \text{fin}] \\ \text{ARG} - \text{ST} \quad \langle [\text{CASE} \quad \text{nom}], \dots \rangle \\ F_{\text{PAST}} \#1, \\ \text{SEM} \quad \left[\begin{array}{l} \text{INDEX} \quad \#2 \\ \text{RESTR} \quad \#A \oplus \left[\begin{array}{l} \text{RELN} \quad t - \text{precede} \\ \text{ARG1} \quad \#2 \\ \text{ARG2} \quad \text{now} \end{array} \right] \end{array} \right] \end{array} \right]$$

Esta regla cambia la ortografía de un verbo a pasado, adecua la información de los atributos; en este caso indica que el verbo está conjugado a través del atributo FORM y agrega una restricción que indicará a la semántica que la acción ocurre en el pasado. Adicionalmente, la regla se encarga de pasar la descripción de tipo *lexeme* a *word*. Por último, la regla agrega una restricción semántica con el objetivo de representar que la acción se ejecuta en el pasado. La siguiente estructura es el resultado de aplicar la regla léxica antes descrita y de aplicar la satisfacción léxica para la palabra *entregó*:

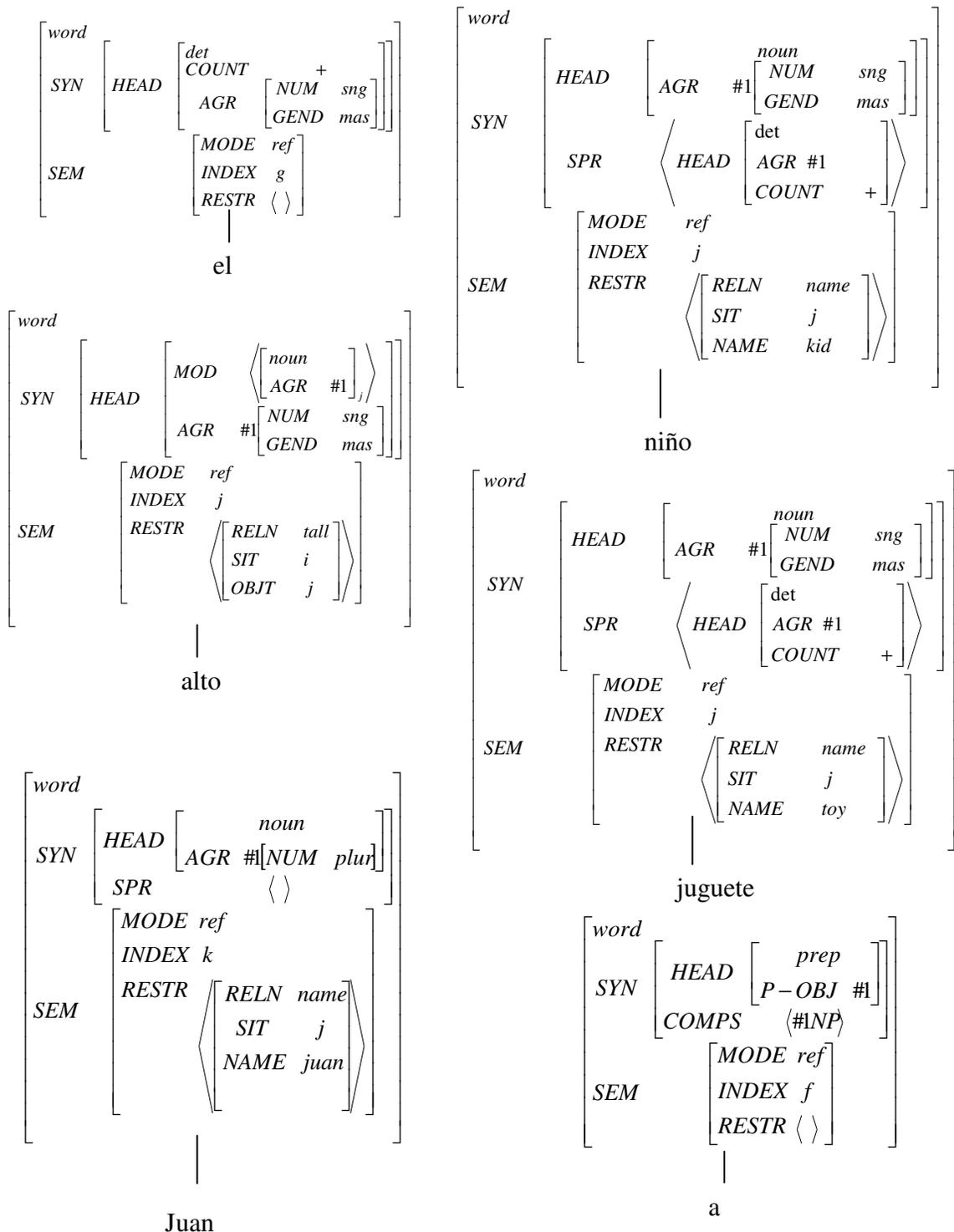
(25)

$$\left[\begin{array}{l} \text{SYN} \quad \left[\begin{array}{l} \text{HEAD} \quad \left[\begin{array}{l} \text{verb} \\ \text{AGR} \quad \text{agr} \end{array} \right] \\ \text{SPR} \quad \langle \text{NP}_i \rangle \\ \text{COMPS} \quad \langle \text{NP}_j, \text{NP}_k \rangle \end{array} \right] \\ \text{SEM} \quad \left[\begin{array}{l} \text{MODE} \quad \text{prop} \\ \text{INDEX} \\ \text{REST} \quad \left[\begin{array}{l} \text{RELN} \quad \text{to_give}^x \\ \text{SIT} \quad w \\ \text{GIVER} \quad i \\ \text{GIVEE} \quad j \\ \text{RECEIVER} \quad k \end{array} \right] \oplus \left[\begin{array}{l} \text{RELN} \quad t - \text{precede} \\ \text{ARG1} \quad x \\ \text{ARG2} \quad \text{now} \end{array} \right] \end{array} \right] \end{array} \right]$$

|
entregó

Las estructuras atributo valor que representan a las demás palabras se presentan a continuación:

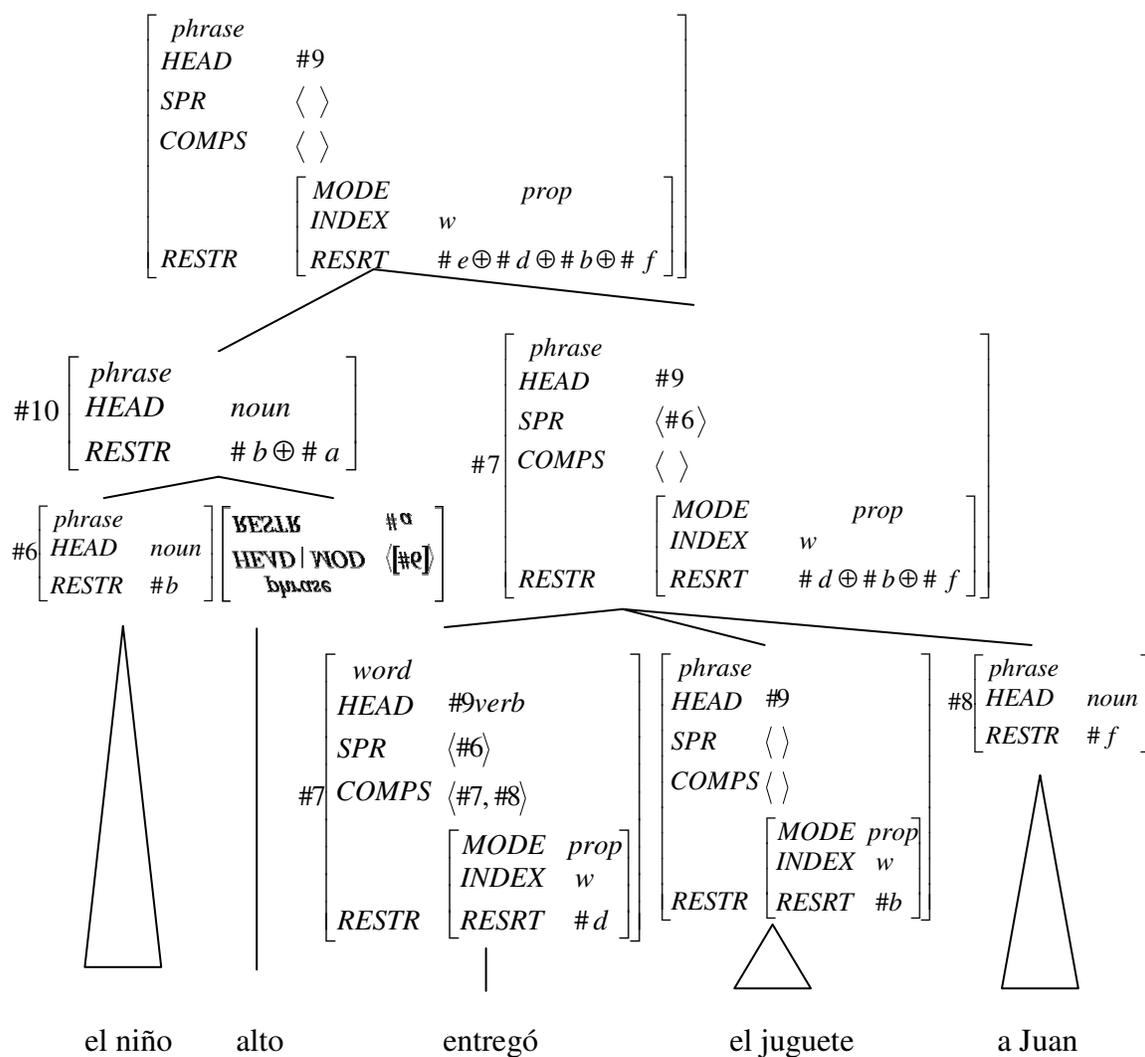
(26)



Estas estructuras definen un conjunto de características lingüísticas de cada una de las palabras. Estas características, en especial los atributos SPR, COMPS y MOD, permiten que las reglas gramaticales de HPSG puedan aplicarse y así obtener una representación de

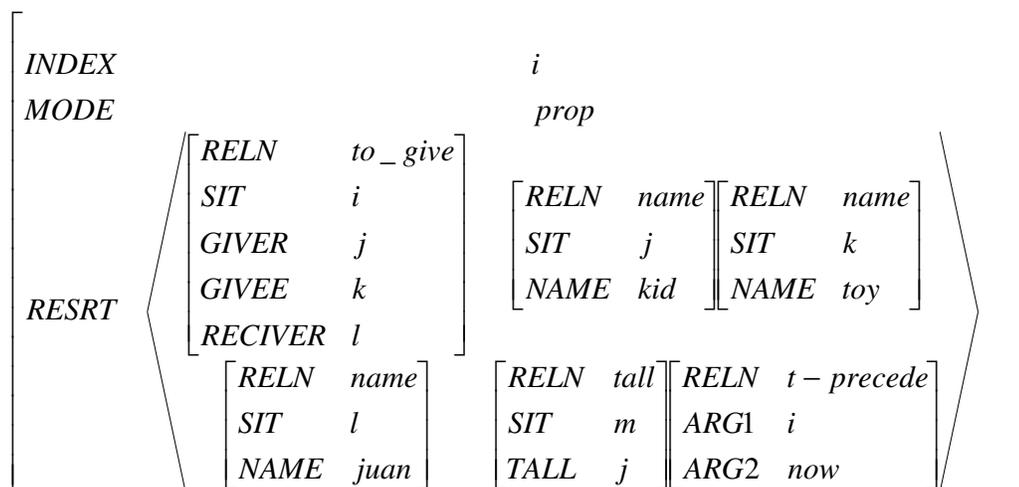
la estructura interna de la oración. En (25) observamos el árbol sintáctico de la oración *el niño alto entregó el juguete a Juan*; la regla HCR agrupa a la frase *a Juan* y la HSF a las de *el juguete* y *el niño*. Una vez que las frase *a Juan* y *el juguete* se encuentran agrupadas la regla de HCR vuelve aplicarse pero con el verbo *entregó* como cabeza de frase; esto resulta en la frase *entregó el juguete a Juan*. La regla HMR es la encargada de agrupar a la palabra *alto* con la frase *el niño*. Por ultimo, la regla HSR aplicada a *entregó* como cabeza hija hace posible arupar a la frases *el niño alto* y *entregó el juguete a Juan*, lo cual resulta en toda la oración. Note que el *HEAD* de toda la oración es el *HEAD* de la frase verbal, es decir, el verbo de la oración.

(27)



La semántica de la estructura que agrupa a toda la oración es la siguiente:

(28)



Esta representación puede interpretarse como la realización la acción i correspondiente a to_give . En donde el objeto j , el niño, es quien da; el objeto k , el juguete, es el dado y quien recibe es l , Juan. Adicionalmente existen dos restricciones, una que define el tiempo en que se realizó la acción, en el pasado, y otra que determina que j es alto.

En este capítulo hemos visto las características principales con las que cuenta HPSG; se han expuestos los principales atributos, las reglas y principios que integran al formalismo. También se hizo hincapié en los tipos lxm que definen de manera abstracta las propiedades que poseen las palabras y a través de los cuales HPSG define el lexicón. En el capítulo cinco se presenta con mayor detenimiento los tipos lxm , en especial aquellos que involucran a los verbos, con el fin de proponer el conjunto de tipos que capturan el comportamiento de los verbos auxiliares del español que es descrito en el siguiente capítulo. Los diferentes elementos aquí expuestos serán de suma importancia en la validación del modelo del sistema de verbos auxiliares del español.

Capítulo 4: Verbos auxiliares

En este capítulo se determinan las propiedades sintácticas que poseen los verbos auxiliares del español. Un modelo sintáctico de los verbos auxiliares del español, como el que se propone en este trabajo, requiere que las características de los verbos auxiliares estén claramente identificadas. Las gramáticas tradicionales (i.g., [Gili, 1991]), recurren a consideraciones que no suficientemente específicas para un análisis computacional como el que se presenta en este trabajo. Estudios más recientes entorno a los verbos auxiliares proponen características muy semejantes a las que se presentan en este trabajo, pero no las aterrizan a conceptos sintácticos [Luna,1991]. Por lo anterior, en este capítulo se presenta un conjunto de pruebas que nos permiten determinar si un verbo es auxiliar o no. Posteriormente, un análisis detallado de estas pruebas nos conduce a las propiedades de los verbos auxiliares. A continuación abordamos con un poco de mayor profundidad las consideraciones hasta ahora expuestas.

El problema de los verbos auxiliares es de actualidad debido a una falta de consenso en la definición de verbo auxiliar. Algunas gramáticas basan la caracterización de los verbos auxiliares en el cambio de significado que presentan[Gili, 1991]. En este trabajo se observa que conjuntamente con el cambio de significado que presentan los verbos auxiliares se da un cambio de comportamiento sintáctico. Otras gramáticas exponen una graduación de los verbos auxiliares; en este trabajo se observará como esta propuesta es limitada bajo un enfoque sintáctico [Alonso y Henríquez, 1967]. Por último, el análisis de este trabajo nos permitirá reinterpretar, en términos sintácticos, las conclusiones de trabajos más recientes [Luna, 1991].

Un concepto muy importante en la caracterización de los verbos auxiliares es el de agente. El agente corresponde al individuo o ente que realiza la acción que indica el verbo. En la oración *el niño alto entregó el juguete a Juan*, quién realizó la acción es el objeto al que hace referencia la frase *el niño alto*; la entidad a la que corresponde esta frase es el agente. En contraste, el paciente es el que recibe la acción del agente; en esta misma oración el paciente es la entidad al que refiere la frase *el juguete*. Tanto agente como paciente son elementos semánticos, los cuales se encuentran relacionados con elementos sintácticos de la oración. Generalmente, el agente coincide con el sujeto del verbo y el

paciente con el primer complemento del verbo. Como se verá más adelante, en la definición de las propiedades de los verbos auxiliares es de suma importancia identificar la relación que se establecen entre el agente del verbo en forma no personal y el verbo auxiliar.

En esta tesis se definen cinco pruebas que permiten identificar si un verbo se comporta como auxiliar. La primera permite verificar que el verbo aparece en una perífrasis donde el primer verbo está conjugado y además, ayuda a identificar el agente de la acción del verbo en forma no personal. En la oración *el niño alto va a entregar el juguete a Juan*, el sujeto del verbo *ir*, es quién *entregó el juguete a Juan*, es decir, es el niño alto es quién realizará la acción de entregar el juguete. La segunda prueba permite verificar que el verbos pueden ir seguido de un verbo que no necesite de agente, como es el caso de los verbos unipersonales. En la oración *va a llover*, no hay un objeto semántico que identifique a lo que llueve. La tercera prueba permite generar la forma interrogativa de una oración que pregunte por el paciente de la oración y observar que no es posible que el verbo aparezca solo. Por ejemplo, para la oración ejemplo la forma interrogativa es *¿qué va a entregar?* y no **¿qué va a?* La cuarta prueba nos muestra que si deseamos preguntar por la acción que representa el verbo en forma no personal se debe sustituir a este verbo por un verbo “comodín”. La oración ejemplo tiene la forma interrogativa de *¿qué va a hacer?* y no **¿qué va a?*, donde *hacer* sustituye a *entregar*. La quinta prueba muestra que la relación del verbo auxiliar y verbo en forma no personal no corresponde a la relación de verbo transitivo y su complemento directo. Esto se logra al tratar de construir la voz pasiva del verbo auxiliar y notar que no es posible; lo que si es posible es que el verbo auxiliar tome como complemento un verbo transitivo pasivizado. Para la oración ejemplo se observa que la oración **el juguete es ido a entregar a Juan por el niño alto* no es correcta gramaticalmente, mientras que la oración *el juguete va a ser entregado a Juan por el niño alto* si lo es.

Este conjunto de pruebas permitirá reclasificar el conjunto de verbos auxiliares propuesto por Gili Gaya compuesto por los siguientes verbos: *ir a, pasar a, echar a, venir a, volver a, haber de, haber que, tener que, deber de y llegar a* que son seguidos por un verbo en infinitivo; *estar, ir, venir, seguir y andar* que son seguidos por un verbo en gerundio; y *llevar, tener, traer, quedar, dejar, estar y ser* que son seguidos por un verbo en

participio [Gili, 1991]. Después de aplicar las pruebas a estos verbos se concluye que de sólo trece de estos verbos se comportan como auxiliar.

Un análisis más profundo de las cinco pruebas arriba mencionadas define las tres propiedades sintácticas que poseen los verbos auxiliares, las cuales son:

- Un verbo auxiliar tiene como complemento a un verbo en forma no personal.
- Un verbo auxiliar comparte su sujeto con el verbo en forma no personal.
- Un verbo auxiliar no tiene un agente.

Estas tres propiedades deberán estar representadas en un modelo de los verbos auxiliares del español.

Una consecuencia interesante de la primera propiedad de los verbos auxiliares son las oraciones anidadas; al no especificarse en la definición de la perífrasis el tipo del verbo en forma no personal es posible que el verbo auxiliar tome otro verbo auxiliar como complemento y así sucesivamente, formando las construcciones que en este trabajo denominamos anidadas; por ejemplo *el niño alto debe haber entregado el juguete a Juan* en donde aparecen los verbos *deber* y *haber*. Aquí, *deber* aparece conjugado pero el segundo auxiliar, *haber*, así como el verbo de contenido, *entregar*, están en forma no personal. Parte importante del análisis del conjunto de verbos auxiliares de Gili Gaya consistirá en analizar las condiciones en las cuales este “anidamiento” se presenta.

En las próximas secciones se aborda con mayor precisión los temas hasta ahora planteados. En la sección 4.1 se presentan los conceptos relacionados a la agentividad como los son agente, paciente, objeto directo y objeto indirecto. En la sección 4.2, se presentan las cinco pruebas del comportamiento de los verbos auxiliares, para continuar con su análisis y concluir con las tres propiedades sintácticas de los verbos auxiliares del español. En la sección 4.3 se presenta un ejercicio hecho con el conjunto de verbos auxiliares propuesto por Gili Gaya y las conclusiones a las que se llegaron sobre la naturaleza de su comportamiento.

La definición de las propiedades sintácticas que posee un verbo auxiliar que se hace en este capítulo permite desarrollar un modelo computacional de este fenómeno. En el capítulo cinco veremos a los tipos de HPSG que capturan las propiedades de los verbos auxiliares del español. Adicionalmente, en capítulo cinco se presentan la definición de tipos que capturan los comportamientos adicionales que presentan los verbos auxiliares y los

comportamientos de los verbos que no son auxiliares. En el capítulo seis se presenta la implementación computacional que se hace del sistema de verbos auxiliares el español en donde las propiedades aquí expuestas están contempladas.

4.1 Construcciones perifrásticas

4.1.1 El verbo

La oración psíquica “es el elemento lingüístico con el cual el hablante decide dividir su discurso para tratar de comunicar aquello que le interesa”⁹ [Gili, 1991]. Esta definición captura en esencia el propósito general de una oración; sin embargo está lejos de ser formal; por lo que se prefiere la definición de *oración gramatical*, que consisten en: “todos los elementos, palabras, frases u oraciones enteras, que relacionen en modo inmediato o mediato con un verbo en forma personal”¹⁰; ésta define con mayor precisión al conjunto de oraciones[Gili, 1991].

Intuitivamente, el verbo es la categoría gramatical que define al conjunto de palabras que dentro de una oración determinan la acción que se ejecuta. Por ejemplo, en la oración *el niño alto entregó el juguete a Juan*, la palabra *entregó* indica que se realizó la acción de dar un objeto a alguien. El verbo se presenta en la oración en formas personales y no personales. Las formas personales son cuando se le asigna una de las seis personas gramaticales y el tiempo de ejecución, es decir, se conjuga; ejemplos de formas personal del verbos *entregar* son *entregó*, *entregan* y *entregará*. Las formas no personal son aquellas en las cuales el verbo no tiene ninguna de las personas gramaticales; por lo que no puede unirse a un sujeto y formar oraciones; estas formas son: el infinitivo, *entregar*; el gerundio, *entregando*; y el participio, *entregado*.

En la busca de precisión en la definición del contenido semántico de la acción que representa el verbo se recurre a modificaciones gramaticales sobre éste. El verbo a través de las desinencias adquiere propiedades de número, persona, modo y tiempo. Los complementos del verbo ayudan a definir su alcance sobre los objetos. Los adverbios ayudan a indicar modificaciones cualitativas y cuantitativas de la acción del verbo. Los tres

⁹ p. 18

¹⁰ p. 23

factores anteriores no agotan las modificaciones posibles, existiendo la posibilidad de modificar a un verbo a través de un verbo adicional, que agregue un matiz diferente a la acción representada por el verbo modificado; este fenómeno es conocido como conjugación perifrástica.

4.2 Conjugación perifrástica

Comparemos las siguientes oraciones en donde aparece el infinitivo escribir:

- (1) a. Tener que entregar un juguete
- b. Estar entregando un juguete
- c. Ir a entregar un juguete

Los tres ejemplos señalan la ejecución de una acción, *entregar un juguete*; pero los tres dan una idea diferente. En (1.a) al concepto de *entregar* se le añade las palabras: *tener que*, las cuales le agregan un matiz de obligación; mientras que en (1.b) *estar* agrega una significación de duración o continuidad del hecho; con *ir a* expresamos voluntad o disposición de ánimo para ejecutarlo. En estos ejemplos notamos que la acción central consiste en la representada por *entregar*, pero un segundo verbo da un pequeño giro a la forma en como se realiza esta acción. Al tipo de construcciones que se encuentra en (1) se le conoce como conjugación perifrástica. Tradicionalmente en la gramática del español a la conjugación perifrástica se le define como el empleo de un verbo auxiliar conjugado seguido de uno en forma no personal, tal como se muestra a continuación

- (2) a. Iba a correr *Infinitivo*
- b. Estaba comiendo *Gerundio*
- c. He caminado *Participio*

Definir de esta forma a la conjugación perifrástica trae consigo dos problemas. En primer lugar, existe una circularidad en la definición porque se definen a las construcciones perifrásticas en términos de los auxiliares, y a los auxiliares en términos de las construcciones. En segundo lugar, estas construcciones están formadas solamente por verbos auxiliares quedando fuera de esta forma el estudio a otros fenómenos muy cercanos a los verbos auxiliares. Observemos los siguientes ejemplos:

- (3) a. Quiero comer una manzana
- b. Oigo cantar a los canarios

- c. Comienzo a trabajar mañana
- d. Paso a contestar la carta

En (3) notamos la aparición de un verbo conjugado seguido de una forma no personal, pero como veremos en el transcurso de esta sección los verbos que aparecen en (3) no son auxiliares; sin embargo en estas oraciones existe un verbo que agrega un matiz diferente a la acción principal y que ayuda a definir los roles de los elementos de la oración, funciones que también desempeña un verbo auxiliar, como veremos más adelante.

Para este trabajo asumimos la posición de que conjugación perifrástica es aquella donde aparece un verbo conjugado seguido por un verbo en forma no personal y sus complementos, donde el verbo conjugado, además de modificar la acción principal, ayuda a definir los roles que desempeñan tanto el sujeto como sus complementos. La forma en como se resuelvan estos roles dependerán de la naturaleza del verbo conjugado.

4.3 Comportamiento de los verbos auxiliares

La definición de los verbos auxiliares se hace tradicionalmente bajo un análisis semántico. Gili Gaya considera a un verbo como verbo auxiliar cuando se ha desprendido de su significado original y adopta uno nuevo, observemos el siguiente ejemplo:

- (4) a. Debo cien pesos
- b. Debo de ir al banco

En (4.a) observamos lo que puede ser considerado el sentido original del verbo *deber*, el de aceptar una deuda; mientras que (4.b) se observa el significado de obligación a realizar una acción. Es evidente que ambos verbos *deber* no significan lo mismo, pero conjuntamente con la pérdida de significado notamos un cambio en el comportamiento del primer verbo *deber*; este deja de aceptar un complemento del tipo objeto directo para aceptar un complemento de categoría verbal en forma no personal, es decir el verbo a pasado de ser transitivo a ser un verbo auxiliar.

Como vemos la respuesta de qué es un verbo auxiliar no puede basarse sólo en un enfoque semántico donde se pone atención al significado del verbo. Se debe tomar en cuenta un enfoque sintáctico en donde se describen la propiedades que hacen posible la aparición de un verbo auxiliar en una oración y las relaciones existentes con los demás miembros de la oración.

4.3.1 Conjugación perifrástica de verbos auxiliares

Antes de avanzar en el comportamiento de las conjugaciones perifrásticas es importante definir los conceptos de *agente* y *sujeto* de una acción verbal. El agente es el encargado de realizar la acción, es la entidad quién ejecuta la acción, mientras que el sujeto es el objeto sintáctico que ayuda a identificar quién es el agente. En español el sujeto comúnmente aparece contenido en la conjugación del verbo, lo que hace que la palabra que distingue al sujeto no sea obligatoria. Por ejemplo, en (5.a) el sujeto está representado por la palabra *Juan*, el agente corresponde a la persona llamada *Juan*. En (5.b) no existe una palabra que defina al sujeto explícitamente como ocurre con el inciso anterior; sin embargo, *como* contiene al sujeto *yo*, primera persona del singular, el cual está expresado en la conjugación del verbo; este sujeto determina que el agente de la acción corresponde a la persona que dice la oración.

- (5) a. Juan come una manzana
b. Como una manzana

Los verbos que tradicionalmente han sido clasificados como auxiliares del español se caracterizan por los siguientes:

1. En las construcciones perifrásticas el verbo auxiliar define al agente del verbo en forma no personal. En (6) el verbo *voy* en su uso nos indica que en el futuro ocurrirá una acción, la de *estudiar español*. El sujeto sintáctico de *voy*, primera persona del singular *yo*, nos ayuda a resolver el agente de la acción de estudiar, y así quién diga la oración es quien estudiará.

(6) Voy a estudiar español

2. Se pueden construir conjugaciones perifrásticas con verbos auxiliares y unipersonales¹¹ como: llover, tronar, relampaguear, granizar, amanecer, anochecer, etcétera. En (7) *va* marca la existencia de un sujeto, tercera persona del singular, pero *llover* no necesita de un agente por lo que esta información no es utilizada.

(7) Va a llover (Verbo auxiliar y verbo unipersonal)

3. No se puede generar una oración interrogativa que pretenda preguntar por el complemento directo utilizando únicamente el verbo auxiliar. En (8.a), al preguntar

¹¹ Verbos que no tienen un agente

por complemento directo de *ir* en *voy a estudiar español* falla, ya que *ir* en esta oración no tiene un objeto directo; el que si lo tiene es *estudiar* como se ve (8.b). Esta clase de comportamiento es descrito en el texto *Los auxiliares españoles* [Fontanella, 1970].

- (8) a. *¿Qué voy a? español
b. ¿Qué voy a estudiar? Español

Antes de continuar con el siguiente inciso es necesario aclarar algunas características de las formas interrogativas. Una forma interrogativa es una transformación de oración declarativa donde la información por la que se pregunta generalmente se encuentra representada por otra palabra. En particular, (8) representa la forma interrogativa por el objeto directo de la oración (6); en (8) la palabra *qué* representa al objeto directo y va al inicio de la oración. El objeto directo de la oración, *estudiar*, como se aprecia no aparece en la oración interrogativa; pero, sí como la contestación de la pregunta.

4. Para preguntar por la acción a la cual modifica el verbo, no es posible realizar la pregunta utilizando únicamente al verbo auxiliar; se necesita de un verbo *comodín* en forma no personal. En (9.a) observamos que no es posible utilizar *ir* para preguntar por la acción que se modifica, sino que necesita de *hacer* para poder realizar esta función:

- (9) a. *¿Qué voy a?, estudiar
b. ¿Qué voy a hacer?, estudiar

En este ejemplo *qué* también reemplaza al objeto directo de la oración; pero en este caso esta palabra aunada al uso del verbo *hacer* permiten preguntar por la acción que representa la oración en forma declarativa (6), la cual corresponde a *estudiar*.

