

# Una gramática básica del español en HPSG

Luis Pineda, Ivan Meza

Universidad Nacional Autónoma de México  
IIMAS, UNAM, AP. 20-726 ADMON. No. 20,  
Del. Alvaro Obregón, 01000 México, D.F., México  
{luis,ivanvladimir}@leibniz.iimas.unam.mx

**Resumen:** En este artículo se presenta una gramática básica del español en HPSG (Sag y Wasow, 1999). Para el diseño de la gramática se parte de un modelo integrado del sistema de verbos auxiliares y el sistema de pronombres clíticos. Para este efecto se introduce un criterio formal para caracterizar a los auxiliares y se presenta un análisis de la perífrasis con y sin auxiliares; se presenta también un modelo innovador del sistema de pronombres clíticos en el que se postula que los grupos de pronombres clíticos constituyen unidades léxicas indisolubles; el modelo permite caracterizar construcciones en que todos los clíticos son argumentos del mismo verbo, el caso de pronombres que representan argumentos correferenciales de verbos diferentes, y también el caso en que los pronombres de un grupo clítico representan argumentos independientes de verbos diferentes; el modelo considera también el caso de los pronombres pleonásticos característicos del español. Se introduce posteriormente una jerarquía de tipos verbales, con la correspondiente estructura de argumentos. En este entorno básico se discuten varios tipos de oraciones como las atributivas, las oraciones con verbos con propiedades ergativas, como *gustar*, la voz pasiva, el complejo grupo de oraciones con *se*, entre las que se encuentran la pasiva refleja y la impersonal activa, las oraciones con *se* personal y las reflexivas. Se aborda también brevemente la estructura de las interrogativas pronominales (e.g. con los pronombres *quién*, *qué*, etc.). Finalmente se introduce una estrategia para abordar algunas de las variaciones del orden de palabras que permite la sintaxis del español con respecto al orden lineal (SVO). Actualmente se cuenta con una implementación en el ambiente de desarrollo LKB (Kopestake, 2002) y la cobertura de la gramática ha sido evaluada con relación al corpus DIME (Villaseñor, Massé y Pineda, 2001), desarrollado en el contexto del proyecto DIME: Diálogos Inteligentes Multimodales en Español (Pineda et. al., 2001, 2002).

**Palabras clave:** Spanish grammar, Auxiliary verbs, clitic pronouns, impersonal constructions, HPSG.

## 1. Introducción

A pesar de su importancia, los sistemas de verbos auxiliares y pronombres clíticos del español han recibido muy poca atención en los estudios de lingüística computacional; al analizar estos fenómenos se asume con frecuencia que el sistema de verbos auxiliares es similar al del inglés y se modela en dichos términos (e.g. Quirino, 2001). Sin embargo, esta posición conduce a dificultades inmediatamente. En inglés, por ejemplo, el verbo *can* es un auxiliar pero su traducción al español, *poder*, no se considera normalmente como tal. Otro ejemplo es el verbo auxiliar inglés *have* cuya traducción al español, *haber*, se tiene más bien como parte de la conjugación compuesta y no como auxiliar. Otra característica muy singular del español es el uso aparentemente idiosincrásico de los verbos copulativos *ser* y *estar* en las oraciones atributivas y pasivas. Para el oído castellano nunca habrá ambigüedad entre *Juan es alto* y

*Juan está alto* o entre *la puerta es abierta* y *la puerta está abierta* en cada contexto de uso, mientras que para el oído extranjero, el uso de una u otra de estas formas es uno de los más grandes misterios de esta lengua, incluso para hablantes de lenguas romances similares como el italiano y el francés.

Estos ejemplos muestran que aunque ambos sistemas verbales son similares su análisis está motivado por tradiciones gramaticales diferentes y no son de ninguna manera idénticos; por lo mismo, el procesamiento automático del español requiere de un modelo propio que capture su naturaleza sintáctica y semántica.

Una asimetría más marcada entre el inglés y el español es el fenómeno de los pronombres clíticos. En español *te lo digo*, *dime* y *hazlo*, por ejemplo, son expresiones comunes, sobre todo en el lenguaje conversacional, que no tienen contraparte directa en

inglés. De hecho y a pesar de que los pronombres clíticos aparecen en varias lenguas romances como el francés y el italiano, su análisis ha sido el foco de pocos estudios en lingüística computacional (e.g. Monachesi, 1993); además, hasta donde sabemos, no se dispone de implementaciones robustas de este fenómeno.

Adicionalmente, en español los verbos auxiliares y los pronombres clíticos interactúan de manera compleja pero sistemática. Por ejemplo, en la oración *me puedes mostrar el catálogo de muebles* el pronombre clítico *me* es realmente el segundo argumento del verbo *mostrar* a pesar de que este último aparece en forma no personal (i.e. como infinitivo) y de que el verbo auxiliar *poder* se encuentra ubicado entre ambos; por lo mismo, la función semántica representada por el verbo está distante a uno de sus argumentos. Esta complejidad aumenta en oraciones con varios auxiliares y más de un verbo de contenido. Por ejemplo, lo mismo que expresa *podría haber visto a Juan comer la manzana* se puede expresar con clíticos de manera natural y directa: *lo podría haber visto comersela*; en este caso *lo* substituye al objeto directo de *ver* (i.e. es lo visto) y también al sujeto de *comer* (i.e. quien come). Más aún, no sólo los verbos de contenido se pueden cliticar sino también los auxiliares; además la cliticación puede darse en conjunto con el fenómeno del léismo en el que el pronombre *le*, que es normalmente un dativo, aparece como acusativo e incluso, por su doble rol como objeto de *ver* y sujeto de *comer*, es simultáneamente nominativo y acusativo: *le podría haber visto comerse la manzana*, *podría haberle visto comerse la manzana*, e incluso, en el uso arcaico, *podría haber vístole comerse la manzana* y *podría haber vístole comerse la manzana*. Esta interacción refuerza la necesidad de contar con un análisis del sistema de verbos auxiliares del español motivado por la naturaleza propia de esta lengua.

Por todo esto, los sistemas de verbos auxiliares y de pronombres clíticos, así como sus formas de interacción, constituyen una parte sumamente importante del núcleo sintáctico del español. Una vez elaborado este núcleo es posible abordar otros fenómenos gramaticales como la voz pasiva, el grupo de oraciones impersonales con *se* entre las que se encuentran la pasiva refleja (*se muestra el catálogo*), la impersonal activa (*se dice*) y su versión con léismo (*se le dice*) así como las reflexivas (*se viste*, *se peina*) que a pesar de ser sumamente conspicuas en esta lengua, presentan un reto singular al análisis sintáctico. También es necesario analizar oraciones con verbos como *gustar*, que tienen propiedades ergativas: en la oración activa *le gustan los chocolates*, los chocolates son lo que recibe la acción de gustar, lo gustado, por lo que *los chocolates* está en acusativo y, sin embargo, es el sujeto de la oración ya que concuerda con el verbo.

Para tratar con estos y otros fenómenos es necesario contar con un análisis de los tipos de verbos del español con sus correspondientes estructuras de argumentos, ya que las formas verbales necesarias para el análisis pueden derivarse de las formas canónicas mediante reglas léxicas. En este artículo proponemos una jerarquía de tipos verbales para el español, con la estructura de argumentos correspondiente, y mostramos como el uso de reglas léxicas permite, en conjunto con el análisis de clíticos y auxiliares, modelar de manera sumamente directa e intuitiva un amplio espectro de tipos oraciones, como las pasivas, las impersonales y las reflexivas.

El diseño de una gramática del español debe también contemplar el análisis de las oraciones interrogativas, tanto las del tipo sí/no, como las interrogativas con los pronombres frontalizados *quién*, *qué*, *a quien*, *dónde*, *cuándo* y *por qué*. El análisis elemental de estas oraciones puede ser similar a los correspondientes del inglés (i.e. *wh-questions*) mediante el uso de las categorías slash (“/”) de las gramáticas de categorías y de unificación, o la regla de *gap* de HPSG. Sin embargo, hay que tomar en cuenta que a diferencia del inglés, en español la frontalización de estos pronombres no es obligatoria. Más aún, en preguntas sí/no, la inversión del sujeto y el verbo tampoco es obligatoria, y ni siquiera común, por lo que el orden normal de las oraciones declarativas e interrogativas puede ser el mismo y su distinción sólo es posible, en esta caso, tomando en cuenta la entonación. Por estas razones el análisis de esta clase de oraciones debe contemplarse, cuando menos en parte, dentro del problema general del orden de los elementos oracionales en español, que en esta lengua se siente como libre.

De acuerdo con Gili Gaya (1991, pp. 81-94) el español, a diferencia del inglés, que sigue un orden lineal en que el determinante sigue al determinado, dando a esta lengua un carácter fundamentalmente analítico, dirigido a quien escucha, el español presenta además un orden envolvente, con un carácter sintético, que toma el punto de vista del hablante, en el que los determinantes pueden ser antepuestos; los complementos circunstanciales de lugar, tiempo, modo o causa, por ejemplo, se pueden frontizar con gran libertad, especialmente en el lenguaje hablado, resaltando de este modo el interés principal del hablante. En la forma lineal *Los chocolates le gustan a María* la atención se centra en el sujeto *Los chocolates*, en *le gustan los chocolates a María* la atención va a la inclinación de María, denotada por la frase verbal *le gustan* y en *a María le gustan los chocolates* el interés se centra en la beneficiaria de dicha inclinación, es decir, María. En español estas tres oraciones son gramaticales, idiomáticas y se pueden observar muy frecuentemente. Más aún, para el hablante de esta

lengua no hay ambigüedad en ninguna de estas formas ya que además de la entonación, los argumentos del verbo se identifican tomando en cuenta el significado de las preposiciones que encabezan cada uno de los elementos, el significado léxico de los términos de las frases preposicionales y del verbo, las restricciones que se establecen entre estos elementos y, finalmente, el orden de los elementos oracionales. Por lo mismo, tratar de deducir el rol como sujeto, complemento directo o circunstancial de cada elemento exclusivamente en función del orden en que éste aparece en la oración, es una estrategia sumamente limitada.

Sin embargo esta libertad está acotada y hay restricciones importantes; por ejemplo, si la oración se enuncia en un solo grupo fónico, la frase verbal nunca va más allá de la segunda posición sin que la oración deje de ser idiomática. Por otro lado, varias formas que difieren del orden lineal (sujeto, verbo, objeto directo, indirecto y circunstancial) corresponden realmente a fenómenos locales, como en el caso de los pronombres proclíticos que preceden al verbo a pesar de fungir como sus complementos, la inversión de los complementos directo e indirecto (*puedes girar noventa grados el cuarto por puedes girar el cuarto noventa grados*) y la posposición del sujeto en la pasiva refleja (*se pueden cambiar los colores*), entre otros, pueden ser tratados con reglas léxicas, y aunque en última instancia pudieran estar relacionados con el problema del orden envolvente, pueden ser analizados de manera independiente.

Sin embargo, la solución general del problema del orden envolvente es muy compleja y requiere muy probablemente del uso de información de otros niveles de análisis, como el prosódico y el pragmático. Por lo mismo, en la presente formulación se presenta una solución parcial, de carácter práctico, que recurre varias estrategias: primero tomamos en cuenta en la identificación de un conjunto de variaciones de carácter local que pueden predecirse de los tipos verbales y la estructura de argumentos, por lo que estos fenómenos se pueden abordar con reglas léxicas y no son realmente problemas de orden; en segundo lugar introducimos en mecanismo de orden local para invertir el orden de los complementos directo e indirecto, como ocurre frecuentemente en esta lengua. En tercer lugar se consideran variaciones con respecto al orden lineal como construcciones en las que un argumento se realiza en una posición diferente a la normal, normalmente frontalizada; este caso se trata con la regla de *gap* de manera análoga a las dependencias a distancia, como en las preguntas pronominales. El siguiente mecanismo consiste en la introducción de versiones alternativas de los esquemas de combinación sintáctica. En HPSG, por ejemplo, se pueden definir versiones de las reglas cabeza específicas para cada

orden; sin embargo, al ser estos esquemas sumamente poderosos, esta opción se limita al máximo y sólo se recurre a ella como último recurso. Asimismo, la regla de *gap* y los esquemas de combinación gramatical adicionales dependientes del orden pueden interactuar de manera compleja, por lo que se deben emplear como sumo cuidado. Finalmente, se recurre al uso de reglas léxicas para aumentar un complemento circunstancial opcional a la lista de argumentos del verbo. Esta regla permite tratar a los complementos circunstanciales como complementos propiamente, y no sólo como modificadores (i.e. como frases adjetivas o adverbiales). La definición de la regla de *gap*, permite frontalizar estos complementos, y su interacción con el sistema de clíticos y otros fenómenos locales, la inversión de complementos directo e indirecto y la posposición del sujeto, permite analizar una cantidad de órdenes de palabras bastante significativa. Sin embargo, a pesar de que el uso de estas estrategias amplía considerablemente la cobertura de la gramática, el fenómeno es sumamente complejo, como ya se ha dicho, y una solución más general requiere del uso de información de otros niveles de representación gramatical como el prosódico y el pragmático.

En el resto de este artículo se presenta el programa esbozado en esta sección: en las secciones 2 y 3 se presentan el sistema de verbos auxiliares y el sistema de pronombres clíticos respectivamente, mientras que la sección 4 se dedica a la interacción entre ambos sistemas. En la sección 5 se presenta nuestra taxonomía de tipos verbales con sus respectivas estructuras de argumentos. En la sección 6 se abordan las oraciones impersonales y otras oraciones con *se*, cuyo análisis se considera problemático; se muestra cómo estas oraciones pueden ser tratadas con reglas léxicas. En la sección 7 se presenta un análisis provisional de las interrogativas pronominales y en la sección 8 la estrategia para el tratamiento del orden de palabras. En la sección 9 se reporta la implementación y el proceso de validación; finalmente, en la sección 10 se presentan las conclusiones generales del presente trabajo.

## 2. El sistema de verbos auxiliares

En estudios tradicionales de la gramática del español no hay aún consenso en la caracterización formal de los verbos auxiliares y diversas gramáticas adoptan diversos criterios con diferentes resultados (e.g., Gili Gaya, 1991). Intuitivamente, un verbo auxiliar en oraciones predicativas se define como el verbo en forma personal que aparece en perífrasis; sin embargo, es común también definir a la perífrasis como una frase que contiene un verbo auxiliar, por lo que existe circularidad entre ambas definiciones. Por lo mismo es necesario contar con una definición de estos conceptos

que apele a nociones gramaticales generales. Tres de estas nociones son:

- (1) Los verbos auxiliares son palabras que tienen una función gramatical fija en lenguaje y normalmente contienen información temporal y aspectual; por lo mismo, no tienen contenido conceptual.
- (2) En oraciones predicativas aparecen en perífrasis en forma conjugada y llevan como complemento una frase verbal en forma no personal (infinitivo, gerundio y participio). En atributivas tienen función copulativa.
- (3) Son verbos que carecen de agente; al carecer de contenido conceptual no hay agente ni acción que realizar; sin embargo, al estar conjugados contienen a un sujeto de manera implícita que a su vez permite determinar al agente de los verbos de contenido de la perífrasis (en las predicativas); más aun, dicha determinación es necesaria ya que los verbos en forma no personal carecen de información acerca de sus posibles sujetos (i.e. no están conjugados).

Tomando como punto de partida estas tres nociones gramaticales es posible definir cinco pruebas sintácticas para determinar si un verbo es auxiliar o no en oraciones predicativas. Estas pruebas son:

1. Son verbos conjugados que encabezan una frase verbal y anteceden a uno o más verbos en forma no personal:
  - (a) *puedes mostrar el catálogo a Juan*
2. Pueden ser modificados por una frase verbal encabezada por un verbo impersonal:
  - (a) *puede llover*
  - (b) *\*tiene llovido*
3. Se requieren todos los verbos de la perífrasis para preguntar por los complementos del verbo de contenido:
  - (a) *¿qué puedes mostrar?*
  - (b) *\*¿qué puedes?*
4. Se requiere un verbo “comodín” (e.g. *hacer*) para preguntar por la acción denotada por el verbo de contenido:
  - (a) *¿qué puedes hacer?*
  - (b) *\*¿qué puedes?*
5. El verbo de contenido puede presentarse en pasiva pero no el verbo auxiliar.
  - (a) *El catálogo puede ser mostrado.*
  - (b) *\*El catálogo es podido ser mostrado.*

La primera de estas pruebas verifica que los verbos auxiliares ocurren en perífrasis (pero hay perífrasis en las cuales no ocurren verbos auxiliares); la segunda garantiza que el verbo auxiliar carece siempre de agente; la tercera verifica que el complemento del verbo auxiliar es una frase verbal, a diferencia de los verbos transitivos que tienen un complemento directo; la cuarta verifica el vínculo necesario que existe entre el verbo auxiliar y la frase verbal que lo complementa; la última refuerza la no agentividad del verbo auxiliar,

ya que si fuera posible pasivizarlo sería agentivo. En estas construcciones, el sujeto del verbo de contenido es también el sujeto de la frase verbal como un todo, aunque sea anónimo (e.g. en *puedes mostrar* quien tiene la posibilidad de mostrar es quien muestra). Desde el punto de vista sintáctico los verbos que satisfacen estas propiedades se conocen en la literatura como de subida de sujeto o *subject-raising verbs* (Sag y Wasow, 1999); por lo mismo, aquí se postula que los verbos auxiliares del español son de este tipo sintáctico.

