Una modelización del mecanismo dinámico de construcción de la significación de expresiones complejas

Pablo GAMALLO OTERO & Michel CHAMBREUIL
Laboratoire de Recherche sur le Langage, Université Blaise Pascal - Clermont 2
34 avenue Carnot. 63000 Clermont-Ferrand. France.
Tel : 33 73 40 63 37
Telecopia : 33 73 40 64 43
Correo electrónico : {gamallo,chambreuil}@lrl.univ-bpclermont.fr

Resumen

El predicado es la entidad semántica fundamental del espacio de interpretación de los lenguajes formales no ambiguos, y por extensión del lenguaje natural. Ahora bien, la rigidez y el estatismo de esta entidad impiden una modelización adecuada del polimorfismo gramatical, de la polisemia léxica y del proceso de desambiguación de las expresiones del lenguaje natural. Este artículo propone, tras un análisis de la organización interna de la noción de predicado, una reestructuración del espacio semántico. En concreto, la hipótesis que se defiende estipula que el objeto central alrededor del cual se estructura el espacio semántico no es el predicado, sino la operación de asignación de una entidad a un papel predicativo. La operación de asignación, junto a las condiciones que se le asocian, se presenta como un mecanismo composicional que permite construir de manera dinámica el significado de las expresiones complejas. La flexibilidad de este mecanismo, por lo demás, se adapta convenientemente al polimorfismo gramatical y a la polisemia léxica de las expresiones combinadas. El objetivo de este artículo es precisamente el de esbozar un modelo formal del mecanismo de construcción de expresiones complejas a partir del análisis de la operación de asignación.

Palabras clave

semántica formal, procesamiento semántico del lenguaje natural, ambigüedad, composicionalidad.

1 Introducción : ambigüedad y proceso de desambiguación

Uno de los problemas esenciales del procesamiento semántico del lenguaje natural reside en que, a diferencia de los lenguajes formales no ambiguos, la significación de las unidades lingüísticas varía en función de los enunciados en los que se insertan.

La ambigüedad[1] está directamente ligada a una propiedad fundamental de las expresiones lingüísticas : la polivalencia o flexibilidad sintáctico-semántica. Toda expresión es polivalente (o flexible) en el sentido que puede reconfigurar su categoría y precisar su contenido conceptual en función del contexto discursivo en el que se integra. La polivalencia de las expresiones se manifiesta tanto en el nivel sintáctico (lo que llamaremos "polimorfismo gramatical") como en el nivel léxico (lo que llamaremos "polisemia léxica"). El proceso de construcción de la significación de expresiones complejas permite caracterizar categorialmente y precisar conceptualmente la polivalencia sintáctico-semántica de las expresiones constituyentes. En otros términos, la interpretación dinámica de expresiones complejas desambigua el contenido de las expresiones constituyentes.

[1] Como el objetivo de este artículo no es el de describir y caracterizar los diferentes tipos o grados de ambigüedad, situados en la escala gradual que va de la homonimia a la indeterminación pasando por la polisemia, no marcaremos ninguna diferencia significativa en el uso de los términos "ambigüedad", "polisemia" o incluso "indeterminación".

1.1 *Polimorfismo gramatical*

Una misma expresión lingüística (o varias formas lingüísticas aparentadas) puede reflejar categorías gramático-syntácticas diferentes y distribuirse en diferentes contextos sintácticos : 'Juan *corre*' / 'la *carrera* de Juan' / 'que Juan *corra*' / '*correr*'...

Uno de los objetivos principales de los enfoques actuales de la semántica cognitiva, especialmente de la gramática cognitiva de R.W. Langacker (1987, 1991), es el de configurar un espacio semántico que pueda dar cuenta de la riqueza y flexibilidad de la polimorfia gramatical. Las entidades básicas de este espacio son abstracciones conceptuales, caracterizadas como mecanismos complejos de estructuración de la información, que representan el contenido semántico de las formas gramaticales y sintácticas del lenguaje. Tales abstracciones son operaciones de conceptualización ligadas a una capacidad particular perceptivo-cognitiva, llamada "imagery", cuya función es la de organizar una misma escena visual de diferentes maneras. Basándose en este tipo de abstracciones conceptuales, Langacker propone una nueva caracterización del sistema de categorías lingüísticas.