5. No se pueden pasivizar a un verbo auxiliar; lo que se hace es construir una perífrasis con voz pasiva. En (10.a) no fue posible construir la voz pasiva con el verbo *ir*; sin embargo, es posible construir la voz pasiva del verbo *estudiar* y que esta oración sea tomada como complemento verbal del ver *ir a* :

- (10) a. *Inglés es ido a estudiar por mi/
b. *?El inglés va a ser estudiado por mi

Estos cinco incisos conforman a lo que hemos llamado en este trabajo como pruebas de los verbos auxiliares. Si un verbo cumple con estos cinco incisos se trata de un verbo auxiliar del español. Es importante notar, que estas cinco pruebas se basan en el tipo de construcciones en las que aparece un verbo auxiliar y no en el significado que presenta el verbo auxiliar en la construcción, tal y como ocurre con las gramáticas tradicionales.

4.3.2 Propiedades sintácticas de los verbos auxiliares

Trabajos más recientes en los verbos auxiliares arrojan conclusiones similares a las enumeradas en la sección anterior. Fontanella y Luna definen un conjunto de construcciones en donde aparecen los verbos auxiliares de manera similar [Fontanella, 1970][Luna, 1991]. Sin embargo, aunque estas construcciones nos ayudan a identificar si un verbo auxiliar se comporta como tal no explican la naturaleza de los verbos auxiliares y no es posible construir un modelo computacional basado en el tipo de construcciones en las que participa. A continuación se presentan tres propiedades de los verbos auxiliares que definen la naturaleza de los verbos auxiliares:

1. Los verbos auxiliares tienen como complemento una frase verbal en forma no personal.
2. Los verbos auxiliares tienen como especificador, sujeto, al especificador del verbo en forma no personal.
3. Los verbos auxiliares no tienen un agente

Estas tres propiedades están basadas en términos sintácticos y es posible capturarlas en un modelo computacional, como se muestra en el siguiente capítulo. La primera propiedad se basa en la misma definición de construcción perifrástica; este comportamiento se puede apreciar en cada una de las pruebas de la sección anterior. La segunda propiedad se basa en la función que desempeña el verbo auxiliar, la cual es establecida por la primera prueba, en la cual un verbo auxiliar toma el sujeto del verbo en forma no personal. Por último, la tercera propiedad se encuentra en la línea de lo sintáctico y semántico; al no tener un agente el verbo no necesita de un especificador, sujeto, que le ayude a determinarlo, pero como el especificador es compartido con el verbo en forma no personal, por la propiedad dos, es posible que este último sea un verbo que tampoco necesite de un sujeto porque no tiene agente. Lo anterior explica porque un verbo auxiliar puede tomar como complemento a un verbo unipersonal, tal y como lo establece la prueba dos.

4.3.3 Construcciones anidadas

Existe un fenómeno que se presenta con los verbos auxiliares que proviene de la definición misma de conjugación perifrástica y que no es exclusivo de los verbos auxiliares. Hablamos de la utilización de varios verbos consecutivos en una sola conjugación perifrástica; este fenómeno es analizado en la literatura tradicional de manera muy superficial. En el presente trabajo llamamos al fenómeno de *construcciones anidadas*. Veamos el siguiente ejemplo:

- (11) a. Vamos a poder jugar todos los días
b. Acaba de empezar a llover
c. Quiere haber comido

En todas estas oraciones aparecen tres verbos. En (11.a) los dos primeros verbos son auxiliares; en (11.b) se trata de dos verbos que no lo son; mientras que en (11.c) se trata de uno que no es auxiliar y otro que sí. Esta sucesión de verbos presenta ciertas restricciones:

- El primer verbo de la sucesión es el único conjugado y los siguientes se encuentran en la forma no personal que solicite el verbo que le precede, como se ve en (11).
- Un verbo aparece solamente una vez en la oración:
(12) a. *Podría poder llover
b. *Estaba estando comiendo
c. *Pude haber podido brincar la cerca
- Algunos verbos auxiliares solamente aparecen en la posición inicial de la oración:
(13) a. Va a poder llover (ir como auxiliar)
b. *Puede ir a llover

En el análisis de los verbos auxiliares será muy importante definir su comportamiento en las oraciones anidadas.

4.4 Análisis de los verbos auxiliares

A continuación se presenta el análisis de los verbos que tradicionalmente han sido considerados como auxiliares; este análisis se basa en las pruebas definidas en la sección 4.2.2 de este capítulo. Los verbos analizados son los propuestos por Gili Gaya más el verbo

poder que resulta de interés para el proyecto DIME (capítulo primero). Este análisis se agrupa en tres partes dependiendo de la forma que toma el verbo no personal.

En el análisis de cada uno de los verbos se presentan los comportamientos que tienen los verbos partiendo de la intuición de lo que significan. Una vez identificado los distintos comportamientos se procede a verificar si alguno de ellos corresponde al de verbo auxiliar aplicando cada una de las pruebas planteadas. Para la etapa de pruebas se ha escogido a la frase *entregar un juguete a Juan* anteponiéndoles el verbo auxiliar, con el fin de observar los contrastes entre los distintos verbo. Adicionalmente, se hace un pequeño resumen de las condiciones necesarias para que el verbo aparezca en una construcción anidada.

4.4.1 Verbos auxiliares + infinitivo

Las construcciones auxiliares que involucran a un verbo en infinitivo son las más comunes en el español de México [Luna, 1991]. El infinitivo en una construcción general agrega a la oración un sentido progresivo dirigido hacia al el futuro. En este tipo de verbos, han sido agrupados un gran número de tipos de verbos como temporales, habituales, incoativos, etcétera. En el presente trabajo dejamos a un lado estas consideraciones para enfocarnos únicamente registrar los comportamientos de un verbo y verificar con las cinco pruebas si se trata de un verbo auxiliar.

4.4.1.1 Poder

El verbo *poder* presenta, como es característica de la mayoría de los verbos auxiliares, un doble significado, lo cual ha sido registrado por la mayoría de los diccionarios como vemos a continuación: “*poder* v tr (Modelo de conjugación 1 1b; precede siempre a otro verbo en infinitivo) **I 1** Tener algo o alguien la capacidad, la fuerza o el derecho de hacer algo: *poder mirar, poder trabajar, poder caminar, poder golpear, poder ordenar, poder salir, descansar* **2** Poder con alguien o con algo Ser uno capaz de comprenderlo, dominarlo o soportarlo: «*Yo puedo con las matemáticas*», «*Veré si puedo con este niño tan travieso*», «*No puedo con el pesado de tu marido*» ... **II 1** Tener algo o alguien la posibilidad, la autorización o el permiso de actuar de cierta manera: *poder entender, poder faltar a clases, poder resistir el cansancio, poder besar a la novia* **2** Haber la posibilidad de que algo suceda: «*Puede estallar una huelga*», «*Pudo haber caído un meteorito*», «*Puede llegar*

Susana el día menos esperado»»[Lara]. En esta definición se notan la existencia de tres verbos sintácticos *poder*, como se aprecia en los siguientes ejemplos:

- (14) a. Puedo con las matemáticas
b. Podía caminar
c. Podré volar

En (a) hablamos de un verbo que requiere una persona que pueda, agente, y algo que de lo que es capaz; se trata de un verbo intransitivo preposicional, es decir, necesita de un complemento, pero este no juega el papel de objeto directo, y además este complemento va ligado a una preposición, *con*. Mientras en (b) vemos que existe alguien que puede caminar y una acción que ese alguien es capaz de realizar, en (c) encontramos que se hace referencia a una posibilidad de una situación, en donde alguien realizará dicha acción. En las dos últimas oraciones es muy difícil saber cuando se quiere dar un matiz de posibilidad o capacidad. Este matiz dependerá mucho de lo factible que resulte la acción a realizar.

El verbo *poder* con el significado de posibilidad se comporta como auxiliar ya que cumple con las cinco pruebas. Observemos los siguientes ejemplos que corresponden al comportamiento registrado por los verbos auxiliares:

- (15) a. Pudo entregar el juguete a Juan
b. Puede llover
c. ¿Qué pudo entregar?
d. ¿Qué pudo hacer?
e. *El regalo es podido mostrar por él/
El regalo pudo ser entregado por el

Estas oraciones representan al comportamiento de verbo auxiliar en una construcción perifrástica. Como vemos el verbo *poder* cumple sin excepción con este comportamiento, por lo que podemos concluir que se trata de un verbo auxiliar cuando presenta el significado de posibilidad. Cuando *poder* presenta un significado de capacidad surge la necesidad de que exista un agente en él que represente a quién es capaz. Como podemos ver en (15.b) la lectura de capacidad queda fuera de alcance, ya que el verbo *llover* no tiene un agente por lo tanto el sentido de capacidad no se presenta.

4.4.1.2 *Ir a*

El verbo *ir* en su significado primario se refiere a que algo o alguien será desplazado de un lugar a otro, como se muestra a continuación.

(16) Voy a Guadalajara

Como auxiliar *ir* pierde este significado para marcar que una situación ocurrirá en el futuro; para lograr este significado necesita de la preposición *a*, así se tiene la construcción: *ir + a + infinitivo*. Una característica peculiar del verbo *ir* es que sólo se presenta en los tiempos de la conjugación de presente e imperfecto de indicativo y subjuntivo. Obsérvese los siguientes ejemplos:

- | | | |
|------|----------------------------------|----------------------------|
| (17) | a. Voy a estudiar inglés | (presente de indicativo) |
| | b. Iba a estudiar inglés | (imperfecto de indicativo) |
| | c. Ojalá vaya a estudiar inglés | (presente subjuntivo) |
| | d. Ojalá fuera a estudiar inglés | (imperfecto subjuntivo) |

En las oraciones anteriores podemos apreciar que el verbo auxiliar añade la intención de realizar una acción con respecto a un tiempo o posibilidad. Sin embargo, en otros tiempos de la conjugación el verbo *ir* sólo puede tener su sentido común de desplazamiento, como se ve en los siguientes ejemplos:

- | | | |
|------|----------------------------------|-----------------------------------|
| (18) | a. Iré a estudiar inglés | (futuro absoluto de indicativo) |
| | b. Fui a estudiar inglés | (pasado absoluto de indicativo) |
| | c. Iría a estudiar inglés | (futuro hipotético de indicativo) |
| | d. Ojalá fuere a estudiar inglés | (futuro absoluto de indicativo) |

Podemos concluir que el verbo *ir* es auxiliar con los tiempos señalados del indicativo y subjuntivo al cumplir con las cinco pruebas como se indica a continuación. El verbo *ir* cumple con el comportamiento de verbo auxiliar como observamos en los siguiente ejemplo:

- (16) a. Va a entregar el juguete a Juan
b. Va a llover
c. ¿Qué vas a entregar?
d. ¿Qué vas a hacer?
e. *El juguete es ido entregar por él/
El juguete va a ser entregado por él

El verbo *ir* también presenta una restricción con respecto a la posición en que aparece en las construcciones anidadas; sólo puede aparecer bajo su función de auxiliar cuando ocupa la primera posición en la cadena de verbos auxiliares que componen la construcción anidada.

- (17) a. Iba a poder entregar el regalo
b. Puedo ir a entregar el regalo

En la oración (a) se hace referencia al significado del verbo *ir* a como auxiliar es decir como marca de una acción futura, mientras que en (b) el único significado que se puede leer es el de desplazamiento.

4.4.1.3 *Pasar a*

Del verbo *pasar a* como verbo auxiliar se tiene los siguientes ejemplos, tomados de Gili Gaya [Gili, 1991]:

- (18) a. En mi poder su carta del 12 de los corrientes, que paso a contestar
b. Con la destrucción de Cartago, los romanos pasaban a ser dueños del Mediterráneo
c. Los atacantes han pasado a dominar la ciudad
d. Con tan buenos negocios habrán pasado a ser ricos

En los ejemplos anteriores se observa que el verbo *pasar* ha perdido su sentido de transcurrir o movimiento para asumir una donde sirve de indicador que una acción comienza como en las oraciones (a) y (b) o de que la acción ha llegado a su fin. El sentido de la marca, de inicio o fin de acción, se debe al tiempo en que sea empleado así para tiempos perfectos *pasar a* significa fin, y en tiempos imperfectos inicio. Esta pérdida de sentido no significa que se trate de un verbo auxiliar y como prueba el siguiente ejemplo, que consiste en la segunda de las pruebas:

- (19) *Paso a llover

Como vemos este ejemplo se trata de una oración gramaticalmente incorrecta, esto se debe a que el verbo *llover*, como habíamos visto con anterioridad, es unipersonal y no necesita de un agente; por el contrario, el verbo *pasar* si requiere de uno, alguien que desde el punto de vista semántico sea el encargado de provocar el inicio o fin la acción. El verbo *pasar* es similar al verbo *poder* con el significado de capacidad, ya que ambos necesitan de un agente.

4.4.2 *Echar a*

El verbo *echar* en su sentido original hace referencia al significado de lanzar, arrojar; bajo este significado *echar* se comportan como un verbo transitivo donde alguien pone en movimiento algo, como en las siguientes oraciones:

- (20) a. Echa la pelota
b. Echaba rayos y centellas

A continuación observemos los siguientes ejemplos, que pretenden reproducir el comportamiento *echar* como de auxiliar; en esta ocasión se ha utilizado el verbo *correr* por la restricción semántica que presenta el verbo:

- (21) a. Echó a correr
b. *Echó a llover
c. ¿Qué echó a andar?
d. *¿Qué echó a hacer?
e. La máquina será echada a andar por mi/
*La máquina echará a ser andada por mi

En estos ejemplos queda de manifiesto tres problemas. Primero, el verbo no acepta verbos unipersonales, aunque en la literatura especializada se encuentra reportada esta oración [Luna, 1991]. Segundo, no exista un verbo comodín de movimiento que nos permita preguntar por la acción (21.d). Por último, el problema que el verbo *echar* puede pasivizarse (21.c) y no puede tomar una oración en voz pasiva. El que este verbo acepte la pasivización nos muestra una mayor dependencia entre los dos verbos participantes que la existente entre un verbo auxiliar y su verbo en forma no personal. Concluimos entonces que *echar a* no es verbo auxiliar. Además notamos en la voz activa de (21.e), *echo andar la máquina*, que *la máquina* es la que va a *andar* (funcionar) y no el sujeto de *echo*, primera persona del singular, por lo que no le ayuda a resolver el agente en *andar*. La versión pronominal, *echarse a*, presenta un comportamiento similar al primero.

4.4.2.1 *Venir a*

En su sentido original *venir* posee un significado contrario al del verbo *ir*; tiene el significado de que algo o alguien parte de un punto con la intención de llegar a donde el hablante se encuentra.

(22) Viene de Guadalajara

Hace referencia a que alguien parte de la ciudad de Guadalajara con destino a donde se encuentra ubicado el hablante; bajo este significado se comporta como un verbo intransitivo que solicita una frase preposicional como complemento, pero que no juega el papel de objeto directo. Un segundo comportamiento adicional del verbo *venir* es cuando aún indica movimiento, como en su sentido original, pero va acompañado de un verbo en forma infinitiva que indica el propósito del desplazamiento, veamos los siguientes ejemplos:

- (23) a. Vino a dejar la carta
b. Vendrá a arreglar la tubería

Estas dos oraciones hacen clara la idea de que el desplazamiento tenía un objetivo; pero el que conserve el sentido de desplazamiento hace que necesite de alguien que lo efectúe, en este caso un agente. El verbo *venir* presenta un comportamiento de auxiliar cuando su significado ha sido identificado como dependiente del tiempo y se la ha marcado como indicador de una acción terminada o como indicador de aproximación, como en los ejemplos siguientes:

- (24) a. Vino a coincidir contigo
b. Viene a costar veinte pesos

En (25) podemos notar el comportamiento del verbo como auxiliar para respaldarlo, observemos los siguiente ejemplos:

- (25) a. Hasta hoy, vino a decir la verdad
b. Hasta hoy, viene a llover
c. ¿Qué vino a entregar?
d. ¿Qué vino a hacer?
e. *El juguete fue venido a entregar por él/
Hasta hoy, el juguete vino a ser entregado por él

En estas oraciones se aprecia dos restricciones muy peculiar del verbo: la aparición de un complemento de tiempo, *hasta hoy* y la necesidad de un tipo de semántica distinta al movimiento para que el verbo se entienda como fin de la acción; es por esto que no se utilizó la oración *entregó el juguete a Juan*. Solamente cuando estas restricciones aparecen es que podemos hablar de un comportamiento como auxiliar; sin embargo, para el modelo

que se presenta en este trabajo estas dos consideraciones quedan fuera de alcance por tratarse de una restricción con un sentido más semántico y pragmático que sintáctico.

4.4.2.2 *Volver a*

El verbo *volver a* es uno de los verbos auxiliares menos conflictivos. Bajo su sentido original significa regreso o retorno como en la siguiente oración:

(26) Vuelvo a casa

Como auxiliar su significado cambia a reiteración o repetición de una acción:

(27) Vuelvo a regresar a casa

Las siguientes oraciones representan el comportamiento del verbo *volver a* como auxiliar:

- (28) a. Volvió a entregar el regalo a Juan
b. Vuelve a llover
c. ¿Qué volvió a entregar?
d. ¿Qué volvió a hacer?
e. *El regalo fue vuelto a entregar por él /
El regalo volvió a ser entregado por él

Las construcciones anidadas en las que aparece el verbo no presentan una restricción; esto se puede apreciar en las siguientes oraciones:

- (29) a. Pudo volver a entregar el regalo
b. Vuelve a poder entregar el regalo

4.4.2.3 *Haber de*

Haber de, haber que, tener que, tener de y deber de representan a los verbos auxiliares que agregan obligatoriedad a la acción que representa el verbo en infinitivo. A *haber* sólo se le asigna un significado de existencia bajo la conjugación irregular *hay* y con su forma no personal de participio (*habido*), el cual es considerado como el significado original del verbo. El verbo *haber de* tiene un significado que agrega obligatoriedad a la acción, esto se puede verificar en las siguientes oraciones que presentan su comportamiento como auxiliar:

- (30) a. Ha de entregar el regalo a Juan
b. Ha de llover
c. ¿Qué ha de entregar?

- d. ¿Qué ha de hacer?
- e. *El regalo es habido de entregar por él /

El regalo habrá de ser entregado por él

Con respecto a las oraciones anidadas el verbo *haber de* presenta la restricción de que solo poder ir en la primer posición, como se aprecia en la siguientes oraciones:

- (31) a. Habrá de poder entregar el regalo
 b. *Puedo haber de encontrar la solución

4.4.3 Haber que

Haber que forma parte de los auxiliares de obligación (sección anterior). *Haber que* presenta la restricción que sólo es válida en la conjugación de la tercera persona. Veamos las siguientes oraciones que reproducen el comportamiento de verbo auxiliar:

- (32) a. Hubo que entregar el regalo a Juan
 b. *Habrá que llover
 c. ¿Qué hubo que entregar?
 d. ¿Qué hubo que hacer?
 e. *El regalo fue habido que entregar por él/
 *El regalo hubo que ser entregado por él

Este comportamiento extraño que presenta el verbo *haber que* se debe a que sólo se conjuga bajo la tercera persona de singular y tiene un aspecto impersonal, es decir aparentemente no tiene agente; pero, si tiene un aspecto impersonal ¿por qué no acepta al verbo unipersonal *llover*?. *Haber que* marca la existencia de una obligación en donde alguien tiene que realizar una acción; quién realice la acción no es lo importante, pero sí que alguien la lleve a cabo; es decir, necesita marcar que la acción debe ser ejecutada por un agente. Es por esto que no puede tomar a verbos unipersonales. El hecho que no acepte la construcción de voz pasiva se debe a que *haber que* no acepta ningún sujeto sintáctico y las voz pasiva solicita la existencia de este (el objeto directo de la voz activa). En conclusión podemos decir que *haber que* no es un verbo auxiliar.

4.4.3.1 *Tener que*

Tener que es el tercer verbo que agrega obligatoriedad a la acción que representa el verbo en forma no personal. *Tener* en su sentido original se refiere al significado de posesión, como en la oración:

(33) Tiene un coche

Bajo el comportamiento de auxiliar el verbo asume el significado de obligación a que se lleve a cabo la acción; como se ve en las siguientes oraciones que representan su comportamiento como verbo auxiliar:

- (34) a. Tienen que entregar el juguete a Juan
b. Tiene que llover
c. ¿Qué tienen que entregar?
d. ¿Qué tienen que hacer ?
e. *El juguete es tenido que entregar por ellos/

El juguete tiene que ser entregado por ellos

El verbo *tener que* no tienen ninguna restricción para aparecer en las construcciones anidadas como podemos ver en las siguientes oraciones:

- (35) a. Vuelven a tener que entregar el juguete a Juan
b. Tienen que volver a entregar el juguete a Juan

Tener de ha sido reportado como verbo auxiliar; mas en la actualidad es un verbo que no aparece en la lengua común y corriente del español. Se tiene únicamente referencia a este verbo en literatura clásica por lo que en el presente trabajo no es analizado.

4.4.3.2 *Deber de*

Como vimos en el ejemplo (4), el verbo *deber* presenta dos significados; uno se refiere a la aceptación de una deuda y otro a la obligación de realizar una acción. Con el significado de obligación *deber* presenta su comportamiento de verbo auxiliar, como vemos en las siguientes oraciones:

- (36) a. Debe de entregar el juguete a Juan
b. Debe de llover
c. ¿Qué debe de entregar?
d. ¿Qué debe que hacer ?

e. *El regalo es debido de entregar por él/

El regalo debe de ser entregado por él

Como vemos en estas oraciones podemos ver que el comportamiento del verbo *deber* es de verbo auxiliar. En las construcciones anidadas *deber* de sólo aparece en la primera posición, como lo muestran las siguientes oraciones:

(37) a. Debe de poder entregar el juguete a Juan

b. *Puede deber de entregar el juguete a Juan

4.4.4 Verbos auxiliares + gerundio

En gerundio da a las frases verbales en que figura un sentido general de acción durativa, cuyos matices dependen de la naturaleza del verbo que le acompaña. Un punto de confusión común, entre de los verbos auxiliares seguidos de un gerundio, es cuando el gerundio desempeña un papel de adverbio, en donde en lugar de que la acción que representa el gerundio este siendo matizada por un verbo auxiliar, es el gerundio el que modifica la acción del verbo que le acompaña; observemos la siguiente oración:

(38) Corro brincando los obstáculos

En esta oración *correr* de ninguna manera se encuentra modificando la forma en que brinco los obstáculos; al contrario *brincar* indica la forma en como se *corre*, es decir tiene un papel de adverbio.

Sobrepasado este punto de confusión en donde no todos los verbos seguidos de un gerundio son verbos auxiliares, presentan un comportamiento de auxiliar los verbos *estar*, *ir*, *venir*, *seguir* y *andar*.

4.4.4.1 *Estar*

El verbo *estar* en su sentido original presenta un sentido de localidad que permite asignar a un sujeto su ubicación. Además *estar* permite establecer una relación entre un sujeto y su estado, observemos las siguientes oraciones:

(39) a. Él está enojado

b. Él está en su casa

En (a) se establece cual es el estado de una tercera persona, mientras que (b) habla sobre la ubicación de esta persona. Existe una controversia sobre la significación que agrega el verbo *estar* cuando actúa como auxiliar. Gili Gaya considera a *estar* como auxiliar cuando

indica duración en la acción que representa el verbo en forma de gerundio, pero esto lo lleva a descartar el significado de auxiliar cuando el gerundio proviene de un verbo de estado como *vivir* o *dormir*[Gili, 1997]. Pero en este trabajo coincidimos con la opinión de José Moreno: “A mí me parece que no hay sino diferencia de grado entre ‘canto’ y ‘estoy cantando’. Si se usa la perífrasis se hace referencia enfática al carácter imperfectivo de la acción, carácter ya señalado en el presente simple”. Por lo anterior consideramos que estar solo es una marca de énfasis sobre la imperfectividad¹² de la acción, es decir sobre el hecho que la acción no ha acabado. Observemos las siguientes oraciones:

- (40) a. Está entregando el juguete a Juan
b. Está lloviendo
c. ¿Qué está entregando?
d. ¿Qué está haciendo?
e. *El regalo es estado entregando por él /
El regalo está siendo entregado por él

Podemos notar que *estar* cumple con el comportamiento de verbo auxiliar; además, el verbo *estar* no impone ninguna restricción en las oraciones anidadas:

- (41) a. Está pudiendo mostrar el catálogo
b. Puede estar mostrando el catálogo

4.4.4.2 *Ir*

A diferencia del significado de futuro que agrega el verbo auxiliar *ir a* (sección 4.4.1.2), *ir* acompañado de un gerundio hace referencia a un aspecto progresivo de la acción, es decir a una acción que se ejecuta paso a paso; de hecho es muy común aparezcan con este verbo auxiliar adverbios o locuciones adverbiales como *poco a poco*, *cada vez* y *lentamente*. Con este significado el verbo no puede interactuar con verbos unipersonales lo que nos indica que necesita a un agente, esto se debe a que *ir* con significado progresivo atribuye esa lentitud a alguien, observemos el siguiente ejemplo:

- (42) a. Juan va regalando su dinero

Es claro que Juan es el responsable que la acción de *regalar* se efectúe; pero además es él el responsable de que esto suceda poco a poco, es decir *ir* tiene un agente.

¹² Que la acción no termina y continua

Existe un significado para *ir* con el cual presenta el comportamiento de auxiliar, es cuando denota aproximación. Veamos las siguientes oraciones que muestran las propiedades de las construcciones perifrásticas¹³:

- (43) a. Juan va entregando las cartas a las 10:00
b. Va lloviendo por ahí de mayo
c. ¿Qué va entregado a las 10:00?
d. ¿Qué va haciendo a las 10:00?
e. *Las cartas son idas entregando por él a las 10:00/

Las cartas van siendo entregadas por él a las 10:00

El comportamiento del verbo *ir* con las construcciones anidadas es el siguiente (donde queda de manifiesto que sólo puede ir en la primera posición de la construcción anidada)

- (44) a. *Puede ir entregando las cartas a las 10:00
b. Irá pudiendo entregar las cartas a las 10:00

4.4.4.3 Venir

Con el verbo *venir* ocurre un fenómeno similar que con *ir* (sección anterior). *Venir* tiene un significado de progresividad de la acción la cual se le atribuye al agente por lo que no es posible construir con este significado el comportamiento de auxiliar. Pero al igual que ocurrió con el verbo *ir* se comporta como auxiliar bajo un significado de proximidad:

- (45) a. Juan viene entregando las cartas a las 10:00
b. Viene lloviendo por mayo
c. ¿qué viene entregado a las 10:00?
d. ¿qué viene haciendo a las 10:00?
e. *Las cartas son venidas entregando por él a las 10:00/

Las cartas vienen siendo entregadas por él a las 10:00

Las oraciones anteriores nos muestran el comportamiento verbo *venir* como auxiliar. Con las construcciones anidadas los verbos *ir* y *venir* se comportan de la siguiente manera:

- (46) a. Pudo venir entregando las cartas a las 10:00
b. Vino pudiendo entregar las cartas a las 10:00
c. *Vino estando entregando las cartas a las 10:00

¹³ Se utiliza la frase *entregando las cartas a las diez* para que la oración tenga más sentido

- d. *Estuvo viniendo entregando las cartas a las 10:00

Como vemos la restricción presente en estos verbos es que no acepta a ninguno de los verbos auxiliares que soliciten un verbo en gerundio; esto se debe a que dos gerundios no pueden ir seguidos.

4.4.4.4 Seguir

Seguir en su sentido original hace referencia a la acción de ir tras alguien o algo. Como auxiliar *seguir* obtiene el significado de continuar con una acción; este comportamiento lo podemos apreciar en las siguientes oraciones:

- (47) a. Sigue entregando el regalo a Juan
b. Sigue lloviendo
c. ¿Qué sigue entregando?
e. ¿Qué sigue haciendo?
f. *El regalo es siguiendo entregado por él/
El regalo sigue siendo entregado por él /

Al igual que ocurrió con *ir* y *venir* no es posible que en una construcción anidada dos gerundios queden cerca como se ve en la siguientes oraciones:

- (48) a. *Sigue viniendo entregando el regalo a Juan
b. Sigue pudiendo entregar el regalo a Juan

4.4.4.5 Andar

Andar en su significado original se refiere a una acción de desplazamiento. Se encuentra muy cercano al significado de caminar (49.a) . Pero, también ha tomado el significado de encontrarse o estar, tomando de esta forma tintes de verbo atributivo (49.b)

- (49) a. El bebé ya anda sólo/El bebé ya camina sólo
b. Andaba por ahí/Estaba por ahí/Me encontraba por ahí

Es precisamente bajo el significado de la oración del inciso (b) que el verbo *andar* se comporta como verbo auxiliar, observemos los siguientes ejemplos:

- (50) a. Anda entregando el juguete a Juan
b. Anda lloviendo
c. ¿Qué anda entregando?
d. ¿Qué andan haciendo?