De acuerdo con estas pruebas hemos encontrado que los siguientes verbos son auxiliares:

- (a) Con modificador en infinitivo: *poder (puedes comer), ir a (vas a comer), venir a (vienes a comer), volver a (vuelves a comer), haber de (habrás de comer), tener que (tienes que comer), deber de (debes de comer), llegar a (llegas a comer), acabar de (acabas de comer) y alcanzar a (alcanzas a comer).*
- (b) Con modificador en gerundio: *Estar (estás comiendo), ir (vas comiendo), venir (vienes comiendo), seguir (sigues comiendo) y andar (andas comiendo).*
- (c) Con modificador en participio: *ser (es querido), estar (está abierta) y haber (has comido).*

Cómo ya se ha mencionado, en la gramática tradicional del español el verbo *haber* en perífrasis no se considera como auxiliar sino como parte de la conjugación; en nuestra caracterización, sin embargo, es un auxiliar y no se requieren consideraciones excepcionales para su tratamiento.

Por otro lado, verbos que no satisfacen las cinco condiciones sintácticas no son auxiliares a pesar de que puedan aparecer en perífrasis; por ejemplo: *querer (quiere comer), pasar a (paso a comer), haber que (hay que comer)*. En el primer caso *comer* es el nombre de la acción de comer y funciona como complemento directo de *quiere*; *pasar a* y *haber que*, por su parte, requieren de un agente que sea compartido con el segundo verbo de la perífrasis (i.e. *\*paso a llover* y *\*hay que llover*), por lo que violan la segunda prueba. En estos casos el sujeto del verbo conjugado es el mismo que el sujeto de la frase infinitival que le sigue (i.e. quien quiere es quien come) y se conocen como verbos de sujeto de control o *subject-control*. Otros verbos que pueden ocurrir en perífrasis son los verbos de voluntad como *mandar* y *prohibir*, y los verbos de los sentidos, como *ver* y *oir*; verbos de estos dos grupos tienen la propiedad de que sus objetos directos son al mismo tiempo el sujeto del verbo que les sigue en la perífrasis y se conocen como verbos de objeto de control o *object-control* (Sag y Wasow, 1999). Por ejemplo, en *prohibió al niño jugar* y *oyó cantar a María*, a quien se prohíbe es quien juega (o no juega) y a quien se oye es quien canta respectivamente. También los verbos

llamados de movimiento y algunos de estado pueden ocurrir en perífrasis pero seguidos de un gerundio en vez del infinitivo. Por ejemplo, *ir corriendo*, *seguir buscando*, *vivir soñando*. Estos verbos son también del subida de sujeto o *subject-raising*, pero cuando preservan su significado léxico no son auxiliares.

En general, todos los verbos que aparecen en perífrasis con la excepción de *haber* pueden preservar su significado léxico original y en estos contextos no son auxiliares: *va a comer* quiere decir, en su acepción como auxiliar, que alguien va a iniciar la acción de comer, y no que alguien se va a trasladar físicamente a un lugar con objeto de comer, como es el caso cuando se emplea el sentido normal del verbo *ir*; incluso los copulativos *ser* y *estar* ocurren con frecuencia con su sentido original de existencia y localidad y la ambigüedad se tiene que resolver. En resumen, en nuestro modelo los verbos auxiliares son todos los verbos del tipo de subida de objeto que pierden su significado léxico original y adquieren un significado gramatical en el contexto en que se usan; sin embargo, el modelo tiene que considerar las construcciones perifrásticas sin auxiliares, ya que estas construcciones interactúan también con el sistema de clíticos, como se verá más adelante.

Pasamos ahora a presentar el modelo de verbos auxiliares en HPSG. La descripción del tipo léxico *subject-raising verb* se presenta en la figura 1. Lo crucial en esta definición es que el especificador o sujeto de un verbo de este tipo y el especificador o sujeto de su complemento son el mismo.

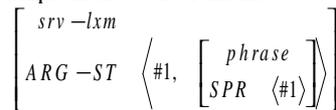


Figura 1: Subject-raising verb

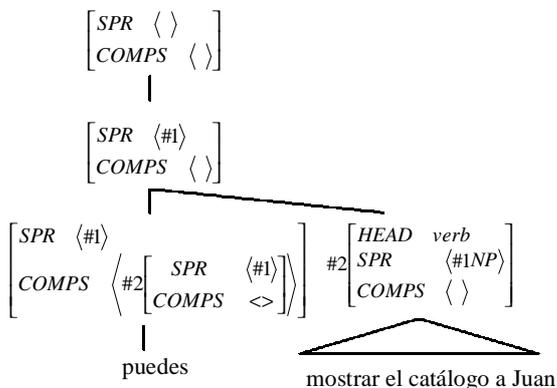


Figura 2: Oración con verbo auxiliar

En la figura 2, *puedes* se combina con su complemento *mostrar el catálogo a Juan* a través de la *Head-Complement Rule*. La frase verbal resultante es objeto de la regla *Subject-PRO-drop Rule* que se ha

adicionado a HPSG para capturar los sujetos contenidos en la conjugación del verbo del español. La representación semántica resultante de este análisis se muestra en la figura 3.

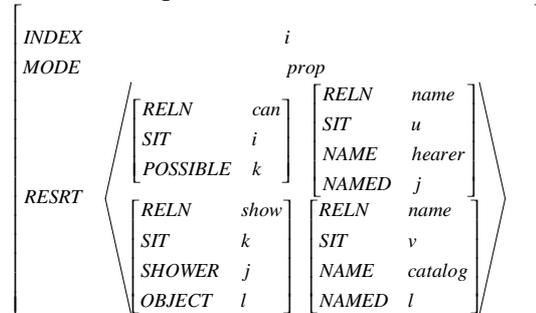


Figura 3: Representación semántica no agentiva

Sin embargo, *puedes mostrar el catálogo a Juan* es una oración ambigua ya que *poder* puede emplearse en su sentido de capacidad y también puede interpretarse como una pregunta acerca de las capacidades que Juan tiene. En este último caso *puedes* es del tipo de sujeto de control; este tipo es similar al tipo de subida de sujeto ya que su sujeto y el de su complemento son también el mismo; sin embargo, a diferencia de los auxiliares, que carecen de agente, los verbos con sujeto de control se realizan de manera agentiva. La descripción del tipo *Subject-control verb* se muestra en la figura 4.

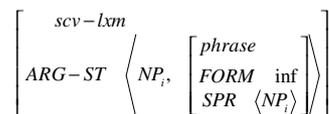


Figura 4: Subject-control verb

El carácter agentivo de estos verbos se captura al requerir no sólo que el sujeto del verbo y del complemento sean el mismo, sino también que haya alguien que pueda realizar la acción nombrada por el verbo. En su sentido auxiliar, por su parte, al no ser *poder* un verbo agentivo, indica sólo la posibilidad de que la acción denotada por el complemento sea realizada.

La diferencia entre el análisis de *poder* como auxiliar y como verbo de contenido puede apreciarse comparando sus representaciones semánticas correspondientes en las figuras 3 y 5; en el caso auxiliar la relación de *poder* se establece entre una situación en la que no hay agente y la situación que es posible (figura 3), mientras que en el caso agentivo *poder* es una situación que se establece entre el agente que puede, que es también el que muestra, y la capacidad que es posible (figura 5). Vale la pena recalcar que este tipo de ambigüedad es mucho menos frecuente en el

inglés pues sus auxiliares, a diferencia del español, han sido casi completamente gramaticalizados.

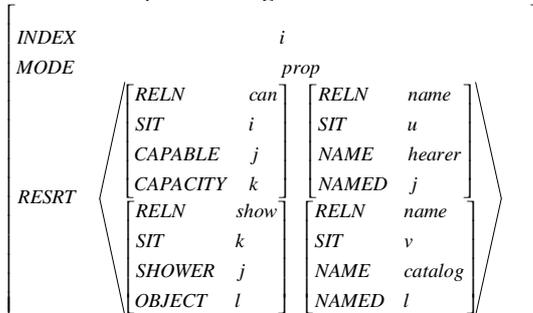


Figura 5: Representación semántica agentiva

Pasamos ahora al análisis de los auxiliares copulativos *ser* y *estar* en las oraciones atributivas. Estos verbos son también del tipo *subject-raising* pero a diferencia de las oraciones con predicación verbal en las que lo que se comparte es el sujeto del verbo de contenido de la perífrasis y el sujeto de la oración en su conjunto, en las atributivas se comparte el sujeto que posee la predicación nominal (e.g. la propiedad adjetiva o nominal especificada como el atributo) con el sujeto de la oración. Como uno de sus sellos distintivos el español tiene dos auxiliares para esta función: *ser* y *estar*. Intuitivamente, *ser* se utiliza para hacer atribuciones permanentes (*el cielo es infinito* pero no *el cielo está infinito*) mientras que *estar* se utiliza para hacer atribuciones transitorias (*La fruta está madura* pero no *la fruta es madura*). Sin embargo, esta distinción no es suficiente: tanto *el niño es alto* como *el niño está alto* son oraciones gramaticales e idiomáticas, pero su significado es diferente. Más aún, no se puede decir *Juan es muerto*, a pesar de que nada puede ser más permanente que dicha condición, y se tiene que decir *Juan está muerto*.

El uso de estos verbos se aclara cuando se toma en cuenta su carácter aspectual; mientras que *ser* es un verbo imperfectivo, *estar* es perfectivo, y estos auxiliares comunican este aspecto a sus atributos. Los predicados nominales con *ser*, como *es azul* y *es infinito* representan a estados o procesos no acotados, que no llegan nunca a un término, por lo que los atributos en estas oraciones se refieren a propiedades que el sujeto tiene como coexistentes con su propia naturaleza. Por otro lado, el carácter perfectivo de *estar* da a la predicación nominal un carácter de hecho acabado al que se llega después de un proceso: *María está guapa* denota el hecho consumado al que María llega después acicalarse. *El niño es alto* significa que el niño tiene esta propiedad de manera permanente, probablemente por que todos en su familia son altos; *el niño está alto* significa en cambio que el niño se encuentra en el proceso de crecer, y estar alto es la culminación de este proceso. Se dice *está muerto* porque este estado se alcanza cuando culmina el

proceso de vivir, y no se puede decir *ser muerto* porque morir no es un proceso, a menos que se esté hablando en sentido figurado, o metafórico: *es hombre muerto*, se refiere al hecho cuando la condena se ha cumplido. Lo que *es* se sabe pero lo que *está* hay que percibirlo: sabemos que un niño es alto porque lo conocemos y sabemos que está en este estado de manera más o menos permanente, pero si después de algún tiempo de no verlo nos sorprende lo que ha crecido decimos que está alto. El verbo *ser* se usa para lo que se sabe mientras que *estar* se usa para enunciar lo que se experimenta.

Presentamos ahora el modelo de las atributivas en HPSG. Como ya se ha dicho *ser* y *estar* son del tipo *srv-lxm*. Por lo mismo, el especificador del predicado nominal se comparte con el sujeto de la oración. Por su parte, la entrada léxica del atributo especifica que éste modifica a su propio especificador, como se muestra en la entrada léxica de *alto* en la figura 6 donde el especificador del adjetivo (i.e. el sujeto que tiene la propiedad de ser alto) está coindexado con la frase nominal que dicho adjetivo modifica, es decir con el valor del atributo MOD del HEAD de esta entrada léxica.

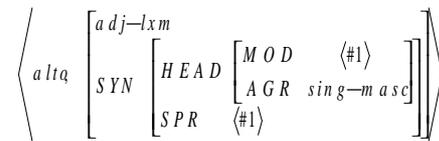


Figura 6: Entrada léxica para alto

El análisis de la oración atributiva procede combinando el auxiliar con el atributo mediante la *HCR*, para formar la predicación nominal, y mediante la *HSR* para formar la oración completa, como se muestra en la figura 7.

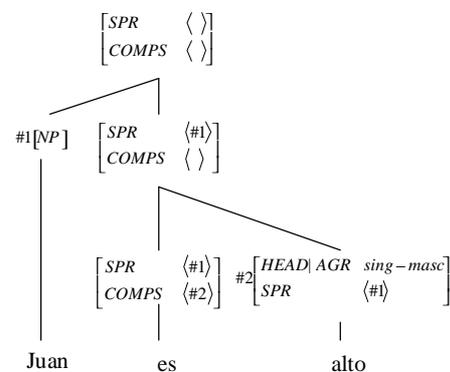


Figura 7: Estructura para oración atributiva

Los análisis sintácticos de las atributivas con *ser* y *estar* son similares y la diferencia de significados se representa en la semántica. Aunque un análisis adecuado de la semántica de las atributivas tiene que expresarse en el contexto de una teoría general de los

tiempos del español, lo que está fuera del alcance de este artículo, es posible adelantar de manera intuitiva la semántica de este tipo de oraciones. En las restricciones de la semántica del predicado nominal se indica si el propiedad nombrada por esta predicación es permanente del sujeto que lo posee, o es una propiedad que el sujeto puede llegar a tener como culminación de un proceso. En la semántica *infinito*, por ejemplo, esta propiedad se marca como condición permanente o imperfectiva, y en la semántica de *madura* como hecho culminado o perfectivo; por otro lado, si el adjetivo puede tomar ambos valores según se combine con *ser* o con *estar*, el valor de esta propiedad queda indeterminado en la entrada léxica. Por otro lado, la semántica de los copulativos especifica, además de la información temporal, su valor aspectual. Para el efectos del análisis, el valor del aspecto del verbo y el adjetivo están coindexados, por lo que si el copulativo y el atributo son compatibles, la oración atributivo tendrá el valor aspectual correspondiente, y si no lo son, la combinación no es posible y la oración no es gramatical. La semántica en nuestro ejemplo se muestra en la figura 8.

Cuando los auxiliares *ser* y *estar* aparecen en pasivas, comunican a sus predicaciones su carácter aspectual de la misma manera que en las atributivas; por lo mismo, la pasiva con *estar* se usa con verbos perfectivos mientras que la pasiva con *ser* requiere verbos imperfectivos: se puede decir *María es querida* pero no *María está querida*; asimismo, se puede decir *la puerta está abierta*, pero no *la puerta es abierta* a menos que el significado de esta última oración se complete con un modificador que complete el sentido de un proceso reiterativo *la puerta es abierta todas las mañanas*. El análisis de estas oraciones se presenta adelante en la sección 6 donde se aborda la voz pasiva.

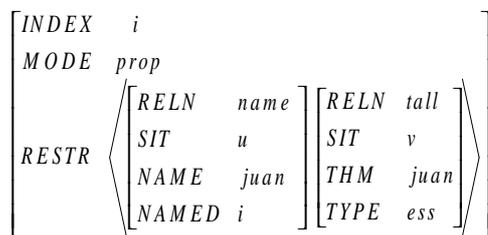


Figura 8: Semántica de atributiva con *ser*

Concluimos esta sección observando que este devenir del estado o proceso de ser que culmina en el hecho de estar, tan peculiar del español, se aprecia también en la complementariedad de estos verbos en la conjugación: los tiempos perfectos del imperfectivo *ser* corresponden a los tiempos imperfectos del perfectivo *estar*: un libro *está escrito* cuando *ha sido escrito*, la casa *está construida* cuando *ha sido construida*, el equipo *está derrotado* cuando *ha sido derrotado*.

### 3. El sistema de pronombres clíticos

Las expresiones perifrásticas mostradas en la sección anterior contienen frecuentemente pronombres clíticos en lugar de los argumentos explícitos del verbo de contenido. Por ejemplo *puedes mostrar el catálogo a Juan* puede realizarse de manera mucho más directa como *se lo puedes mostrar* o *puedes mostrárselo*. Más aún, este tipo de pronombres ocurre también con mucha frecuencia en construcciones en las que no intervienen verbos auxiliares. En general, el sistema de pronombres clíticos, ampliamente utilizado en el español como en otras lenguas romances, ha sido objeto de una gran cantidad de estudios lingüísticos (e.g. véase Miller y Sag, 1995, y las referencias ahí contenidas); sin embargo, a pesar de su importancia y la atención recibida, el nivel de representación de los pronombres clíticos es sujeto de controversia. Por un lado hay evidencia de que son fenómenos de carácter estrictamente morfológico, mientras que por otro se arguye que los pronombres clíticos son objetos léxicos independientes y por lo mismo pertenecen al nivel de representación sintáctico.