El polimorfismo gramatical es también el objeto de teorización de la teoría de propiedades (Chierchia G. & Turner R. 1988), cuyo enfoque no es de naturaleza cognitiva sino lógico-formal. La teoría de propriedades no asocia a las formas verbales temporales (por ejemplo a la forma de presente de indicativo 'corre') y a las diferentes formas nominalizadas ('que Juan corra', 'correr', 'carrera'...) el mismo tipo de entidad en el espacio semántico. Una forma verbal temporal denota una función predicativa, mientras que una forma nominalizada denota una entidad individual. La teoría de propiedades llama "modos de ser" las dos manifestaciones conceptuales de una predicación : la función predicativa (que es una estructura predicativa incompleta o no saturada) y la entidad individual nominalizada (que corresponde a una estructura predicativa saturada). En el marco nocional de la gramática cognitiva, estos dos modos de ser representan en realidad dos conceptualizaciones abstractas particulares.

El espacio semántico de la teoría de propiedades, sin embargo, no está adaptado a la modelización de la polisemia léxica.

1.2 *Polisemia léxica*

Una misma expresión léxica puede vehicular informaciones diferentes pero aparentadas. Todo lexema posee un grado más o menos elevado de polisemia. Por poner algunos ejemplos, el verbo 'cargar' puede significar : *desplazar objetos* ('cargar sacos en el camión'), *llenar un espacio* ('cargar el camión de sacos') ; el nombre 'botella' puede significar : *objeto que sirve como recipiente* ('llena esta botella'), *substancia contenida en un recipiente* ('prueba esta botella').

Uno de los intentos de teorización formal de la polisemia léxica más en voga corre a cargo de J. Pustejovsky (1995). En su modelo, una entrada léxica se asocia a una compleja estructura informativa compuesta de un conjunto de facetas conceptuales, facetas que representan las diferentes significaciones posibles de la entrada. El mecanismo composicional opera precisamente en el interior de la estructura informativa para seleccionar una faceta conceptual (es decir un siginificado posible) de la entrada léxica.

El enfoque de Pustejovsky hereda, sin embargo, la rigidez y las limitaciones del sistema de categorías semánticas montagoviano. En efecto, la rigidez de la semántica de Montague no permite en ningún caso modelizar los fenómenos de polimorfismo gramatical (Gamallo P. 1998).

1.3 *La construcción dinámica de la significación de expresiones complejas : la desambiguación*

Existen numerosos trabajos sobre el procesamiento sintáctico-semántico de enunciados complejos desde el punto de vista de un agente-receptor en comprensión. Estos trabajos analizan la relación entre en el proceso de construcción de la estructura sintáctica del enunciado y el acceso a las informaciones léxicas de las expresiones constituyentes. Entre estos trabajos, se pueden distinguir esencialmente dos tipos de enfoques :

El primer enfoque (basado en el modelo "garden path" de Frazier L. 1989) se caracteriza por dar una mayor preeminencia a la sintaxis. Inspirándose en el generativismo chomskiano, este enfoque postula que el proceso de construcción de la estructura sintáctica de una oración se efectúa tomando en cuenta únicamente informaciones de naturaleza morfo-sintáctica. La resolución de los fenómenos de ambigüedad se relegan de esta manera a una segunda etapa, donde se da acceso a la información léxica de las expresiones constituyentes.

El otro enfoque (McRae y alii, 1997), por el contrario, presenta el procesamiento de la oración como un mecanismo que da acceso simultáneamente a informaciones de naturaleza morfo-sintáctica, léxica y, eventualmente, pragmático-discursiva : la situación de enunciación, los conocimientos compartidos por el locutor y el/los oyente(s).... Este enfoque rechaza por tanto la autonomía de la sintaxis. La construcción de la estructura sintáctica depende de informaciones léxico-pragmáticas. En otros términos, la interpretación de un enunciado se efectúa en el proceso mismo de construcción de su estructura sintáctica. Es lo que se conoce por el procesamiento "on line".