- f. *Los juguetes son andados repartidos por ustedes
/Los juguetes andan siendo repartidos por ustedes

Como *andar* hace referencia a un significado de estado, no es posible que aparezcan simultáneamente *estar* y *andar* en una construcción anidada, como se observa en (51.a). Además, sigue siendo imposible que aparezcan dos verbos en gerundio seguidos, como muestra (51.b). Con respecto a la posición *andar* no impone ninguna restricción, como se observa en (c) y (d):

- (51) a. *Anda pudiendo estar herido
b. *Viene andando entregando el regalo
c. Anda teniendo que entregar el regalo
d. Puede andar entregando el regalo

4.4.5 Verbo auxiliar + participio

El participio se comporta como un adjetivo además de su función verbal. Éste adjetivo nos indica que al sujeto que modifica se la efectuado una acción. Es importante notar que el participio agrega un sentido de perfectividad¹⁴. Observemos las siguientes oraciones:

- (52) a. Las casas construidas
b. El caballo corrido

En (a) se hace referencia a un tipo de casas las cuales en el pasado se construyeron y en (b) a un caballo que el pasado corrió. Un punto importante es que el comportamiento de adjetivo involucra que el participio tome concordancia, como habíamos visto las formas no personales del verbo no tienen concordancia; un participio bajo el comportamiento de adjetivo se ve obligado a tomar la concordancia, este punto nos ayudará a distinguir cuando un verbo está actuando como adjetivo y no como complemento de un verbo auxiliar.

Otro punto que nos ayuda a distinguir sobre la naturaleza del participio es la posición que puede tomar en una oración. La definición de conjugación perifrástica establece que primero va el verbo auxiliar y a continuación el verbo en forma no personal; mientras que el orden que presentan los adjetivos es muy variado. Observemos los siguientes ejemplos:

- (53) a. El caballo corre atado por la boca

¹⁴ Acción terminada o que ha llegado a su fin

b. El caballo corre por la boca atado

Mientras que el inciso (a) indica que el orden del verbo *correr* es similar al de un verbo auxiliar; (b) nos demuestra que en realidad se trata de un verbo transitivo, donde *atado* juega un papel de adjetivo que describe la forma en que se encuentra el caballo. Estas dos características son muy común en los verbos copulativos, es decir, aquellos que tienen la función gramatical de unir una frase nominal con una frase adjetiva, con el fin de establecer la relación entre el sujeto que representa la frase nominal y la cualidad que describe la frase adjetiva. Ejemplo de estos verbos son: *ser* y *estar*. Observemos los siguientes ejemplos:

- (54) a. Juan es alto
b. Las paletas están frías
c. La tortuga es cambiada de lugar

En estas oraciones el verbo se encarga de atribuir una cualidad al sujeto. Como vemos existe concordancia entre adjetivo y sujeto, punto que distingue a los verbos copulativos de los auxiliares.

4.4.5.1 *Haber*

Haber es por excelencia el verbo auxiliar; es tanta la gramaticalización que existe en el verbo que hoy en día este verbo auxiliar es considerado como parte fundamental de los tiempos verbales. El significado que toma el verbo *haber* es el de indicar una acción terminada, situación que es relativa al tiempo al en que se conjuga. Observemos las siguientes oraciones que corresponden al comportamiento de verbo auxiliar:

- (55) a. Ha entregado el juguete a Juan
b. Ha llovido
c. ¿Qué ha entregado?
d. ¿Qué ha hecho?
e. *El juguete fue haber entregada por él/
El juguete ha sido entregado por él

Como vemos cumple con todas las propiedades de la conjugación perifrástica lo que nos garantiza que cumpla con las propiedades de verbo auxiliar. Su comportamiento en las estructuras anidadas no presentan ninguna restricción:

- (56) a. Podría haber eliminado la pila
b. Habría podido eliminar la pila

4.4.5.2 *Llevar, tener, traer y dejar*

Llevar, tener, traer y dejar son recogidos por Gili Gaya como verbos auxiliares; pero como vemos a continuación presentan concordancia y orden libre con su complemento por lo que no pueden ser considerados de esta forma:

- (57) a. Llevo andados tres kilómetros
b. Llevo andadas tres leguas
c. Llevo tres leguas andadas
- (58) a. Tengo estudiada esta cuestión
b. Tengo estudiado este problema
c. Tengo esta cuestión estudiada
- (59) a. Traía curados muchos toros
b. Traía curadas muchas vacas
c. Traía muchas vacas curadas
- (60) a. Dejaron dicho su plan
b. Dejaron dicha su idea
c. Dejaron su plan dicho

4.4.5.3 *Ser y estar*

Ser y estar seguidos de un participio son otros verbos considerados por muchas gramáticas como verbos auxiliares y que como veremos a continuación en realidad no lo son[Gili, 1991][Luna, 1991]. Como se ve en las siguientes oraciones, el verbo *ser* rompe con el orden que exigen los verbos auxiliares y además exige concordancia.

- (61) a. La noticia es divulgada
b. Divulgado es el evento por mi
c. Divulgada es la noticia

El verbo *estar* presenta un comportamiento similar al de ser:

- (62) a. Juan estaba convencido de su mala suerte
b. María estaba convencida de su mala suerte
c. María estaba de su mala suerte convencida

Los ejemplos en (61) y (62) son claros ejemplos de que los verbos *ser* y *estar* no presentan un comportamiento de auxiliar desde el punto de vista que plantea este trabajo.

Con estos dos últimos verbos se cierra el análisis de los verbos auxiliares. Como resultado del análisis hecho en este capítulo hemos logrado los siguientes puntos:

1. Se cuenta con un conjunto de pruebas sintácticas que permiten identificar a un verbo auxiliar del español.
2. A través de las pruebas sintácticas se han clasificado de un total de veintidós verbos de los cuales solamente trece son verbos auxiliares.
3. Se han identificado tres propiedades sintácticas que poseen los verbos auxiliares.

En el siguiente capítulo se presenta el modelo computacional de los verbos auxiliares del español. En el siguiente capítulo se describe el modelo para uno de los verbos auxiliares, en donde las tres propiedades identificadas juegan un papel muy importante. Adicionalmente, en el siguiente capítulo se presenta el modelo computacional para el sentido original de los verbos auxiliares y para aquellos verbos que no cumplieron con las pruebas sintácticas.

Capítulo 5: Modelo en HPSG

En este capítulo se presenta el modelo computacional del sistema de verbos auxiliares del español bajo el formalismo Head-driven Phrase Structure Grammar (HPSG)[Sag y Wasow, 1991]. En primer lugar, se presentan la caracterización que se hace de los diferentes tipos de verbos en HPSG. Se comienza con la descripción de los tipos transitivos e intransitivos para seguir con la de los verbos con complemento de tipo frase. Estos últimos se basan en la distinción de verbos de *control* y *raising* hecha por Noam Chomsky [Sag y Wasow, 1999]. En particular, el tipo *subject raising verb (srv-lxm)* cumple las propiedades que hemos identificado en los verbos auxiliares del español; por lo que permite definir el modelo del comportamiento de los verbos auxiliares en HPSG. A continuación se presenta un esquema general de los tipos de verbos del español que cuenta con una caracterización en HPSG.

El concepto de oración gramatical se basa en la presencia del verbo; una gramática que modele al lenguaje deberá considerar los diferentes comportamientos que presentan los verbos. El formalismo HPSG en su jerarquía de tipos propone una clasificación de los diferentes tipos de verbos basada en las clases de complementos que tienen. Para los verbos que requieren que su complemento o complementos jueguen un rol de objeto directo e indirecto existe el tipo *tr-vr* que engloba a los llamados verbos transitivos. Por ejemplo, un verbo transitivo es *entregar* en la oración *el niño alto entregó el juguete a Juan*, porque necesita de un objeto directo e indirecto. Los verbos transitivos son muy importantes ya que son muy numerosos y porque poseen propiedades características, como el hecho de que son los únicos que aparecen en construcciones de voz pasiva. Pero, si el verbo no necesita de un complemento o el rol del complemento no es de objeto directo, HPSG propone al tipo *iv-lxm* que representa a los verbos intransitivos. Un ejemplo de verbo intransitivo es *ir* en la oración *el niño alto va a su casa*; en esta oración el complemento no representa es un objeto directo, es decir, no es lo ido. Generalmente los verbos auxiliares toman como complemento a una frase verbal cuya cabeza de frase es un verbo transitivo o intransitivo en una de sus formas no personales (infinitivo, gerundio y participio).

El formalismo HPSG define dos tipos de verbos que requieren una frase verbal como complemento; dichos tipos son *subject raising lexeme (srv-lxm)* y *subject control lexeme*

(*scv-lxm*). Mientras que el tipo *srv-lxm* captura aquellos verbos cuyo complemento es una frase verbal con la cual comparte su sujeto, el tipo *scv-lxm* captura aquellos verbos cuyo complementos es una frase verbal con la cual comparte el agente. Como se puede observar, los verbos de control tienen un agente, por lo que de acuerdo a las propiedades de los verbos auxiliares del español, este tipo de comportamiento no corresponde al de verbo auxiliar del español; sobretodo por la prueba en donde un verbo unipersonal puede seguir a un verbo auxiliar. Por otro lado, el tipo *srv-lxm* cumple con las propiedades de los verbos auxiliares: tiene como complemento a una frase, comparte el sujeto con el verbo en forma no personal y no tiene un atributo que se refiera al agente. En suma, los verbos auxiliares del español corresponden al tipo *srv-lxm* de HPSG. A diferencia de otros idiomas donde los verbos auxiliares se encuentran representados por un subtipo con más restricciones de la de *srv-lxm*, debido a que presentan un conjunto de características adicionales, por ejemplo en el inglés las propiedades NICE (negación, interrogación, contracción y ellipsis)[Sag y Wasow, 1991]¹⁵.

El cambio de significado que es señalado en las gramáticas tradicionales del español se ve reflejado en el sistema de verbos auxiliares del español por medio de dos entradas léxicas. Por ejemplo, el verbo *volver* significa originalmente el regreso a un lugar, como en *el niño alto volvió a casa*; pero además indica, como verbo auxiliar, que una acción se ejecutó con anterioridad como en la oración *el niño alto volvió a entregar el juguete a Juan*. Cada una de las acepciones se representa con una entrada léxica distinta. En particular, para capturar este doble significado del verbo *volver* debe existir una entrada léxica del tipo *iv-lxm*, que representa el significado original, y del tipo *srv-lxm* que representa el comportamiento de verbo auxiliar. Adicionalmente, la oración con el significado de verbo auxiliar antes mencionada es ambigua, debido a que la frase *a entregar el juguete a Juan* puede interpretarse como marca de propósito, es decir, que el niño alto regreso con el propósito de entregar el juguete a Juan; en este caso *volver* se comporta como verbo del tipo *scv-lxm*, ya que, quién entregó y regreso a un sitio son la misma persona, en consecuencia el mismo agente.

¹⁵ p. 295

Los verbos atributivos son verbos que se confunden frecuentemente como verbos auxiliares. Aunque presentan semejanzas con los verbos auxiliares, estos son más restrictivos con el tipo de oraciones en las que toman parte. Un verbo atributo establece una relación de atribución entre un adjetivo y un sustantivo. En la oración *El niño está alto*, se le atribuye a *el niño* la cualidad de *alto*. La semántica de esta oración con la frase de *el niño alto* son similares; la diferencia es sintáctica: mientras que la frase no es una oración completa por la falta del verbo, la oración atributiva sí lo es. La confusión de los verbos atributivos y los auxiliares surge cuando un verbo se transforma en un adjetivo y se forman oraciones donde un verbo atributivo va seguido de un verbo en su forma de participio. Por ejemplo, la oración *el niño alto está dedicado a sus estudios* podría tomarse como ejemplo del comportamiento del verbo *estar* como auxiliar; sin embargo, en esta oración se presenta una consideración que no contemplan los verbos auxiliares, ya que existe concordancia obligada entre sujeto y verbo en participio, género singular y número masculino. A pesar de las diferencias entre verbos auxiliares y atributivos, HPSG representa el comportamiento de los verbos atributivos como un subtipo del tipo *srv-lxm_j*, sin que esto signifique que los verbos atributivos son verbos auxiliares.

En las siguientes secciones se describen los conceptos aquí presentados con mayor detalle, utilizando los elementos que integran a HPSG. En la sección 5.1 se introducen los tipos de verbos transitivos e intransitivos de HPSG. Adicionalmente, en esta sección se incluyen a los verbos de complemento de frase y la discusión que lleva a concluir que los verbos auxiliares del español son del tipo *srv-lxm*. En la sección 5.2 se presenta la definición de los tipos que ayudarán a capturar el comportamiento de los verbos auxiliares. En la sección 5.3 se presenta un resumen de las entradas léxicas para los verbos auxiliares y los no auxiliares, así como de sus comportamientos adicionales.

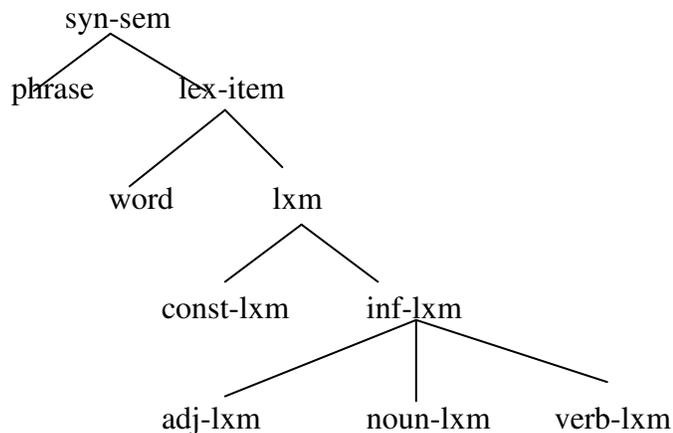
La definición de tipos y de entradas léxicas de HPSG, que se presentan en este capítulo, componen el modelo del sistema de verbos auxiliares del español propuesto en esta tesis. Este modelo será utilizado en el capítulo seis para realizar su implementación computacional, en la cual las definiciones hechas son de suma importancia. Un aspecto muy importante de este capítulo es la definición de los verbos auxiliares en función de un elemento sintáctico como lo es el tipo *srv-lxm*.

5.1 La jerarquía de verbos en HPSG

5.1.1 El tipo *verb-lxm*

El tipo encargado de reproducir las propiedades de los verbos en HPSG es *verb-lxm*. La posición que ocupa éste en la jerarquía de tipos se muestra en la siguiente figura:

(1)



La jerarquía de tipos impone sobre el tipo *verb-lxm* las siguientes restricciones:

- Posee los atributos SYN y SEM en su estructura atributo valor, por ser hijo del tipo *syn-sem*. Estos atributos son utilizados para representar la información sintáctica y semántica.
- Por ser un tipo hijo del *lxm-item* tiene como atributo a la lista de argumentos ARG-ST. Este atributo le permitirá solicitar su especificador (SPR) y complementos (COMPS) después de la satisfacción léxica.
- La descripción del tipo podrá definir su modo semántico, debido a que por *default* no tiene modo. Lo anterior por la restricción que impone el tipo *lxm*.
- El tipo *verb-lxm* podrá ser tomado por reglas léxicas del tipo inflexión, por ser hijo del tipo *infl-lxm*. Esto nos permitirá definir reglas de inflexión para las conjugaciones del verbo.

Las restricciones heredadas no hacen una distinción entre *verb-lxm* y los tipos *adj-lxm* y *noun-lxm* que también son hijos del tipo *infl-lxm* y por lo tanto heredaron las mismas restricciones. Pero, debemos recordar que un tipo está formado por un nombre (identificador) en este caso *verb-lxm*, un nodo padre del cual hereda restricciones (*infl-lxm*)

y la descripción de una restricción. A continuación se muestra la restricción que tiene el tipo *verb-lxm*:

(2)

$$verb - lxm \left[\begin{array}{l} SYN \left[\begin{array}{ll} HEAD & verb \\ SPR & \langle [] \rangle \end{array} \right] \\ ARG - ST \ / \langle NP, \dots \rangle \\ SEM \ [MODE \ prop] \end{array} \right]$$

Esta restricción captura las propiedades que caracterizan a un verbo: su categoría gramatical es de verbo, requiere de una frase sustantiva, es decir tiene un sujeto, y su modo semántico es de proposición. La descripción tiene las siguientes restricción que modelan las propiedades de un verbo:

- Todos los verbos tienen un atributo HEAD del tipo *verb*.
- Todos los verbos deben tener un SPR.
- Por omisión la estructura de argumentos de un verbo debe tener como primer elemento una frase sustantiva.
- La semántica de un verbo siempre tendrá el atributo MODE con el valor *prop*.

El comportamiento de todos los verbos en HPSG se modela mediante un conjunto de tipos hijos de *verb-lxm*. En este conjunto se definen tipos para modelar los verbos transitivos, verbos intransitivos y verbos con complementos verbal en forma no personal. A continuación, se describen los tipos que representan estas diferentes clases de verbos.

5.1.2 Verbos transitivos

Los verbos transitivos se caracterizan porque requieren de uno o más complementos, para completar sus significado, los cuales toman roles sintácticos de objeto directo e indirecto; donde el objeto directo es el afectado por la acción verbal y el indirecto denota al que se beneficia con ésta. Este tipo es importante porque agrupa a una gran cantidad de verbos, además acepta a la voz pasiva y preguntas interrogativas por sus complementos. Estas propiedades los distinguen de los verbos auxiliares, por lo que son de utilidad para clasificar el significado original de los verbos auxiliares. El tipo *tv-lxm* representa a todos los verbos transitivos, este tipo tiene como hijos los siguiente subtipos:

(3)

Tipo	Restricción
<i>stv-lxm</i>	[ARG-ST <[_i , [_j]>]
<i>dtv-lxm</i>	[ARG-ST <[_i , [_j , NP _k]>]
<i>ptv-lxm</i>	[ARG-ST <[_i , [_j , PP _k]>]

Todos estos tipos toman como primer argumento una NP, debido a que es el valor por omisión en *verb-lxm*; esta NP representa al sujeto de la oración. El verbo *comer* es un ejemplo de verbo del tipo *stv-lxm*; a continuación se presenta una oración en voz activa y pasiva empleando al verbo *comer*:

- (4) a. Juan come la manzana (Voz activa)
b. La manzana es comida por Juan (Voz pasiva)

Como se nota el verbo *comer* toma como complemento a una frase sustantiva. La entrada léxica para *comer* como transitivo es:

(5)

$$\left\langle \textit{comer}, \left[\begin{array}{l} \textit{stv-lxm} \\ \textit{ARG-ST} \langle [\] , NP \rangle \end{array} \right] \right\rangle$$

En esta entrada léxica *comer* no especifica nada acerca de su primer argumento, especificador, debido a que toma el valor por omisión, pero si define que su segundo argumento, su primer complemento, es una frase sustantiva. El tipo *dtv-lxm* captura el comportamiento de verbos que son seguidos por dos frases sustantivas. Esta situación no es común en el español porque el objeto indirecto se suele marcar con las preposiciones: *a* o *para*; pero sí es común en el inglés, por ejemplo:

- (6) a. The tall kid give him the toy

El tipo *ptv-lxm* al igual que *dtv-lxm* tiene dos complementos, pero en *ptv-lxm* el segundo complemento es acompañado de una preposición. Este tipo verbo captura el comportamiento de los verbos ditransitivos del español. El verbo *entregar* es un ejemplo de este tipo de verbo:

- (7) a. El niño alto entregó el regalo a Juan (Voz activa)
b. El regalo es entregado a Juan por el niño (Voz pasiva)

La entrada léxica para el verbo *entregar* es la siguiente:

(8)

$$\left\langle \text{entregar} \left[\begin{array}{l} \text{ptv} - \text{lxm} \\ \text{COMPS} \langle \text{NP}, \text{PP}[\text{P} - \text{OBJ} \text{ NP}] \rangle \end{array} \right] \right\rangle$$

El atributo P-OBJ tiene el objetivo de identificar al elemento con el cual se encuentra agrupada la preposición, es decir, establece la restricción que el elemento al cual está ligada sea una frase sustantiva.

5.1.3 Verbos intransitivos

Como hijos del tipo *verb-lxm* se encuentran *iv-lxm* y *piv-lxm* que representan a los verbos intransitivos. Los complementos en verbos intransitivos no pueden ser objeto directo e indirecto; esto repercute en que los fenómenos que intervienen con los verbos intransitivos son distinto a los transitivos (ejemplo: voz pasiva). Las restricciones de estos tipos se ilustran a continuación:

(9)

Tipo	Restricción
<i>iv-lxm</i>	[ARG-ST <[_i >]
<i>piv-lxm</i>	[ARG-ST <[_i ,PP _j >]

Un ejemplo de verbo intransitivo del tipo *iv-lxm* del español es *dormir*, como vemos en la oración siguiente no tiene complemento y por lo tanto es imposible construir la voz pasiva con este tipo:

- (10) a. Juan duerme
b. *Dormido por Juan

La entrada léxica para *dormir* es la siguiente:

(11)

$$\langle \text{dormir} \text{ [iv-lxm]} \rangle$$

Es importante notar que el verbo *comer* también se comporta como un verbo intransitivo cuando refiere a la ejecución de la acción de comer (ejemplo: Juan come). El verbo *comer* presenta ambigüedad léxica que es reflejada en dos comportamientos sintácticos diferentes;

esto se soluciona de manera sencilla agregando las dos entradas léxicas en nuestro lexicón¹⁶ que representen a cada uno de los comportamientos.

Los verbos de tipo *piv-lxm* llevan un complemento, pero a diferencia de los verbos transitivos este complemento no juega un papel de objeto directo y por lo tanto algunas construcciones, como la voz pasiva, no son posibles. El verbo *viajar* es un ejemplo de este tipo, como muestran las siguientes oraciones:

- (12) a. Juan viaja a Chiapas
 b. *Chiapas es viajado por Juan

La entrada léxica para *viajar* es la siguiente:

(13)

$$\left\langle \text{viajar} \left[\begin{array}{l} \text{piv-lxm} \\ \text{COMPS} \langle \text{PP}[\text{P-} \text{OBJ} \text{ NP}] \rangle \end{array} \right] \right\rangle$$

5.1.4 Semántica de verbos transitivos e intransitivos

En HPSG se garantiza que los roles semánticos desempeñados por los elementos que integran a la oración sean asignados correctamente. Con el fin de cumplir esta tarea se utilizan el atributo SEM y un conjunto de restricciones estándar para representar la semántica de verbos transitivos e intransitivos. Los subíndices *i*, *j* y *k* que aparecen en (3) y (9) representan al valor del atributo INDEX del elemento correspondiente; esos índices son tomados por la semántica del verbo para completar sus restricciones. A continuación presentamos una tabla que contiene la descripción de la restricción semántica que agregan los verbos transitivos e intransitivos:

(14)

Tipo	Restricción semántica
stv-lxm	[AGT <i>i</i> , THM <i>j</i>]
dtv-lxm	[AGT <i>i</i> , THM <i>j</i> , PCT <i>k</i>]
ptv-lxm	[AGT <i>i</i> , THM <i>j</i> , PCT <i>k</i>]
iv-lxm	[AGT <i>i</i>]

¹⁶ Otra posibilidad es estudiar este cambio de comportamiento y crear una regla léxica que se reproduzca el fenómeno. Con esta opción solamente se captura una entrada léxica.

piv-lxm	[AGT i, THM j]
---------	----------------

El atributo AGT se relaciona con el agente de la acción, THM con el tema de la acción y PCT es el beneficiario de la acción. Como vemos todos los atributos AGT toman como valor el índice *i* como este índice proviene del primer elementos del la estructura de argumentos (ARG-ST), garantiza que el sujeto identifique al agente. El referirnos a estos atributos de esta forma supone mayor generalidad que ayuda para la definición de reglas.

5.1.5 Complementos no personales

El formalismo HPSG cuenta con dos estructuras para los verbos que requieren un verbo en forma no personal como complemento se tratan de *subject raising verb* y *subject control verb*.

5.1.5.1 Subject Raising Verb

El tipo *subject raising verb* (*srv-lxm*) es el tipo de verbo en HPSG que se encarga de caracterizar aquellos verbos que cumplen con las siguientes características:

- El verbo tiene como complemento una frase
- El verbo comparte su especificador con el verbo en forma no personal.
- El verbo no tiene agente quien realice su acción.
- El verbo *volver a* es un ejemplo de este tipo de verbo. En (15) se aprecia que el verbo tienen como complemento una frase y que el sujeto de este verbo define al sujeto de la acción de *entregar*, el cual a la vez ayuda a definir el agente de esta acción; pero este agente no es utilizado por *volver* ya que este verbo señala únicamente la repetición de la acción de *entregar*, lo cual no necesita de un agente.
- El niño alto volvió a entregar el juguete a Juan

La definición de este tipo se hace de la siguiente forma:

(15)

$$\left[\begin{array}{l} \text{srv-lxm} \\ \text{ARG-ST} \left\langle \#1, \left[\begin{array}{l} \text{phrase} \\ \text{SPR} \langle \#1 \rangle \end{array} \right] \right\rangle \end{array} \right]$$

Como se ve más adelante este tipo sirve de base para capturar una variedad de comportamientos, por lo que se evita especificar la categoría de la frase que toma como segundo argumento. En particular, para los verbos auxiliares esta frase es de categoría verbal.

5.1.5.2 Subject Control Verb

El tipo *subject control verb* (*scv-lxm*) representa a los verbos con las siguientes características:

- El verbo tiene como complemento una frase verbal en forma no personal.
- El verbo comparte su especificador con el verbo en forma no personal.
- La acción denotada por el verbo tiene agente quien realice su acción.

El verbo *volver* también presenta un comportamiento de control cuando en lugar de marcar la repetición de una acción significa el regreso a un lugar con el fin de realizar algo. La misma oración (15) tiene esta lectura, es decir, es ambigua. En ese ejemplo vemos que tiene como complemento una frase, que el sujeto ayuda a definir el sujeto de la frase verbal y que el agente de *volver*, es decir, quien vuelve es el mismo que quien entrega. La compartición del agente entre el *volver* y *entregar* es lo que diferencia a un verbo *scv-lxm* de un *srv-lxm*. El tipo *scv-lxm* se define de la siguiente manera:

(16)

$$\left[\begin{array}{c} scv-lxm \\ ARG-ST \left\langle NP_i, \left[\begin{array}{c} phrase \\ FORM \quad inf \\ SPR \quad \langle NP_i \rangle \end{array} \right] \right\rangle \end{array} \right]$$

La estructura atributo valor es aparentemente similar a la de *srv-lxm*; sin embargo, existe una pequeña diferencia, el índice que marca al especificador y al primer argumento de la lista ARG-ST nos indica que la NP debe de estar indexada. Es decir, la NP debe estar representada por una restricción semántica, porque la estructura semántica de *scv-lxm* hace uso del índice para definir al agente de la acción que representa el verbo de este tipo. En otras palabras, un verbo de control requiere necesariamente de un agente, quien también es el agente de la acción denotada por la frase verbal que toma por complemento.

5.2 Auxiliares del español en HPSG

Los tipos de verbos que tienen una frase verbal como complemento en HPSG (*srv-lxm* y *scv-lxm*) cumplen con dos de las propiedades de los verbos auxiliares, la primera, que consiste en tener como complemento a un verbo en forma no personal, y la segunda, en compartir su especificador con su complemento. Sin embargo, existe una diferencia entre estos tipos: *scv-lxm*, requiere que el verbo en forma no personal tenga agente, mientras que el tipo *srv-lxm* no, la cual es una propiedad de los verbos auxiliares. Por lo mismo, el tipo que representa a los verbos auxiliares del español en HPSG es *srv-lxm*. Este tipo permite modelar de manera integral a los verbos auxiliares, como se ve en los siguientes incisos:

- El tipo *srv-lxm* ayuda a definir el agente en el verbo que le sigue, si este lo necesita. Esto se logra porque tanto su especificador como el de su complemento debe ser el mismo; por lo tanto, si el complemento necesita de un agente el auxiliar se encargará de pedirlo y definirlo a través de su conjugación.
- El tipo *srv-lxm* acepta formas no personales sin agente; esto nos permitirá que el verbo en forma no personal sea unipersonal.
- Existe una regla léxica exclusiva de los verbos transitivos que les permite formar la voz pasiva y la forma interrogativa para preguntar por el complemento directo. Una vez que estas reglas son aplicadas sobre algún verbo transitivo, un verbo del tipo *srv-lxm* puede tomarlo como complemento.
- Como el tipo *srv-lxm* es declarado como hijo de *verb-lxm* y las reglas léxicas para preguntar por el complemento directo y para formar la voz pasiva sólo se aplican a los verbos transitivos, el *srv-lxm* no es afectado por estas reglas.

En conclusión, el comportamiento de los verbos auxiliares del español es capturado por el tipo *srv-lxm* de HPSG. A continuación se presentan las consideraciones generales que integran al modelo de verbos auxiliares del español. Este modelo está integrado por un conjunto de tipos hijos de *srv-lxm* y uno de entradas léxicas que definen los comportamientos que tienen los verbos auxiliares del español. Con el propósito de facilitar la presentación de las siguientes secciones se sigue una estructura similar a la presentada con el análisis de cada uno de los verbos en el capítulo anterior (Sección 4.2).

5.2.1 Verbo auxiliar + infinitivo

Debido a que este conjunto de verbos van seguidos de un infinitivo y donde una preposición puede aparecer entre el verbo auxiliar y el infinitivo, se crearon los tipos llamados *inf_srv-lxm* y *pp_inf-srv-lxm*. Estos nuevos tipos son hijos de *srv-lxm*, y solamente especifican con mayor precisión el tipo de verbo que toma como complemento un verbo auxiliar del español. La definición de *inf_srv-lxm* es la siguiente:

(17)

$$\text{inf_srv-lxm} \left[\begin{array}{l} \text{srv-lxm} \\ \text{ARG-ST} \langle [], [\text{HEAD} \mid \text{FORM inf}] \rangle \end{array} \right]$$

Esta definición permitirá a los verbos declarados como *inf_srv-lxm* tomen como complemento a las estructuras que tengan un valor FORM de *inf*. El atributo FORM se utiliza en los verbos para marcar la forma del verbo; el valor *inf* corresponde al infinitivo. La definición de *pp_inf_srv-lxm* es la siguiente:

(18)

$$\text{pp_inf_srv-lxm} \left[\begin{array}{l} \text{srv-lxm} \\ \text{ARG-ST} \langle [], [\text{HEAD} \left[\begin{array}{l} \text{Phrase} \\ \text{FORM prep} \\ \text{P-OBJ} [\text{FORM inf}] \end{array} \right] \rangle \rangle \end{array} \right]$$

Esta definición requiere como complemento una frase preposicional que tenga como P-OBJ una forma infinitiva; esto nos garantiza que la preposición este ligada a un verbo en infinitivo.