Aquí se presenta un análisis que toma en cuenta ambas tendencias y se propone que los pronombres clíticos, ya sea de manera independiente o agrupados secuencialmente, constituyen unidades morfológicas indisolubles, independientemente de que aparezcan en forma proclítica o enclítica. Sin embargo, estos grupos pueden ser sujetos de una operación sintáctica, y pertenecen a este último nivel de representación gramatical. De esta observación se derivan varias restricciones que se imponen a las construcciones con pronombres clíticos, con base en las cuales en este trabajo se presenta un modelo del sistema de pronombres clíticos que satisface todos los fenómenos que hemos observado en el español, y permite, hasta donde sabemos, expresar los sistemas clíticos de otras lenguas romances como el francés y el italiano, aunque en este artículo no se abordarán estas aplicaciones del modelo.

A continuación se presenta un modelo en HPSG que permite analizar oraciones incluyendo pronombres clíticos atómicos o en grupos y en forma enclítica y proclítica. Los elementos esenciales del modelo son:

- Las entradas léxicas de los verbos que pueden recibir pronombres clíticos como argumentos.
- Las reglas léxicas que modifican a los verbos en el proceso de cliticización.
- Una regla gramatical para combinar una secuencia de uno o más pronombres clíticos con su verbo.
- Las entradas léxicas de las *palabras clíticas* secuencias posibles de clíticos en el lenguaje.

A continuación se muestran los elementos del modelo a través del análisis de *muestras el catálogo*

a Juan así como de sus formas clitizadas *se lo muestras* y *muéstraselo*. La entrada léxica de *muestras* contiene un especificador y dos complementos como se aprecia en la figura 9.

$$\left[ \begin{array}{l} SPR \quad \langle NP \rangle \\ COMPS \langle NP \ NP \rangle \end{array} \right]$$

Figura 9: Entrada léxica de *muestras*

En esta forma básica la palabra *muestras* debe ser seguida por su complemento acusativo y dativo (qué y a quién); sin embargo, cuando ocurre en construcciones con pronombres proclíticos la forma en la figura 9 se modifica como se indica en la figura 10.

$$\left[ \begin{array}{l} SPR \quad \langle NP \rangle \\ COMPS \quad \langle \rangle \\ CL-LIST \langle NP, NP \rangle \end{array} \right]$$

Figura 10: Forma clitizada de *muestras*

Esta transformación se logra mediante una regla léxica llamada de *clitización proclítica* cuyo efecto es consumir a los complementos del verbo en el rasgo *COMPS* e incluirlos en lista de clíticos en el rasgo *CL-LIST*. De este modo el verbo se transforma en una frase verbal que pospone la combinación con sus argumentos. Por otro lado, la secuencia de pronombres “*se lo*” se considera una unidad morfológica indisoluble y se combina con el verbo en una sola operación mediante la regla de *Head Proclitic Rule (HPR)* que se propone en el presente trabajo. La entrada léxica de la secuencia de clíticos se muestra en la figura 11 y la regla *HPR* en la figura 12; el análisis sintáctico de nuestra oración de ejemplo se muestra en la figura 13.

$$\left\langle se \ lo, \left[ \begin{array}{l} clitic \\ CL-LIST \langle noun[CASE \ dat], \ noun[CASE \ dat] \rangle \end{array} \right] \right\rangle$$

Figura 11: Entrada léxica de compuesto clítico

En la figura 12, el atributo *CL-LIST* contiene los argumentos del verbo clitizado en el orden requerido para su combinación; asimismo, en la figura 13 se puede apreciar como la combinación de una palabra o complejo clítico con una frase verbal clitizada produce una frase verbal en la que tanto la lista de complementos como la lista de clíticos están saturadas. El último paso del análisis para la combinación del sujeto con la frase verbal en la figura 13 se logra mediante la aplicación de la regla *Head-PRO Drop Rule*, como ya se ha visto.

$$\left[ \begin{array}{l} SPR \quad \langle \#2 \rangle \\ COMPS \quad \langle \rangle \\ CL-LIST \quad \langle \rangle \end{array} \right] \rightarrow \left[ \begin{array}{l} clitic \\ CL-LIST \quad \#1 \end{array} \right] H \left[ \begin{array}{l} verb \\ SPR \quad \langle \#2NP \rangle \\ COMPS \quad \langle \rangle \\ CL-LIST \quad \#1 \end{array} \right]$$

Figura 12: Head Proclitic Rule

La definición de la *HPR* previene la inserción de material léxico entre la palabra clítica y la frase verbal con verbo clitizado, ya que estos dos constituyentes deben combinarse directamente. Se puede decir *al estudiante se ayuda*, o con repetición del acusativo, *al estudiante se le ayuda* pero no *se al estudiante ayuda* y tampoco *se le al estudiante ayuda*. En esta unidad entre del complejo clítico y el verbo se apoya la hipótesis de que el fenómeno es estrictamente morfológico; sin embargo, aunque en nuestro modelo ésta es una propiedad de la forma final, no lo es así su proceso de combinación que utiliza una regla sintáctica.

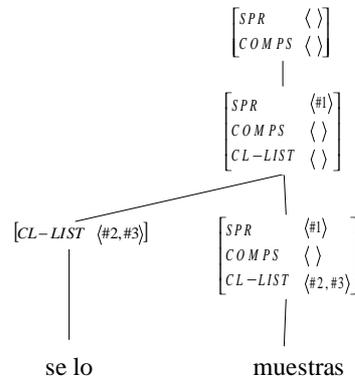


Figura 13: Estructura de oración proclítica

Pasamos ahora a mostrar el análisis para la forma enclítica *puedes mostrárselo*. En este caso el verbo de contenido está en forma no personal (i.e., infinitivo) y los pronombres *se* y *lo* son una sola forma léxica. En nuestro análisis, la entrada básica del verbo *mostrar* se modifica mediante una regla léxica para consumir sus dos argumentos y formar una frase verbal saturada en una sola operación. En la figura 14 se muestra dicha regla para clitizar el verbo *mostrar* con dos argumentos enclíticos. La forma resultante se puede combinar con el sujeto, en este caso omitido, de manera similar a la construcción en la figura 13.

$$\left\langle mostrar, \left[ \begin{array}{l} SPR \quad \langle NP \rangle \\ COMPS \langle NP \ NP \rangle \end{array} \right] \right\rangle \Rightarrow \left\langle mostrarme, \left[ \begin{array}{l} SPR \quad \langle NP \rangle \\ COMPS \quad \langle \rangle \end{array} \right] \right\rangle$$

Figura 14: Regla de clitización enclítica

La efectividad del análisis puede apreciarse de manera más clara considerando la semántica de la oración. En el caso enclítico la regla léxica incluye de manera explícita los argumentos del verbo, como se indica en la figura 15.

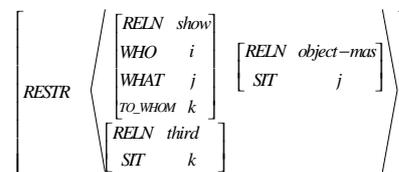


Figura 15: Semántica de frase verbal enclítica

Una característica sumamente distintiva de los sistemas de clíticos en diferentes lenguas es el orden estricto pero idiosincrásico que se da entre los pronombres; en español el orden es: *se, te, me, le, lo, la*, donde *lo* y *la* nunca aparecen juntos. Por ejemplo *se lo das*, pero no *lo se das* o *me lo dices* pero no *lo me dices*, etc. Adicionalmente este orden se conserva independientemente de que los pronombres se presenten de manera pro o enclítica; por ejemplo: *dáselo* pero no *dálose*, *dímelo* pero no *dílo*. Asimismo, en español todos los pronombres son proclíticos o enclíticos: *se lo das* o *dáselo* pero no *se daslo* (*lo das se* se descarta de antemano ya que *se* siempre precede a *lo* ya sea en pro o en enclíticos). El orden en el caso proclítico se captura en la definición de la entrada léxica de la palabra clítica como se muestra en la figura 11; los diferentes ordenes se capturan definiendo una palabra clítica para cada orden posible. Por su parte, para el caso enclítico, imponemos la restricción de que las formas enclíticas permitidas corresponden exclusivamente a las palabras clíticas del lexicon. Por lo mismo, el orden en el caso enclítico corresponde al del caso proclítico. Por ejemplo, la regla de cliticación en la figura 14 puede asignar el *me* dativo y el *lo* acusativo en la morfología ya que la palabra “*me lo*” está en el lexicon. Asimismo, la regla preserva la relación entre los argumentos explícitos y los pronombres clíticos correspondientes en la semántica de la frase verbal enclítica, como se muestra en la figura 15. Adicionalmente, la semántica de la frase verbal saturada en la combinación de “*se lo*” con *muestras* produce para el caso proclítico la misma estructura que la regla léxica para el caso enclítico, como se puede apreciar comparando el segundo nivel en la figura 13 con la figura 14 (i.e. ambas requieren de un sujeto para completar su predicación).

Concluyendo, el presente análisis postula la existencia de un conjunto de secuencias de pronombres clíticos que se consideran como unidades léxicas: “*se*”, “*te*”, “*me*”, “*le*”, “*lo*”, “*la*”, “*se te*”, “*se me*”, “*se lo*”, “*se la*”, “*te me*”, “*te lo*”, “*te la*”, “*me le*”, “*me lo*”, “*me la*”, etc., un conjunto de reglas léxicas para cliticar los verbos, y una regla de combinación sintáctica para formar complejos clíticos. A pesar de que esta lista de palabras clíticas es parcial y no incluye los órdenes de clíticos posibles, ya sea el peninsular o en las diversas variedades latinoamericanas, los plurales y las combinaciones de clíticos singulares y plurales posibles, se puede apreciar que el número de estas secuencias es finito y bajo; por lo mismo, este mecanismo permite modelar con un costo computacional aceptable, centrado principalmente en procesamiento del lexicon, las complejidades del fenómeno de clíticos debidas a su orden rígido e idiosincrásico.

## 4. Interacción entre sistemas

Los sistemas de verbos auxiliares y pronombres clíticos no son independientes; en español es posible que exista un verbo auxiliar o una frase verbal encabezada por un verbo auxiliar entre dichos pronombres y el verbo de contenido de la oración. Este fenómeno se conoce como *clitic climbing* o *subida de clíticos* y ha sido sujeto a numerosos estudios (e.g. Monachesi 1993, Miller y Sag, 1995). A continuación se muestra cómo en el modelo propuesto, los sistemas verbos auxiliares y pronombres clíticos interactúan sistemáticamente y permiten el análisis gramatical de oraciones con ambas construcciones, incluyendo subida de clíticos.

En *se lo puedes mostrar*, por ejemplo, “*se*” y “*lo*” son argumentos de *mostrar* a pesar del material intermedio *puedes*, y para interpretar correctamente esta oración es necesario co-indexar la realización explícita de cada pronombre con el argumento verbal que le corresponde. Para nuestro ejemplo, el análisis debe permitir establecer que *lo* es lo mostrado y *se* es a quien se muestra. Para capturar este fenómeno es necesario extender la definición de los verbos que toman como complemento una frase verbal, ya sean estos subida de sujeto, sujeto de control y también de objeto de control, ya que estos tres tipos de verbos son los que aparecen en perífrasis. Asimismo, es necesario considerar que una palabra clítica contiene normalmente clíticos de un sólo verbo, pero también hay que tomar en cuenta dos casos adicionales: cuando los clíticos son argumentos compartidos por dos o más verbos, y cuando la palabra clítica está formada por argumentos de verbos diferentes.

Primero consideramos el caso en que todos los clíticos son argumentos del mismo verbo. En *puedes mostrar el catálogo a María* los objetos directo e indirecto de *mostrar* (i.e. *el catálogo* y *María*), pueden substituirse por pronombres clíticos dando lugar a las formas *puedes mostrarselo* y *se lo puedes mostrar*; en ambos casos la palabra clítica “*se lo*” contiene tan sólo argumentos del verbo *mostrar*, ya que el argumento de *puedes* es toda la frase verbal *mostrarselo*.

Sin embargo, en la oración *ve a María comerse la manzana*, sus formas clíticas *ve a María comersela* y *se la ve comer* contienen un argumento compartido: *María* es el sujeto de comer y *se* es un pronombre reflexivo (i.e. dativo ético) que funciona como complemento de *comer* y está correferenciado con *María* (de ahí la reflexividad), y la oración subordinada quiere decir que *María* se comió a la manzana en su propio beneficio. Sin embargo, *a María* es también el objeto directo de *ver*, es decir, es lo visto, por lo que *se* es un argumento que se sale del alcance local de *comer* y se comparte con *ver*. En este caso, el pronombre

clítico *se*, que surge de la clitización de *comer*, subsume a los otros dos argumentos correferenciados.

La relación correferencial entre el objeto de *ver* y el sujeto de *comer* permite que clíticos de diferentes dominios locales se subsuman y pasen al frente de la oración, pero si la correferencia no existe o el objeto directo de *ver* se realiza explícitamente (sin subsumirse) los clíticos tienen que permanecer en sus dominios locales. Por ejemplo, no se puede decir *ve a María se comersela*, ni *ve a María se comerla*, ni *ve a María se la comer* ni *tampoco se ve a María la comer*; se obtiene el mismo resultado si se substituye a *María* por un clítico acusativo: *ve la se comerla*, *vela se la comer*, *se ve la la comer*. Sin embargo, si se puede decir *la ve comersela*, *a María ve comersela*. Lo interesante es que en los casos no gramaticales el objeto directo de *ver*, es decir *a María*, se realiza explícitamente (es decir no se subsume) y crea una barrera que los clíticos de *comer* (*se la*) no pueden pasar. El fenómeno se complica aún más si consideramos oraciones como *a María la ve comersela* e incluso *se la ve a María comer* en las que *a María* se realiza explícitamente, pero como veremos más adelante, estas últimas formas son gramaticales porque en español los argumentos pueden ocurrir de manera pleonástica y además pueden intervenir fenómenos de orden de palabras. Pero, en ausencia de estos fenómenos, los clíticos correferenciales tienen que subsumirse, o el material se tiene que realizar ya sea como clítico o de manera explícita en sus dominios locales. Si no se toma en cuenta esta condición de localidad la oración *la ve comersela* no sería gramatical ya que hay un *la* antes que un *se* por lo que se violaría la restricción de orden; sin embargo, en este caso, dicha violación no existe porque *la* y *se* pertenecen a dominios locales diferentes: el primer *la* es un argumento de *ver* mientras que *se* y el segundo *la* son argumentos de *comer*.

Pasamos ahora al caso en que la palabra clítica contiene argumentos (no correferenciales) de verbos diferentes. Ilustramos este caso con la oración *puede ver a María poner el sombrero sobre la mesa*; sus formas clíticas son *puede verle ponerlo* y *se lo puede ver poner* con el pronombre *se* acusativo. En la segunda oración el pronombre *le* acusativo, un caso claro de leísmo, substituye a *a María*, el objeto directo de *ver*. El cambio de *le* a *se* de la segunda a la tercera forma es posible ya que el clítico *se* personal, a diferencia del *se* reflexivo que deriva del reflexivo latino *se*, es una duplicación de *le* que el español derivó, aparentemente, por razones fonológicas: no se puede decir “*le lo*” pero el equivalente “*se lo*” es correcto. En la última oración el compuesto clítico “*se lo*” se deriva de la concatenación de dos clíticos independientes que surgen de la clitización de dos verbos diferentes. Aquí designamos a esta operación como composición

clítica, o simplemente composición. Aunque esta operación permite combinar clíticos compuestos con muy pocas restricciones, sólo aquellas combinaciones que resulten en palabras clíticas del lexicon pueden ser generadas, ya que de otra manera no pueden ser canceladas por la regla de clitización proclítica, ni por ninguna regla de cancelación léxica, como veremos más adelante. En caso de que la composición de clíticos de dominios diferentes se vea bloqueada por esta última razón, los clíticos tienen que aparecer en sus dominios locales correspondientes. Por ejemplo, se puede decir *puede verla ponerse*, pero si queremos combinar el *la* de *ver* con el *selo* de *poner* no es posible decir *puede verlaselo poner*, ni *la se lo puede ver*, y lo más que podemos hacer es decir *puede ver ponerse*.

Como conclusión, los clíticos se insertan en sus dominios local mediante la clitización de verbos pero pueden moverse con gran libertad a cualquier posición de la perífrasis; estos movimientos son posibles siempre y cuando no existan conflictos de alcance entre los dominios locales, y los clíticos pueden subsumirse si son correferenciales, o combinarse si no lo son; la restricción global de este proceso combinación se da en última instancia por conjunto de palabras clíticas en el lexicon, en las que se codifica el orden, y que se requieren para cancelar a cualquier compuesto clítico introducido directamente, o producido mediante las operaciones de subsumisión y composición. A continuación presentamos el modelo de este sistema en HPSG.