Teniendo en cuenta que el polimorfismo gramatical y la polisemia léxica son dos fenómenos de naturaleza semántica, la desambiguación de expresiones polimórficas y polisémicas se debe llevar a cabo en un proceso de construcción de la significación en el que se dé acceso simultáneamente a informaciones morfo-sintácticas y léxicas, es decir en un procesamiento "on line".

El objetivo de la parte constructiva de este artículo será el de modelizar en el espacio lógico-semántico el proceso de construcción "on line" de la significación. Para ello, analizaremos, en el interior de la estructura de argumentos de un predicado, el mecanismo de saturación de los papeles de un predicado, en particular, nos centraremos en la operación de asignación de una entidad a un papel predicativo.

2 Un esbozo[2] de modelización del procesamiento "on line" de expresiones complejas

Basaremos la modelización de este procesamiento en la concepción formal de un predicado como una operación de saturación de sus papeles internos. Tal concepción es el eje central en torno

[2] En Gamallo Otero P. (1998), se realiza una elaboración detallada de esta modelización.

al cual se caracterizan las entidades del espacio semántico en las versiones actualizadas de la teoría de situaciones (FRACAS 1994, Chambreuil M. y alii 1998).

2.1 *La operación de saturación de los papeles de un predicado*

Un predicado puede ser analizado como una operación que, asignando un conjunto de entidades a sus papeles, da como resultado una entidad integradora de orden superior. En el proceso de resolución de esta operación, se pueden distinguir, por tanto, tres tipos de objetos : (i) el operador predicativo y sus papeles, que conforman la operación (o función) predicativa ; (ii) las asignaciones de entidades a los papeles predicativos : los argumentos de la función ; (iii) y la entidad integradora, valor de la función.

La entrada verbal 'descargar', por ejemplo, podría denotar[3] en el espacio semántico una función predicativa con, al menos, dos asignaciones :

(1)

$$\lambda(X)^{\text{sujeto}}\ \lambda(Y)^{\text{objeto}}\ (\mathit{DESCARGAR}\ ;\ (X)^{\text{sujeto}}, (Y)^{\text{objeto}})$$

Notése que en (1), los objetos lambda (resultantes de las operaciones de λ-abstracción) no son parámetros de entidades, sino asignaciones de parámetros a papeles predicativos. La interpretación de la oración 'Chema descargó el camión' puede asociarse por tanto al proceso de saturación de los papeles de la función predicativa de (1). La aplicación de esta función a las asignaciones $(\mathit{chema})^{\text{sujeto}}$ y $(\mathit{camión})^{\text{objeto}}$, donde *chema* y *camión* representan las entidades denotadas respectivamente por 'Chema' y 'el camión' , da como resultado, no un valor de verdad V o F (como en la lógica de predicados), sino una entidad individualizada, *descargar-chema-camión*, donde se integran las entidades asignadas. La entidad integradora designa el evento en el que Chema descarga el camión. Como en la teoría de propiedades, una expresión verbal temporal puede asociarse a la función predicativa, función que representa el "modo de ser" composicional del verbo, mientras que una nominalización verbal se asocia al resultado de la función, es decir a la entidad individualizada como el evento designado por la expresión. Esta doble caracterización —objeto composicional y evento— permite tratar el polimorfismo de las expresiones de naturaleza verbal.

Sin embargo, queda todavía un problema por resolver. Un mismo lexema verbal puede asociarse a un gran número de funciones predicativas. Por ejemplo, el lexema 'descargar' en las oraciones 'Chema descargó el camión de sacos' y 'Chema descargó el camión sobre la empalizada' se asocia a dos funciones predicativas diferentes : *DESCARGAR1* (acción de modificar un espacio-lugar) y *DESCARGAR2* (desplazamiento de un objeto). Se trata por tanto de un lexema polisémico que contiene, al menos, dos significados, es decir dos procesos de composición diferentes. Como cada significado representa en realidad un proceso composicional, la elección de uno u otro (es decir la desambiguación) se efectúa antes de que se inicie el proceso composicional de saturación de los papeles predicativos. En otros términos, antes de comenzar el proceso de interpretación de la oración, es necesario elegir uno de los significados composicionales del verbo. La desambiguación

[3] A lo largo del artículo, los términos "denotar" y "asociar" son empleados indistintamente para referirse a la relación entre una expresión o índice lingüístico y el objeto semántico correspondiente.

no se realiza por tanto al mismo tiempo que se precisa el significado de las expresiones que se combinan en el interior de la oración : no hay procesamiento "on line".