5.2.1.1 Poder

En el capítulo anterior pudimos ver que existen dos significados y tres comportamientos sintácticos diferentes para el verbo *poder*:

- (19)
- a. Puedo con las matemáticas (Habilidad)
 - b. Puedes entregar el juguete a Juan (Habilidad)
 - c. Puedes entregar el juguete a Juan (Posibilidad)

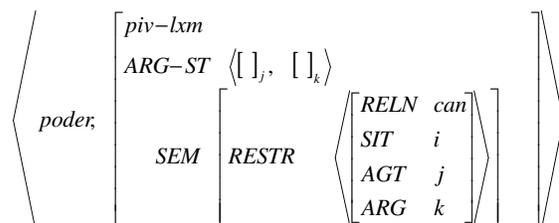
El ejemplo (20.a) corresponde a un verbo con una frase preposicional, existen dos estructuras en HPSG que requieren de una preposición: *piv-lxm* y *stv-lxm*, pero como se

observa en (21) no es posible formar la voz pasiva de (21.a); por lo tanto es un verbo con intransitivo.

(20) *Las matemáticas son podidas por mi

El tipo que le corresponde a *poder* es de *piv-lxm*; su entrada léxica es la siguiente:

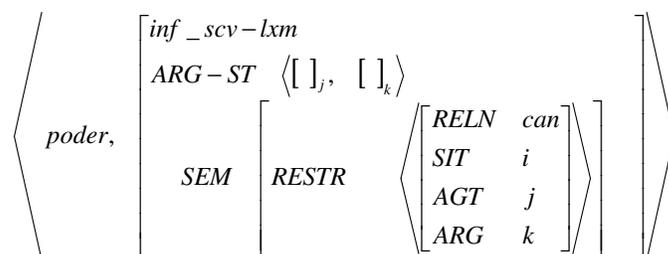
(21)



Esta entrada léxica hereda todas la propiedades sintácticas del tipo *piv-lxm* al declararse de este tipo. Adicionalmente, la entrada léxica agrega la restricción semántica *can* que necesita dos argumentos: *AGT* que identifica al agente de la acción y *ARG* que identifica a la capacidad o habilidad del sujeto; en el ejemplo (20.a) se trata de *las matemáticas*.

Pasamos ahota al análisis de la oración (20.b). En esta oración *poder* tiene el significado de capacidad. Bajo este significado el verbo requiere a un agente que tome el rol de “el capaz”; vale la pena destacar que significado no es de auxiliar, dado a que este requiere de un agente. *srv-lxm* es el tipo permite representar el significado de *capacidad*, ya que necesita de un agente. La entrada léxica es la siguiente¹⁷:

(22)



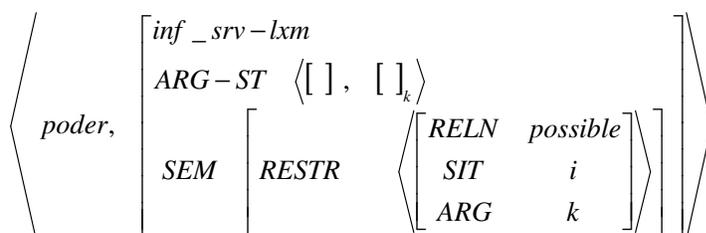
El primer punto que debemos observar es que se trata de una entrada léxica del tipo *scv-lxm*. Notamos que en la restricción semántica que se agrega es igual a la restricción a la que se presenta en (22); esto se debe que ambas representan la misma restricción semántica en donde hay un capaz (*AGT*) y una capacidad (*ARG*). Sin embargo, en (22) *ARG* se

¹⁷ El tipo *inf_scv-lxm* se declara de forma similar a *inf_srv-lxm*, para que cumpla una función similar.

relaciona con la materia de las matemáticas, es decir, una NP; mientras que (23) con una acción verbal, *entregar*.

Como se puede notar la oración (20.b) y (20.c) son las mismas en cuanto a las palabras que las componen, es decir, son ambiguas. Pero, mientras que (20.b) presenta el significado de capacidad, (20.c) hace referencia al significado de posibilidad. Esto se debe a que el verbo *poder* es ambiguo. *Poder* bajo el significado de posibilidad es precisamente el tipo de verbo al que está orientado este trabajo; se trata de un verbo auxiliar como quedo demostrado en el capítulo anterior. La entrada léxica se define de la siguiente forma:

(23)



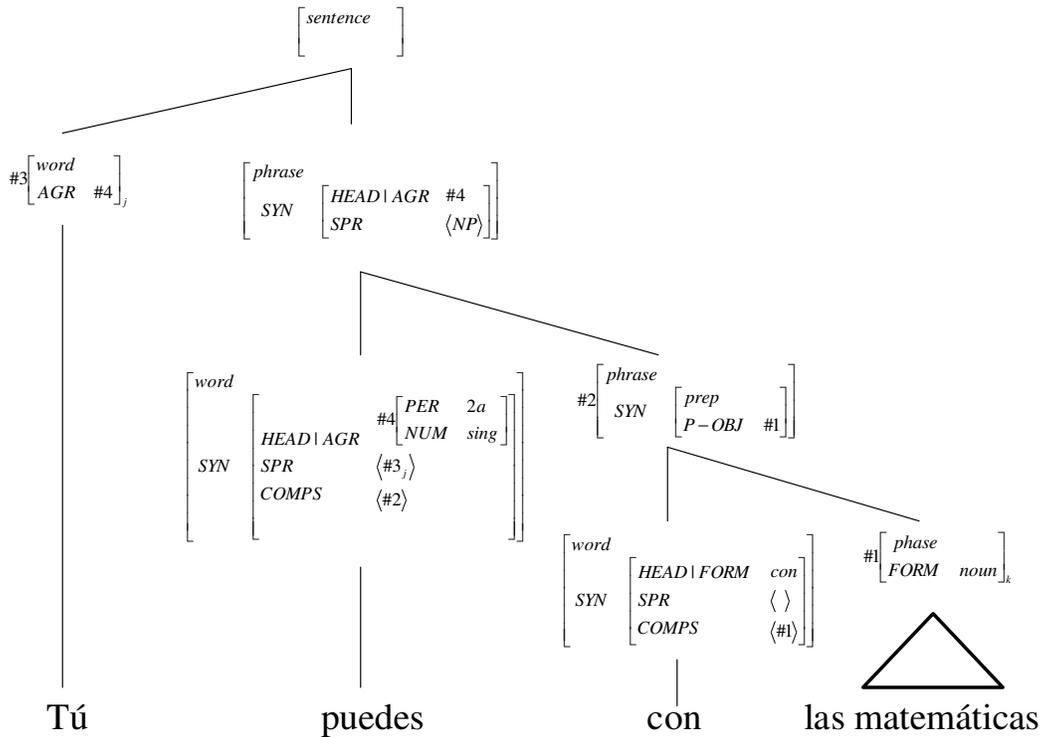
Esta entrada léxica para *poder* es de tipo *inf_srv-lxm*. La restricción semántica que en ella se agrega, se trata de la relación *possible*, la cual toma como argumento una situación posible (ARG). Nótese que a diferencia de (22) y (23) en la restricción semántica de (24) no hay ningún elemento que se refiera al agente (AGT), ya que el tipo *srv-lxm* no tiene uno.

A continuación presentamos un conjunto de árboles sintácticos y semánticas, los cuales corresponden a la oración (18.a) y a *tú puedes mostrar el catálogo*, oración perteneciente al corpus DIME. Un ejemplo de este tipo no se repetirá con los siguientes verbos auxiliares, a menos que presenten un caso especial, ya que los árboles que se forman y las semánticas resultantes son muy similares.

El árbol sintáctico del verbo *poder* como tipo *piv-lxm* es mostrado en (25). En la parte inferior vemos las entradas léxicas una vez que cumplen con la satisfacción léxica. La primer regla en aplicarse es la de HCR; la preposición *con* toma a su complemento que consiste en la frase sustantiva *las matemáticas*. La regla se vuelve aplicar, pero en esta ocasión se hace sobre *puedes* como frase cabeza; *poder* requiere como complemento una frase preposicional cuyo atributo P-OBJ sea una frase sustantiva; en este caso *con las matemáticas*. Por último la regla HSR se aplica sobre esta frase para que sature su atributo SPR y así se agrupe en una misma estructura toda la oración. El HEAD de la oración es el HEAD de *puedes* y sus atributos COMS y SPR se encuentran saturados. Los índices de la

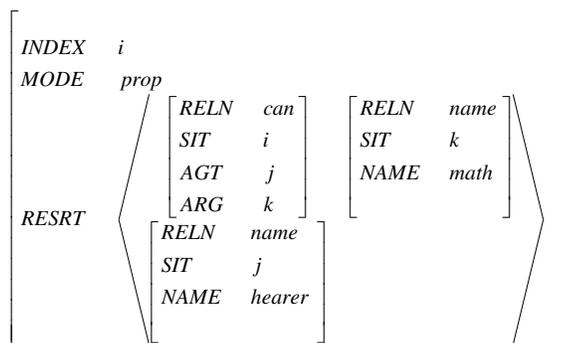
semántica se ligan con el valor correspondiente cuando se solicita la unificación entre la descripción del complemento en COMPS y el complemento mismo, a través de la regla HCR. Ya que *poder* solicita que el INDEX de su especificador y complemento sean el valor del atributo AGT y ARG respectivamente.

(24)



El atributo SEM que corresponde a esta oración es el siguiente:

(25)

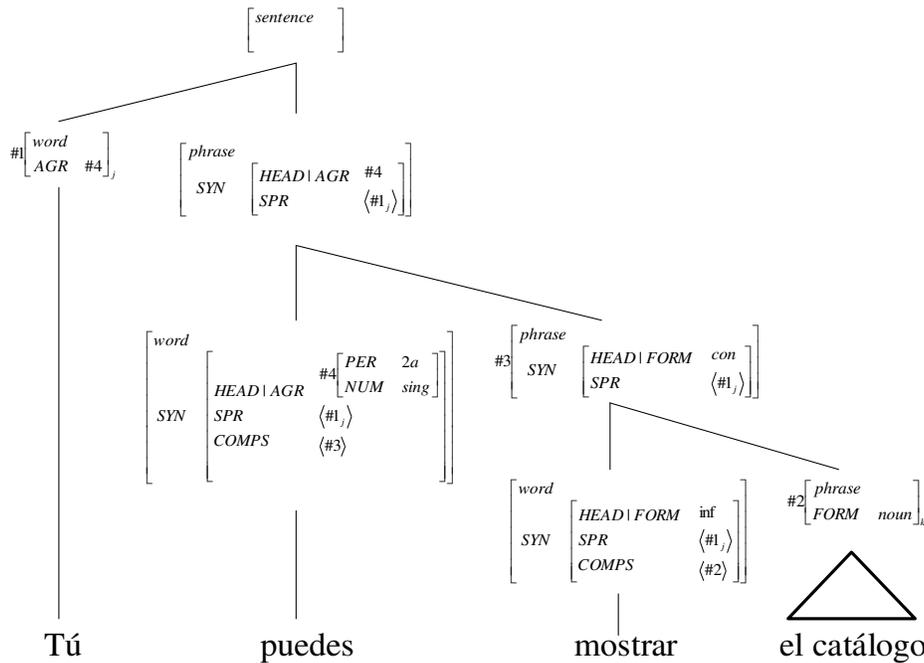


En esta estructura se aprecia como se han determinado los argumentos de la semántica. La relación de *can* es la que representa a toda la semántica, porque el valor *i* de INDEX corresponde a esta situación. En esta relación hay dos argumentos, uno que representa al

agente (AGT) y otro a la capacidad (ARG); el valor del primero se refiere al escucha, segunda persona del singular (tú); el segundo se refiere a las matemáticas.

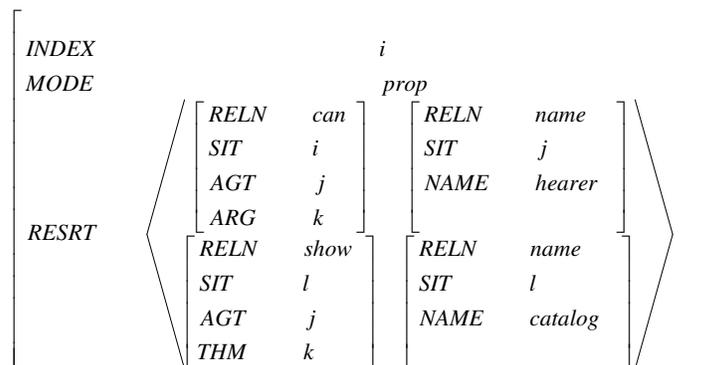
El árbol sintáctico para el verbo *poder* comportándose como verbo de control es el siguiente:

(26)



El árbol se formó como sigue: la primera regla en aplicarse es la de HCR que permitió a *mostrar* tomar su complemento (*el catálogo*); una vez que tomó su complemento se aplica de nuevo la misma regla, ahora teniendo como cabeza a la estructura de *puedes*, para que esta tome a su complemento (*mostrar el catálogo*); por último *puedes* toma su especificador agrupándose de esta forma toda la oración. La semántica correspondiente a esta estructura es la siguiente:

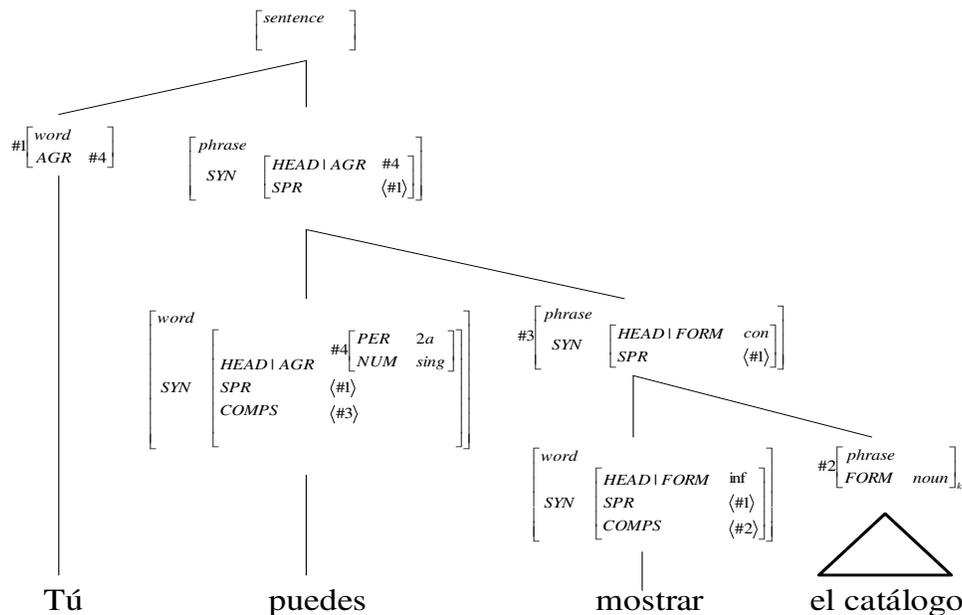
(27)



En (28) se ve que la relación *can* sigue siendo la principal y que es similar a la del ejemplo (26), pero se diferencian en el tipo de restricción que representa el índice del complemento (ARG); en esta estructura el argumento ARG se refiere a una acción de mostrar (*show*), la cual tiene dos atributos. El atributo que representa al agente (AGT) y del objeto mostrado (OD). El objetivo principal del tipo *scv-lxm* es el de garantizar que tanto el verbo de control como el verbo en forma no personal tengan el mismo agente, esto se cumple dado que ambos atributos AGT comparten al índice *j*, que se refiere al escucha, segunda persona del singular (tú).

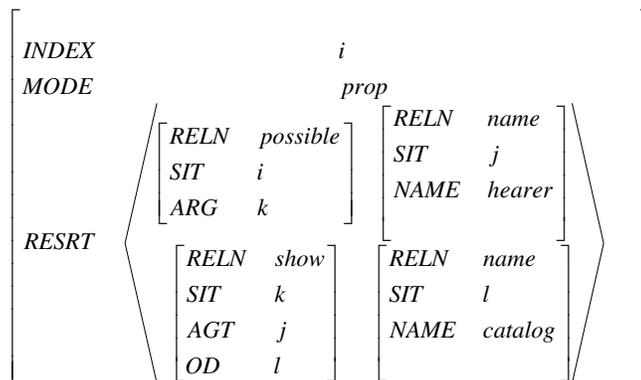
La estructura sintáctica de la oración cuando *poder* es un verbo auxiliar se muestra a continuación:

(28)



La aplicación de las reglas generan este árbol es similar a la que se aplicó en (27); la diferencia radica en la semántica y en cómo se resolvieron los índices de ésta. La estructura semántica resultante se muestra a continuación:

(29)



La restricción semántica que agrega *poder* es diferente a la de (28) y (26); en primer lugar ésta se refiere a la situación de *possible* y ya no a la de *can*, además esta sólo requiere de un argumento (ARG), el cual hace referencia a la relación de *show*. Como esta restricción sólo tiene un argumento y este ya está resuelto, el argumento que corresponde al agente de *mostrar*, el índice *j*, no es compartido. Esta semántica cumple con las condiciones que establecen los verbos auxiliares del español al no tener agente la restricción que representa a la acción de *poder*.

Con el fin de clarificar un poco más la diferencia entre las lecturas de *poder* como capacidad y posibilidad se presenta como quedarían representadas las semánticas a manera de predicado lógico:

- (30) a. *Poder* de capacidad:
can(listener,show(listener,catalog))
b. *Poder* de posibilidad:
possibility(show(listener,catalog))

El primer predicado denota la situación en donde el escucha tiene la capacidad de mostrar el catálogo; mientras que la segundo a la situación en donde es posible que el escucha muestre el catálogo.

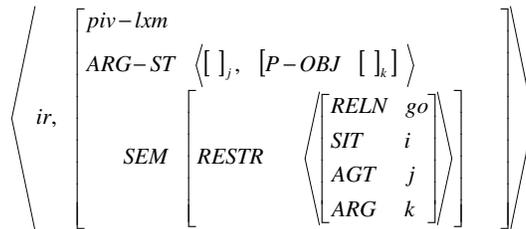
5.2.1.2 *Ir a*

El verbo *ir* presenta dos significados: su significación original de desplazamiento y como un auxiliar que marca la ocurrencia de una situación el en futuro. Con el sentido de desplazamiento el verbo *ir* se comporta como intransitivo que lo acompaña la preposición *a* como se muestra a continuación:

- (31) a. Juan va a la casa (voz activa)
 b. *La casa es ida por Juan (voz pasiva)

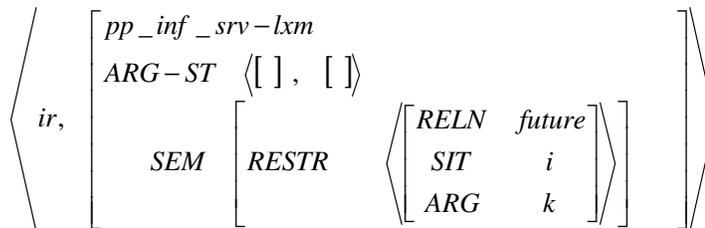
La entrada léxica que captura estas oraciones es:

(32)



Con el significado de marca de futuro el verbo *ir* se comporta como un auxiliar, tal y como se observó en el capítulo anterior. La entrada léxica correspondería a la siguiente:

(33)



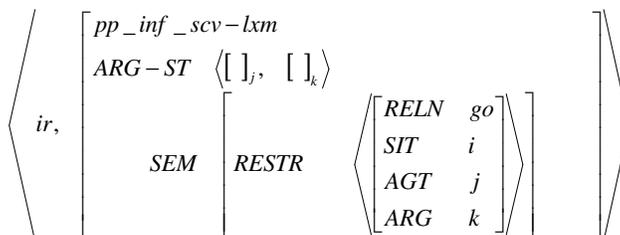
Se declara como *pp_nf_srv-lxm* para que acepte un complemento en infinitivo y cumpla con el comportamiento de *subject raising verb*, es decir de verbo auxiliar del español.

Solamente nos queda por capturar el comportamiento de la siguiente oración:

Juan va a comer (significado de desplazamiento)

Esta oración hace uso del significado de desplazamiento, en donde Juan se dirige con el propósito de comer. En esta oración indudablemente el comportamiento de *ir* es de *subject control verb*, debido a que el que se dirige y el que come es Juan. La entrada léxica que lo representaría es:

(34)



Esta última oración muestra un comportamiento diferente del verbo *ir*; el significado sigue siendo de desplazamiento pero no existe un argumento que identifique al destino; ya que este argumento ha sido sustituido por el objetivo o propósito de la acción de desplazamiento (ARG). Generalmente el propósito u objetivo de una acción se marca con las preposiciones *a* y *para*, como se muestra en la siguiente oración (considerando un sentido de desplazamiento):

(35) a. El niño alto va a la casa para entregar el regalo a Juan

b. El niño alto va a la casa a entregar el regalo a Juan

En estas oraciones el que se dirige y el que entrega es la misma persona, *el niño alto*. Lo que ha ocurrido con el verbo *ir* de las oraciones (35), es que han perdido su complemento; pero además, debido a que *ir* utiliza a la preposición *a* para marcar su destino y es la misma para marcar el propósito de la acción, se han unido ambas funciones e *ir a* se comporta como un verbo de control.

5.2.1.2.1 Preposición de control de sujeto (*Subject control preposition*)

En (36) se observa como se comparte el agente entre los verbos *ir* y *entregar* a través de las preposiciones *a* y *para*. En este caso *a* y *para* cumplen con la función sintáctica de unir a dos frases y permitir que los agentes sean los mismo. Este fenómeno es mucho más general, observemos los siguientes ejemplos:

(36) a. Tú puedes entregar el regalo a Juan para ir a comer

b. Juan envía la carta para estar tranquilo

c. La carta es endosada por Juan para ser enviada

d. *Llueve para crecer las plantas/

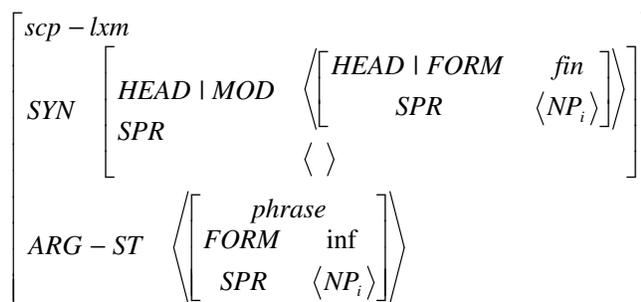
Llueve para que crezcan las plantas

La preposición *para* en estas oraciones marca el propósito de la acción realizada. Siempre va seguida de un verbo en infinitivo si quiere compartir al agente como en los incisos (a), (b) y (c). Cuando se cambia de agente en la segunda acción verbal se recurre al empleo de una cláusula relativa marcada por *que*, (d). Las cláusulas relativas están fuera del alcance de este trabajo; pero la preposición *para* bajo el comportamiento que presentó en las primeras tres oraciones de (37) puede ser declarada con el mismo espíritu que *srv-lxm* y *scv-lxm*. Existe suficiente evidencia que la preposición como marca de propósito es un fenómeno más complejo, por ejemplo: *La carta es endosada por Juan para estar tranquilo*, en donde

Juan es el agente de la segunda acción y no *la carta*; pero en este trabajo no abordaremos dicha complejidad; solamente nos enfocamos al comportamiento presentado en (37) en donde los agentes de ambas oraciones son compartidos.

El tipo *scp-lxm* (*subject control preposition lexeme*) se declararía como hijo del tipo *mkp-lxm*¹⁸ que captura el comportamiento de las preposiciones para que adquiriera los atributos de preposición. La restricción del tipo queda como se muestra en (37); esta restricción garantiza que la preposición *para* va a tomar un complemento en forma infinitiva (FORM inf) cuyo agente será el mismo que el de una oración a la cual es capaz de modificar, esta última oración deberá estar conjugada (FORM fin).

(37) Tipo *scp-lxm*



Esta restricción pide que ambos verbos tengan un agente dejando fuera de consideración a oraciones donde intervienen verbos unipersonales¹⁹:

- (38) a. *Llueve para dormir
 b. *Duerme para llover

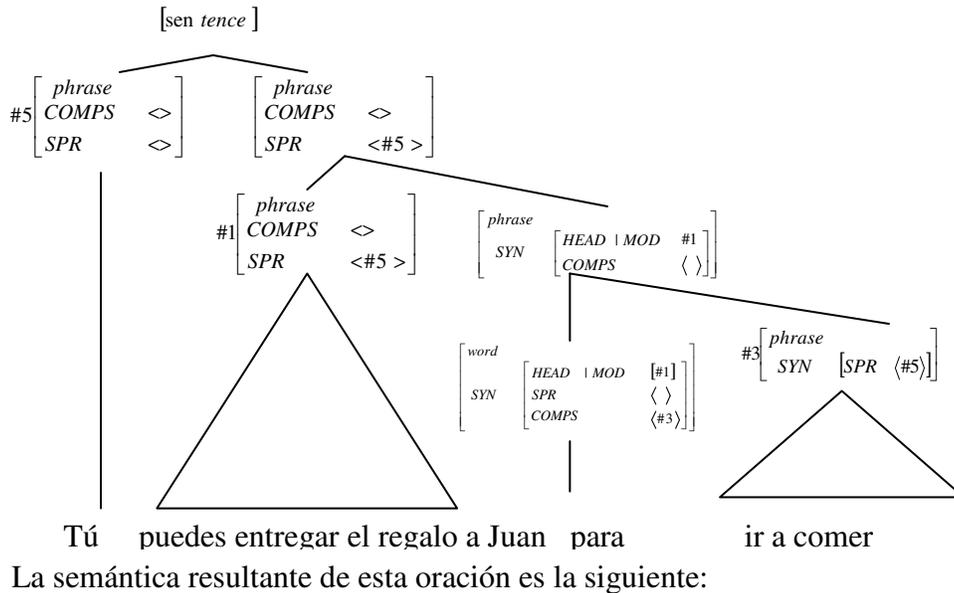
El árbol sintáctico común para una oración como la presentada en (37.a) se puede observar en (40). Para formarse este árbol, los verbo *poder* y *cumplir* debieron haber tomado sus complementos por la regla de HCR. Una vez que esto fue posible la preposición *para* toma a su complemento, en este caso a la frase: *cumplir con su tarea*. En este momento se puede

¹⁸ Existen dos tipos de preposiciones, aquellas que se utilizan como complementos de un verbo (*pdp-lxm*), por ejemplo *a* in *voy a la casa*; y las preposiciones que sirven para agregar información a la oración cuya presencia es opcional (*mkp-lxm*), por ejemplo *en* en la oración: *voy a la casa en coche*.

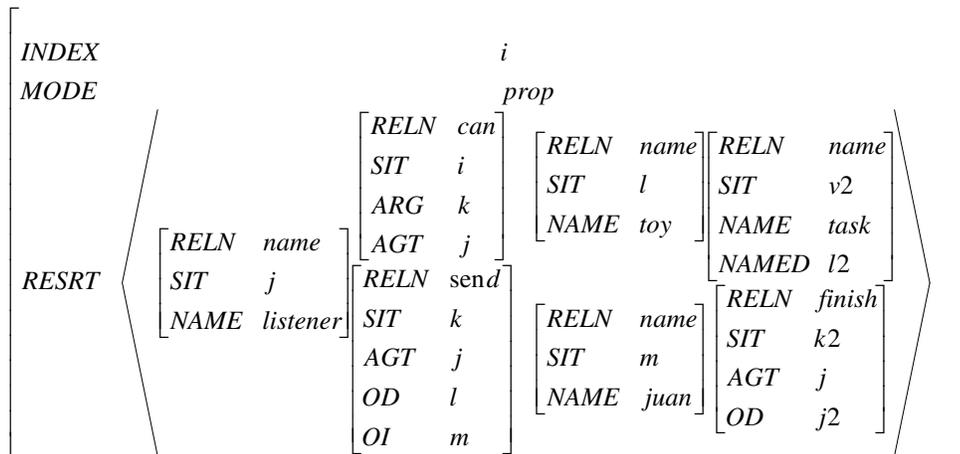
¹⁹ Existe evidencia que es posible formar oraciones como: *el cielo se nubla para llover*, donde el segundo verbo no tiene agente; pero quedan fuera de alcance debido a que provienen de texto literarios y de literatura especializada en meteorología.

aplicar la regla HMR, donde *para* modifica a la estructura que representa a la frase: *puede mostrar el catálogo*. A esta última frase le hace falta saturar su especificador por lo que se aplica la regla HSR.

(39)



(40)



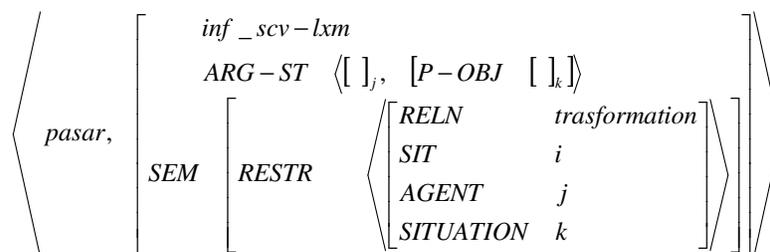
En esta estructura es importante notar que existen tres relaciones que tienen a un agente como parte de su restricción, *can*, *show* y *finish*; las tres representan un verbo de la oración (35.a). La forma en como se resuelven los índices en *can* y *show* fue analizada en (27). La declaración de *para* como *scp-lxm* nos garantiza que el agente en el verbo que modifica (*poder*) sea el mismo que el de la frase verbal que toma como objeto (*ir*), en este caso el

agente es identificado con el índice j , es decir, la segunda persona del singular, el cual corresponde al agente.