Para capturar el fenómeno de subida de clíticos es necesario extender la definición de los verbos que toman como complemento una frase verbal, ya sean estos *Subject-raising* o *Control-raising*; para este propósito se extiende la definición de los verbos auxiliares para permitir que no sólo el especificador del auxiliar y su complemento se compartan, sino también para que la lista de clíticos del auxiliar se comparta con la de su complemento, como se muestra en la figura 16.

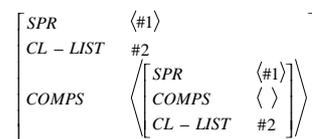


Figura 16: Entrada léxica de *puedes*

Además es necesario que la lista de clíticos pase hacia “arriba” en el proceso de clitic climbing. Para este propósito se define la lista de clíticos como un atributo de la valencia verbal, similar a la lista de complementos. El valor de este atributo es la lista de clíticos propiamente. En el proceso de combinación sintáctica y al igual que el resto de los atributos de valencia (i.e. el sujeto o especificador y los complementos verbales) el valor de estos atributos para

el nodo madre de la estructura sintáctica es el mismo que el de la hija “cabeza” a menos que la regla de combiación sintáctica defina explícitamente el valor de este atributo (i.e. el principio de valencia). Subsecuentemente, del mismo modo que la reglas *head-specifier (HSR)* y *head-complement (HCR)* saturan al especificador y a los complementos respectivamente, la regla *head-proclitic (HPR)* satura a la lista de clíticos. Para el caso de verbos no clítizados la lista de clíticos es simplemente vacía, por lo que el mecanismo funciona de manera general.

A continuación ilustramos el modelo con el análisis de la oración *se lo puedes mostrar*; en esta oración el auxiliar *puedes* toma como complemento la forma clítizada de *mostrar*. Esta combinación se logra mediante la aplicación de la *HCR* y la estructura resultante se combina con la forma clítica “*se lo*” mediante la aplicación de la *HPR*; finalmente el análisis de la oración completa se logra mediante el uso de *Subject-PRO Drop Rule* al igual que en los ejemplos anteriores. El árbol sintáctico se muestra en la figura 17. Esta construcción sintetiza las propiedades de ambos modelos de manera simple y con gran generalidad.

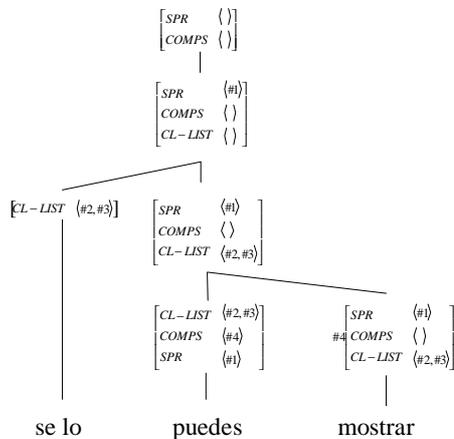


Figura 17: Oración con subida de clíticos

La generalidad del modelo puede ser mejor apreciada si se consideran oraciones con varios auxiliares y verbos de tipo de sujeto de control que pasan la lista de clíticos con el mismo mecanismo, y no hay restricciones en el número de verbos que puede haber en la perífrasis. El mecanismo funciona para oraciones como *se lo pude haber mostrado*, *se lo hubiera podido haber mostrado*, *se lo hubiera querido haber podido mostrar*, etc. Todos estos verbos toman como complemento a una frase verbal que contiene a la lista de clíticos, la cual simplemente se pasa al siguiente verbo de la perífrasis, hasta que se cancela por la regla de combinación proclítica.

En español es también posible que quien se clítice sea cualquiera de los verbos auxiliares: *hubiera querido habérselo podido mostrar*. La enclisis es incluso posible en los participios aunque la gramática tradicional prescribe que los argumentos clítizados de los participios deben ser proclíticos por lo que *hubiera queridoselo haber podido mostrar* y *hubiera querido haber podídoselo mostrar* no se consideran gramaticales. De modo similar la enclisis del verbo en forma finita de la perífrasis no se usa por lo que *hubieraselo querido haber podido mostrar* tampoco es gramatical. Sin embargo, estas restricciones obedecen más a la costumbre lingüística que a una restricción expresiva de los sistemas de auxiliares y clíticos, ya que estas oraciones tienen un significado claro. En general, en la enclisis la lista de clíticos no sube hasta el frente de toda la frase verbal, sino que “cae” en medio de la secuencia de auxiliares. En este caso, la intuición lingüística española es que estos pronombres nunca pueden realizarse como unidades léxicas independientes y tienen que ser enclíticos: en la ortografía no se puede escribir *hubiera querido haber se lo podido mostrar*. La razón es probablemente de orden fonético ya que en la curva melódica del español la parte tensiva de la oración se da en el verbo, por lo que todo lo que le sigue es siempre enclítico y, de hecho, esta es también probablemente la razón por la que en español los proclíticos siempre están diferenciados y se sienten como palabras completas y no como inflexiones, como en el caso enclítico. En la perífrasis, en particular, el acento de intensidad principal se da en el verbo que está al frente, por lo que los clíticos posteriores son siempre enclíticos.

Para capturar la enclisis en nuestro modelo es tan sólo necesario agregar una regla léxica para la clítización de los auxiliares y verbos de sujeto de control que no aparezcan como participio. En la definición de esta regla se considera que al estar ya el clítico como sufijo del auxiliar, la lista de clíticos de su complemento no debe “subir” o “envolver” al auxiliar, por lo que si el auxiliar se clítiza su lista de clíticos es vacía. En nuestro análisis, la regla léxica que clítiza al auxiliar (i.e. *habérselo*) cancela a los pronombres que vienen en la lista de clíticos de su complemento (i.e. el *se lo* de *hubiera querido habérselo podido mostrar*). Esta regla léxica se muestra en la figura 18 y el análisis de la oración *hubiera querido habérselo podido mostrar* en la 19.

Este análisis ofrece también una predicción interesante: en toda oración sólo se puede clítizar un auxiliar y todos los pronombres deben ser enclíticos y respetar el orden. En español se puede decir *hubiera querido habérselo podido mostrar*, pero no se puede decir *hubiera querido haberse podido mostrarlo*. Esta intuición refuerza la generalidad de nuestro modelo: al ser los grupos posibles de clíticos palabras en el

lexicón, y al haber una palabra clítica para cada orden posible, solo pueden insertarse grupos completos en el proceso de selección léxica, ya sea que se realicen al frente de la oración como proclíticos, o como sufijos enclíticos en auxiliares o en verbos de contenido.

$$\langle \#2, [COMPS \#1] \rangle \Rightarrow \left\langle F_{clit-verb}(\#2), \left[ \begin{array}{l} CL-LIST \langle \rangle \\ COMPS \#1([CL-LIST \langle [NP-dat] [NP-acc] \rangle]) \end{array} \right] \right\rangle$$

Figura 18. Regla léxica de cliticación de auxiliares

Pasamos ahora a mostrar el modelo para la subida de clíticos con subsumisión de argumentos; para ilustrar el análisis consideramos la oración *ve a María comerse la manzana* y su forma clítica *se la ve comer donde se* es un pronombre reflexivo. En esta oración *a María* es el objeto directo del verbo *ver*, lo que se ve, y el sujeto de *comer*, es decir quien se come la manzana; *la manzana* es el objeto directo de *comer*, y el *se* de *comerse* es un dativo de ético o de participación correferencial con *María*. En la forma clítica, *se la ve comer* el reflexivo *se* designa a un argumento de *ver* y a dos de *comer* y estos tres argumentos hacen referencia a *María*. Para efectos del modelo, la cliticación del objeto directo y el dativo de *comer* se produce por la regla léxica para la cliticación proclítica y los dos argumentos se ponen en la lista de clíticos sin alterar sus correferencias con la estructura de argumentos, al tiempo que se satura la lista de complementos del verbo, como ya se ha visto.

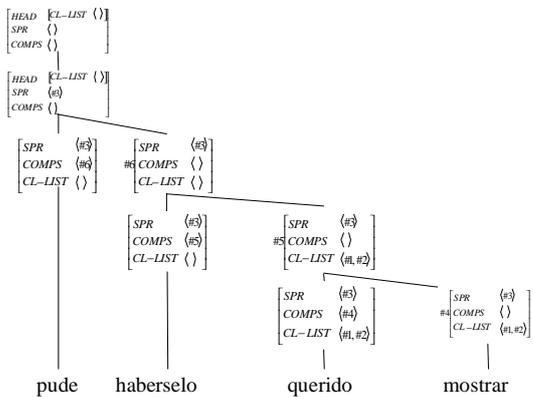


Figura 19. Oración con auxiliar cliticado

Para completar el modelo de la oración hay que considerar como se combina la lista de clíticos de *comer* con la de *ver*. Como ya se ha dicho, *ver* es un verbo de objeto de control que tiene la característica que su objeto directo es también el sujeto del verbo que le sigue en la perífrasis; asimismo, al ser uno de los verbos constitutivos de la perífrasis tiene como uno de sus complementos a la frase verbal que lo modifica. Es decir, *ver* tiene como complemento la frase verbal *comérsela*. La relación entre el objeto directo de *ver* (i.e. lo que se ve) y el sujeto de *comer* se establece coindexando el primer complemento de *ver* con el

especificador de su segundo complemento, es decir con el sujeto de *comer*. Estas correferencias se establecen de antemano en la entrada léxica de los verbos de tipo de control, como se muestra en la figura 20.

$$\left[ \begin{array}{l} ocv-lxm \\ CL-LIST \langle [1] \oplus \#a \rangle \\ SPR \langle \#1 \rangle \\ COMPS \left\langle \begin{array}{l} \#2, \#3 \\ \left[ \begin{array}{l} phrase \\ CL-LIST \langle \#a \rangle \\ SPR \langle \#2 \rangle \end{array} \right] \end{array} \right\rangle \\ ARG-ST \langle \#1, \#2, \#3 \rangle \end{array} \right]$$

Figura 20. Verbos de tipo de objeto de control

La Figura 20 muestra la forma básica de los verbos de objeto de control. Esta forma es neutra respecto a la frase verbal que lo modifica; ésta puede ocurrir con sus argumentos realizados de manera explícita o con sus argumentos cliticados. Por ejemplo, *ver* puede combinarse con *comer la manzana*, *poner el sombrero al niño* o *poner el sombrero en la mesa* y también con sus correspondientes formas clíticas *comerla*, *ponérselo* y *ponerlo*. En ninguno de estos casos, los pronombres clíticos del complemento están correferenciados con el objeto directo del verbo de control y su propio sujeto ya que no son reflexivos. Sin embargo, en el caso de *se la ve comer*, donde *se* es el dativo de participación de nuestro ejemplo principal, es necesario establecer la correferencia reflexiva. Para este efecto se utiliza la regla léxica en la Figura 21

La función de esta regla es transformar la forma de la entrada léxica de los verbos de objeto de control para el caso en que su complemento contiene un pronombre reflexivo. Operativamente, la regla clitiza al objeto del verbo control y lo coindexa con el primer elemento de la lista de clíticos de su segundo complemento, subsumiendo de este modelo al objeto directo de *ver* en la lista de complementos de *comer*. Una vez establecidas estas referencias, el complemento cliticado se elimina de la lista de complementos del verbo de objeto de control, y la lista de clíticos del verbo que le sigue pasa a ser la lista de clíticos del propio verbo de objeto de control, como se puede apreciar en la Figura 21.

$$\left\langle \#1, \left[ \begin{array}{l} ocv-lxm \\ CL-LIST \langle \rangle \\ COMP \langle [1], [CL-LIST \#a(1,1)] \rangle \end{array} \right] \right\rangle \Rightarrow \left\langle \#1, \left[ \begin{array}{l} CL-LIST \#a \\ COMP \langle [CL-LIST \#a] \rangle \end{array} \right] \right\rangle$$

Figura 21. Regla de subsumisión clítica

Con la entrada léxica del verbo de objeto control modificada para la operación de subsumisión, el análisis de la oración clítica con complementos reflexivos es directo y simple. El verbo de control se combina con su complemento por medio de la HCR y el análisis se completa cancelando la palabra clítica con la HPR, como se muestra en la Figura 22.

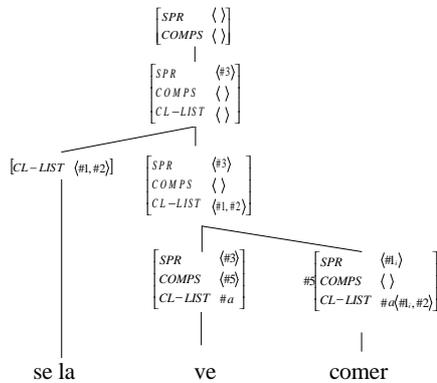


Figura 22. Estructura de la oración con subsunción de clíticos

Pasamos ahora al modelo de composición clítica. La operación de composición se define directamente en la entrada léxica básica de los verbos de objeto de control como se muestra en la figura 20. Como se puede apreciar, la lista de clíticos de los verbos de este tipo es la suma de una lista de clíticos subespecificada (que corresponde a la lista de clíticos del propio verbo una vez que éste se clítice) con la lista de clíticos del verbo que le sigue, es decir, con la lista de clíticos del segundo complemento del propio verbo de objeto de control.

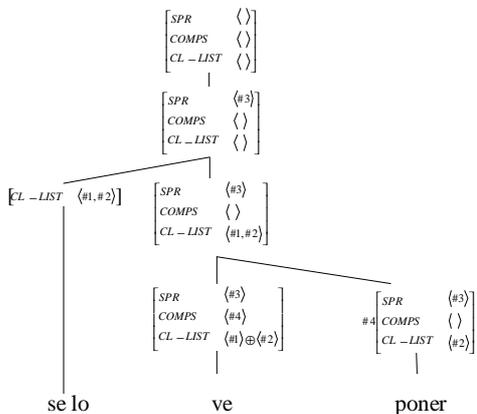


Figura 23. Estructura de la oración con composición de clíticos

Si alguno de estos verbos no está clítizado (i.e. el verbo de control y el que le sigue) su lista de clíticos es vacía, por lo que el verbo no contribuye a la lista de clíticos que sube; de hecho, si ambas listas están vacías ya que todos los argumentos de ambos verbos se realizan de manera explícita, lo que sube es simplemente la lista vacía, ya saturada.

Ilustramos este análisis con la oración *ve a María poner el sombrero* (en la mesa) y su forma clítica *se lo ve poner*. En este caso *se* substituye *a María* y su caso es acusativo, como ya se ha dicho. El análisis de la

oración *se lo ve poner* se desprende directamente de estas consideraciones, como se muestra en la figura 23.

Concluimos esta sección con el análisis de los pronombres complementarios que ocurren de manera redundante en español. De hecho, los pronombres clíticos anticipan en muchas ocasiones a los complementos que substituyen, y la presencia explícita de dichos complementos tiene por función especificar y elaborar al pronombre, de la misma manera que presencia explícita del sujeto en la oración española tiene por función especificar o elaborar al sujeto contenido en la conjugación, especialmente cuando la indeterminación puede dar lugar a ambigüedad de tipo referencial. Los pronombres pleonásticos pueden ocurrir con gran libertad, y además aparecen en prácticamente cualquier posición con respecto al resto de los elementos de la oración, con las restricciones para el orden de palabras que se discuten abajo en la sección 8. Sin embargo, estos pleonasmos no aumentan la complejidad de nuestro modelo de clíticos, y de hecho su análisis es hasta cierto punto trivial. Lo único que se requiere es dar una regla de clitización alternativa que al clitizar un argumento e incluirlo en la lista de clíticos, no lo remueva de la lista de complementos, cuidando tan sólo que el complemento y el clítico están coindexados. El complemento pleonástico se cancela con la HCR, pero esta operación se da de modo independiente a las operaciones que afectan al pronombre clítico que le corresponde. En la figura 24 se muestra la regla para clitizar de manera redundante al objeto directo de un verbo, y en la 25 se muestra como la entrada léxica resultante de esta regla participa en el análisis de la oración *se la ve a María comer*:

$$\langle \#1, [COMPS \langle \#2 \rangle] \rangle \Rightarrow \langle \#1, [CL-LIST \#2_i, COMPS \langle \#2NP_i \rangle] \rangle$$

Figura 24. Regla de clitización con pronombres pleonásticos

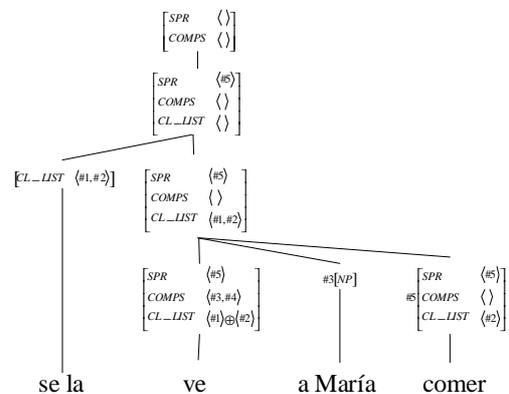


Figura 25. Estructura de oración pleonástica

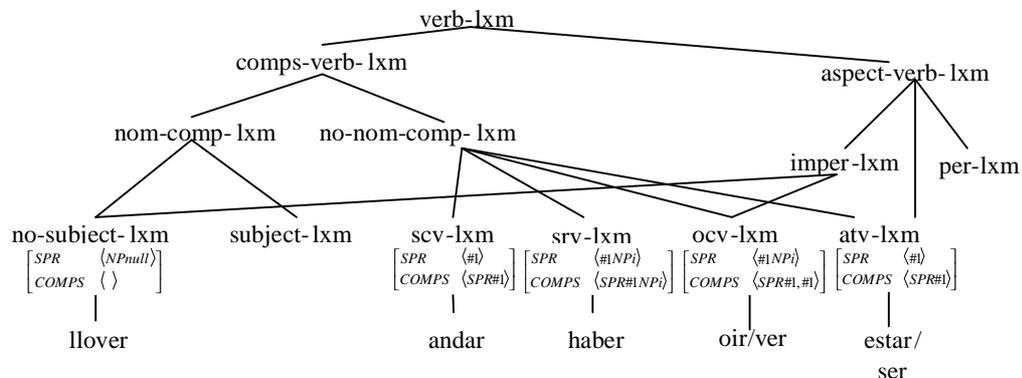


Figura 26. Jerarquía de tipos (I)

## 5. Jerarquía de tipos verbales y estructura de argumentos

El español es una lengua con gran riqueza en su estructura léxica; esta estructura puede ser capturada de manera natural a través de los formalismos gramaticales orientados hacia el lexicón, como es el caso de HPSG. La expresión de la jerarquía de lexemas verbales, en particular, permite modelar de manera económica las diferentes formas que los verbos adquieren a través de la inflexión, que en español marca para el verbo el modo indicativo, subjuntivo e imperativo (10 tiempos simples o compuestos para el indicativo y 7 para el subjuntivo) y un verbo puede llegar a tener más de 20 inflexiones verbales diferentes.