Este análisis de la noción de predicado nos lleva al postulado siguiente : el eje del mecanimo composicional no es el proceso de saturación de los papeles predicativos ; la composicionalidad se configura más bien a un nivel inferior en el análisis de la predicación, concretatemente en el interior de la operación de asignación. El análisis de esta operación nos permitirá :

- caracterizar las sub-categorías básicas de entidades individuales de nuestro modelo, sub-categorías por medio de las cuales podremos tratar el fenómeno del polimorfismo gramatical ;
- caracterizar un mecanismo composicional basado en la precisión gradual del significado de las expresiones (polisémicas) combinadas.

2.2 *La operación de asignación de una entidad a un papel predicativo*

La operación de asignación puede definirse como una función con dos argumentos — l'entidad asignada al papel predicativo y la entidad que individualiza la predicación— que los combina e integra en una entidad de orden superior : la entidad compleja resultado de tal combinación. Formalmente, una operación de asignación es un caso particular de función predicativa. La asignación de *chema* al papel sujeto del operador predicativo *DESCARGAR* puede representarse como una función cuyos argumentos son $(descargar)^{sujeto\downarrow}$ y $(chema)^{sujeto\uparrow}$, y cuyo resultado es la entidad *descargar-chema*, que integra en una unidad individualizada la entidad *chema*, asignada al papel sujeto, y el evento *descargar*. La entidad compleja *descargar-chema* es el evento asociado a la expresión verbal 'Chema descargó...' (los tres puntos representan el hecho que puede haber otras asignaciones que saturen la expresión verbal). La flecha descendente ↓ simboliza el hecho que la entidad que desempeña ese papel "busca" los elementos constituyentes que van a venir a integrarse a su estructura : la entidad *descargar* es el evento donde se integra *chema*. La flecha ascendente ↑ simboliza el hecho que la entidad que desempeña ese papel "busca" la entidad que pueda integrarla : la entidad *chema* se integra en el evento *descargar*.

En el caso de los predicados verbales, las operaciones de asignación representan en realidad los objetos semántico-composicionales asociados a las funciones sintácticas. De esta manera, el sujeto puede caracterizarse a partir de la función predicativa (2) :

(2)
$$\lambda(X)^{sujeto\downarrow}\,\lambda(Y)^{sujeto\uparrow}\,(SUJETO\ ;\ (X)^{sujeto\downarrow},\ (Y)^{sujeto\uparrow})$$

La aplicación de esta función a las asignaciones $(descargar)^{sujeto\downarrow}$ y $(chema)^{sujeto\uparrow}$, donde *descargar* y *chema* representan las entidades denotadas respectivamente por 'descargar' y 'Chema', da como resultado la entidad compleja *descargar-chema*, asociada a 'Chema descargó...'.

2.2.1 *Caracterización de dos sub-categorías de entidades individuales*

Los diferentes papeles de los operadores de asignación caracterizan dos tipos de entidades individuales :

- la entidad asignada al papel descendente ↓ se conceptualiza como el todo integrador con respecto a sus entidades constituyentes. A este tipo de entidad le llamaremos "dominio". La entidad

descargar denotada por el verbo 'descargar' en 'Chema descargó...' se conceptualiza como el dominio de la expresión. La anotaremos *$descargar*.
- la entidad asignada al papel ascendente ↑ se conceptualiza como un constituyente inserto en el dominio. A este tipo de entidad, le llamaremos "átomo". La entidad *chema* asociada al nominal 'Chema' en 'Chema rompe...' se conceptualiza como el átomo de la expresión. La anotaremos *#chema*.

El resultado de la función de asignación es una entidad compleja categorizada como un dominio o como un átomo. En la expresión verbal 'Chema descargó...', el resultado de la operación es una entidad caracterizada como un dominio complejo : *$descargar-chema*. La caracterización de la categoría de la entidad compleja depende de las propiedades composicionales del tipo de función de asignación. La función de sujeto verbal no posee las mismas propiedades composicionales que, por ejemplo, la función de modificador nominal.