5.2.1.3 Pasar a

El verbo *pasar* no es un verbo auxiliar como quedo asentado en el capítulo anterior; esto se debió a que *pasar* requiere de un agente. El hecho de que necesite de un agente, indica que la descripción del verbo debe tener un agente. Para *pasar* en construcciones perifrásticas el tipo correspondiente es el de *scv-lxm*. La entrada léxica es la siguiente:

(41)

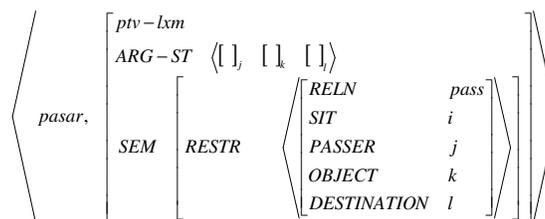


El significado original para el verbo *pasar*, el cual se refiere al movimiento de un objeto, el verbo se comporta como un verbo ditransitivo, posee objeto directo e indirecto, como muestran los siguientes ejemplos:

- (42) a. Juan paso la pelota a María (voz activa)
 b. La pelota es pasada a María por Juan (voz pasiva)

La entrada léxica para este comportamiento es:

(43)



5.2.1.4 Echar a

El verbo *echar* es uno de los verbos más difíciles de clasificar, debido a lo restrictivo de su comportamiento respecto al tipo de verbos que toma y a las ambigüedades sintácticas que presenta. El problema mayor lo presenta debido a que el agente de *echar* no coincide con el agente del verbo en infinitivo, pero además es posible generar un orden diferente al que muestran las construcciones perifrásticas:

- (44) a. Echás a andar la máquina
 b. Echás a la máquina andar

Un modelo que permite manejar este comportamiento es el que se propone con el tipo *orv-lxm*, donde uno de los complementos del verbo realiza la función de objeto directo de un verbo y de sujeto de otro. Como se trata de una descripción *raising* es posible que el segundo verbo involucrado no tenga un agente lo que coincide con el hecho que *echar* permita la utilización de verbos unipersonales. La lista de argumentos de un *orv-lxm* es de la siguiente forma:

(45)

$$\left[ARG-ST \left\langle [NP]_j, [\#1]_k, \left[\begin{array}{l} phrase \\ SPR \langle \#1 \rangle_i \end{array} \right] \right\rangle \right]$$

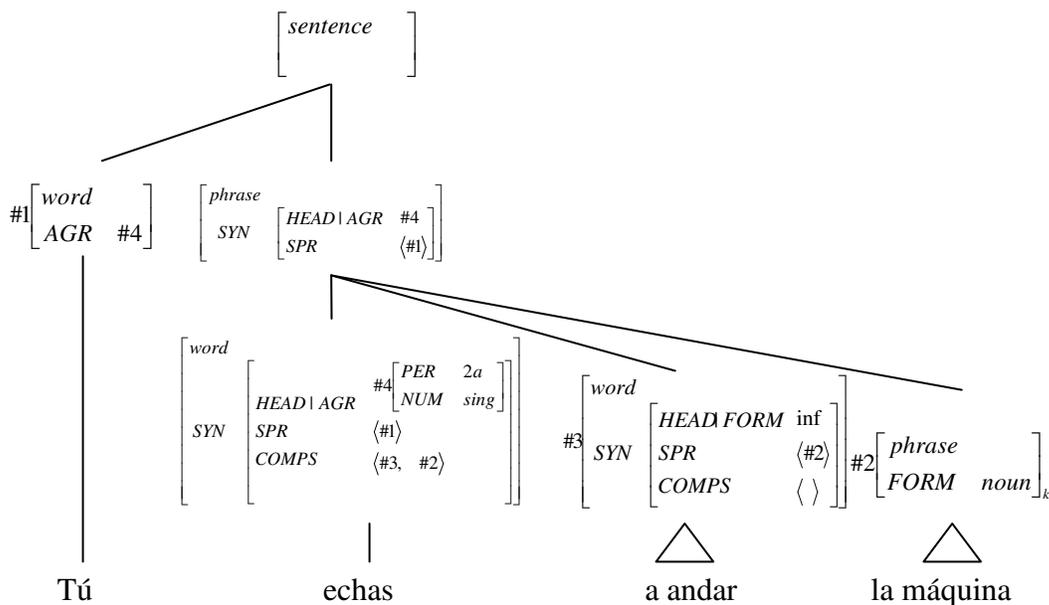
En el español adicionalmente se tendría que capturar también el siguiente orden, en donde se encuentran intercambiados el sujeto y el objeto directo :

(46)

$$\left[ARG-ST \left\langle [NP]_j, \left[\begin{array}{l} phrase \\ SPR \langle \#1 \rangle_k \end{array} \right], [\#1]_i \right\rangle \right]$$

La estructura sintáctica para una oración como (45.b)²⁰ es la siguiente:

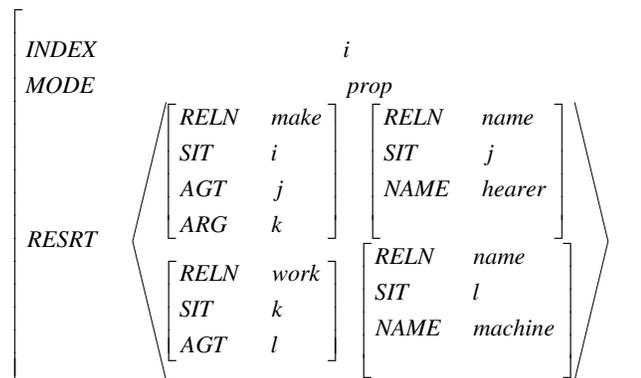
(47)



²⁰ Se a agregado el sujeto *tú*.

La oración se parsea con la aplicación de dos reglas. El verbo *echar* pide dos complementos uno es un verbo (*andar*) y el otro su objeto directo (*la máquina*), el especificador de *andar* y el objeto directo (*la máquina*) unifican, por lo tanto es posible aplicar la regla HCR. Por último, sólo queda aplicar HSR para que el verbo *echar* tome su sujeto y defina su agente. La semántica resultante es la que se muestra en (49). Esta estructura debe leerse como la situación en donde alguien hace funcionar un artefacto, en donde ese alguien es el escucha y el artefacto es una máquina.

(48)



Como se puede observar los agentes correspondientes a los dos verbos son diferentes, el primero se refiere a la segunda persona del singular, el escucha; mientras que el segundo a la máquina.

5.2.1.5 Venir a

El verbo *venir* presenta un comportamiento similar a *ir*, existe la versión intransitiva (50.a), una como auxiliar (50.b) y una como verbo de control (50.c) el cual proviene de un razonamiento similar al que nos llevó definir *scp-lxm*.

- (49)
- a. Viene de Chiapas
 - b. Vino a coincidir con Juan
 - c. Vino a comer /Vino para comer

Las entradas léxicas son semejantes a las que se declararon en *ir*.

5.2.1.6 Volver a

Como se mencionó en el capítulo anterior *volver* es uno de los verbos auxiliares como se muestra en (51.b). Por otro lado, por las mismas causas que *ir* y *venir*, *volver* es un verbo

de control como se observa en (51.c). Y adicionalmente, *volver* también presenta un comportamiento de verbo intransitivo (51.a).

- (50) a. Vuelve a casa
 b. Vuelve a entrar a la casa
 c. Vuelve a comer/Vuelve para comer

Las entradas léxicas para *volver* se definen de forma similar a las declaradas en *ir*.

5.2.1.7 *Haber de*

El verbo *haber* bajo su significado original presenta un comportamiento impersonal, es decir, no tiene agente; además sólo se puede conjugar en tercera persona del singular bajo la conjugación irregular de *hay*. Este tipo de verbos, dentro de los cuales también se encuentran los unipersonales, se declaran como del tipo *no_subj-verb-lxm*, el cual es hijo de *verb-lxm*. La restricción para este tipo de verbos es la siguiente:

(51)

$$\left[ARG-ST \left\langle \left[SYN \mid HEAD \mid AGR \left[\begin{array}{ll} NUM & sng \\ PER & 3rd \\ GEND & null-gnd \end{array} \right] \dots \right] \right\rangle \right]$$

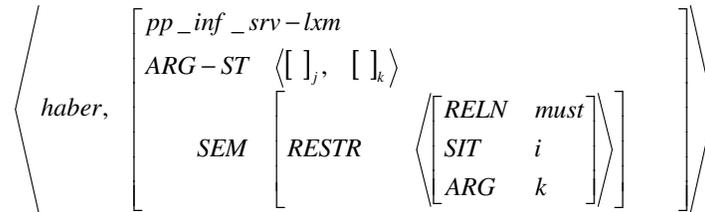
Esta estructura nos dice que el sujeto de este verbo debe estar en tercera persona del singular, esto evita que se combine con su sujeto, ya que los sujetos poseen un género femenino (*fem*) o masculino (*masc*) y nunca *null-gnd*. Existe el inconveniente de que el sujeto representado en el SPR seguirá presente por lo que necesitamos que se sature; esto se logra quitando el especificador a través de una regla gramatical que discutiremos con más detalles en el próximo capítulo (*Subject PRO Drop Rule*) y que es general para el sujeto omitido del español. La entrada léxica para *haber* como verbo impersonal es:

(52)

$$\left\langle haber \left[\begin{array}{l} no_subj-verb-lxm \\ ARG-ST \langle [] []_j \rangle \\ SEM \left[RESTR \left\langle \left[\begin{array}{ll} RELN & exist \\ SIT & i \\ AGT & j \end{array} \right] \right\rangle \right] \end{array} \right] \right\rangle$$

Como verbo auxiliar *haber de* tiene la siguiente entrada léxica:

(53)

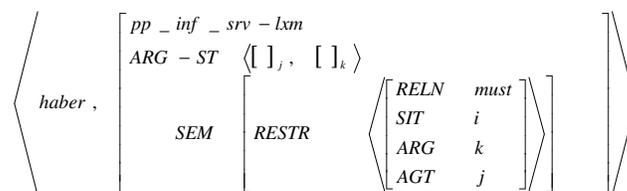


5.2.1.8 *Haber que*

Como se observó el verbo *haber que* no es un verbo auxiliar y denota la obligación de realizar una acción. Esta perífrasis se presenta solamente bajo una conjugación impersonal, por lo que esperaríamos que *haber* no tuviera un agente. Pero por el contrario, el verbo requiere de un agente en el verbo en infinitivo; este sólo sirve para marcar la existencia “alguien” que tiene la obligación de realizar la acción marcada por el verbo en infinitivo. La situación de necesitar un agente en ambos verbos para que al final este no sea definido parecería algo complicado de capturar; pero en realidad HPSG no se encarga de relacionar los índices con objetos reales del mundo, por lo que el comportamiento del verbo *haber que* es capturado por el tipo *scv-lxm*; y será un módulo ajeno al parseo el encargado de asignar que esta tercera persona del singular realmente no es alguien en específico.

Otra situación que es de mucho cuidado es la aparición de la palabra *que*. Esta es considerada como un pronombre o como conector de frases; pero en este trabajo proponemos que se comporta como una preposición cuando acompaña a *haber*. Esta aseveración se basa en la consideración que hace HPSG al definir categoría gramatical como aquellas palabras que en un contexto determinado se comportan de manera semejante. Como pudimos ver en el capítulo anterior, el contexto en el que se desempeña la palabra *que* es similar al de las preposiciones *a*, en *ir a*, o *de*, en *venir de*. No dudamos que este uso proviene de una simplificación del uso de la palabra *que* como pronombre o conector; pero para fines de este trabajo consideramos acertada esta distinción. La entrada léxica para este verbo es:

(54)



5.2.1.9 Tener que

El verbo *tener que* es un auxiliar y también es uno de los verbos que marca la obligación a ejecutar una acción. *Tener* en su sentido original refiere a la posesión de un objeto, en donde existe alguien que posee y el objeto poseído; su comportamiento es de verbo intransitivo, como se muestra a continuación:

- (55) a. Juan tiene una casa (Voz activa)
 b. *Una casa es tenida por Juan (Voz pasiva)

Debido a que este verbo no va acompañado de una preposición y es intransitivo no existe la posibilidad de capturar el comportamiento de *tener* con los tipos de HPSG, ya que para verbos intransitivos están los tipos *iv-lxm* y *piv-lxm* y con ninguna de estas definiciones cumple el verbo *tener*. Por lo que proponemos la creación del tipo *siv-lxm* cuyo ARG-ST sería:

(56)

$$[ARG-ST \langle [NP]_j [NP]_k \rangle]$$

La entrada léxica es la siguiente:

(57)

$$\left\langle \begin{array}{l} \text{tener,} \\ \left[\begin{array}{l} \text{ siv-lxm} \\ \text{ ARG-ST } \langle []_j, [P-OBJ []_k] \rangle \\ \text{ SEM } \left[\begin{array}{l} \text{ RESTR } \left\langle \left[\begin{array}{l} \text{ RELN } \text{ have} \\ \text{ SIT } \text{ i} \\ \text{ AGT } \text{ j} \\ \text{ ARG } \text{ k} \end{array} \right] \right\rangle \end{array} \right] \end{array} \right. \end{array} \right\rangle$$

Por último, la entrada léxica que define *tener que* como auxiliar es:

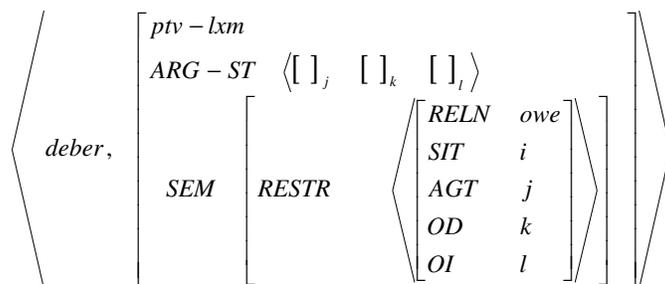
(58)

$$\left\langle \begin{array}{l} \text{tener} \\ \left[\begin{array}{l} \text{ SEM } \left[\begin{array}{l} \text{ RESTR } \left\langle \left[\begin{array}{l} \text{ VCL } ? \\ \text{ VBC } F \\ \text{ ZIL } i \\ \text{ BETV} \text{ mmm} \end{array} \right] \right\rangle \end{array} \right] \\ \text{ ARG-ST } \langle []^1, []^k \rangle \\ \text{ HPSG-TYPE } \text{ siv-lxm} \end{array} \right. \end{array} \right\rangle$$

5.2.1.10 Deber de

El verbo *deber* presenta dos comportamientos sintácticos, uno de verbo auxiliar y de verbo ditransitivo. La entrada léxica como auxiliar es similar a las declaradas anteriormente y la de ditransitivo corresponde a la siguiente:

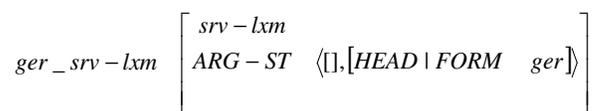
(59)



5.2.2 Verbo auxiliar + gerundio

Al igual que se hizo con los complementos infinitivos y siguiendo la idea de generalizar los tipos según la forma no personal del complemento, para los verbos auxiliares que son seguidos de un verbo en gerundio declaramos el siguiente tipo:

(60)

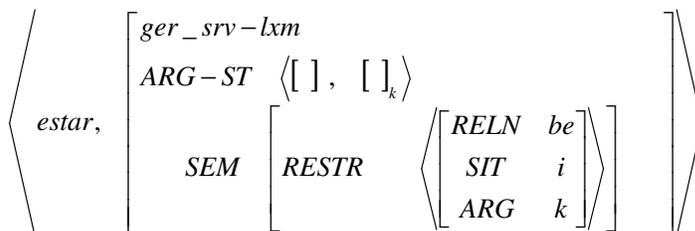


Este tipo solamente agrega la condición que los verbos de este tipo piden como complemento una frase verbal en gerundio (FORM ger). Para los verbos seguidos de un gerundio no es necesario formar el tipo que acepte una preposición intermedia, porque esta situación no se presentó con estos verbos.

5.2.2.1 *Estar, ir, seguir, venir y andar*

Este conjunto de verbos son verbos auxiliares ya que cumplieron con las cinco pruebas sintácticas definidas en el capítulo anterior. En (62) se muestra la entrada léxica del verbo *estar* como auxiliar; las entradas léxicas para los demás verbos se declaran de manera similar:

(61)

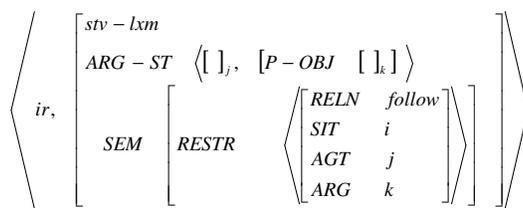


En lo referente a su sentido original los verbos *ir* y *venir* fueron discutidos en la sección 5.2.1 de este capítulo y el verbo *estar* se aborda con mayor detallan en la sección siguiente; por lo que queda definir la entrada léxica para el significado original de los verbos *seguir* y *andar*. Los ejemplos siguientes muestran este comportamiento:

- (62) a. Sigo a Juan (Voz activa)
 b. Juan es seguido por mi (Voz pasiva)
 c. El niño anda (Voz activa)
 d. *Es andado por el niño (voz pasiva)

Como se ve *seguir* se comporta como un verbo transitivo mientras que *andar* como un verbo intransitivo, al no existir la posibilidad de crear la voz pasiva con este verbo. Las entradas léxicas correspondientes son las siguientes:

- (63) a. Seguir



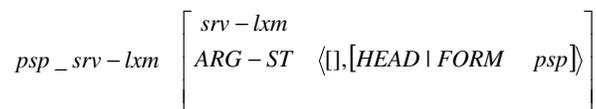
- b. Andar

$\langle andar \quad [iv - lxm] \rangle$

5.2.3 Verbo auxiliar + participio

El siguiente tipo es el encargado de definir al conjunto de verbos auxiliares que van seguidos por un participio:

- (64)

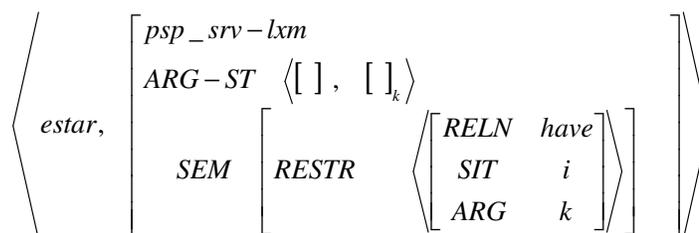


5.2.3.1 Haber

Como se mencionó, las gramáticas tradicionales del español consideran al verbo *haber* como parte de la conjugación del verbo, en los llamados tiempos perfectos; esto tiene consideraciones serias ya que la conjugación es un fenómeno morfológico y no sintáctico. Lo anterior implica que el verbo *haber* en las conjugaciones compuestas tendría que ser generado por una regla de inflexión. Los análisis hechos para considerar al verbo *haber*

como parte de la conjugación son bastante acertados; pero en este trabajo exploramos la posibilidad de definir al verbo *haber* como auxiliar. El factor que más influye en esta consideración es de aspecto técnico y se debe a que la mayoría de ambientes de desarrollo de HPSG poseen una morfología débil ya que es imposible definir una regla que genere las conjugaciones para estos tiempos. Por otro lado, el declarar a *haber* como verbo auxiliar permite generar una semántica que represente la perfectividad tal y como ocurre en las conjugaciones compuestas. La entrada léxica para *haber* como auxiliar es la siguiente:

(65)



5.3 Otros usos del tipo *srv-lxm*

La definición del tipo *srv-lxm* hace posible representar otros fenómenos sintácticos. En particular, son de interés para este trabajo los fenómenos de voz pasiva y verbos atributivos. Estas construcciones juegan un papel muy importante en el análisis hecho en el capítulo anterior al ser alternativos al comportamiento de verbo auxiliar.

5.3.1 Voz pasiva

En una oración en voz pasiva el objeto directo de un verbo juega el papel de sujeto sintáctico, el verbo principal está en la forma no personal de participio y además aparece el verbo *ser*; por ejemplo:

- (66) a. El juguete es entregado por él
 b. Los juguetes son entregados por él

En la oración (67.a) el sujeto es *el juguete*, sin embargo es el objeto directo del verbo *entregar*, es decir, lo que es entregado. El pronombre *él* juega el papel de agente de la acción pero aparece en una posición de complemento precedido por la preposición *por*. En las dos oraciones podemos notar que el verbo *entregar* presenta concordancia con su sujeto al igual que el verbo *ser*.

El modelo de esta construcción se divide dos partes: primero se asegura que el verbo en su forma de participio sea capaz de tomar como sujeto a su objeto directo y a su objeto directo como sujeto. Esto lo logramos con una regla de inflexión para generar las terminaciones: *-ado*, *edo* e *ido* del participio, esta regla se define de la siguiente forma:

(67)

$$\left\langle w, \begin{bmatrix} tr-lxm \\ SYN | HEAD | FORM [FORM inf] \\ ARG-ST \langle \#1 \oplus \#a \rangle \end{bmatrix} \right\rangle \Rightarrow \left\langle F_{pass}(w), \begin{bmatrix} pass-lxm \\ SYN \begin{bmatrix} HEAD [FORM pass] \\ SPR \langle [] \rangle \end{bmatrix} \\ ARG-ST \langle \#a \oplus [P-OBJ \#1] \rangle \end{bmatrix} \right\rangle$$

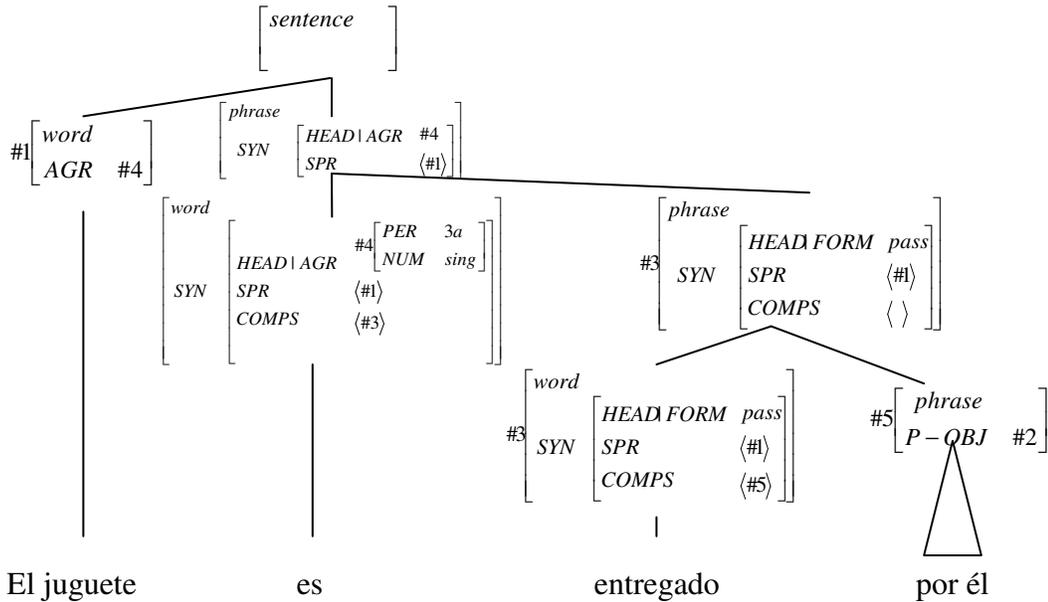
Esta regla crea la voz pasiva de un verbo. Toma al verbo en forma infinitiva cambia su escritura a la de pasivo (F_{pass}); genera una estructura en donde el primer complemento (#1) juega ahora el papel de objeto de una preposición que ocupa el último lugar en la lista de argumentos; el segundo complemento ahora juega el papel de especificador y este nuevo verbo tiene el valor *pass* en su atributo FORM. Con esta regla tenemos a un verbo que solicita como sujeto a su objeto directo, pero como es una forma no personal no puede tomarlo directamente. Para lograr que el verbo pasivo tome a su sujeto nos servirá la definición léxica que a continuación se presenta del verbo *ser*:

(68)

$$\left\langle ser, \begin{bmatrix} pass_srv-lxm \\ ARG-ST \langle [] , [SEM | INDEX \#2]_k \rangle \\ SEM \begin{bmatrix} INDEX \#2 \\ RESTR \langle \rangle \end{bmatrix} \end{bmatrix} \right\rangle$$

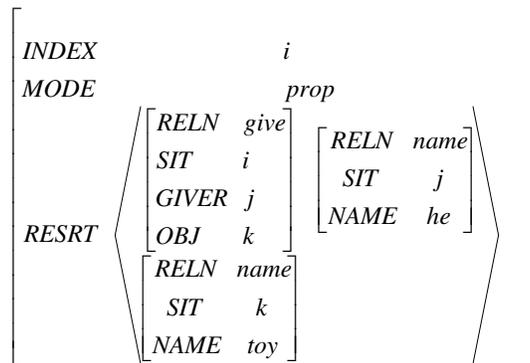
El tipo *pass_srv-lxm* toma como complemento a un verbo cuyo atributo FORM tenga el valor *pass*. Este verbo *ser* nos va a permitir que el verbo en forma pasiva tome su sujeto sintáctico. Además se pide que el índice del verbo pasivo sea el mismo que el del verbo *ser* con esto se garantiza que la restricción principal sea la del verbo pasivo. A continuación se muestra un árbol sintáctico correspondiente a una oración pasiva:

(69)



La semántica que corresponde a la oración es la siguiente:

(70)



Como vemos los roles que juegan cada elemento de la restricción son los mismos de la oración en voz activa. Existe una relación principal de *entregar* (*give*); ésta tiene dos argumentos el que entrega en este caso, *Juan*, y lo que es entregado, *el juguete*.

El haber declarado al verbo *ser* como un verbo *srv-lxm* no indica que es un verbo auxiliar. En primer lugar este verbo tiene concordancia y, en segundo lugar, porque el verbo *ser* requiere de la aplicación de una regla de inflexión que cambie el orden del ARG-ST del verbo; situaciones que nunca se presentaron con un verbo auxiliar del español. En conclusión, un verbo auxiliar de español es del tipo *srv-lxm*; pero no todo verbo *srv-lxm* es un verbo auxiliar del español.

5.3.2 Oraciones atributivas

Como se menciona en el capítulo anterior una gran cantidad de los verbos seguidos de un participio que se consideran como auxiliares cuando en realidad se trata de verbos atributivos o de estado²¹. Un verbo atributivo o de estado relaciona a un elemento con una característica que posee o estado en el que se encuentra. Las siguientes oraciones son ejemplos de oraciones atributivas y de estado:

- (71) a. Juan es alto
 b. María es alta
 c. Los niños están altos
 d. Las niñas están altas

Los elementos que intervienen en esta oración son dos una NP, en (72.a) es *Juan*, y un adjetivo, en la misma oración es *alto*. Como podemos notar existe concordancia entre ambos elementos, además todas estas oraciones se pueden reducir a la siguientes frases nominales, que nos proporcionan un contenido semántico semejante:

- (72) a. Juan alto
 b. María alta
 c. Los niños altos
 d. Las niñas altas

El modelo propuesto en HPSG para capturar a lo adjetivos es el siguiente :

(73)

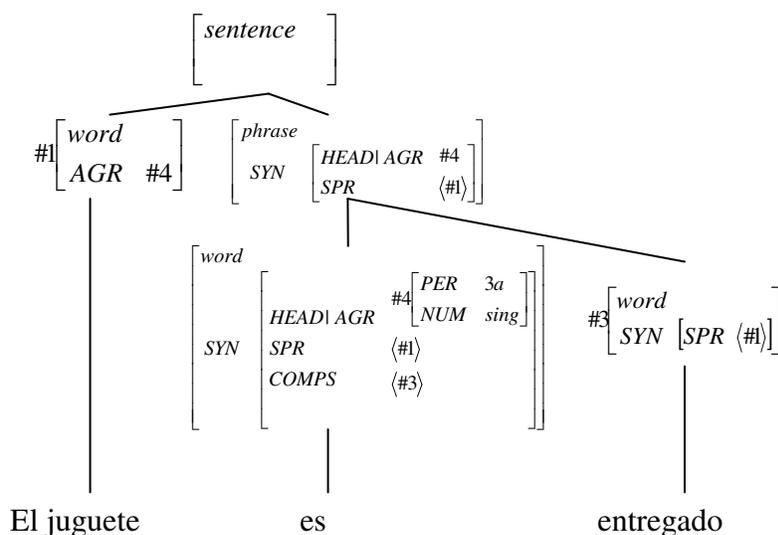
$$adj - lxm \left[\begin{array}{l} SYN \left[\begin{array}{l} HEAD \left[\begin{array}{l} PRED \quad + \\ MOD \quad [noun] \\ AGR \quad \#1 \end{array} \right] \\ SPR \quad \langle [] \rangle \end{array} \right] \\ ARG - ST \quad / \langle NP[AGR \quad \#1] \rangle \end{array} \right]$$

Esta definición brinda versatilidad a los adjetivos. En primer lugar les permite modificar a una palabra del tipo *noun* como ocurre en (73). Pero además, los adjetivos poseen como especificador a una NP; esta última restricción permite a los adjetivos ser tomados por una declaración semejante a la que se hizo del verbo *ser* en (69). En (75) se presenta el árbol sintáctico para la oración *el regalo es mostrado* y su correspondiente semántica en (76).