Por otro lado el español presenta varios tipos de oraciones que se desvían de la sintaxis normal debido a la naturaleza léxica de los verbos, por lo que deben tratarse de manera individual, como el caso ya visto de la pasiva en tiempo imperfecto con verbo perfectivo que requiere necesariamente de un complemento circunstancial para completar el sentido imperfectivo de un proceso repetitivo (e.g. *la puerta es abierta por el portero*).

Otra idiosincrasia se presenta en oraciones con verbos con propiedades ergativas como *gustar* en los que los roles semánticos del sujeto y objeto directo están invertidos; en estas oraciones el sujeto designa al recipiente de la acción mientras que el objeto directo designa al agente. Por ejemplo, *los chocolates le gustan a María* es una oración activa donde *los chocolates* es el sujeto sintáctico, a pesar de ser un sujeto paciente. Más aún, gracias a la libertad del orden de palabras del español, este sujeto puede posponerse (*a María le gustan los chocolates*) y *los chocolates* sigue siendo el sujeto ya que concuerda en número y persona con el verbo (i.e. tercera del plural). El sujeto está también contenido en la conjugación y puede omitirse si no necesita ser desarrollado o determinado, y el beneficiario cliticarse (*le gustan*), e incluso de manera

redundante (*le gustan a María*); asimismo, si el complemento indirecto ocurre sin redundancia la realización del clítico es obligatoria: se puede decir *le gustan* pero no *a María gustan*, a menos que el dativo siga al verbo (*gustan a María*). La representación de este sistema de interacciones en la sintaxis no se podría hacer sin incurrir en complejas excepciones; sin embargo, la representación de estos fenómenos en el lexicón es simple y el procesamiento de estas estructuras tiene un costo computacional bajo, como se mostrará al concluir esta sección y más adelante cuando abordemos el problema del orden de palabras.

La orientación léxica del análisis es también esencial para el modelo del sistema de clíticos. Como ya se ha visto, las reglas de cliticación enriquecen significativamente la productividad del lexicón. Adicionalmente, el grupo de oraciones con *se* (i.e. reflexivas, pasiva refleja e impersonal activa) dan lugar a numerosas construcciones idiosincráticas cuyo análisis presenta retos muy complejos a la sintaxis, pero que pueden ser modeladas de manera natural si se representan en el lexicón.

Pasamos ahora a mostrar la jerarquía de lexemas verbales y la estructura de argumentos correspondiente. Como se puede apreciar en la figura 26, la estructura es realmente una lattice (i.e. no sólo los nodos padres pueden tener más de hijo sino también los hijos pueden tener más de un padre); la raíz de esta estructura es el tipo *verb-lxm* al cual pertenecen todos los verbos; a partir de éste surgen los tipos *comps-verb-lxm* y *aspect-verb-lxm*. El primero es la raíz de la jerarquía verbal propiamente y captura la estructura de argumentos; el segundo codifica el carácter perfectivo o imperfectivo de los tipos verbos, que se codifica a este nivel dado al impacto del valor de los atributos aspectuales no sólo en las oraciones predicativas sino también en las atributivas, como ya se ha visto.

El tipo *comps-verb-lxm* se divide a su vez en *nominal-comp-lxm* y *no-nominal-comp-lxm*. En el

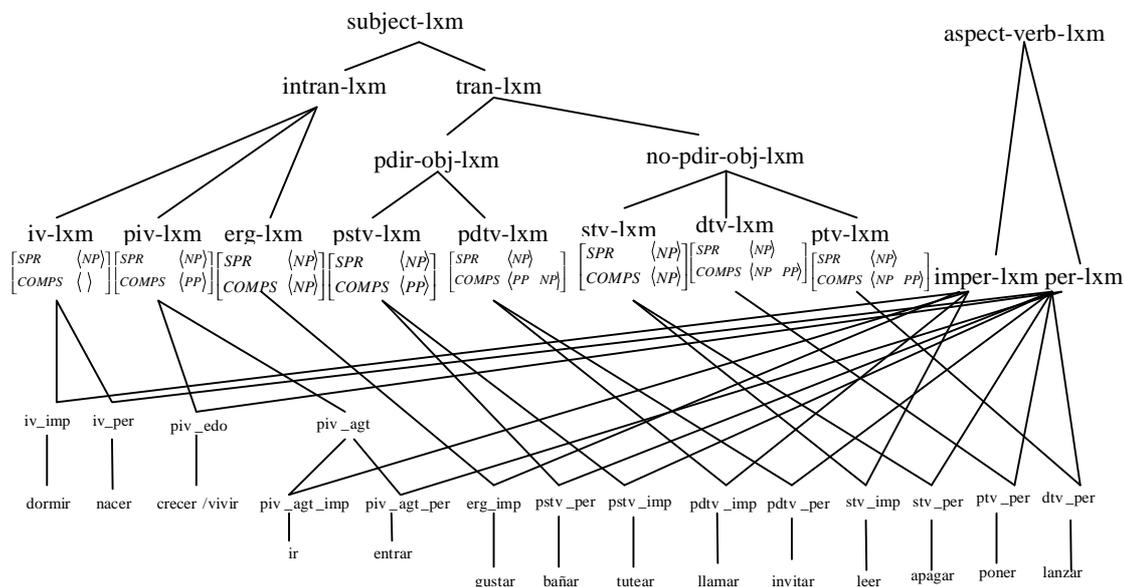


Figura 27. Jerarquía de tipos

primero se agrupan aquellos verbos que toman como complementos frases nominales (i.e. una frase nominal propiamente o una frase preposicional donde el objeto de la preposición es una frase nominal). Por otro lado, en el tipo *no-nominal-verb-lxm* están los verbos que toman como complemento una frase verbal o una frase adjetiva.

El tipo *no-nominal-comp-lxm* se divide en cuatro tipos: *srv-lxm*, *scv-lxm*, *ocv-lxm* y *atv-lxm*. Los dos primeros se trataron en la sección dos en relación con el sistema de verbos auxiliares. El tercero agrupa aquellos verbos que tienen dos complementos, el primero verbal y el segundo nominal, pero donde el segundo complemento actúa como sujeto del primero. A este tipo pertenecen los verbos que nombran a los procesos de la percepción sensorial. En *oigo cantar a los pájaros*, por ejemplo, *los pájaros* es no sólo el segundo complemento de *oigo* sino también el sujeto de *cantar*. Finalmente, el tipo *atv-lxm* contiene a *ser* y *estar* tanto en su uso copulativo en las atributivas como en su uso como auxiliares en la formación de la voz pasiva.

El tipo *nominal-comp-lxm* se divide en aquellos verbos que necesitan un sujeto agente (*subject-lxm*) y los impersonales (*no-subject-lxm*). El tipo *no-subject-lxm* agrupa a verbos como *llover* que carecen de agente pero determinan un sujeto en la conjugación de la tercera persona del singular. Por su parte, el tipo de *subject-lxm* concentra a varios tipos verbales como se muestra en la figura 27. Este tipo se divide en los verbos intransitivos (*intr-verb-lxm*) y los transitivos (*tran-verb-lxm*).

En el tipo *intr-verb-lxm* se encuentran verbos que no tienen complemento obligatorio con carácter imperfectivo (*iv\_imp*) como *dormir*, y con carácter perfectivo (*iv\_per*) como *nacer*; dentro de este tipo se encuentran también los llamados verbos de estado (i.e. intransitivos e imperfectivos que normalmente requieren una frase preposicional) de tipo *piv\_edo* como *vivir* y *crecer*. Estos verbos denotan estados o procesos y, por lo mismo, la actividad de sus agentes se encuentra sumamente disminuida, al grado que se sienten casi como pasivas ya que el sujeto más que llevar a cabo una acción está en un estado o proceso (*María vive en Cuernavaca*), y casi como atributivas, ya que el sujeto tiene la cualidad marcada por el predicado casi nominal (*María crece contenta*). Dentro de los intransitivos también se encuentran los verbos de tipo *piv-lxm* que requieren como complemento una frase preposicional, tanto con aspecto imperfectivo (*piv\_agt\_imp*) como *ir* (*María va al cine*) como con aspecto perfectivo (*piv\_agt\_per*) como *entrar* (*María entra a la casa*). Al carecer de objeto, los intransitivos no pueden ponerse en pasiva, incluso si tienen algún otro tipo de complemento; no se puede decir, por ejemplo, *el camino es ido por mí* ni *se va el camino* (i.e., las versiones pasiva y pasiva refleja *voy por el camino* respectivamente). Finalmente, el grupo de los verbos intransitivos incluye también el tipo de verbos ergativos *erg-lxm*, de carácter imperfectivo y con las propiedades mencionadas arriba.

Los verbos transitivos se dividen en los verbos cuyo objeto directo es una frase preposicional con la preposición *a* (*pdir-obj-lxm*) y aquellos en el que no la llevan (*no-pdir-obj-lxm*). En español los complementos

directos de persona van precedidos de la preposición *a* mientras que los complementos directos de cosa van sin preposición; este recurso expresivo se desarrollo en la lengua para evitar confusión entre el sujeto y el objeto directo cuando estos intercambian su posición en el orden oracional, como se verá más adelante cuando abordemos este problema; la *a* antepuesta identifica al objeto directo frontalizado en el caso de que sea agentivo, y como los objetos de cosa no pueden ser el sujeto en oraciones activas pueden anteponerse sin que surja ambigüedad. El tipo *pdir-obj-lxm* se divide a su vez en los tipos *pstv-lxm* y *pdv-lxm*, que toman uno y dos argumentos respectivamente. Ejemplos de estos tipos son *tutear* (*Juan tutea a María*) e *invitar* (*Juan invita a los niños a la fiesta*). Estas oraciones se pueden poner en pasiva, y la preposición *a* en la activa evita toda ambigüedad posible respecto a la identidad del objeto directo que se convierte en sujeto pasivo sin preposición: *María es tuteada por Juan, los niños son invitados a la fiesta*.

Por su lado, los verbos transitivos que toman como primer complemento una frase nominal (*no-pdir-obj-lxm*) se dividen tres tipos; primero están los que tienen un solo complemento (*stv-lxm*) como *comer* en *como la fruta*; el segundo caso corresponde a los verbos ditransitivos con dos complementos nominales (*dtv-lxm*) como *dar* en *doy el regalo a Juan*. Por último están los verbos que tienen como complementos una frase nominal y una frase preposicional; por ejemplo, *poner* en *pone el plato en la mesa*. Al contener un objeto directo, todas estas oraciones pueden ponerse en pasiva; sin embargo, como ya se ha dicho, hay ciertas restricciones ya que los auxiliares y los verbos deben ser compatibles en su carácter aspectual.

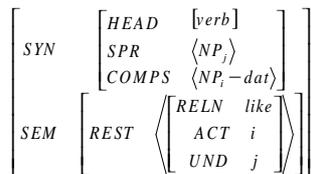


Figura 28: Entrada léxica de *gustar*

Para mostrar el poder expresivo de la maquinaria que hemos construido hasta este punto, a continuación se presenta el análisis de una oración con el verbo *gustar* y con un clítico replicado: *los chocolates le gustan a María*. El análisis se logra mediante la cliticación proclítica del verbo *gustar* a partir de su lexema original, cuya definición se muestra en la figura 28. En esta entrada léxica, el sujeto sintáctico ( $NP_i$ ) se encuentra coindexado con el recipiente de la acción en la semántica (*UNDerger*), mientras que el objeto sintáctico en dativo ( $NP_i$ ) está coindexado con el actor pasivo (*ACTor*) en la semántica, dando origen a la construcción ergativa.

En el contexto de la oración *los chocolates le gustan a María* el complemento dativo se repite, lo cual es también una característica distintiva del sistema de clíticos del español en relación a otras lenguas con clíticos como el francés y el italiano, en las que esta repetición no es aceptable. Para lograr la combinación es necesario que la regla de cliticación tome en cuenta dicha duplicidad y sea posible consumir a *a María* como un complemento, al tiempo que la forma *le* quede en la lista de clíticos, tomando en cuenta, además, que estos dos constituyentes deben estar coindexados. La regla léxica para este efecto se muestra en la figura 29.

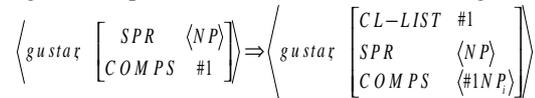


Figura 29: Regla de cliticación proclítica con repetición de clítico

Con esta entrada léxica es posible combinar *gustan* (se omite aquí la regla léxica para la inflexión en presente del indicativo para la tercera persona del singular) con el argumento dativo *a María* mediante la regla *HCR*, consumir el clítico mediante la regla de combinación pro-clítica (*Head-proclitic rule*) y, finalmente, consumir el sujeto *los chocolates* con la regla *HSR*, como se muestra en la figura 30.

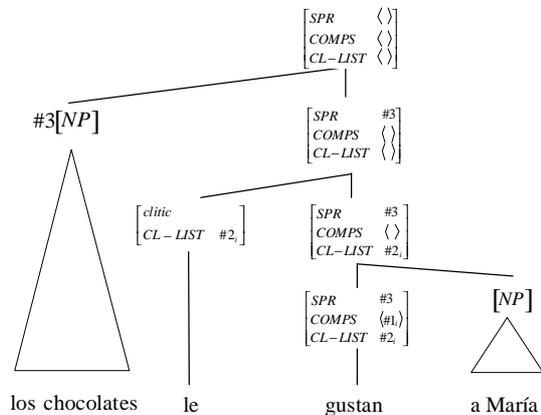


Figura 30: Análisis de la construcción ergativa

Este análisis es simple y tiene un costo computacional razonable, como lo demuestra nuestra implementación; sin embargo, es aún incompleto: es necesario todavía capturar las formas con sujeto pospuesto *le gustan los chocolates a María* y *le gustan a María los chocolates*, que son gramaticales e idiomáticas, además de las formas sin sujeto *le gustan a María* o simplemente *le gustan*, y sin clíticos *gustan a María* y *a María gustan*. Sin embargo, por el momento posponemos el análisis de estos casos hasta la sección 7 en la que se abordan algunas estrategias para tratar el problema del orden de palabras en el presente marco de trabajo

Tipo de oración	Tipo de origen	ejemplos
Pasiva	Transitivos imperfectivos	<i>Juan es querido</i>
Pasiva (Modificación adverbial)	Transitivos perfectivos	<i>La puerta es abierta en la mañana</i>
Pasiva refleja	Transitivos	<i>La puerta se abre por la mañana</i>

Tabla 1. Oraciones pasivas

## 6. Pasiva refleja, impersonales y reflexivas

Equipados con el sistema de verbos auxiliares, clíticos y la estructura de argumentos es posible capturar otros fenómenos gramaticales que se presentan con frecuencia en el español como la pasiva refleja, las oraciones impersonales y las oraciones reflexivas. Las variaciones más importantes de estos tipos de oraciones tienen en común que utilizan la forma léxica *se*. En la pasiva refleja (*se abre la puerta*) la partícula *se* es simplemente un signo sintáctico de esta forma oracional (Gili Gaya, pp. 129), pero en la impersonal activa (*se vive bien en Cuernavaca*), *se* es ya un pronombre, aunque su referente no está determinado y por lo mismo es impersonal. Sin embargo, cuando el verbo es agentivo (*se le dice, se le puede decir*) el pronombre *se* encuentra ya en el límite con su interpretación personal. A pesar que *le* es normalmente un dativo puede también tomar el caso acusativo dando origen al fenómeno conocido como leísmo, y la interpretación de *se* es que hay alguien determinado aunque anónimo, quien dice, y *le* se refiere a quien recibe la acción de decir, es decir a quien se le dice. En caso de que la oración se dé sin leísmo (*se lo dice*) la determinación del *se* es aún más clara, aunque sigue siendo anónimo y por lo mismo impersonal. Independientemente del grado de determinación del *se*, el carácter anónimo de su referente clasifica a estas oraciones en la impersonal activa.