Los objetos composicionales no solamente se asocian, por tanto, a las marcas de funciones sintácticas verbales, sino también a toda marca de relación morfo-sintáctica : modificadores nominales (adjetivos, complementos preposicionales, relativas...), modificadores verbales (adverbios, complementos preposicionales...), etc.

2.2.2 *Caracterización interna de los objetos composicionales*

Los mecanismos de composición semántica de los que disponemos son las funciones que designan las operaciones de asignación a los papeles predicativos. A estas funciones se le pueden asociar condiciones semánticas que restringen la naturaleza de las entidades combinadas. Distinguimos dos tipos de condiciones semánticas :
- las condiciones categoriales que restringen la sub-categoría de las entidades combinadas : dominio o átomo.
- las condiciones conceptuales que delimitan la clase conceptual específica de las entidades combinadas : eventos, situaciones estáticas, objetos físicos, localizaciones, seres humanos, dimensiones físicas...

Las condiciones categoriales se precisan en gran parte por medio de marcas lingüísticas de naturaleza morfo-syntáctica. Las condiciones conceptuales, en cambio, se precisan por medio de marcas de naturaleza léxica.

En la expresión compleja 'Chema descargó...', las marcas morfo-sintácticas relativas a la función de sujeto pueden asociarse al objeto composicional siguiente :

(3)
$\lambda(\$X)^{\text{sujeto}\downarrow}\ \lambda(\#X)^{\text{sujeto}\uparrow}(\mathit{SUJETO}\ ;\ (\$X)^{\text{sujeto}\downarrow},\ (\#X)^{\text{sujeto}\uparrow})$

Este objeto especifica, con respecto a (2), las condiciones categoriales de las entidades combinadas, en concreto pone en relación un dominio con un átomo para obtener como resultado un dominio complejo integrador.

Las expresiones léxicas 'descargar' y 'Chema', por su parte, añaden a (3) las restriciones conceptuales adecuadas : (i) la operación *SUJETO* pasa a ser *AGENTE*, operación específica que designa la integración de una entidad agentiva en una acción ; (ii) el dominio *$X* pasa a ser una acción agentiva *$ACTION*, que queda todavía indeterminada con respecto al hecho de si se trata de un desplazamiento de objetos, o bien de una acción de modificación de un espacio-lugar ; (iii) el átomo *#X* pasa a elaborarse como un ser agentivo *#ENT-AGENTIVA*.. El objeto composicional específico así constituido es :

(4)
$\lambda(\$ACTION)^{agente\downarrow}\ \lambda(\#ENT\text{-}AG)^{agente\uparrow}\ (AGENTE\ ;\ (\$ACTION)^{agente\downarrow},\ (\#ENT\text{-}AG)^{agente\uparrow})$

Este objeto representa la función que se aplica a las asignaciones $(\$descargar)^{agente\downarrow}$ y $(\#chema)^{agente\uparrow}$, para construir la entidad integradora *$descargar-chema*, entidad que pertenece, por un lado, a la categoría de los dominios, y por otro lado, a la clase conceptual de las acciones agentivas, clase todavía indeterminada.

2.3 *Polimorfismo gramatical y categorización semántica*

En la separación entre categorización semántica y clasificación conceptual se encuentra la clave para poder caracterizar y modelizar el polimorfismo gramatical. Por ejemplo, las expresiones '...descargó...' y 'la descarga...' se asocian a dos entidades que se categorizan respectivamente como un dominio y como un átomo : el dominio *$descargar* y el átomo *#descargar*. Estas dos entidades, a pesar de categorizarse de diferente manera, pertenecen, sin embargo, a la misma clase conceptual : las dos designan una acción de descarga (Gamallo P. 1998).

2.4 *La desambiguación de expresiones polisémicas y la construcción dinámica de la significación*

El procesamiento "on line" de expresiones complejas se basa en la evolución progresiva del discurso. El resultado del procesamiento de una expresión puede aportar nuevas condiciones conceptuales que restringen y precisan el procesamiento de la siguiente expresión. O a la inversa, si la primera expresión procesada da lugar a una entidad todavía polisémica, las restricciones aportadas por el procesamiento de la siguiente expresión pueden permitir la determinación de una significación precisa de la primera expresión. La dinámica del procesamiento "on line" se basa precisamente en la herencia de restricciones en el proceso de elaboración de los objetos composicionales que combinan las entidades constituyentes.