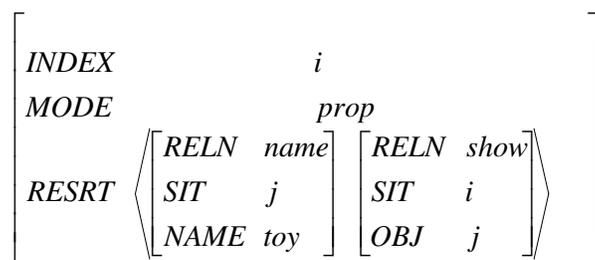
²¹ La distinción se refiere al aspecto perfectivo o imperfectivo que otorgan a la oración.

Como podemos ver esta semántica no es semejante a la de un verbo auxiliar. No hay dos acciones verbales, solamente existe una, la cual caracteriza a un objeto (*juguete*). La aclaración en este punto es la misma que con las oraciones pasivas, el hecho de haber declarado a los verbos atributivos como *srv-lxm* no los convierte en verbos auxiliares, como vimos los elementos que intervienen en su sintaxis son diferentes y la semántica que representan también lo es.

(74)



(75) Semántica



En este capítulo se han presentado un conjunto de tipos y entradas léxicas que integran al sistema de verbos auxiliares del español; en el apéndice C y D se pueden observar en su totalidad el conjunto de tipos y entradas que integran al sistema que se propone en este trabajo. Adicionalmente, en este capítulo se presentaron algunos modelos de fenómenos alternativos al de los verbos auxiliares como los son la voz pasiva y verbos atributivos. En el siguiente capítulo se aborda el diseño y la programación de este modelo en el sistema de desarrollo LKB.

Capítulo 6: Modelo en LKB

En este capítulo se presenta el diseño y programación del sistema de verbos auxiliares del español en el ambiente de desarrollo Linguistic Knowledge Building (LKB)[Copestake, 2000]. LKB es un ambiente de desarrollo para gramáticas basadas en restricciones, en este trabajo, basadas en el formalismo HPSG. LKB permite la declaración de estructuras atributo valor de manera semejante a las que se utilizan en HPSG, pero en un formato de caracteres de texto. Aunque la programación en LKB de una especificación en HPSG debiera ser sencilla, las restricciones de LKB en la morfología, la utilización exhaustiva de las estructuras atributo valor y la forma como se encuentra definida la unificación, hacen que la programación sea un tanto complicada. Sin embargo, como se puede apreciar en las siguientes secciones, se logró implementar al sistema de verbos auxiliares en sistema de desarrollo LKB, a través de la definición de las entradas léxicas y tipos presentados en el capítulo anterior. A continuación se explican algunos de los elementos que integran a LKB.

El sistema LKB reproduce los elementos que integran a una gramática en HPSG por medio de un sistema de archivos. Los archivos principales son los correspondientes a la jerarquía de tipos, a las reglas gramaticales, reglas morfológicas y entradas léxicas. En el archivo de la jerarquía de tipos, llamado archivos de tipos, se definen a todos los tipos que intervienen en la gramática y las relaciones de tipo padre e hijo que poseen. En el archivo de tipos se han declarado todos aquellos tipos que integran el sistema de verbos auxiliares del español.

Las reglas HCR, HSR, HMR y SPDR se encuentran declaradas en el archivo de reglas gramaticales. Un punto muy importante en LKB es que no se puede generalizar tanto como se hacía en HPSG. En particular, con las reglas gramaticales es necesario definir explícitamente a los elementos que intervienen. Por ejemplo, de la regla HCR en LKB se cuentan con tres versiones, dependiendo del número de elementos que intervienen, desde ninguno hasta dos. Con las reglas HSR y HMR sucede algo similar.

En el archivo de reglas morfológicas se encuentran las reglas que cambian la ortografía de una palabra. Por ejemplo, para el sistema de verbos auxiliares es indispensable contar con las reglas de conjugación y las reglas de las formas no personales del español. Las pocas opciones para modificar la ortografía de una palabra hacen que un

archivo de excepciones crezca de manera desbordada. A este archivo se le tuvieron que incluir la mayoría de las conjugaciones de los verbos auxiliares.

Por último, el archivo de entradas léxicas contiene las descripciones de todas las palabras que la gramática puede reconocer. En este archivo se especifica la ortografía de la palabra, a la cual se le conoce como símbolo, su tipo y alguna restricción adicional. A este archivo se le agregaron las descripciones de los verbos auxiliares y las entradas con las que comparten el símbolo pero que tienen diferente tipo (i.e., verbos de control, atributivos, transitivos e intransitivos)

En dos sentidos la implementación aquí presentada no refleja al sistema de verbos auxiliares. En primer lugar, como fue mencionado con anterioridad, la morfología de LKB obligó a declarar a los verbos auxiliares con una morfología irregular y, aún más grave, a utilizar algunos trucos, como cambiar la ortografía de una palabra, para representar restricciones del tiempo en que aparecen algunos verbos auxiliares. En segundo lugar, el modelo para capturar el anidamiento de los verbos auxiliares no es lo suficientemente general para que no sea dependiente de las entradas léxicas. A pesar de estos contratiempos se logró trabajar con LKB y así implementar el sistema de verbos auxiliares del español bajo el ambiente.

En las siguientes secciones se describe con mayor detalle los elementos que integran a LKB y el resumen de la declaración de los elementos que integran el sistema de verbos auxiliares del español. En la sección 6.1 se exponen los archivos que integran a LKB, su funcionamiento y su relación con HPSG. En la sección 6.2 se presenta la implementación del sistema de verbos auxiliares; en la primera parte se define el conjunto de tipos que integran el sistema y en la segunda se presentan algunas de las entradas léxicas que componen el sistema. En la sección 6.3 se presenta la implementación de los fenómenos de verbos unipersonales, de la preposición de control de sujeto (*scp-lxm*), de anidamiento de verbos de complemento de frase, de la voz pasiva y de los verbos atributivos, los cuales fueron de suma importancia en el análisis del comportamiento de los verbos auxiliares del español. Tanto en la sección 6.2 y 6.3 los ejemplos de las entradas léxicas van a acompañados de la salida sintáctica y semántica de LKB después del análisis sintáctico de las oraciones.

En este capítulo se implementa el modelo de los verbos auxiliares del español propuesto en el capítulo anterior y bajo el ambiente de desarrollo de LKB que corresponde al formalismo de HPSG presentado en el capítulo tres de esta tesis. Los ejemplos de las salidas de LKB, tanto sintácticas como semánticas, que aquí se presentan corresponden a la validación computacional del modelo del sistema de verbos auxiliares del español y serán tomados como referencia para elaborar las conclusiones de los resultados esperados en el siguiente capítulo.

6.1 Sistema de Archivos de LKB

El sistema de archivos de LKB son el conjunto de archivos en los cuales se encuentra definida la gramática del un lenguaje en particular. El sistema de archivos está confirmado por los siguientes archivos:

- Sistema de tipos
- Entradas léxicas
- Reglas léxicas
- Reglas gramaticales
- Reglas morfológicas

A continuación se describe la función que cumple cada uno de estos archivos

6.1.1 Archivos del sistema de tipos

Los archivos que integran al sistema de tipos contienen la descripción de la jerarquía de tipos de HPSG; en estos archivos se describen las diferentes estructuras atributo valor y su relación entre sí. La sintaxis de este archivo pretende ser lo más cercana a la utilizada a las descripciones de matrices atributo valor (AVM) de HPSG, pero con un formato de texto. Por ejemplo la definición del tipo *verb-lxm* que se presentó en el capítulo anterior es de la siguiente forma (Sección 5.1.1):

(1)

```
verb-lxm := infl-lxm &
[ SYN [ HEAD verb,
  SPR <[]>,
  ARG-ST </1 np-phrase, ... >,
  SEM [ MODE prop ] ].
```

Esta declaración en LKB define al tipo *verb-lxm*, cuyo padre, como se indica , es el tipo *infl-lxm*. En seguida del padre se presenta la descripción de la estructura atributo valor que especifica a las restricciones que debe cumplir el tipo *verb-lxm*; a estos dos elementos se le separa por el símbolo $\&$, el cual representa a la operación de unificación en LKB, en este caso entre el tipo *infl-lxm* y la descripción que sigue. En el apéndice C se muestra la gramática que describe el formato de los archivos utilizados por LKB.

La descripciones que integran a los archivos del sistema de tipos de LKB además de cumplir con las restricciones que impone HPSG a la jerarquía de tipos deben satisfacer las siguientes condiciones:

- Todos los tipos deben estar definidos: ningún tipo puede ser especificado si su padre no es especificado; tampoco se podrá definir un identificador como su propio padre.
- El orden en que se definen los tipos no es de importancia para LKB. Sólo en el caso de definir dos tipos con el mismo identificador la última definición es la que se toma en cuenta.
- Solamente un tipo contendrá a toda la jerarquía de tipos, generalmente se trata del identificador ***top***, por lo que es un error definir a un tipo sin ningún padre.
- Herencia consistente: las restricciones que son especificadas por el padre de un tipo no deberán ser violadas por sus hijos.
- Máxima introducción de atributos: Un atributo deberá ser introducido en el tipo de jerarquía más alta, para que puedan ser utilizados por sus descendientes. No es permitido introducir un atributo en dos tipos incompatibles.
- Estructuras finitas: dado que todos los tipos se expanden para su utilización, no es valido crear estructuras infinitas, ya que no se podrá obtener la máxima expansión de éstas.

El archivo principal del sistema de tipos es: `types-spanish.tdl`; sin embargo, existen otros archivos que contienen la declaración de tipos más específicos como aquellos relacionados con la reglas gramaticales o léxicas.

6.1.2 Archivo de entradas léxicas

El archivo de entradas léxicas concentra al lexicón, es decir al conjunto de palabras que integran nuestra gramática. Las entradas léxicas consisten en estructuras de atributo valor que definen la relación entre la cadena de caracteres que identifican a una palabra, el símbolo, y su descripción lingüística; esta descripción pretende ser lo más básica y abstracta posible. En LKB el atributo donde se guarda el símbolo ORTH y está al mismo nivel que los atributos SYN y SEM. El archivo de entradas léxicas presentan la misma sintaxis que el sistema de tipos y debe satisfacer las mismas condiciones que se le impusieron a las descripciones que integran estos archivos. El archivo utilizado para guardar las entradas léxicas es: `lexicon-spanish.tdl`. El siguiente ejemplo muestra cómo luce una entrada léxica en LKB:

(2)

```
juan_1 := pn-lxm &
[ ORTH <! "juan" !>,
  SYN [ HEAD [ AGR [ GEND masc] ] ],
  SEM [ RESTR <! [ NAME juan ] !> ] ].
```

Como se puede apreciar la forma en como se declaraban las entradas léxicas en HPSG y en LKB varían. En LKB la forma de representar una entrada léxica es a través de estructuras atributo valor, situación que no se presentaba en HPSG (sección 3.1.6). El ejemplo corresponde a la entrada léxica para la palabra *juan*, la cual se relaciona con el identificador `juan_1` que define como se conoce a esta descripción en LKB. Esta entrada léxica declara a la palabra como tipo *pn-lxm*, que corresponde a los nombres propios. Además se agregan aquellas características propias de la palabra; en particular, en este ejemplo se define el género y parte de su restricción semántica.

6.1.3 Archivos de reglas léxicas y morfología

El formalismo HPSG hace una distinción entre las reglas léxicas que no modifican el tipo de un lexema y aquellas que si lo hacen. A las primeras las denomina reglas derivacionales, mientras que a las segundas las llama de inflexión. En LKB la distinción que se hace entre las reglas léxicas se refiere a si son capaces de modificar la ortografía de la palabra o no. LKB utiliza dos archivos, uno de reglas morfológicas, para aquellas reglas que cambian la ortografía de las palabras, y el otro denominado solamente como de reglas léxicas.

El archivo de reglas léxicas se encarga de modificar la información lingüística de una entrada léxica sin tocar su ortografía de la palabra. Las reglas aquí declaradas cambian la categoría gramatical o convierten estructuras del tipo *lexeme* (sin valores en SPR y COMPS) al tipo *word* (con valores en SPR y COMPS), es decir aquí se encuentra implementado el principio de Argument Realization (Sección 3.1.6).

Con el fin de lograr modularidad en los archivos del sistema de tipos se dividen los tipos utilizados por las reglas léxicas y las reglas morfológicas en un archivo separado; pero estos seguirán cargándose como parte del sistema de tipos; este archivo es: `lexical-types-spanish.tdl`, y el archivo que contiene a todas las reglas léxicas es: `lrules-spanish.tdl`.

Las reglas morfológicas además de modificar la información lingüística de una palabra, pueden cambiar la ortografía de estas. Un ejemplo de una regla morfológica es la siguiente:

(3)

```
3rd-sing-pres-verb-spanish_infl_rule :=
%suffix (ar a) (er ) (ir )
pres-verb-lex-rule-spanish &
[ NEEDS-AFFIX true,
  ARGS [ FIRST [ AFFIX 3rd-sing-verb-spanish ]]].
```

Esta regla de inflexión se encarga de crear la conjugación de la tercera persona del singular para los verbos regulares. Como podemos notar esta regla no se parece en mucho a las que se definieron en HPSG (Sección 3.1.7); en primer lugar toda la regla es una estructura de atributos y no se distingue en ninguna parte a la madre e hijas; sin embargo sí están contenidas ahí. El atributo `ARGS` describe la estructura a la que se le aplica la regla (la hija de la regla) y la estructura en la cual se encuentra contenida el atributo `ARGS` define la estructura resultante (la madre de la regla). Aunque esta estructura se parece a las antes definidas para LKB, existen una diferencia significativa: la aparición de la opción `%suffix`. Esta opción ayuda a crear la nueva ortografía de la palabra; por ejemplo, la opción `ar` representa al sufijo a buscar en la entrada léxica y “a” representa el sufijo por cual cambiarlo, es decir, `ar` representa terminación “ar” de los verbos, como *cantar*; mientras que “a” representa por qué sustituir al sufijo para que nos quede la conjugación de la tercera persona del singular en presente, en este caso *canta*. La pareja de sufijos por buscar y cambiar se encierra entre paréntesis y puede haber tantos casos como sean necesarios; en

el ejemplo anterior se incluyen los casos para las terminaciones *er* e *ir*. El atributo `NEEDS-AFFIX` con el valor `true`, especifica a LKB que esta definición hace uso de la opciones de sufijo o prefijo. El conjunto de reglas morfológicas utilizadas en este trabajo se encuentran definidas en el archivo: `inflr-spanish`; los tipos a los que se hacen referencia en este archivo en `lexical-types-spanish.tdl`, el mismo utilizado por las reglas léxicas.

Como podemos observar las posibilidades del módulo de morfológico de LKB son muy pobres. Para capturar la morfología de un lenguaje es necesario un número mayor de transformaciones que las ofrecidas por LKB; esto limita en mucho el trabajo que se puede realizar con las conjugaciones e inflexiones y conlleva a una carga excesiva en el número de irregularidades para las reglas de inflexión.

6.1.4 Archivo de irregularidades

Una forma de capturar aquellas inflexiones que la morfología de LKB no puede reproducir es a través de un archivo de irregularidades; éste consiste en una lista de las palabras que no cumplen con una regla morfológica. Este archivo está formado por la lista de palabras que debería resultar de aplicar la regla léxica, la regla que se le debió haber aplicado y la ortografía de la palabra que la representa en el lexicon. A continuación se representa la lista de conjugaciones irregulares del verbo *ser* para el tiempo del presente:

(4)

```
soy    1ST-SING-PRES-VERB-SPANISH  ser
eres   2ND-SING-PRES-VERB-SPANISH  ser
es     3RD-SING-PRES-VERB-SPANISH  ser
somos  1ST-PLUR-PRES-VERB-SPANISH  ser
sois   2ND-PLUR-PRES-VERB-SPANISH  ser
son    3RD-PLUR-PRES-VERB-SPANISH  ser
```

El conjunto de excepciones utilizadas en este trabajo están incluidas en el archivo `irreg-both.tdl`.

6.1.5 Archivo de reglas gramaticales

Las reglas gramaticales se expresan en LKB a través de estructuras atributo valor. Esta representación permite utilizar a la unificación como el mecanismo de combinación de los cabezas de frase con sus complementos y especificadores, eliminando cualquier

interpretación procedural del formalismo gramatical. [Sag y Wasow]²². Para lograr esto se han creado nuevos atributos para las reglas gramaticales.

En la sección de reglas léxicas ya se había encontrado un nuevo atributo, el cual era *ARGS*, en este se colocaban las entradas léxicas y de ahí se toman sus valores para cambiar su ortografía (4). En las reglas gramaticales se procede de manera similar; se agregan nuevos atributos para identificar a los elementos hijas y madre de la regla. Estos nuevos elementos son *H* para el representar a la estructura hija cabeza y *NH* para las hijas no cabeza de regla. La estructura resultante, es decir la madre, está contenida dentro de la misma regla, como se aprecia en el siguiente ejemplo:

(5)

```
head-complement-rule-1 := binary-COMPS-rule &
[ SYN [ SPR #spr,
  COMPS < > ],
  H [ SYN [ SPR #spr,
    COMPS < #1 > ] ],
  NH1 #1 ].
```

Esta estructura representa a la regla de HCR cuando sólo tiene un complemento; el sistema LKB necesita ser más específico, por lo que no podemos generalizar la regla como lo hacíamos en HPSG. En LKB se necesitan tantas reglas para representar a la regla HCR como el número mayor de complementos que acepte nuestra gramática (generalmente dos) más el caso de cero complementos. En esta regla, la parte de abajo a partir del atributo *H* puede ser considerado como la parte derecha de la regla en HPSG, es decir las estructuras hijas, y la parte de arriba como la madre de esta, es decir la estructura resultante. El atributo *H* corresponde a la cabeza de las estructuras hijas y *NH1* a la estructura no cabeza de la regla; en particular, para esta regla representa el complemento. En la regla podemos ver que la información contenida en el *COMPS* de la hija cabeza debe unificar con la estructura de la hija no cabeza, que es la condición que impone HCR. Además, el atributo *SPR* de la hija cabeza y madre se unifican a través de la etiqueta *#spr*, esto para cumplir con el principio de valencia (VP). Una repercusión de este modelo es que toda la historia de aplicación de las reglas estará contenida en la estructura atributo valor resultante.

²² P. 355

El archivo utilizado en la programación del modelo de verbos auxiliares es: `grules-spanish.tdl`. Al igual que ocurrió con las reglas léxicas, las reglas gramaticales también usan un archivo para definir los tipos que utilizan, esta definición se hace en el archivo: `grule-types-spanish.tdl`.

6.2 Implementación del sistema de verbos auxiliares

La implementación del sistema de verbos auxiliares del español se llevo a cabo en tres etapas. La primera consistió en declarar los tipos de datos que capturan el modelo de los verbos auxiliares en HPSG. La segunda, en definir las entradas léxicas correspondientes. Para la última etapa se realizaron un conjunto de pruebas para validar el modelo. En esta sección se aborda ampliamente la etapa uno, ya que es el núcleo de la definición de los verbos auxiliares; cabe mencionar que algunos de los tipos que se declaran en esta etapa tienen el objetivo de modelar el sentido original de los de los verbos auxiliares. La etapa dos será ejemplificada con las entradas léxicas más representativas; el conjunto completo de entradas léxicas se encuentran definidas en el Apéndice E. Adicionalmente se muestran algunas de las pruebas realizadas con el sistema de verbos y sus resultado.

6.2.1 Definición de tipos

El tipo esencial para capturar el fenómeno de verbos auxiliares del español es el de *srv-lxm*. En la gramática que nos sirvió de base [Simoës, 1999] este tipo se declara como sigue:

(6)

```
;;; Declaracion de srv-lxm
srv-lxm-spanish := verb-lxm-spanish & srv-sem &
[ ARG-ST < #1 & [SYN [HEAD [ CASE case-cat]]] ,
[ SYN [SPR < #1 > ]] > ].
```

Esta declaración tiene una restricción que garantiza que el primer argumento de un verbo de este tipo sea el mismo que el especificador de su segundo argumento; tal y como se había definido este tipo en el capítulo anterior. Adicionalmente, esta declaración define al tipo `srv-lxm-spanish` como hijo del tipo `verb-lxm-spanish`, lo que hace posible que herede el comportamiento de un verbo. Aprovechando una de las ventajas que ofrece la jerarquía de tipo de HPSH, el tipo `srv-lxm-spanish` se define también como hijo del tipo `srv-sem`, lo que garantiza que los argumentos de la semántica de un verbo de este tipo se

interrelacionen; durante el análisis de HPSG este paso se daba sobre entendido al indexar el complemento. El tipo *srv-sem* se define de la siguiente forma:

(7)

```

srv-sem := verb-sem &
[ SEM [ RESTR <! srv_pred & [ ARG #3 ] !> ],
  ARG-ST < [], [ SEM [INDEX #3 ] ] > ].

```

En esta definición se observa que el índice del segundo argumento debe coincidir con el valor del atributo ARG. Este índice en un *srv-lxm* representa a la acción del verbo en forma no personal. Las restricciones del tipo *srv_pred* que aparece en esta descripción contiene a los atributos adecuados para un *srv-lxm*; en particular este tipo de restricción no contiene un atributo que represente a un agente, ya que el tipo *srv-lxm* no requiere de uno.

El tipo *pp_srv-lxm*, a diferencia del tipo *srv-lxm*, declara a su segundo argumento antecedido por una preposición, con el fin de definir aquellos verbos auxiliares que son seguidos por una preposición antes del verbo en forma no personal. La declaración de este tipo es la siguiente:

(8)

```

;; srv-lxm con preposicion
pp_srv-lxm-spanish := verb-lxm-spanish & pp_srv-sem &
[ ARG-ST < #1 & [SYN [HEAD [ CASE case-cat]]],
  [SYN PP-spanish & [ HEAD [ P-OBJ <[SYN [SPR < #1
                                     > ]]> ]]] > ].

```

Como vemos el segundo complemento debe corresponder a una sintaxis de frase preposicional por el tipo *PP-spanish*; además el especificador del objeto de la preposición es el mismo que el verbo declarado como de este tipo. Por otro lado, en LKB es necesario definir un conjunto de tipos *srv-lxm* según el complemento que presentan, con el fin de simplificar la declaración de las entradas léxicas. Estos tipos se definen de la siguiente forma:

(9)

```

;; Forma no personal en infinitivo
inf_srv-lxm-spanish := srv-lxm-spanish &
[ SYN [ HEAD verb-spanish & [ AUX true ] ],
  ARG-ST < [ ], [ SYN [ HEAD verb-spanish & [FORM inf ] ] ] > ].

;; Forma no personal en gerundio
ger_srv-lxm-spanish := srv-lxm-spanish &
[ ARG-ST < [ ], [ SYN [ HEAD verb-spanish & [FORM ger ] ] ] > ].

```

```
;; Forma no personal en participio
perf_srv-lxm-spanish := srv-lxm-spanish &
[ ARG-ST < [ ], [ SYN [ HEAD verb-spanish & [FORM psp ] ] ] > ].
```

Para los verbos que van seguidos de una preposición es necesario generalizar dos cosas, el tipo de preposición y la forma del verbo no personal. Esto se logra con las las siguientes declaraciones de tipos:

(10)

```
;; Forma no personal en infinitivo y preposición
pp_inf_srv-lxm-spanish := pp_srv-lxm-spanish &
[ ARG-ST < [ ], [ SYN [ HEAD [P-OBJ <[ SYN [ HEAD
verb-spanish & [ FORM inf ] ] ] ] ] ] > ]].
```

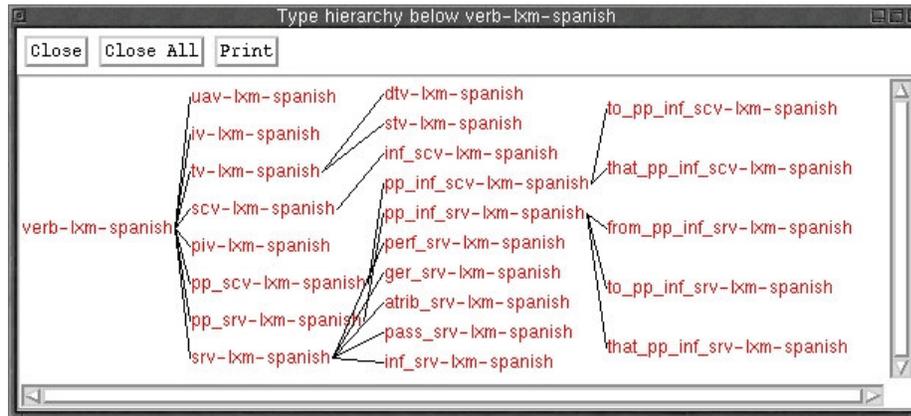
```
;; Forma no personal en infinitivo y preposición "a"
to_pp_srv-lxm-spanish := pp_inf_srv-lxm-spanish &
[ ARG-ST < [ ], [ SYN [ HEAD [ FORM f_to ]]] ] > ].
```

```
;; Forma no personal en infinitivo y preposición "de"
from_pp_srv-lxm-spanish := pp_inf_srv-lxm-spanish &
[ ARG-ST < [ ], [ SYN [ HEAD [ FORM f_from ]]] ] > ].
```

```
;; Forma no personal en infinitivo y preposición "que"
that_pp_srv-lxm-spanish := pp_inf_srv-lxm-spanish &
[ ARG-ST < [ ], [ SYN [ HEAD [ FORM f_that ]]] ] > ].
```

Solo se citan las definiciones para verbos seguidos de un infinitivo porque en los verbos seguidos de un gerundio o participio no intervienen preposiciones. Con estas definiciones de tipos es posible modelar el sistema de los verbos auxiliares del español. Sin embargo, aún falta modelar aquellos comportamientos sintácticos adicionales que presentan los verbos. Para esto la gramática base [Simoës, 1999] cuenta de manera consistente con el conjunto de verbos transitivos e intransitivos; adicionalmente se han agregado definiciones para verbos de control, las cuales son semejantes a las presentadas para los verbos auxiliares, verbos atributivos y la voz pasiva; estas últimas dos son descritas en la sección 6.3. La parte de jerarquía tipos correspondiente a los verbos del español presentada en este trabajo, tal y como la interpreta LKB, se presenta en el siguiente esquema:

(11)



En esta imagen se encuentra representada los distintos tipos de verbos que intervienen directamente con los verbos auxiliares; las restricciones a las cuales se encuentran unidas se pueden consultar en el apéndice D. A continuación, se presentan las entradas léxicas de manera resumida para los verbos auxiliares.

6.2.2 El verbo *poder*

Con el conjunto de tipos que se han declarado es posible realizar las primeras pruebas que muestren la forma en que está funcionando el sistema de verbos auxiliares. En particular nos interesa comprobar el comportamiento del verbo *poder*, ya que es el verbo que originó la necesidad de que se hiciera este análisis en el proyecto DIME. Para esto se trabaja con las oraciones ejemplo de este verbo analizadas en el capítulo anterior:

- (12) a. Puedes con las matemáticas
b. Puedes mostrar el catálogo

6.2.2.1 Entradas léxicas para ejemplo

Antes de poder parsear las oraciones necesitamos definir las entradas léxicas de cada palabra y sus comportamientos adicionales en caso de tenerlos. A continuación presentamos como se definieron estas entradas léxicas:

(13)

```
;; Poder: (control) (posible que muestres el catalogo)
poder_1 := inf_scv-lxm-spanish &
[ ORTH <! "poder" !>,
  SEM [ RESTR <! [ RELN r_can ] !> ] ].
```

```

;; Poder: (auxiliar) (tienes capacidad que muestres el catalogo)
poder_2 := inf_srv-lxm-spanish &
[ ORTH <! "poder" !>,
  SEM [ RESTR <! [ RELN r_possible ] !> ] ].

;;Poder: (Intransitivo) (tienes capacidad de estudiar matematicas)
poder_3 := piv-lxm-spanish &
[ ORTH <! "poder" !>,
  SYN [ HEAD [ MOD-VERB [ PODER true ] ] ],
  SEM [ RESTR <! [ RELN r_can ] !> ] ].

;; Mostrar: Transitivo con un solo complemento.
mostrar_2 := stv-lxm-spanish &
[ ORTH <! "mostrar" !>,
  SEM [ RESTR <! [ RELN r_show ] !> ] ].

;; El: Artículo él
el_1 := def_det-lxm-spanish &
[ ORTH <! "el" !>,
  SYN [ HEAD [ AGR [ GEND masc,
                    NUM sing ] ] ],
  SEM [ RESTR <! [ RELN r_the ] !> ] ].

;; Catálogo
catalogo_1 := cn-lxm-spanish &
[ ORTH <! "catalogo" !>,
  SYN [ HEAD [ AGR [ GEND masc ] ] ],
  SEM [ RESTR <! [ RELN r_catalog ] !> ] ].

```

Como se observa se han declarado tres verbos *poder*, cada entrada léxica corresponde a los comportamientos que presenta el verbo; así tenemos uno intransitivo (*piv-lxm-spanish*), uno de control (*inf_scv-lxm-spanish*) y otro de auxiliar (*inf_srv-lxm-spanish*). Existe una entrada para el verbo *entregar* que lo declara como transitivo (*dtv-lxm-spanish*). Antes de continuar con este ejemplo es importante definir como se enfrenta el problema de sujeto omitido del español.