En su forma personal *se* es un pronombre referencial, para la tercera persona del singular y el plural; es un pronombre personal proclítico y recibe el análisis que se muestra en las secciones 3 y 4; sin embargo, en su forma personal *se* es también un pronombre reflexivo (*se viste, se lava las manos*) y de hecho ésta es fue su función original en el lenguaje. Asimismo, aparece en oraciones recíprocas (*los niños se tutean*). Con estos verbos *se* no puede repetirse; si *los niños se substituye* por un pronombre en este último ejemplo, se tiene que decir *se tutean unos a otros* pero no se puede *se se tutean*. De manera similar, se tiene que decir *uno se hace tonto* en vez de *se se hace el tonto*.

Por todas estas interacciones, el grupo de oraciones con *se* ha sido siempre problemático en la gramática tradicional del español y, hasta donde sabemos, su análisis computacional ha resultado sumamente elusivo. Afortunadamente, la maquinaria que hemos construido provee de una análisis simple y directo de un amplio espectro este grupo de oraciones. A continuación se presenta primero el análisis de la pasiva y la pasiva refleja; seguimos con las formas impersonales para concluir con las personales y las reflexivas y recíprocas.

Al igual que el inglés el español permite la voz pasiva mediante el uso de un sujeto paciente y el verbo de contenido en participio precedido del auxiliar *ser* y *estar*. La forma *Juan abre la puerta* tiene la forma pasiva *la puerta es abierta por Juan* (primera de pasiva). La omisión del agente ablativo da origen a las llamadas segundas de pasiva (*la puerta es abierta*). Esta forma oracional se captura fácilmente mediante una regla léxica de manera similar al análisis de la pasiva del inglés en HPSG y otros formalismos similares. También, como en la oraciones atributivas, el valor aspectual de los auxiliares *ser* y *estar* debe ser compatible con la frase encabezada por el participio, permitiendo oraciones como *la puerta está abierta* y *María es querida*, pero bloqueando *la puerta es abierta* y *María está querida*. A diferencia de las atributivas que no tienen verbo de contenido, el carácter perfectivo o imperfectivo del verbo en pasivas es parte de sus restricciones semánticas, y la oración pasiva se acepta o no dependiendo de la compatibilidad del valor aspectual del participio y del verbo auxiliar.

Sin embargo, y como ya hemos mencionado, la modificación de un participio perfectivo por un complemento circunstancial de causa, tiempo, modo o instrumento, transforma al hecho consumado en su reiteración no acotada, dándole un carácter imperfectivo a toda la predicación participa, como en *la puerta es abierta por la mañana*. Para permitir esta modificación se define una regla léxica que produce una forma del participio con aspecto imperfectivo, pero con un complemento circunstancial obligatorio. Con estos elementos, este tipo de oraciones recibe el análisis que se ejemplifica en la figura 31. La semántica es similar a la ilustrada para las atributivas en la figura 8. En este caso, el participio se siente casi como un adjetivo que marca una acción reiterativa como una

Tipo de oración	Tipo de origen	ejemplos
Activa impersonal	Transitivos	<i>Se abre la puerta</i>
Activa impersonal	Transitivos y leísmo	<i>Se llama a Juan, se le llama a Juan</i>
Activa impersonal	Intransitivos	<i>Se vive tranquilo en Cuernavaca</i>

Tabla 2. Oraciones activas impersonales

propiedad dinámica del sujeto pasivo. En español, como ya hemos dicho, esta forma expresiva, que marca el devenir que se inicia o se completa en un hecho, se debe a los sentidos complementarios de los auxiliares *ser* y *estar*.

Sin embargo, y a diferencia del inglés, en español la pasiva se usa poco y resulta en un estilo pesado y afectado tanto en la lengua escrita como en el lenguaje hablado. Como recurso expresivo alternativo, el español cuenta con la oración pasiva refleja en la que el verbo está en la voz activa y se precede de la partícula *se*. De acuerdo con Gili Gaya (pp. 126-127), la pasiva refleja tiene probablemente su origen en la reflexiva a través de una serie de derivaciones que van de la reflexiva directa donde el agente y paciente es el mismo (*yo me lavo*), a la indirecta donde el agente recibe la acción de manera indirecta y por lo mismo va en dativa (*me lavo las manos*), hasta el llamado dativo ético o de interés donde el agente ya sólo participa de la acción pero no la recibe (*tu te haces una casa*). Esta participación del sujeto se puede alejar gradualmente del *se* reflexivo y por analogía extenderse a verbos intransitivos de estado y finalmente a transitivos llegando a la pasiva refleja. En *la puerta se abre*, *la puerta* es ya un sujeto pasivo que recibe la acción y *se* es simplemente un signo de esta forma; asimismo, la frase verbal se puede frontalizar (*se abre la puerta*) captando la atención hacia el proceso genérico que se expresa y dando a la oración el carácter imperfectivo de repetición sin que sea necesario reiterarlo como en las segundas de pasivas con verbos perfectivos.

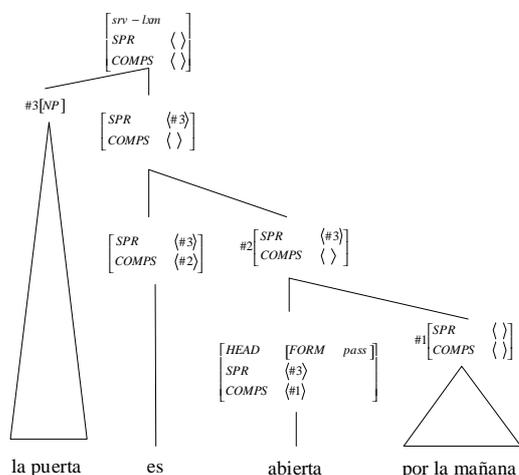


Figura 31. Oración pasiva

En nuestro análisis la pasiva refleja se modela con una regla léxica que toma verbos transitivos y convierte al objeto directo en el sujeto de la oración pasiva refleja. El signo de pasiva *se* se trata como un clítico y la regla léxica de pasiva refleja también clitiza al verbo; *se* está en tercera persona y debe concordar con el sujeto, por lo que verbo debe estar conjugado en tercera persona del singular o plural. Sin embargo, este *se* es no referencial, y así se marca en la semántica; para este efecto se asigna el valor *NON-REF* al atributo del modo (*MOD*) en la semántica del clítico. La regla léxica se muestra en la figura 32 y el análisis de *se abre la puerta* en la 33.

$$\left\langle \#1, \left[ \begin{array}{l} SPR \quad \langle NP \rangle \\ COMPS \quad \langle \#2, \#3... \rangle \end{array} \right] \right\rangle \Rightarrow \left\langle F_{3rd}(\#1), \left[ \begin{array}{l} SPR \quad \langle \#2 \& agr \rangle \\ COMPS \quad \langle \#3... \rangle \\ CL-LIST \quad \langle \{se \& agr\} \rangle \end{array} \right] \right\rangle$$

Figura 32: Regla léxica de pasiva refleja

Pasamos ahora al análisis de la impersonal activa. Hasta antes del siglo XV (Gili Gaya, pp. 61), las oraciones *se tutean los niños* o *se ayudan los estudiantes* eran ambiguas ya que podían interpretarse como pasivas (*los niños son tuteados, los estudiantes son ayudados*) o como reflexivas, en el sentido de que los niños se tutean unos a otros, y los estudiantes se ayudan mutuamente.

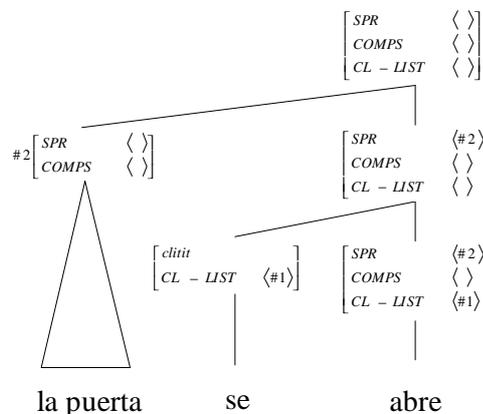


Figura 33: Pasiva refleja

Sin embargo, cuando el uso de la preposición *a* se consolidó como obligatorio para los complementos directos esta ambigüedad quedó resuelta: los reflexivos en singular y plural pasaron a ser *se tutea a los niños* y

Tipo de oración	Tipo de origen	ejemplos
Reflexiva directa	Transitivos	<i>Te vistes</i>
Recíprocas	Transitivos	<i>Se ayudan los estudiantes</i>
Recíprocas con pronombre <i>uno</i>	Transitivos e intransitivos con preposición	<i>Uno se hace tonto</i>
Pseudoreflejas	Transitivos e intransitivos	<i>Me voy, te estás</i>
Dativo ético y de participación	No ditransitivos	<i>Me pones la mesa</i>

Tabla 3 Reflexivas y recíprocas

*se ayuda a los estudiantes*; al ser *se* el único pronombre reflexivo para la tercera persona singular y plural se derivaron por analogía las formas *se tutea a los niños* y *se ayuda a los estudiantes*, quedando la oración en activa y con el objeto directo claramente diferenciado por la preposición, o alternativamente, por el pronombre dativo asumiendo funciones de acusativo (léismo): *se les tutea, se les ayuda*, o en singular *se le tutea, se le ayuda*. Esta forma se conoce como impersonal activa y es ya común para una gran variedad de verbos tanto intransitivos como transitivos y permite centrar la atención en la acción (*se ayuda al estudiante*), como en el paciente (*se le ayuda*), dejando al agente como indeterminado y anónimo.

A diferencia de la pasiva refleja, la activa impersonal puede tomar también verbos intransitivos y de estado: *se vive tranquilo en Cuernavaca*. Aquí el pronombre *se* está indeterminado completamente, con lo que la expresividad de la pasiva refleja se extiende a todo tipo de construcciones impersonales. La entrada léxica para este tipo de construcciones se genera a partir de la forma básica del verbo en el lexicon, y *se* se trata como un clítico, y se marca como no referencial al igual que en la pasiva refleja. El análisis para nuestro ejemplo se muestra en la figura 34.

En su forma personal *se* es un pronombre personal en tercera persona del singular y el plural y, exceptuando los reflexivos, funciona dentro del sistema de pronombres clíticos tal y como se discute en las secciones 3 y 4. En *se come las manzanas, se puede comer las manzanas, se las come y se las puede comer* la forma *se* es ya un pronombre personal pues el verbo es transitivo y tiene un carácter definitivamente agentivo. El sujeto puede permanecer anónimo si no hay interés de revelar su identidad, pero en este caso no se marca como no referencial, sino como referencial (el modo semántico es *REF*) pero con su valor (el índice semántico) no especificado. Por otro lado, *se* es anafórico cuando la oración se precede de un contexto lingüístico apropiado (e.g. *María está hambrienta, se come las manzanas*), o indexical, cuando la oración se dice en relación a un contexto espacial y temporal determinado (e.g. *María se come las manzanas* cuando

*María se está comiendo las manzanas*). Sin embargo, si el mismo verbo se usa como intransitivo (e.g. *se come bien en España*) la oración es impersonal activa.

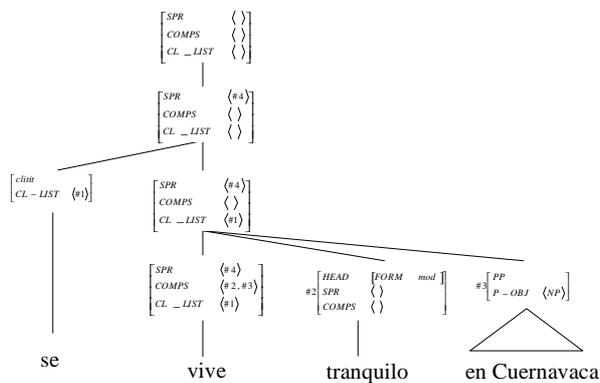


Figura 34: Impersonal activa

En las reflexivas *se* es también un pronombre personal. Este pronombre aparece en las reflexivas directas (*se viste, se baña, se moja*), en las indirectas (*se lava las manos*), como pronombre dativo (*se lo puedes mover a la izquierda*) y además en oraciones recíprocas (*uno se puede peinar todas las mañanas*); en todos estos casos y como ya se ha dicho, si *se* se repite, una de las ocurrencias correferenciales tiene que ser substituida por el pronombre *uno*. En este caso, ambas ocurrencias del pronombre son argumentos del verbo, y así debe quedar reflejado en la semántica. En nuestro análisis, los pronombres reflexivos son también considerados como clíticos y las formas verbales reflexivas son generadas por medio de una regla léxica. Esto es, el análisis de los reflexivos utiliza al sistema de clíticos (y de auxiliares) de manera simple y directa. La regla léxica que genera los verbos reflexivos también los clítiza, pero en este caso coindexando los argumentos correferenciales. En el caso de las reflexivas directas (*te vistes*), el objeto directo se incluye como un argumento en la semántica del verbo de manera similar a los pronombres enclíticos, produciendo una frase verbal que requiere un sujeto para convertirse en una oración completa. Adicionalmente, la regla léxica coindexa este objeto implícito con el sujeto, por lo que en la semántica el

agente y paciente de la acción son el mismo. En la figura 35 se presenta la regla léxica para las reflexivas directas; el análisis se ejemplifica con *te vistas* en la figura 36. Para las reflexivas indirectas y reciprocas se definen reglas léxicas similares, que clitizan y coindexan los argumentos de manera apropiada para caso.

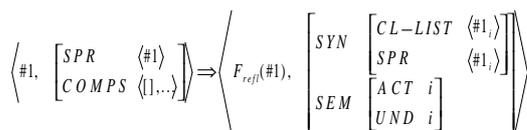


Figura 35: Regla léxica de reflexivas directas

Pasamos ahora al caso de las pseudoreflejas como *me muero*, *Juan se muere*, *se hace el listo*, *nos hacemos tontos*. En estas oraciones al ser el verbo intransitivo o de estado no hay acción que recaiga sobre el objeto paciente, y el sujeto tan sólo participa del estado o proceso nombrado por el verbo (morirse), o de la modificación nombrada por el complemento circunstancial (¿cómo se hace?). En este caso, la naturaleza sumamente idiosincrática de estas construcciones no permite generalizar una regla léxica, por lo que se recurre a la inserción directa en el lexicon los verbos que dan origen a esta construcción.

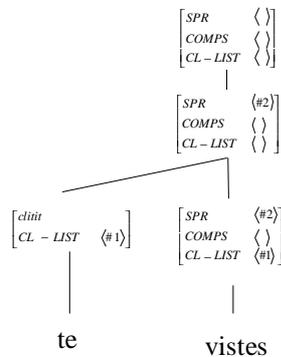


Figura 36: Estructura de reflexiva

## 7. Preguntas pronominales

Pasamos ahora al análisis de las preguntas pronominales del español, que corresponden a las *wh-questions* del inglés. Esta forma interrogativa pregunta o solicita alguno de los complementos del verbo: en la oración declarativa *Juan muestra el catálogo de muebles a María en la cocina a las ocho de la noche con mucho interés por que quiere vender*, los pronombres *quién*, *qué*, *a quién*, *dónde*, *cuándo*, *cómo* y *por qué* solicitan información acerca del complemento correspondiente. *Quién*, es un pronombre nominativo en *¿quién muestra el catálogo a María?*, *qué* es un acusativo en *¿qué muestra Juan?*, *a quién*

dativo en *¿a quién muestra Juan el catálogo?* y *dónde*, *cuándo*, *cómo* y *por qué* pueden considerarse ablativos que preguntan por el lugar, tiempo, modo y causa del suceso. En la pregunta pronominal el pronombre casi siempre se frontaliza (aunque esto no es obligatorio en español), y es necesario contar con un mecanismo que permita formar una representación semántica de la información explícita en la pregunta. El material solicitado se marca en el análisis con una variable en la estructura de argumentos; esta variable corresponde al complemento por el que se pregunta y la respuesta se formule en términos de los valores posibles para dicha variable en un contexto de interpretación determinado. Intuitivamente, *¿qué muestra Juan?* se interpreta como una función predicativa que con una variable que necesita tomar un valor para completar su significado (i.e.  $\lambda x. Juan' muestra' x$ ).