6.2.2.2 Sujeto omitido del español

Para resolver el fenómeno de sujeto omitido la gramática Simoes, la cual sirvió de base para esta implementación, tiene la siguiente regla gramatical *Subject-Pro Drop Rule* (SPDR); el término *Pro* se refiere generalmente a sujetos sintácticos que no aparecen en la oración [Simoes, 1999]. Esta regla permite saturar al especificador cuando este no aparezca en la oración. Ésta regla se define de la siguiente forma:

(14)

```

subject_pro-drop-rule-spanish := unary-grule-headed &
[ SYN [ SPR < > ],

```

```

SEM [ RESTR [ LIST < #sprfirst . #restr >,
              LAST #semplast ] ],
H   [ SYN [ HEAD verb-spanish,
          SPR < [ SEM [ RESTR [ LIST [ FIRST #sprfirst,
                                  REST #rest ],
                                  LAST #rest ] ] ] >,

          COMPS < > ],
      SEM [ RESTR [ LIST #restr,
                    LAST #semplast ] ] ] ].

```

Esta regla sólo tiene un argumento (H), la cual es marcada además como HEAD. Este argumento debe cumplir con las siguientes restricciones:

- El atributo HEAD debe ser del tipo `verb-spanish`, es decir, debe tratarse de una frase verbal.
- Su complemento debe estar vacío (`COMPS <>`), es decir, debe haber saturado sus complementos.
- Su atributo SPR debe tener una restricción semántica, en este caso esta marcado con la etiqueta `sprfirst`.

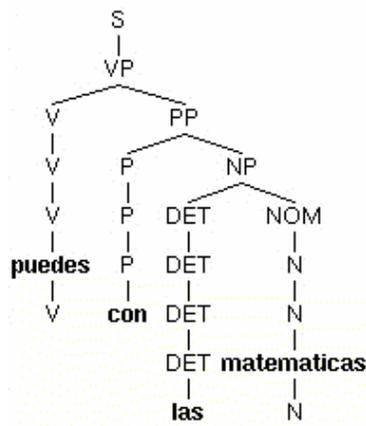
Una vez que una estructura cumple con estas restricciones se procede a crear a la madre de la siguiente forma:

- El SPR es saturado (`SPR <>`)
- Se le agrega a la restricción semántica de la oración la restricción del especificador, esto a través de la operación de concatenación (`#sprfirst.#restr`)

6.2.2.3 Poder

A continuación presentamos el árbol sintáctico que resulta de la oración (12.a):

(15)



Como se ve el resultado es similar al del análisis del capítulo anterior. Sin embargo, en este árbol no representan las estructuras intermedias porque son gigantescas debido a que todo está expandido, y además, porque hay una cantidad mayor de nodos porque aparecen aquí tanto reglas léxicas como gramaticales. Lo importante es obtener una semántica similar al del capítulo anterior, la semántica que se presenta a continuación es la resultante del proceso de parseado²³:

(16)

```

sem-struct & [ MODE prop,
              INDEX #0,
              RESTR < [ RELN    r_hearer,
                       SIT     sit-index,
                       INSTANCE #1],
                       [ RELN    r_t-overlap,
                       SIT     sit-index,
                       ARG1    #0,
                       ARG2    now],
                       [ RELN    r_can,
                       SIT     #0,
                       THM     #2,
                       ACT     #1 ],
                       [ RELN    r_with,
                       SIT     sit-index,
                       INSTANCE ref-index ],
                       [ RELN    r_the,
                       SIT     sit-index,
                       INSTANCE #2],
                       [ RELN    r_math,
                       SIT     sit-index,
                       INSTANCE #2 ]>]

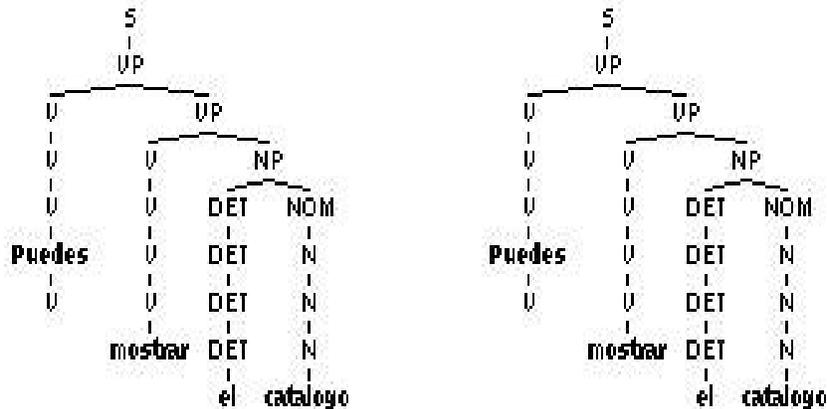
```

Aunque en esta estructura aparecen representados cada elemento de la oración es importante observar que los índices de la relación *can* se satisfacen como habíamos establecido. Debemos recordar que bajo este significado *poder* necesita resolver dos argumentos en su semántica; el capaz, agente (ACT²⁴), y el atributo que marca la cosa de la cual uno es capaz, que corresponde al atributo THM. Ambos atributos están relacionados con los elementos de la oración que habíamos previsto en el análisis de HPSG.

La ambigüedad de la oración (14.b), se refleja en LKB al momento de parsearla y obtener como resultado dos árboles sintácticos:

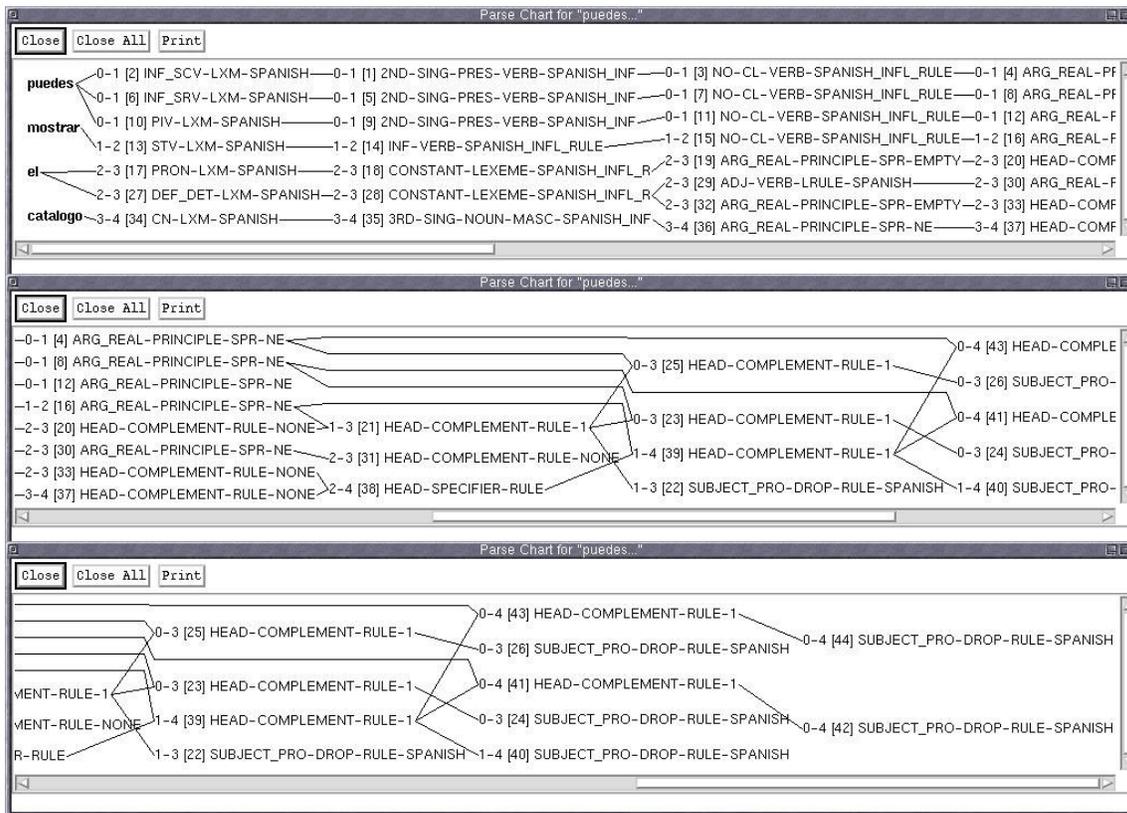
²³ Las semánticas que se presentan aquí han sido modificadas para una mejor comprensión.

(17)



Como observamos la diferencia entre ambos árboles no puede ser apreciada; sin embargo son estructuras diferentes. En la siguiente figura que representa al proceso de parseo y el cual esta dividido en tres partes por cuestiones de tamaño, se aprecia como se construyen las estructuras finales partiendo de dos posibles entradas léxicas para *poder*, una como un *svr-lxm* y otra como un *scv-lxm*.

(18)



²⁴ El atributo ACT corresponde al atributos AGT.

Este esquema representan a todas la estructuras que se crearon durante el proceso de parseo. Cada nodo proviene de la aplicación de una regla. Solamente dos estructuras de las presentadas en este esquema agrupan a toda la oración, es decir, son estructuras sintácticas que representan a la oración. Una de esa estructuras corresponde al sentido de control del verbo *poder* y la otra al de verbo auxiliar. Las semánticas que tienen estas estructuras son las siguientes:

(19) a. Semántica de control

```

sem-struct & [ MODE      prop,
                INDEX    #0,
                RESTR    < [ RELN      r_hearer,
                            SIT        sit-index,
                            INSTANCE   #1],
                            [ RELN      r_t-overlap,
                              SIT        sit-index,
                              ARG1      #0,
                              ARG2      now],
                            [ RELN      r_can,
                              SIT        #0,
                              ARG        #2,
                              ACT        #1 ],
                            [ RELN      r_show,
                              SIT        #2,
                              UND        #3,
                              ACT        #1 ],
                            [ RELN      r_the,
                              SIT        sit-index,
                              INSTANCE   #3 ],
                            [ RELN      r_catalog,
                              SIT        sit-index,
                              INSTANCE   #3 ]>]

```

b. Semántica de auxiliar

```

sem-struct & [ MODE      prop,
                INDEX    #0,
                RESTR    < [ RELN      r_hearer,
                            SIT        sit-index,
                            INSTANCE   #1],
                            [ RELN      r_t-overlap,
                              SIT        sit-index,
                              ARG1      #0,
                              ARG2      now],
                            [ RELN      r_possible,
                              SIT        #0,
                              ARG        #2 ],
                            [ RELN      r_show,
                              SIT        #2,
                              UND        #3,
                              ACT        #1 ],
                            [ RELN      r_the,
                              SIT        sit-index,
                              INSTANCE   #3 ],

```

```
[ RELN      r_catalog,
  SIT       sit-index,
  INSTANCE  #3 ]>]
```

Como se aprecia, ambas semánticas cumplen con el análisis que se desarrolló en el capítulo anterior. La diferencia fundamental está en la restricción del verbo *poder*. Una de estas alude a la relación de posibilidad (*possibility*) con un solo argumento; mientras que la otra a la relación de capacidad (*can*) con dos argumentos. La relación de posibilidad solamente tiene un argumento (ARG) y corresponde a un comportamiento como auxiliar. La relación de capacidad tiene dos argumentos, donde uno juega el rol de agente (ACT) y el otro corresponde a la situación de la que alguien es capaz (ARG).

6.2.3 Otros verbos auxiliares

El ejemplo del verbo *poder* nos da una idea del funcionamiento del sistema de verbos auxiliares. Sin embargo todavía quedan otros verbos por verificar. Como la mayoría sigue un comportamiento similar se ha escogido a tres verbos más para ejemplificar como funciona el modelo de los verbos auxiliares en el sistema LKB. Los verbos escogidos son *ir a*, para los verbos seguidos de una preposición; *estar* para los verbos con un complementos en gerundio y *haber* como verbo seguido de participio. Todos estos ejemplos son acompañados de nuestra oración ejemplo que hemos venido utilizando en este trabajo *entregar el juguete a Juan*.

6.2.3.1 *Ir a*

Las entradas léxicas para el verbo *ir* son las siguientes:

(20)

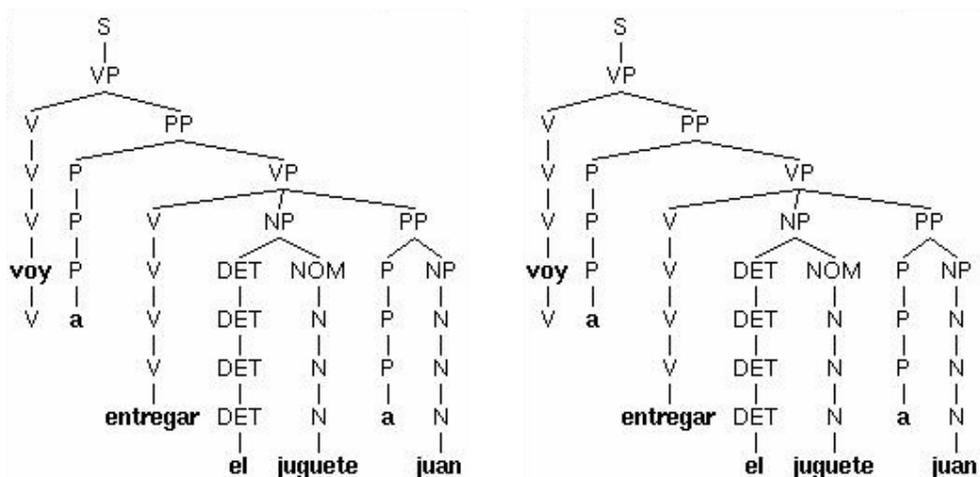
```
;; Ir : Voy a comer (auxiliar) (En el futuro comeré)
ir_2 := to_pp_inf_srv-lxm-spanish &
[ ORTH <! "ir_srv" !>,
  SEM [ RESTR <! [ RELN r_future] !> ] ].

;; Ir : Voy a comer (control) (Me dirijo para que yo coma)
ir_3 := to_pp_inf_scv-lxm-spanish &
[ ORTH <! "ir" !>,
  SEM [ RESTR <! [ RELN r_go] !> ] ].
```

Estas entradas léxicas nos permiten persear la oración: *voy a entregar el juguete a Juan*.

Los árboles resultantes son:

(21)



Los árboles son dos, porque uno se refiere a la acción de desplazarse con el propósito de *entregar el regalo a Juan*, bajo el comportamiento de verbo de control, y el otro a una acción que se efectuara en el futuro, bajo el comportamiento de auxiliar. Las diferencias semánticas se pueden apreciar en la siguiente semánticas.

(22) a. Semántica de control

```
sem-struct & [ MODE prop,
                INDEX #0,
                RESTR < [ RELN r_speaker,
                        SIT sit-index
                        INSTANCE #1],
                        [ RELN r_t-overlap,
                        SIT sit-index,
                        ARG1 #0,
                        ARG2 now],
                        [ RELN r_go,
                        SIT #0,
                        ARG #2,
                        ACT #1 ],
                        [ RELN r_give,
                        SIT #2,
                        ACT #1,
                        ARG1 #3,
                        ARG2 #4],
                        [ RELN r_toy,
                        SIT sit-index
                        NSTANCE #3],
                        [ RELN r_juan,
                        SIT sit-index
                        INSTANCE #4]>]
```

(23) a. Semántica de auxiliar

```
sem-struct & [ MODE prop,
                INDEX #0,
                RESTR < [ RELN r_speaker,
                        SIT sit-index
```

```

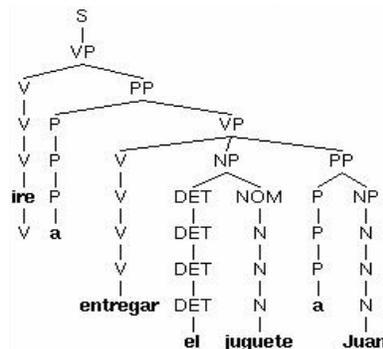
    INSTANCE #1],
  [ RELN    r_t-overlap,
    SIT     sit-index,
    ARG1    #0,
    ARG2    now],
  [ RELN    r_future,
    SIT     #0,
    ARG     #2 ],
  [ RELN    r_give,
    SIT #2,
    ACT     #1,
    ARG1    #3,
    ARG2    #4],
  [ RELN    r_toy,
    SIT     sit-index
  NSTANCE #3],
  [ RELN    r_juan,
    SIT     sit-index
  INSTANCE #4]>]

```

Se pueden apreciar las diferencias que fueron marcadas en el verbo *poder*. Uno de los verbos posee un agente (ACT), como verbo de control, y el otro no, cuando es auxiliar.

Una peculiaridad que fue señalada en el capítulo anterior es que el verbo *ir* como auxiliar presenta restricciones con respecto al tiempo en que puede aparecer. Esta problemática fue rebasada con un pequeño truco. Este consistió en modificar la ortografía de este verbo por la de *ir_srv*, como se puede apreciar en la entrada léxica correspondiente. Como el verbo es totalmente irregular en el archivo de irregularidades se incluyeron las excepciones solamente para los tiempos permitidos de este verbo auxiliar (presente y pasado imperfecto). Con esto garantizamos que no se pueda construir los otros tiempos verbales para una interpretación de futuro (*ir* como auxiliar). Así que el parseo de la oración: *ire a entregar el juguete a Juan*, solamente genera la interpretación de control como muestra el árbol y su semántica:

(24) a. Árbol sintáctico



b. Semántica

```
sem-struct & [ MODE      prop,
                INDEX     #0,
                RESTR < [ RELN      r_speaker,
                        SIT         sit-index ,
                        INSTANCE    #1],
                        [ RELN      r_t-future,
                        SIT         sit-index,
                        ARG1        #0,
                        ARG2        now ],
                        [ RELN      r_go,
                        SIT         sit-index ,
                        ARG         #2,
                        ACT         #1 ],
                        [ RELN      r_to,
                        SIT         sit-index,
                        INSTANCE    ref-index],
                        [ RELN      r_give,
                        SIT         #2,
                        ACT         #1,
                        ARG1        #3,
                        ARG2        #4],
                        [ RELN      r_toy,
                        SIT         sit-index
                        INSTANCE    #3],
                        [ RELN      r_juan,
                        SIT         sit-index
                        INSTANCE    #4]>]
```

6.2.3.2 *Estar*

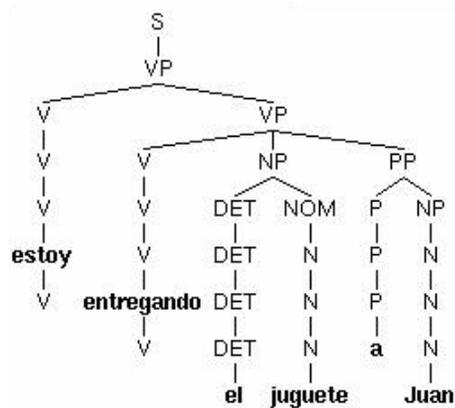
La entrada léxica del verbo *estar* como auxiliar es la siguiente:

(25)

```
; Estar: Esta comiendo (auxiliar)
estar_2 := ger_srv-lxm-spanish &
[ ORTH <! "estar" !>,
SEM [ RESTR <! [ RELN r_be ] !> ] ].
```

Su entrada léxica como verbo atributivo será discutida más adelante. Con esta entrada léxica podemos capturar el comportamiento del verbo *estar* como auxiliar, en la oración: *estoy entregando el juguete a juan*. El árbol resultante se muestra en (26).

(26)



La semántica de esta oración satisface el comportamiento que hemos visto con los verbos auxiliares, la semántica se presenta a continuación:

(27)

```
sem-struct & [ MODE      prop,
                INDEX    #0,
                RESTR    < [ RELN      reln,
                            SIT        sit-index ,
                            INSTANCE  #1],
                            [ RELN      r_t-precede,
                              SIT        sit-index ,
                              ARG1      #0,
                              ARG2      now ],
                            [ RELN      r_be,
                              SIT        sit-index ,
                              ARG        #2 ],
                            [ RELN      r_give,
                              SIT        #2,
                              ACT        #1,
                              ARG1      #3,
                              ARG2      #4],
                            [ RELN      r_toy,
                              SIT        sit-index
                              NSTANCE  #3],
                            [ RELN      r_juan,
                              SIT        sit-index
                              INSTANCE  #4]> ]
```

6.2.3.3 Haber

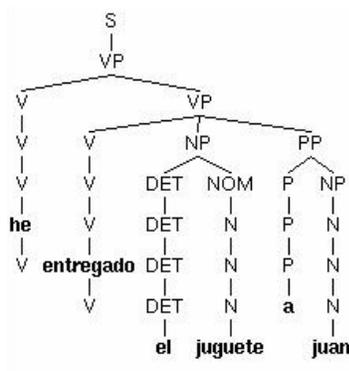
La entrada léxica para *haber* es la siguiente:

(28)

```
;; Haber: he comido
haber_1 := perf_srv-lxm-spanish &
[ ORTH <! "haber" !>,
  SEM [ RESTR <! [ RELN r_have ] !> ] ].
```

Con esta entrada léxica podemos capturar el comportamiento de auxiliar del verbo *haber* de la oración: *he entregado el juguete a Juan*. La figura (21) muestra al árbol se forma durante el parseo de la oración. Volvemos a recalcar que este comportamiento sería preferible que se capturará desde la morfología del sistema, desafortunadamente en este aspecto LKB es muy débil.

(29)



Sin embargo ,es posible generar una semántica equivalente al fenómeno morfológico donde se resalta la perfectividad con que se realiza una acción, como se muestra a continuación:

(30)

```
sem-struct & [ MODE      prop,
                INDEX    #0,
                RESTR <  [ RELN      r_speaker,
                          SIT        sit-index,
                          INSTANCE   #1],
                          [ RELN      r_t-overlap,
                          SIT        sit-index,
                          ARG1       #0,
                          ARG2       now ],
                          [ RELN      r_have_perf,
                          SIT        #0,
                          ARG        #2],
                          [ RELN      r_give,
                          SIT        #2,
                          ACT        #1,
                          ARG1       #3,
                          ARG2       #4],
                          [ RELN      r_toy,
                          SIT        sit-index
                          INSTANCE   #3],
                          [ RELN      r_juan,
                          SIT        sit-index
                          INSTANCE   #4]>]
```

6.2.3.4 Reglas morfológicas

En los tres ejemplos anteriores y en el de *poder* se observó como el sistema reconoció a los verbos conjugados y en las formas no personal de infinitivo, gerundio y participio; cuando las entradas léxicas solamente se hacen utilizando la cadena que representa el verbo en infinitivo. Lo anterior lo logró el sistema por medio de la aplicación de reglas morfológicas (sección 5.2.1.4).

Para la conjugación de los verbos en los tiempos del indicativo²⁵ se utilizaron treinta, reglas de conjugación, que en su mayoría se encontraban implementadas en la gramática de Simoes [Simoes, 1999] y dos más para el participio y gerundio de los verbos. Estas dos últimas reglas son las que se ejemplifican a continuación:

(31)

```
;; Gerundio
gerundive-verb-spanish_infl_rule :=
%suffix (ar ando) (er iendo) (ir iendo)
verb-inflectional-lrule-spanish &
[ NEEDS-AFFIX true,
  SYN [ HEAD [ FORM ger ] ],
  ARGS [ FIRST [ AFFIX gerundive-verb-spanish ]]].

;;; Participio
past-part-verb-spanish_infl_rule :=
%suffix (er ido) (ir ido) (ar ado)
verb-inflectional-lrule-spanish &
[ SYN [ HEAD [ FORM psp ] ],
  NEEDS-AFFIX true,
  ARGS [ FIRST [ AFFIX past-part-verb-spanish ]]].
```

Estas reglas permiten cambiar la ortografía de los verbos en infinitivo por la correspondiente en gerundio o participio utilizando la opción de: `%suffix`; además cambian algunas propiedades sintácticas, en este caso la forma del verbo (atributo FORM). La regla morfológica para el infinitivo no es necesario agregarla porque los verbos son capturados en su forma de infinitivo.

6.3 Otros fenómenos lingüísticos

Existen un conjunto de fenómenos adicionales que se encuentran ligados al comportamiento de los verbos auxiliares o a la problemática de definir a los verbos

²⁵ Los tiempos implementados son: presente, pasado absoluto, futuro condicional, pasado imperfecto y futuro simple. Los tiempos del subjuntivo no fueron implementados.

auxiliares. Estos fenómenos tuvieron que ser implementados para ver su desempeño y su relación con los verbos auxiliares. Uno de estos fenómenos es el de los verbos de control, de los cuales ya se apreció su comportamiento en ejemplos anteriores. Los otros fenómenos a los que nos referimos son: verbos unipersonales, anidamiento en las perífrasis, *Subject control preposition*, voz pasiva y verbos atributivos. A continuación, se presenta como fueron abordados cada uno de estos fenómenos para su implementación.

6.3.1 Verbos unipersonales

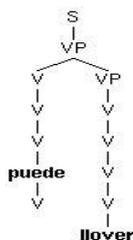
Los verbos unipersonales son muy importantes en el comportamiento de los verbos auxiliares; éstos al no poseer un agente que provoquen la acción pueden ser el complemento de verbos auxiliares mas nunca de un verbo de control. Esto repercute que las oraciones en las que aparecen los verbos unipersonales no sean ambiguas. El modelo de verbos unipersonales ya se encontraba implementado en la gramática Simoes[Simoes, 1999]; sin embargo, este modelo era deficiente por lo que se optó por sustituirlo por la siguiente definición:

(32)

```
no-subj-verb-lxm-spanish := verbal-lxm-spanish &
[ SYN [ SPR <[] > ],
  ARG-ST <[ SYN [HEAD [AGR [ GEND null-gend,
    NUM sing,
    PER 3rd ] ] ] ] > ].
```

Esta definición nos garantiza que verbos unipersonales sólo sean conjugados en tercera persona, además como el especificador que solicitan no es posible que se presente, por el tipo de concordancia que solicita (*GEND null-gend*), nunca podrá agruparse con este; pero sí podrá actuar la regla *subject_pro-drop-rule-spanish* (sección 6.2.2.2) para saturar a este especificador. Esta definición y la de verbos auxiliares hacen posible que se pueda parsear la oración: *puede llover*, sin ambigüedad. El árbol y la semántica resultante son las siguientes:

(33) a. Árbol sintáctico



b. Semántica

```
sem-struct & [ MODE      prop,  
                INDEX    #0,  
                RESTR    < [ RELN      reln,  
                            SIT        sit-index,  
                            INSTANCE  ref-index],  
                            [ RELN      r_t-overlap,  
                              SIT        sit-index instloc ],  
                              ARG1      #0,  
                              ARG2      now ],  
                            [ RELN      r_possible,  
                              SIT        #0,  
                              ARG        #1 ],  
                            [ RELN      r_rain,  
                              SIT        #1,  
                              ACT        null-index]>]
```

Como vemos la única semántica resultante corresponde a la semántica de verbo auxiliar y no la de un verbo de control. En este ejemplo la semántica representa la posibilidad de que llueva.

6.3.2 Anidamiento de perífrasis

El fenómeno anidamiento que presentan las perífrasis y en consecuencia los verbos auxiliares es sumamente complicado de modelar en HPSG. Es por esto que el modelo que se presenta aquí no había sido expuesto en el capítulo anterior; más que todo porque rompe con el espíritu de generalizar lo más posible un fenómeno en HPSG y se vuelve muy específico para cada verbo. Cabe mencionar que otros modelos fueron explorados pero estos requerían que el número de verbos en una oración fuera finito (uno o dos).

Como los verbos que construyen perífrasis piden un complemento en cierta forma y no les importa el tipo de verbo, la anidación está garantizada con la definición que hasta ahora hemos hecho; sólo faltaría delimitar los alcances de este anidamiento. En particular, aquellas restricciones relacionadas con que un verbo aparece solamente en una ocasión en la oración y que algunos de los verbos auxiliares, como *ir a*, se restringen a la primera posición del anidamiento en la oración.

El modelo propuesto consiste en construir una estructura de atributos donde se guarde si la estructura ha sido modificada por un verbo auxiliar en particular. Un verbo que participa en una perífrasis tendrá que verificar en su complemento que el atributo que lo representa no indique una modificación por un verbo semejante a él y además checará que

ninguno de los verbos que van en primera posición haya modificado a la oración. La estructura capaz de guardar esta información es la siguiente:

(34)

```
mod-verb := feat-struct &
[ PODER boolean,
  IR      boolean,
  PASAR   boolean,
  ECHAR   boolean,
  VENIR   boolean,
  VOLVER  boolean,
  HABER_DE boolean,
  HABRA_QUE boolean,
  TENER_QUE boolean,
  DEBER   boolean,
  SEGUIR  boolean,
  ANDAR   boolean,
  ESTAR   boolean,
  HABER   boolean
]
```

Esta estructura será valor del atributo AUX-MOD, el cual es un atributo contenido en el atributo HEAD de todos los verbos; es decir, se trata de una propiedad cabeza. Se marcará con *true* cuando ya haya sido modificado por un verbo de los que se encuentra listado y con un *false* si esta no puede ser modificado por un verbo. Con esta modificación la entrada léxica para *poder* como verbo auxiliar queda de la siguiente forma:

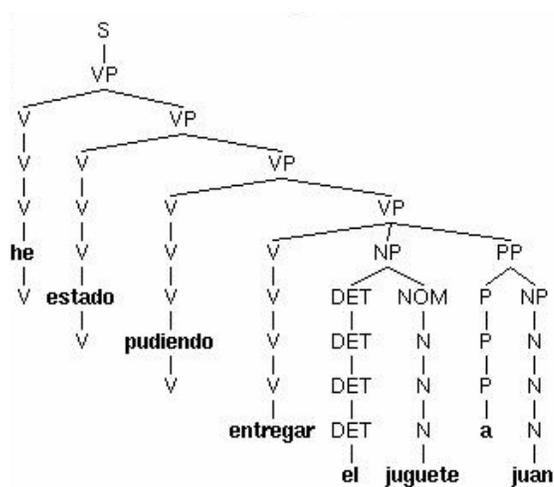
(35)

```
poder_2 := inf_srv-lxm-spanish &
[ ORTH <! "poder" !>,
  SYN [ HEAD [ MOD-VERB [ PODER true,          IR #ir,
                        PASAR #pasar,         VENIR #venir,
                        VOLVER #volver,        HABER_DE #haber_de,
                        HABRA_QUE #habra_que, TENER_QUE #tener_que,
                        DEBER #deber,          SEGUIR #seguir,
                        ANDAR #andar,         ESTAR #estar,
                        HABER #haber ]],
      COMPS < [SYN [HEAD [ MOD-VERB [ PODER false,
                        IR #ir&false, PASAR #pasar, VENIR #venir,
                        VOLVER #volver, HABER_DE #haber_de&false,
                        HABRA_QUE #habra_que&false, TENER_QUE #tener_que,
                        DEBER #deber&false, SEGUIR #seguir&false,
                        ANDAR #andar&false, ESTAR #estar&false,
                        HABER #haber]]]] >],
      SEM [ RESTR <! [ RELN r_possible ] !> ] ].
```

Como podemos apreciar esta entrada léxica se ha complicado demasiado. Notamos que el atributo PODER se fija a *true* porque es su entrada léxica, por lo tanto el valor PODER de su complemento debe ser *false*, ya que en caso contrario indicaría que la frase que toma

como complemento ya ha sido modificada por el verbo *poder*. Todos los demás atributos de MOD-VERB se unifican al de su complemento para así no perder ninguna modificación hecha tal vez por otro verbo. Los verbos que solo pueden ir en primera posición se solicitan que sean *false* así si alguno se encuentra con su atributo con *true* los valores no unificarán. Una oración anidada como *he estado pudiendo entregar el juguete a Juan* es bien parseada por el sistema; a diferencia de la oración **He estado pudiendo estar entregando el juguete a Juan*; donde aparecen dos veces el verbo *estar*. El árbol y una de las semánticas resultantes de la primera oración se muestran a continuación:

(36) a. Árbol sintáctico



b. Semántica como verbo de control

```

sem-struct & [ MODE      prop,
               INDEX     #0,
               RESTR < [ RELN      r_speaker,
                       SIT        sit-index
                       INSTANCE    #1],
                       [ RELN      r_t-overlap,
                       SIT        sit-index,
                       ARG1       #0,
                       ARG2       now ],
                       [ RELN      r_have,
                       SIT        #0,
                       ARG        #2],
                       [ RELN      r_be,
                       SIT        #2,
                       ARG        #3 ],
                       [ RELN      r_can,
                       SIT        #3,
                       ARG        #4,
                       ACT        #1 ],
                       [ RELN      r_give,
                       SIT        #4,
                       ACT        #1,
                       ARG1       #5,
  
```

```

ARG2      #6],
[ RELN    r_toy,
  SIT     sit-index
  NSTANCE #5],
[ RELN    r_juan,
  SIT     sit-index
  INSTANCE #6]>]

```

6.3.3 *Subject control preposition*

En el análisis de los verbos de control juega un papel muy importante el análisis de las preposiciones *para* y *a* (sección 5.2.1.2.1). Aquellos verbos intransitivos que originalmente van seguidos por la preposición *a* y un complemento suelen formar perífrasis al perder a su complemento e ir acompañados por una de las preposiciones arriba mencionada, ya no con el propósito de acompañar al complemento, sino con el de marcar el objetivo de la acción. Esta observación marcó la existencia de un tipo singular de preposiciones que se encarga de asignar a dos frases verbales distintas el mismo especificador. A este tipo de preposiciones se les denominó como *Subject control preposition*. La definición de este tipo de preposición se muestra a continuación:

(37)

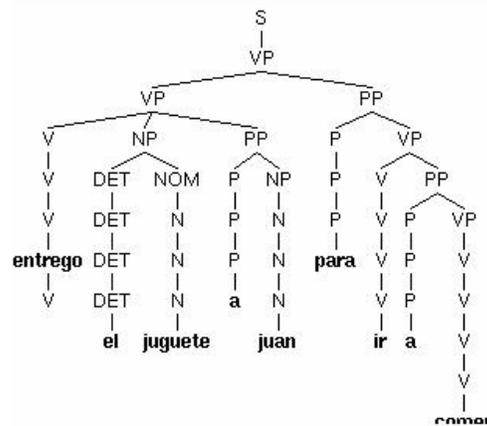
```

scp-lxm-spanish := const-lxm & prep-sem &
[ SYN [ HEAD prep-spanish & [ MOD < #syn & [SYN [HEAD verb-spanish
&[FORM fin ],
SPR <[SEM[ INDEX #1]]> ]]> ],
SPR < >,
COMPS < [] >],
ARG-ST <[SYN [ HEAD verb-spanish & [FORM inf],
SPR <[SEM [ INDEX #1]] >]]>].

```

Esta definición garantiza que las preposiciones declaradas como de este tipo tomnr una frase verbal en infinitivo como complemento y adicionalmente modifican a otra frase verbal conjugada. Además los especificadores de ambas frases se refieren al mismo individuo. Esta definición permite parsear la oración: *entregó el juguete a Juan para ir a comer*. El árbol obtenido y su semántica se muestran a continuación:

(38) a. Árbol sintáctico



b. Semántica

```
sem-struct & [ MODE prop,  
INDEX #0,  
RESTR < [ RELN r_speaker,  
SIT sit-index,  
INSTANCE #1],  
[ RELN r_t-overlap,  
SIT sit-index,  
ARG1 #0,  
ARG2 now],  
[ RELN r_give,  
SIT #0,  
ACT #1,  
ARG1 #2,  
ARG2 #3],  
[ RELN r_toy,  
SIT sit-index  
INSTANCE #2],  
[ RELN r_juan,  
SIT sit-index  
INSTANCE #3]>  
[ RELN r_finish,  
SIT sit-index,  
ACT #1 ]> ]
```

Los resultados de este proceso de parseo muestran que la oración no es ambigua debido a que el verbo *ir* necesita tener un agente, descartándose con esto la lectura de este verbo como un verbo a auxiliar.

6.3.4 Voz pasiva

La voz pasiva es un fenómeno que ayudo a la definición del comportamiento de verbo auxiliar. Los verbos auxiliares no pueden tomar la voz pasiva, pero además el tipo *srv-lxm* que caracteriza a los verbos auxiliar define a un verbo *ser* que captura a la voz pasiva. Recordemos del capítulo anterior que el fenómeno de voz pasiva se dividió en dos etapas,

la primera consiste en generar la entrada léxica para el pasivo y, la segunda, consiste en declarar a un verbo *ser* que soporte el comportamiento de la voz pasiva (sección 5.3.1). Se proponen dos reglas léxicas dependiendo del género que posea el objeto directo, en este caso el primer argumento:

(39)

```
;Participio para voz pasiva masculina
masc-pass-verb-spanish_infl_rule :=
%suffix (er ido) (ir ido) (ar ado)
passive-lrule-spanish &
[ SYN [ HEAD [ FORM pass ] ],
  NEEDS-AFFIX true,
  ARG-ST< [SYN [HEAD [AGR [NUM sing, GEND masc]]]],...>,
  ARGS [ FIRST [ AFFIX masc-pass-verb-spanish ]]].

; Participio para voz pasiva femenina
fem-pass-verb-spanish_infl_rule :=
%suffix (er ida) (ir ida) (ar ada)
passive-lrule-spanish &
[ SYN [ HEAD [ FORM pass ] ],
  ARG-ST< [SYN [HEAD [AGR [NUM sing, GEND fem]]]],...>,
  NEEDS-AFFIX true,
  ARGS [ FIRST [ AFFIX masc-pass-verb-spanish ]]].
```

Ambas reglas generaran el participio para la voz pasiva; una para el género masculino y la otra para el femenino; recordemos que la concordancia en voz pasiva es muy importante. Adicionalmente, la forma verbal se define como del tipo *pass* para indicar que forman voz pasiva y así no sean tomadas por otro verbo que no sea el verbo *ser* de voz pasiva. Fue necesario declarar un nuevo tipo de verbo *srv-lxm* que solamente acepta como complemento a formas verbales del tipo *pass*, con el objetivo de capturar el comportamiento del verbo *ser*. A continuación se presenta la declaración del tipo y la entrada léxica para *ser*:

(40)

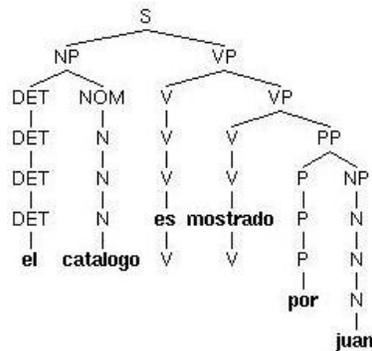
```
; Tipo seguido de un complemento de la forma pass
pass_srv-lxm-spanish := srv-lxm-spanish &
[ ARG-ST < [ SYN NP-acc-spanish ], [ SYN [ HEAD verb-spanish &
[FORM pass ] ] ] > ].

; ser: el catálogo es mostrado por Juan
ser_1 := pass_srv-lxm-spanish &
[ ORTH <! "ser" !>,
  SEM [ RESTR < > ] ].
```

Esta entrada léxica conjuntamente con las reglas léxicas nos permiten parsear la siguiente oración: *el regalo es entregado por él*. Como se ve en (41) su semántica dista mucho de la

que hemos visto en los verbos auxiliares. Lo que hacen los elementos que intevienen en la voz pasiva es romper la barrera de esta construcción (estructura superficial) de oración para obtener una semántica como si se tratase de un verbo en voz activa (estructura profunda). A continuación se presenta el árbol y la semántica que arroja LKB:

(41)a. Árbol sintáctico



b. Semántica

```

sem-struct & [ MODE      prop,
                INDEX    #0,
                RESTR    < [ RELN      r_the,
                            SIT        sit-index
                            INSTANCE   #1],
                            [ RELN      r_toy,
                            SIT        sit-index,
                            INSTANCE   #1 ],
                            [ RELN      r_t-overlap,
                            SIT        sit-index,
                            ARG1       #0,
                            ARG2       now ],
                            [ RELN      r_give,
                            SIT        #2,
                            UND        #1,
                            ACT        #3,
                            [ RELN      r_third,
                            SIT        sit-index
                            INSTANCE   #3]>]
  
```

6.3.5 Verbos atributivos

Por último un fenómeno que llevó a confundir a algunos verbos como verbos auxiliares es el de verbos atributivos, el cual ha sido discutido y diferenciado de los verbos auxiliares en el capítulo anterior (Sección 5.3.2). En la utilización de los verbos atributivos se ligan dos frases una sustantiva y otra adjetiva por medio de un verbo. Entre ambas frases debe existir concordancia. Gracias a la forma en como se definen los adjetivos en HPSG es posible capturar al fenómeno de los verbos atributivos con el tipo *srv-lxm*. El tipo que captura al

comportamiento de verbos atributivos y la entra léxica para *ser* como atributivo se presenta a continuación:

(42)

```
;; Verbos atributivos o de estado
atrib_srv-lxm-spanish := srv-lxm-spanish &
[ ARG-ST < [ ], [ SYN [ HEAD adj-spanish ] ] > ].

;; Ser para atributivas
:begin :instance.
ser_2 := atrib_srv-lxm-spanish &
[ ORTH <! "ser" !>,
SEM [ RESTR < > ] ].
```

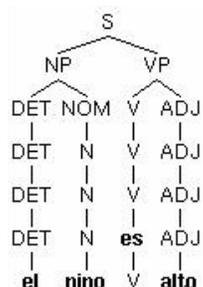
Una vez capturado el comportamiento de los verbos atributivos debemos garantizar que el verbo pase de comportarse como tal para convertirse en adjetivo. Esto lo logramos con la siguiente regla léxica:

(43)

```
;; Adjetivización para los verbos
adj-verb-lrule-spanish := word &
[ SYN [ HEAD adj-spanish & [ FORM normal,
AGR #agr,
MOD < [ SYN [ HEAD noun-spanish
& [AGR #agr],
COMPS < > ] ] > ],
SPR <[]>,
SEM [ INDEX #1, RESTR <! thm_pred & [RELN r_be, SIT #1] !>
],
ARG-ST < [ SYN NP-spanish & [HEAD [AGR #AGR]],
SEM [INDEX ref-index] ] >,
ARGS < word & [SYN [HEAD [FORM pass],
SPR < [SYN[HEAD [AGR #agr]]]>]]> ].
```

Aunque la regla se ve muy complicada, el único objetivo que tiene es tomar una estructura de verbo y transformarla a un adjetivo. Con estas descripciones somos capaces de generar las estructuras correspondientes de los verbos atributivos y poder parsear la oración: *el niño es alto*, la cual genera el siguiente árbol (44.a) y su correspondiente semántica (44.b).

(44) a. Árbol sintáctico



b. Semántica

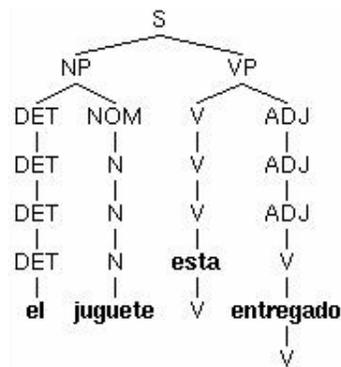
```

sem-struct & [ MODE      prop,
                INDEX    #0,
                RESTR <  [ RELN      r_the,
                          SIT        sit-index,
                          INSTANCE    #1,
                          [ RELN      r_kid,
                            SIT        sit-index,
                            INSTANCE    #1 ],
                          [ RELN      r_t-overlap,
                            SIT        sit-index,
                            ARG1      #0,
                            ARG2      now ],
                          [ RELN      r_tall,
                            SIT        #2,
                            THM        #1]>]

```

En este ejemplo *alto* es un adjetivo declarado como tal desde la entrada léxica; pero adicionalmente podemos parsear la oración: *el regalo está entregado*, donde *entregado* es un adjetivo que proviene del verbo *entregado*. El árbol resultante y su semántica son mostradas en (47).

(45) a. Árbol sintáctico



b. Semántica

```

sem-struct & [ MODE      prop,
                INDEX    #0,
                RESTR <  [ RELN      r_the,
                          SIT        sit-index,
                          INSTANCE    #1],
                          [ RELN      r_toy,
                            SIT        sit-index,
                            INSTANCE    #1 ],
                          [ RELN      r_t-overlap,
                            SIT        sit-index,
                            ARG1      #0,

```

```

ARG2      now ],
[ RELN    r_be,
SIT       #0,
ARG       #2],
[ RELN    r_give,
SIT       #2,
THM      #1 ]> ]

```

Los tipos aquí definidos y las entradas léxicas descritas conforman un ejemplo de lo que hemos llamado durante este trabajo como implementación del sistema de verbos auxiliares del español; en los apéndices D y E se encuentran listados en su totalidad los tipos y las entradas que conforman al sistema. Los ejemplos aquí presentados dan una idea de cómo funciona la implementación y sobretodo validan el modelo propuesto en este trabajo. Para presentar una idea más general de la capacidad del sistema en el apéndice F se muestra una lista de algunas de las oraciones que actualmente son parseadas por el sistema.

Capítulo 7: Conclusiones

En este capítulo se presentan las conclusiones a las que se arribaron durante la elaboración esta tesis. En este trabajo se realizó un análisis de los verbos auxiliares del español desde un punto de vista sintáctico. Los resultados de este análisis permitieron definir tres aspectos fundamentales de los verbos auxiliares del español. Se creó un conjunto de cinco pruebas que permiten identificar si un verbo es auxiliar o no; se identificaron tres propiedades básicas del comportamiento sintáctico de los verbos auxiliares y se definió el concepto de verbo auxiliar en término de *subject raising verb*, un tipo de comportamiento sintáctico identificado por Noam Chomsky para el inglés[Sag y Wasow, 1999]²⁶. Estos resultados permitieron crear el modelo del sistema de verbos auxiliares del español en el formalismo Head-driven Phrase Structure Grammar (HPSG) [Sag y Wasow, 1999]. El modelo final del sistema de verbos auxiliares del español consiste en un conjunto de tipos que capturan las propiedades de los verbos auxiliares, y en un conjunto de entradas léxicas, que representan a los verbos. A continuación se presentan de manera breve los pasos seguidos en el análisis y los resultados obtenidos.

Como primer paso del análisis se recurrió a la literatura especializada, la cual reportaba diferentes observaciones sobre el contexto sintáctico en el que aparecían los verbos auxiliares. Del análisis de estas observaciones se definieron cinco pruebas que permiten establecer si un verbo se comporta como auxiliar o no. Estas pruebas se utilizaron en el conjunto de verbos auxiliares del español propuestos por Gili Gaya [Gili, 1991]. Este análisis mostró que dicho conjunto no está compuesto únicamente por verbos auxiliares, como fue el caso de los verbos *echar y pasar*, y que es incompleto, ya que no considera a verbos como *poder*.

Como segundo paso, se realizó un análisis de las pruebas en términos de la información sintáctica que interviene en estas pruebas. A través de este análisis se pudo identificar tres propiedades sintácticas de los verbos auxiliares del español. Estas propiedades son: Un verbo auxiliar tiene como complemento un verbo en forma no

²⁶ P. 280

personal; un verbo auxiliar comparte el sujeto, especificador, con su complemento, verbo en forma no personal, y un verbo auxiliar no posee un agente.

Como tercer paso, se buscó un tipo en HPSG que capturara las propiedades de los verbos auxiliares. El tipo identificado con estas propiedades es *srv-lxm*. Este tipo proviene del análisis que hace Noam Chomsky para el inglés de los verbos que tienen como complemento un verbo [Sag y Wasow, 1999]²⁷. En este análisis se señalan las propiedades de este tipo, las cuales resultan similares a las de los verbos auxiliares del español. Sin embargo, en el inglés los verbos auxiliares son un subtipo de verbos *raising* debido a que poseen propiedades adicionales, como lo son las propiedades NICE. El resultado de identificar al tipo que captura las propiedades de los verbos auxiliares del español fue el de poder establecer una definición para estos en términos del tipo. Esta definición queda de la siguiente forma: los verbos auxiliares del español son del tipo *srv-lxm* con un complemento de frase verbal en forma no personal.

Adicionalmente, los tipos que se encuentran definidos en HPSG hicieron posible poder capturar comportamientos adicionales que presentan los verbos auxiliares, como es el caso de los comportamientos transitivo e intransitivo. Además, fue posible representar el comportamiento ambiguo que presentan los verbos auxiliares del español con tipos el tipo *scv-lxm*.

Una vez definido el tipo que captura el comportamiento de los verbos auxiliares del español se definieron tipos hijos de *srv-lxm* que restringían la forma no personal del verbo complemento. Después, se crearon las entradas léxicas de los verbos auxiliares propuestos por Gili Gaya que habían pasado las cinco pruebas. La definición de los tipos y las entradas léxicas corresponden al sistema de verbos auxiliares del español. Con el fin de complementar al sistema se agregaron la definición de tipos y entradas léxicas de los comportamientos adicionales de los verbos auxiliares y de los verbos que fueron considerados como verbos auxiliares por Gili Gaya, pero que en este trabajo no lo son. Con las definiciones en HPSG del sistema de verbos auxiliares y los fenómenos complementarios fue posible realizar la implementación computacional en el ambiente de desarrollo Linguistic Knowledge Building (LKB) [Copestake, 2000]. Esta implementación

²⁷ p. 280

y el conjunto de oraciones que son parseadas validan el modelo del sistema de verbos auxiliares del español propuesta en esta tesis.

En las siguientes secciones se expone de manera más amplia los resultados obtenidos. En la sección 7.1 se presentan los resultados cualitativos y cuantitativos del análisis sintáctico y la sección 7.2 incluye aspectos de la implementación y su validación. En resumen, el objetivo de modelar el sistema de verbos auxiliares y realizar su implementación y validación computacional se cumplió en esta tesis.

7.1 Análisis sintáctico

El análisis sintáctico hecho a los verbos auxiliares del español permitió establecer las bases para realizar el modelo sintáctico de los verbos auxiliares del español. Este análisis culminó con tres puntos muy importante para el desarrolló de este trabajo. El primero, consiste en un conjunto de cinco pruebas que determinan el comportamiento de un verbo como auxiliar o no. El segundo, en un conjunto de tres propiedades que definen el comportamiento sintáctico de estas palabras. Y por último, se relacionaron estas propiedades a un concepto que dentro de la literatura especializada: el tipo *subject raising verb*. A continuación se presenta un resumen de los resultados obtenidos en cada uno de estos aspectos.

7.1.1 Pruebas sintácticas

El observar el comportamiento de los verbos auxiliares arrojó como resultado un conjunto de cinco pruebas que nos permiten identificar si un verbo se comporta como verbo auxiliar o no. Estas pruebas se basan únicamente condiciones sintácticas, a diferencia de las gramáticas tradicionales [Gili, 1991]. Cabe destacar que trabajos más recientes en los verbos auxiliares llegan definir a los verbos auxiliares basados en pruebas similares a las que presentamos en este trabajo [Luna, 1991]. Las pruebas son las siguientes:

1. Un verbo auxiliar aparece en construcciones perifrásticas donde la conjugación del verbo auxiliar ayuda a definir el agente del verbo en forma no personal si este requiere de uno. Por ejemplo, en *el niño alto va a entregar el juguete a Juan* quien entrega el juguete es el niño alto.
2. Un verbo auxiliar aparece en construcciones perifrásticas donde intervienen verbos unipersonales, aquellos que no tienen un agente. Por ejemplo, *va llover*.

3. Para preguntar por el objeto directo del verbo en forma no personal es necesario que el verbo auxiliar vaya acompañado por el verbo en forma no personal. Por ejemplo, la forma correcta es *¿qué va a entregar?* y no *¿qué va a?*.
4. Para preguntar por la acción que representa el verbo en forma no personal es necesario que el verbo auxiliar vaya acompañado por un verbo comodín como *hacer*. Por ejemplo, la forma correcta *¿qué va a hacer?* Y no *¿qué va a?*
5. Un verbo auxiliar no puede formar la voz pasiva, pero si puede tomar como complemento la voz pasiva de un verbo transitivo. Por ejemplo, la forma correcta es *el regalo va a ser entregado por él* y no *el regalo es ido a entregar por él*.

Si un verbo es capaz de aparecen en las cinco construcciones que describen estas pruebas podemos concluir que se trata de un verbo auxiliar. El conjunto propuesto por Gili Gaya fue sometido a este conjunto de pruebas, la siguiente muestra el resultado obtenido de la aplicación de estas reglas:

Verbo	Prueba 1	Prueba 2	Prueba 3	Prueba 4	Prueba 5
Poder ²⁸	4	4	4	4	4
Ir a	4	4	4	4	4
Pasar a	4	7	4	4	4
Echar a	4	7	4	4	7
Venir a	4	4	4	4	4
Volver a	4	4	4	4	4
Haber de	4	4	4	4	4
Haber que	4	7	4	4	4
Tener que	4	4	4	4	4
Deber de	4	4	4	4	4
Estar	4	4	4	4	4
Ir	4	4	4	4	4
Venir	4	4	4	4	4

²⁸ El verbo *poder* no forma parte del conjunto original pero es de suma importancia para el proyecto DIME.

Seguir	4	4	4	4	4
Andar	4	4	4	4	4
Haber	4	4	4	4	4
Llevar	7	7	4	4	7
Tener	7	7	4	4	7
Traer	7	7	4	4	7
Dejar	7	7	4	4	7
Estar	7	7	4	4	7
Ser	7	7	4	4	7

Tabla 2. Resumen de pruebas

En esta tabla están marcadas con el símbolo 4 aquellas pruebas que fueron cumplidas y con 7 las que no. La línea doble separa los diferentes tipos de complementos, el primer grupo corresponde a los seguidos por un infinitivo, el segundo a los de gerundio y el tercero por un participio. En resumen, del conjunto original de veintiún verbos solamente doce cumplieron con la condición de auxiliar y adicionalmente se verificó que el verbo *poder* cumple con este comportamiento, lo que nos deja trece verbos auxiliares. Por otro lado, es importante destacar que todos los verbos seguidos por el gerundio cumplieron las pruebas y que solamente uno seguido por un participio cumplió con las pruebas, el verbo *haber*.

7.1.2 Propiedades sintácticas

La mayoría de los estudios recientes sobre los verbos auxiliares proponen una definición basada en un conjunto de pruebas similares a las mostradas en la sección anterior. Sin embargo, no es posible construir un modelo computacional basándose en las construcciones en las que aparecen los verbos auxiliares. Por lo anterior, se realizó un análisis de las pruebas que nos permitió definir un conjunto de tres propiedades sintácticas que definen el comportamiento de los verbos auxiliares estas pruebas son:

1. Un verbo auxiliar tiene como complemento una frase verbal en forma no personal.
2. Un verbo auxiliar comparte su sujeto con el de su complemento.
3. Un verbo auxiliar no tiene agente

7.1.3 El tipo *srv-lxm*

El conjunto de propiedades de los verbos auxiliares permitió identificar un tipo dentro del formalismo HPSG que captura a estas propiedades. Este tipo, *srv-lxm*, permite modelar las tres propiedades que poseen los verbos auxiliares. Este tipo en HPSG nos permitió modelar al sistema de verbos auxiliares en este formalismo y aún más importante nos permitió definir a los verbos auxiliares del español en términos de un objeto lingüístico con mucha solidez como lo es el concepto de *subject raising verb*. La definición es la siguiente:

Defición: Los verbos auxiliares del español son verbos del tipo *srv-lxm*.

Estos tres puntos presentados en la sección 7.1 resumen los alcances lingüísticos que se lograron durante la realización de esta tesis. A continuación se resume a la implementación y validación del sistema de verbos auxiliares del español.

7.2 Implementación

Una vez que se contó con una definición formal de los verbos auxiliares del español, se procedió a crear el modelo del sistema. Este modelo consiste en un conjunto de tipos y entradas léxicas que capturan el comportamiento de los trece verbos identificados como verbos auxiliares del español. Adicionalmente, se crearon un conjunto de tipos y entradas léxicas que capturan el fenómeno alternativos a los verbos auxiliares como la voz pasiva, verbos atributivos, verbos unipersonales y *Subject Raising Preposition*. A continuación, se resumen aspectos generales de la implementación de sistema de verbos auxiliares y los fenómenos alternativos.

7.2.1 Sistemas de verbos auxiliares

El sistema de verbos auxiliares implementado en este trabajo consiste en un conjunto diversos de tipos y treintitrés entradas léxicas, para representar el total de trece verbos. El aumento en el número de entradas léxicas se debe a que se consideran los comportamientos adicionales que presentan los verbos auxiliares. La siguiente tabla ilustra los diversos comportamientos que presentan el conjunto original de Gili Gaya más el verbo *poser*, donde están contenidos los trece verbos auxiliares implementados en este trabajo.

Verbo	Auxiliar	De control	Transitivo	Intransitivo	Atributivo
Poder	4	4		4	
Ir a	4	4		4	
Pasar a		4	4		
Echar a ²⁹		4	4		4
Venir a	4	4		4	
Volver a	4	4		4	
Haber de	4				
Haber que		4			
Tener que	4			4	
Deber de	4		4		
<hr/>					
Estar	4				
Ir	4				
Venir	4				
Seguir	4		4		
Andar	4			4	
<hr/>					
Haber	4				
Llevar			4		4
Tener				4	4
Traer			4		4
Dejar			4		4
Estar					4
Ser					4

Tabla 3. Resumen de comportamientos

Los tipos y las entradas léxicas que capturan se encuentran descritos bajo el formato de LKB en el apéndice E y F.

La validación del sistema se puede observar en los diferentes ejemplos que fueron mostrados en el capítulo seis, donde se observa que el modelo es capaz de reproducir la

²⁹ El verbo *echar* cumple con un comportamiento de *object raising verb*.

estructura interna de las oraciones donde intervienen los verbos auxiliares. En el apéndice E se presenta una lista más amplia de oraciones parseadas por el sistema LKB que validan el modelo presentado en este trabajo.

7.2.2 Otros fenómenos

Adicionalmente al modelo de los verbos auxiliares fue posible implementar un conjunto de otros fenómenos que giran entorno a los verbos auxiliares. Estos fenómenos son: la voz pasiva, verbos atributivos, verbos unipersonales y *Subject Raising Preposition*

La voz pasiva se implemento a través de una regla léxica llamada *pass-verb-spanish_infl_rule*, que transforma un verbo transitivo a un verbo donde han sido cambiado el orden de sus argumentos, y el tipo hijo de *srv-lxm* que toma como complemento con el nuevo orden en sus argumentos, como *ser* y *estar*, a este nuevo tipo se le denominó *pass_srv-lxm*. Los verbos atributivos se implementaron con un tipo hijo de *srv-lxm*, al cual se denominó *atrib_srv-lxm*. Para los verbos unipersonales fue necesario crear un nuevo tipo hijo de *verb-lxm*, al cual se denominó *no-subj-verb-lxm*. Finalmente, para las preposiciones *Subject Raising Preposition* fue necesario declarar un nuevo tipo de preposiciones que reproducen el fenómeno (*srp-lxm*). Al igual que con la sección anterior en el apéndice D y E se pueden observar a todos los tipos y las entradas léxicas involucradas. Así también, en el apéndice F se muestra una lista de oraciones parseadas en donde intervienen estos fenómenos; el análisis exitoso de estas oraciones valida el modelo desarrollado en esta tesis.

En conclusión los resultados presentados en estas secciones nos permiten apreciar que se cumplió con el objetivo de: modelar el sistema de verbos auxiliares y realizar su implementación y validación computacional, el cual se planteo para el desarrollo de esta tesis. Por otro lado, un resultado muy satisfactorio de la implementación del modelo es su utilización como parte de la gramática del proyecto DIME, sobretodo enfocada a parsear aquellas oraciones en donde interviene el verbo *poder*.