Para modelar de estas preguntas recurrimos aquí al mecanismo estándar con el que se cuenta para este propósito en HPSG y formalismos similares. Este consiste en el uso de las categorías incompletas, o con "gaps", que requieren de un argumento para completar su significación (Sag y Wasow, 1999). Por ejemplo, *qué* en *¿qué muestra Juan?* es una variable de tipo frase nominal (i.e. que toma como valores posible sujetos sustantivos) y *muestra Juan* es una oración con un gap, también de tipo frase nominal, que necesita combinarse con un objeto del mismo tipo (i.e. frase nominal) para completar su significado. En HPSG la regla sintáctica, llamada de gap, realiza precisamente esta última combinación.

Para el análisis de estas oraciones se utilizan reglas léxicas que modifican la estructura de argumentos del verbo removiendo el argumento que corresponde al pronombre interrogativo de la lista de complementos, al tiempo que dicho argumento se incluye en una estructura adicional llamada la lista de gaps. El efecto es similar a la de cliticización del verbo pero a diferencia de este último fenómeno que tiene carácter local, el gap puede correlacionar complementos en posiciones distantes entre sí (i.e. la posición donde se esperaría el complemento en el orden lineal en la oración declarativa, y la posición donde se realiza finalmente en la interrogativa), por lo que tiene un carácter global, mucho menos acotado. La regla léxica también considera que los pronombres tienen el caso que corresponde a sus complementos, por lo que la regla léxica asigna dicho caso a la categoría asociada al gap. Asimismo, la regla asigna el modo interrogativo a la semántica del verbo, y consecuentemente a la semántica de la oración. La regla léxica para los verbos que aparecen en preguntas con el pronombre acusativo *qué* se muestra en la figura 37.

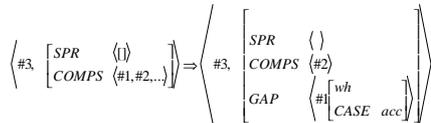


Figura 37: Regla léxica para verbos en oraciones interrogativas

Finalmente, si el verbo está cliticado la reglas léxicas de gaps y clítico se serializan para producir la forma verbal apropiada. En la figura 38 se ilustra el análisis de la oración pronominal con *qué* y con verbo cliticado *¿qué le muestras a María?*.

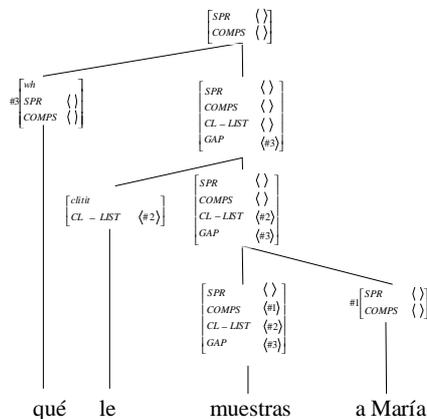


Figura 38: Estructura de la pregunta pronominal

El verbo *mostrar* tiene dos complementos en su forma original (lo y a quien se muestra), como se puede apreciar en la estructura de argumentos en la figura 19; la regla léxica para esta forma interrogativa remueve el objeto directo de la lista de complementos y lo incluye como un gap, en la lista que se define para este propósito. Por su parte, el complemento indirecto se cliticiza (de manera redundante) por la regla clítica, como ya se ha visto. Adicionalmente, al estar el pronombre interrogativo en la posición del sujeto, no es necesario incluirlo en la forma léxica como el especificador de la frase verbal, simplificando considerablemente el análisis. Con estos elementos el análisis procede de manera natural combinando *muestras* con *a María* con la HCR, combinando el clítico *le* con la frase verbal *muestras a María* para finalmente combinar consumir el gap, al combinar la frase verbal saturada pero con un gap con el pronombre acusativo *qué*. En la semántica, la restricción correspondiente al objeto paciente queda subespecificada, y la semántica de la oración es la función  $\lambda x$ . y *muestra x*, que tiene como argumentos las cosas que pueden ser mostradas (i.e. el catálogo, los chocolates, etc.). En esta función el sujeto queda indeterminado, pero puede tomar un valor a nivel pragmático en un contexto de interpretación específico (e.g. la persona con la que se está platicando en un

momento dado, o el programa con el que se está interactuando).

Si el pronombre está por un complemento circunstancial el análisis es muy similar. En *¿cuándo se lo muestra a María?*, la regla léxica que genera la entrada léxica del verbo tan sólo agrega un complemento marcado para el caso ablativo en la lista del gap; los complementos obligatorios del verbo no se modifican ya que el circunstancial es normalmente un modificador. El verbo puede también cliticarse, y el sujeto eliminarse de la lista SPR, ya que, de nueva cuenta, el pronombre interrogativo ocupa la posición del sujeto. Para la semántica, se aumenta simplemente un elemento a la lista de argumentos, el cual corresponde a la variable de la función que la pregunta representa: en este caso la función toma modificadores temporales como argumentos y (e.g. a las ocho de la noche, diariamente, etc.) y regresa valores de verdad (i.e. si el evento ocurre o no en dicho estado o situación temporal). En la figura 39 se muestra el análisis de *¿cuándo se lo muestra a María?*

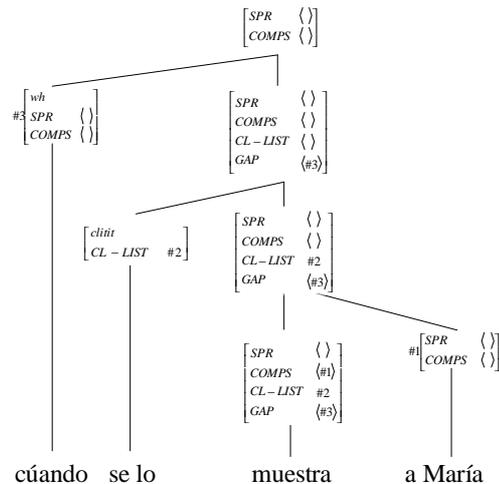


Figura 39: Pregunta con pronombre ablativo

## 8. Orden de los elementos oracionales

Concluimos este artículo con una introducción al problema del orden de palabras en español. A diferencia del inglés y el francés que tienden a la construcción lineal, también llamada analítica, en la que el determinante sigue al determinado (sujeto, verbo, objeto directo, indirecto, seguido de otros complementos), el español presenta adicionalmente una construcción envolvente en la que se realiza el carácter sintético de la lengua y los elementos más salientes en la intención del que habla se pueden anteponer en la oración. Por lo mismo, además del orden lineal, el español admite la frontalización de la frase verbal, de

los argumentos obligatorios del verbo, así como de cualquier complemento circunstancial. Esta flexibilidad tiene sus límites que dependen de factores sintácticos como fonéticos. Si la oración se emite en un solo grupo fonético, el verbo no puede ir más allá de la segunda posición en la oración sin que la oración deje de ser aceptable, pero si la oración se divide en más de un grupo fónico, las restricciones de orden se relajan y prácticamente cualquier orden es posible, siempre y cuando el grupo fónico de la frase verbal contenga al verbo en la primera o segunda posición del grupo. Por otro lado, si los complementos del verbo se expresan como pronombres clíticos, el verbo puede ir pospuesto (como en las oraciones proclíticas) sin que la oración deje de ser gramatical.

Una construcción común en español es la frontalización de los complementos circunstanciales; en *por la tarde a María le gusta tomarse su chocolate*, la frontalización del complemento circunstancial enfoca la atención en la situación temporal, cuándo a María le gusta tomarse su chocolate. Sin embargo, la frontalización no tiene que ser completa (e.g. *a María por la tarde le gusta tomarse su chocolate*) y realmente existe gran flexibilidad para la localización de estos elementos (con la entonación apropiada). Al indicar modificaciones de lugar, tiempo, modo, causa o instrumento, cualquiera de estos elementos puede moverse hacia adelante en la oración para ser resaltado. Por otro lado, si estos complementos son modificadores de los constituyentes mayores de la oración, como los genitivos, que en español se marcan con la preposición *de* (e.g. *los chocolates de María*), tienden a permanecer al final del constituyente, o a moverse hacia al frente pero sin salirse de las fronteras del elemento oracional al que pertenecen, y también sin salirse del grupo fónico en que dicho elemento se realiza, a menos que el grupo sea muy largo y el elemento tenga que partirse; sin embargo, en este último caso, este modificador se realiza como un grupo fónico independiente, evitando ser confundido como miembro de un elemento oracional diferente.

Los argumentos obligatorios del verbo pueden también moverse hacia adelante en la oración aunque de modo mucho más restringido y local, ya que estos deben permanecer bajo el control del verbo. La frase verbal también puede frontalizarse dejando al sujeto en posición pospuesta. También es frecuente que los objetos directo e indirecto se inviertan, aumentando el número de ordenes posibles. Adicionalmente, los adverbios pueden también quedar pospuestos, junto a la frase verbal, o frontalizados en la oración. Por todo esto, se ve que el orden de palabras en español es realmente muy flexible y su solución muy compleja. Por otro lado, no es suficiente dar un esquema de combinación para cada orden posible, ya que al tener los complementos la misma forma y usar las misma

preposiciones pueden surgir numerosas ambigüedades sintácticas, además de que el número de hipótesis de parseo crece muy rápidamente, y hace esta solución impráctica desde el punto de vista computacional.

Con estos antecedentes podemos formular el problema del orden de palabras en función de la inferencia necesaria para decidir la función sintáctica de cada uno de los elementos oracionales, y de este modo identificar la predicación que la oración expresa. En un orden lineal perfecto esta inferencia podría realizarse exclusivamente en función del orden en que estos elementos aparecieran en la oración: el primero correspondería al sujeto, el segundo al verbo, el tercero al objeto directo, etc. Sin embargo, este ideal analítico no es alcanzado por ninguna lengua, y es necesario atender a los factores que permiten realizar esta inferencia no sólo correctamente, sino también de manera sumamente eficiente, en tiempo real.

En español hay tres factores que permiten realizar esta inferencia. El primero es el significado de las preposiciones antepuestas a cada complemento. En segundo, la entonación y en tercer lugar el orden de colocación de los elementos en la oración.

A continuación, analizamos brevemente el significado y uso de las preposiciones con relación al problema del orden de palabras. En español el significado de la preposición, en conjunto con el significado léxico del término de la frase preposicional que forma a casi todos los complementos oracionales, es una de las pistas principales para identificar la función sintáctica del complemento. Como ya se ha dicho, el uso obligatorio de la preposición *a* para el objeto directo de persona o ente abstracto con alto nivel de determinación es uno de los factores principales. De acuerdo con Gili Gaya, este uso se desarrolló en la lengua precisamente para evitar la ambigüedad entre el sujeto y el objeto cuando estos invertían su posición. No existe ambigüedad alguna entre *Juan ama a María* y *a María ama Juan*, ni entre *el bien vence al mal* y *al mal vence el bien*. Por otro lado, los complementos de cosa van sin preposición pero al no ser estos agentivos no pueden ser el sujeto en oraciones activas; por lo mismo, no existe ambigüedad entre *Juan le da los chocolates a María* y *los chocolates le da Juan a María*. Asimismo, todo complemento indirecto lleva una de las preposiciones *a* o *para* para indicar el beneficiario de la acción, como *a María* en el último ejemplo. Aquí tampoco hay ambigüedad pues ya el sujeto y objeto ya están determinados por la presencia o ausencia de la preposición *a* en el objeto. Existiría ambigüedad sólo en el caso que tanto el paciente como el beneficiario fueran personas como en *Juan da a María a Pedro*. Sin embargo, esta oración ya no es gramatical gracias a que la sabiduría de la lengua hace que el objeto no sea ya más persona; de hecho se llega

a oír *Juan le da la María a Pedro*, cómo cuando el padre da a su hija en matrimonio en el lenguaje campesino. Como se ve, este sólo recurso expresivo de la lengua hace que el sujeto, el objeto directo e el indirecto puedan permutar su posición sin que se pierda u oscurezca su función sintáctica gracias a las restricciones mutuas que se establecen por el presencia o ausencia de la *a* en el complemento directo. Asimismo, cuando *a* aparece en complemento circunstancial tienen un sentido de dirección espacial o temporal: *voy a las ocho*, *voy para Cuernavaca* por lo que estos complementos tampoco se pueden confundir. Asimismo, las diferentes preposiciones en circunstanciales identifican lugares (*en*, *donde*), tiempos (*a las ocho*, *entre las tres y las cuatro*), modos (*como*) causas (*porque* y *para que*), instrumento (*con*), etc., y su función semántica queda casi siempre determinada.

Por lo mismo, conociendo el significado de la preposición *a* en cada complemento y tomando en cuenta el significado del material léxico en la frase preposicional, así como el sentido del verbo, es posible identificar de que complemento se trata: *a Juan* puede ser directo o indirecto pero nunca circunstancial y *a las ocho* será siempre circunstancial y las oraciones *Juan dio los chocolates a María a las ocho*, *Juan dio a María los chocolates a las ocho*, *a María dio Juan los chocolates a las ocho*, *a María dio los chocolates Juan a las ocho*, *a las ocho Juan dio los chocolates a María* y *a las ocho a María dio Juan los chocolates* no son ambiguas. Aunque gramaticales no todas estas oraciones son igualmente idiomáticas, y para ser aceptables tienen que pronunciarse con una entonación apropiada en relación con un contexto lingüístico o espacio-temporal bien definido. Por lo mismo una solución general al problema de la colocación de los elementos oracionales tiene que tomar en cuenta a la prosodia, lo cual está fuera del alcance del presente trabajo.

Como se ve la contribución del orden de los elementos oracionales a la determinación del tipo de complementos es en español sumamente reducida. Por lo mismo, un esquema en que la inferencia se haga a partir de esta fuente de información únicamente será siempre muy limitado. En particular, en formalismos gramaticales como HPSG en que la posición de los argumentos y modificadores se especifica en los esquemas de combinación (*HSR*, *HMR*, *HCR*, etc.) es posible crear versiones de estos esquemas para cada orden posible; sin embargo, esta multiplicación tiene que hacerse con mucho cuidado ya que los diversos tipos de complementos pueden tener la misma forma sintáctica y podrían producirse muchas hipótesis que deberían ser de inmediato descartadas.

Por otro lado, muchos fenómenos donde el orden de los argumentos se invierte aparentemente, son de orden local y se pueden modelar con reglas léxicas, como es el caso de las construcciones con clíticos. Por lo mismo, todos los fenómenos de clíticos y en general las idiosincrasias de las oraciones con *se* no son parte del problema del orden de palabras.

Tomando en cuenta estas consideraciones y las limitaciones implícitas de formalismos orientados más bien al orden lineal, como HPSG, aquí se propone una solución limitada dentro del nivel sintáctico del problema del orden de colocación de los elementos oracionales. La solución va en niveles, y primero tomamos en cuenta los fenómenos locales que ya se han analizado: oraciones con clíticos, reflexivas, pasiva refleja, impersonal activa, las oraciones ergativas, etc., que como ya se dijo no son problemas de orden y reciben soluciones independientes. En segundo lugar, aumentamos una regla léxica para invertir los complementos directos e indirectos de verbos ditransitivos; en tercer lugar, permitimos que la regla de gap mueva complementos realizados hacia el frente en la oración; en cuarto lugar introducimos una variante de la regla de *Head-Specifier Rule* para invertir al sujeto y al verbo, así como una variante de la *Head-Modifier Rule*, para que los modificadores (adverbiales y circunstanciales) puedan preceder al elemento oracional que modifican. Por último, aumentamos una regla léxica para introducir un complemento opcional adicional al final de la lista de complementos obligatorios del verbo en la estructura de argumentos. Este complemento es circunstancial por lo que lleva siempre una preposición que indica el tipo de modificación que dicho complemento indica. Por otro lado, no puede ser cliticizado, ya que los pronombres clíticos substituyen al complemento directo e indirecto, excepto en el caso del dativo ético que ya se ha tratado en el fenómeno de subsumisión de clíticos. Sin embargo este complemento opcional si puede ser substituido por una "gap" por lo que puede topicalizarse y moverse hacia el frente de la oración. La interacción de estos cinco mecanismos permite capturar un gran número de ordenes posibles de manera simple.

Esta estrategia es, como ya se ha dicho, limitada; sin embargo, permite extender la cobertura de la gramática y analizar de manera simple construcciones con varias interacciones. En la figura 40 se muestra la regla léxica para invertir los complementos directo e indirecto, y en la 41 la regla sintáctica que invierte al sujeto con la frase verbal. En la figura 42 se ilustra el uso de estas reglas en el análisis de *puedes girar noventa grados la cocina*, con objeto directo e indirecto invertidos, y que aparece en el corpus DIME.

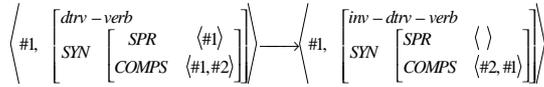


Figura 40: Regla léxica de inversión de complementos

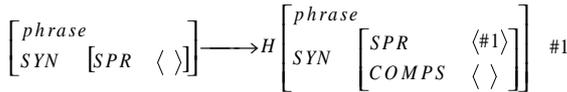


Figura 41: HSR invertida

Para ilustrar el uso de la regla de gap y su interacción con la regla del sujeto y el verbo presentamos a continuación el análisis de las oraciones *a María los chocolates le gustan* y *a María le gustan los chocolates*, que son variaciones de orden de la ergativa *los chocolates le gustan a María*, analizada ya en la sección 5.

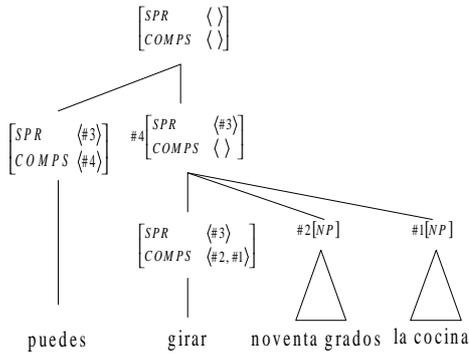


Figura 42: Inversión de complementos directo e indirecto

Como se recordara, las oraciones ergativas no tienen realmente complemento directo, por lo que *le* y *a María* son el complemento indirecto que se presenta de manera redundante; por lo mismo, la regla clítica pone al pronombre en la lista de clíticos y la realización explícita del complemento indirecto se introduce como un gap; en la entrada léxica del verbo ambas instancias están coindexadas. En la primera versión, menos idiomática, el verbo ergativo y el clítico se combinan con la *Head-Proclitic Rule*, y este compuesto toma su sujeto mediante la *HSR*; finalmente, la oración con gap consume al complemento circunstancial frontalizado mediante la regla de gap, de manera similar a los pronombres interrogativos, como se muestra en la figura 43.

La segunda variante es mucho más idiomática, a pesar de que el sujeto *los chocolates* va pospuesto. Para tratar con estos casos se introduce la regla de inversión de sujeto pospuesto para la *HSR*. En este caso, el compuesto clítico se forma de la manera usual, y esta frase verbal incompleta toma su sujeto pospuesto mediante la nueva versión de la *HSR*, para finalmente consumir el complemento indirecto *a María* con la

regla de gap; el análisis de *A María le gustan los chocolates* se muestra en la figura 44.

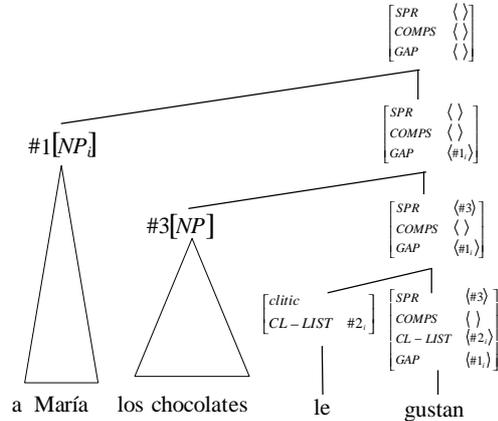


Figura 43: Estructura de gustar con gap

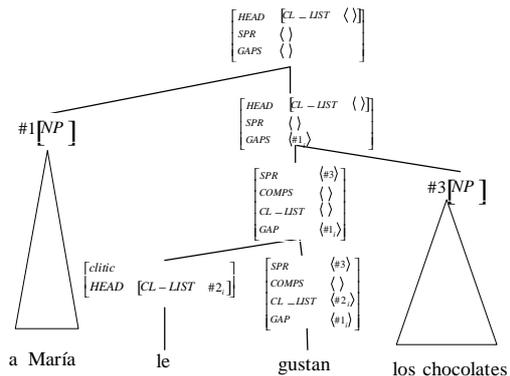


Figura 44: Estructura de gustar con gap y HSR invertida

Pasamos ahora a ilustrar el quinto y último mecanismo con el que contamos para tratar el orden de palabras. Este consiste en una regla léxica para agregar un complemento circunstancial al final de la lista de complementos del verbo; este complemento es una frase preposicional cuya semántica depende de la frase nominal que representa al término de la preposición con la modificación de tiempo, lugar, modo, causa o instrumento que marca la preposición propiamente.

Por ejemplo, en *da a Juan los chocolates a las ocho* el complemento circunstancial *a las ocho* consiste de la preposición *a* que indica que la modificación es temporal, y su término, la frase nominal *las ocho*, indica el tiempo propiamente. Esta regla se ilustra en la figura 45.

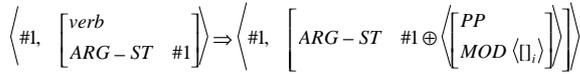


Figura 45: Regla léxica para incluir un complemento circunstancial opcional

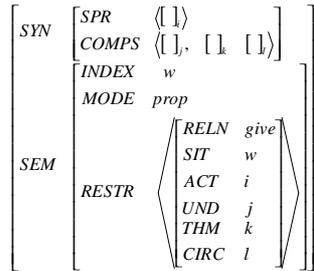


Figura 46: Semántica del verbo con un complemento circunstancial opcional

La semántica del verbo con este argumento adicional se ilustra en la figura 46, donde el complemento sintáctico y el argumento temporal en la semántica se encuentran coindexados.

Con estos elementos es posible analizar la oración *da a Juan los chocolates a las ocho* mediante la combinación del verbo con sus argumentos a través de la HCR, y de la frase verbal resultante con el sujeto a través de la *Subject-PRO-drop Rule*, como se muestra en la figura 47.

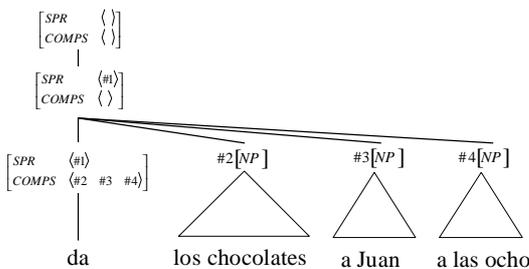


Figura 47: Análisis de la oración con complemento circunstancial opcional

Con estas herramientas es posible analizar oraciones en las que no sólo los complementos directo e indirecto se encuentran frontalizados, sino también el nuevo complemento circunstancial pasa al frente de la oración como en *a las ocho da a Juan los chocolates*. En esta última oración el complemento opcional se marca como gap, y se consume por la regla correspondiente de la manera usual, como se ilustra en la Figura 48. El análisis procede de la manera usual combinando el verbo con sus complementos explícitos por medio de la *Head-Complement Rule*, consumiendo el sujeto omitido mediante la *Subject-PRO-drop Rule* y consumiendo el complemento circunstancial marcado como gap con la regla de gap.

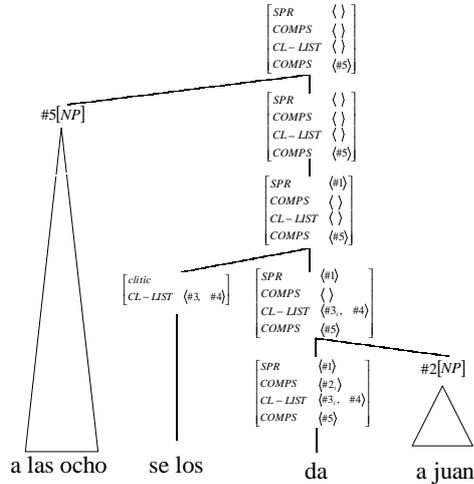


Figura 48: Análisis de la oración con complemento circunstancial opcional frontalizado

En caso de oraciones con alguno o ambos de los complementos directo e indirecto se encuentren clitizados, el análisis procede aplicando de manera secuencial las reglas léxicas de inserción de complemento opcional, de clitización y de gap para generar la entrada verbal correspondiente. El uso de las tres reglas léxicas es necesario para analizar, por ejemplo, *a las ocho se los da a Juan*. Esta oración contiene además el complemento indirecto clitizado pero realizado pleonásticamente. El análisis correspondiente se muestra en la Figura 49.

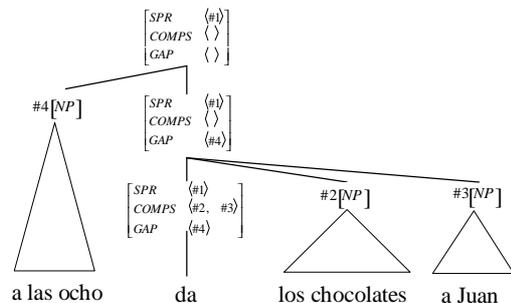


Figura 49: Análisis de la oración con complemento opcional, clíticos y gap.

## 9. Implementación y verificación

La cobertura de esta gramática se ha verificado empíricamente con expresiones del Corpus DIME (Villaseñor, Massé y Pineda, 2001). Este corpus consiste de diálogos orientados hacia la solución de problemas de diseño, en particular el diseño de cocinas; los diálogos fueron recopilados en un escenario del mago de Oz, dentro del contexto del proyecto DIME (Pineda 2001, 2002). Adicionalmente, se han generado ejemplos de estructuras que no aparecen en el corpus, pero que son necesarios para validar la cobertura de la gramática; en particular, en el Corpus DIME no se dan

ejemplos de las construcciones con subsumisión combinación clítica, ya que todas las oraciones con subida de clíticos que aparecen en el corpus contienen pronombres que representan argumentos de un solo verbo. Además del análisis gramatical, se cuenta con una implementación de la gramática en el ambiente de desarrollo de HPSG, LKB (Kopetake, 2002); como punto de partida para este esfuerzo tomamos la gramática del español desarrollada por Quirino (2001) que corresponde en términos generales a la gramática inglesa presentada en el texto de Sag y Wasow (1999).

El corpus DIME está compuesto por 6121 elocuciones, de las cuales aproximadamente un 55% no admiten un análisis gramatical exhaustivo (e.g. interjecciones, actos del habla expresados en una palabra o frase hecha, oraciones elípticas, etc.). De las expresiones parseables un total de 748 incluyen formas clíticas, de las cuales 688 forman parte del sistema de pronombres clíticos propiamente; de éstas 677 contienen un solo pronombre y tan sólo 11 llevan dos. No hay ningún caso de clíticos con verbos de sujeto u objeto de control o auxiliares cliticados, y todos los casos que se presentan en el corpus reciben un análisis adecuado con nuestro modelo básico. Las 60 formas clíticas restantes corresponden a las oraciones con *se*: pasiva refleja, impersonal activa, reflexivas, recíprocas, pseudoreflejas y dativo de interés. La pasiva refleja ocurre 39 veces con 16 verbos diferentes y dos auxiliares; la forma activa impersonal *se le llama* ocurre dos veces, y no hemos explorado si hay las pasivas reflejas que puedan ser también analizadas como impersonal activas, dada la convergencia de los fenómenos. Hay 17 instancias de pseudoreflejas con 7 verbos diferentes. Todas estas formas, excepto dos que ocurren en oraciones subordinadas, reciben un análisis adecuado con nuestra gramática. Como dato singular en todo el corpus no se presenta ninguna oración en voz pasiva.

## 10. Conclusiones

En este artículo se presenta una gramática básica del español en HPSG. La gramática se estructura alrededor de un núcleo sintáctico constituido por el sistema de verbos auxiliares y el sistema de pronombres clíticos, sobre el cual se modelan un conjunto significativo de las formas oracionales del español. En particular se abordan las formas ergativas y el complejo grupo de oraciones con *se*, como la pasiva refleja, la impersonal activa, las reflexivas y también las pseudoreflejas. Se aborda también brevemente el modelo de las preguntas pronominales y se presentan una estrategia simple para enfrentar el problema del orden de los elementos oracionales del español. Se muestra cómo con un número limitado de esquemas de composición se pueden abordar varios ordenes que aparecen frecuentemente. En particular, se muestra como

numerosas construcciones que podrían considerarse como alteraciones del orden normal corresponden realmente a fenómenos acotados, como el caso de los pronombres clíticos y sus formas pleonásticas, o las construcciones ergativas, que realmente reciben un análisis ortogonal a los problemas de orden.

En general, se destacan los resultados relacionados con el modelo de la perífrasis, con y sin auxiliares, y el sistema de clíticos. Hasta donde sabemos nuestra propuesta es original en el ámbito de la gramática tradicional y también desde el punto de vista de la lingüística computacional. En particular, la definición de verbos auxiliares como verbos de subida de sujeto que adquieren una función gramatical produce un conjunto de verbos auxiliares diferente a las propuestas tradicionales y en particular provee un criterio formal para considerar a *ser*, *estar* y *haber* como auxiliares y no sólo como copulativos o parte de la conjugación compuesta; otros verbos de subida de sujeto, como *poder* y algunas frases verbales pueden o no ser auxiliares dependiendo de si preservan o no su significado léxico original. Cuando estos verbos así como los verbos de sujeto de control ocurren en perífrasis con su significado original no son auxiliares, pero interactúan con el sistema de clíticos. En resumen, las formas verbales que funcionan como auxiliares frecuentemente preservan su significado léxico original por lo que surgen ambigüedades que deben ser resueltas. Nuestro análisis captura cualquiera de estas acepciones.

Con relación al fenómeno de los pronombres clíticos, se presenta una solución en la que se toma una posición intermedia en la controversia acerca del nivel de representación de este fenómeno. Por un lado se considera a las secuencias de uno o más pronombres clíticos como estructuras léxicas indisolubles que se representan la morfología; por el otro, dichas unidades pueden ser objeto de operaciones sintácticas por lo que también forman parte del nivel de representación sintáctico. Asimismo, el presente análisis propone una solución motivada desde el punto de vista lingüístico que da cuenta del orden rígido e idiosincrásico de los pronombres clíticos y del fenómeno de subida de clíticos en el que el sistema de clíticos interactúa con perífrasis, con y sin auxiliares. En particular nuestra propuesta tiene como consecuencia la existencia de dominios clíticos locales y las operaciones de subsumisión y composición; asimismo, tanto las operaciones de subsumisión y composición como los principios de cancelación sintáctica y léxica, cubren con gran generalidad y de manera sumamente simple los fenómenos de los clíticos del español que hemos observado. Más aún, investigaciones preliminares nos permiten aventurar, aunque por el momento sólo de modo tentativo, que los principios sobre los que está construido nuestro modelo son válidos para otras

lenguas romances. Adicionalmente el modelo se presta para una implementación simple y poco costosa en recursos computacionales como lo muestra nuestra propia implementación en LKB.

Sánchez (eds.), SMCC, Aguascalientes, Ags. México, Septiembre, 2001.

### **Agradecimientos**

Se agradece el apoyo del CONACYT proyecto C092-A y a los miembros del grupo DIME del Departamento de Ciencias de la Computación del IIMAS, UNAM.

### **Referencias**

**A. COPESTAKE.** (2002) "The LKB System". Stanford University, <http://www.csl.stanford.edu/~aac/lkb.html>

**GILI GAYA, S.** (1991). "Curso Superior de Sintaxis Española". Biblograf, S. A., 103-153.

**MEZA, I. & PINEDA, L.** (2002) "The Spanish auxiliary verb system in HPSG". En Proceedings of CICling-2002 Alexander Gelbuck (ed.), Lecture Notes in Computer Science 2276, Springer-Verlag, pp. 200-209.

**MILLER P. H. AND SAG, I. A.** (1995) "French Clitic Movement Without Clitics or Movement". CSLI Technical Report, Stanford.

**MONACHESI, P.,** (1993). "Object clitics and clitic climbing in Italian HPSG grammar". Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics. ACL Conferences, Utrecht, ACL.

**PINEDA, L. A., MASSÉ, A., MEZA, I., SALAS, M., SCHWARZ, E., URAGA, E AND VILLASEÑOR, L.** (2001). "El Proyecto DIME". En Proceedings of Second International Workshop on Spanish Language Processing and Language Technologies SLPLT2, pp. 41-46, Universidad de Jaén, Spain, September 14-15.

**PINEDA, L. A., MASSÉ, A., MEZA, I., SALAS, M., SCHWARZ, E., URAGA, E AND VILLASEÑOR, L.** (2002). "The Dime project". En Proceedings of MICA2002, Lectures Notes in Artificial Intelligence 2313, pp.166-175.

**SAG I. & WASOW T.** (1999). "Syntactic Theory: A Formal Introduction", CSLI Publications, Stanford.

**QUIRINO, A. P.** (2001), "Spanish clitics: A computational model", Universität Bielefeld, Master's thesis, <http://coral.lili.uni-bielefeld.de/~aquirino/thesis.html>

**VILLASEÑOR, L., MASSÉ, A. & PINEDA, L. A.** (2001). "The DIME Corpus", Memorias 3º. Encuentro Internacional de Ciencias de la Computación ENC01, Tomo II, C. Zozaya, M. Mejía, P. Noriega y A.