Analicemos brevemente el procesamiento de la expresión 'la harina', una vez procesada la expresión compleja 'Chema cargó...'. La combinación entre 'Chema cargó...' y 'la harina' se lleva a cabo a partir de la explotación de las marcas sintácticas de objeto directo, marcas que determinan la elaboración del objeto composicional siguiente :

(5)
$\lambda(\$X)^{odirecto\downarrow}\ \lambda(\#X)^{odirecto\uparrow}\ (ODIRECTO\ ;\ (\$X)^{odirecto\downarrow},\ (\#X)^{odirecto\uparrow})$

En cuanto a las condiciones léxicas, el resultado del procesamiento de 'Chema cargó...' y la explotación del nombre 'harina' precisan el contenido conceptual de las entidades combinadas, en concreto las condiciones de 'harina' determinan que el átomo asignado designe una substancia física y no un espacio localizador. Tal restricción impide que el evento de descargar sea clasificado como una acción de modificación de un espacio-lugar. Por consiguiente, la descarga se precisa como un desplazamiento de objetos. La función composicional definida en (5) se elabora en consecuencia como una operación de integración de un átomo designando una substancia física en un dominio designando un desplazamiento directivo :

(6)
$\lambda(\$DESPL)^{desplazado\downarrow}\ \lambda(\#OBJ\text{-}FIS)^{desplazado\uparrow}\ (DESPLAZADO\ ;\ (\$DESPL)^{desplazado\downarrow}\ (\#OBJ)^{desplazado\uparrow})$

La aplicación de este objeto composicional a las asignaciones *($descargar-chema)*$^{desplazado\downarrow}$ y (*#harina*)$^{desplazado\uparrow}$, da como resultado el dominio complejo *$descargar-chema-harina*, entidad no polisémica denotada por 'Chema descargó la harina...', que pertenece a la clase conceptual correspondiente a las acciones de desplazamiento de objetos.

3 Conclusión

La originalidad de nuestro modelo formal del espacio semántico radica en la caracterización de los objetos composicionales a partir del análisis de la operación de asignación de una entidad a un papel predicativo. Tal caracterización viene motivada por el hecho de que permite dar cuenta, en un marco teórico coherente, del polimorfismo gramatical de las expresiones lingüísticas y del proceso de desambiguación (o de determinación) de la polisemia (o indeterminación) léxica. La brevedad de este artículo nos impide, no sólo mostrar detalladamente la coherencia interna de nuestra modelización, sino también profundizar de manera crítica en las hipótesis y nociones fundamentales en las que basamos tal modelización. No en vano, la reflexión crítica que nos ha llevado a formular este conjunto de hipótesis ha sido y es el producto de un estudio pluridisciplinar donde se abordan diferentes tipos de fenómenos lingüísticos desde el menos tres enfoques complementarios : el enfoque de la semántica formal (o logicista), el enfoque de la semántica cognitiva, y el enfoque ligado al procesamiento informático del lenguaje natural.

Referencias bibliográficas

Chambreuil M., Ben Gharbia A., Bernigot C., Gamallo Otero P., Panissod C., Reinberger M-L. (1998). *Théories sémantiques pour la langue naturelle*. Paris, Hermès (edición prevista para junio de 1998).

Chierchia G. & Turner R. (1988). "Semantics and Property Theory". *Linguistics and Philosophy*, vol 11, n° 3, (pp. 261-302).

FRACAS (1994). *Describing the Approches* (D8). Informe interno del proyecto "Framework for Computational Semantics".

Frazier L. (1989). "Against lexical generation of syntax". En Lexical representation and Process, William Marsllen-Wilson, ed. Massachusetts, (pp. 505-528).

Gamallo Otero P (1998). *Construction conceptuelle d'expressions complexes : traitement de la combinaison "nom-adjecif"*. Tesis doctoral de la universidad Blaise Pascal Clermont 2.

Langacker R. (1987, 1991). *An introduction to Cognitive Grammar.* Vol I y II, Standford University Press.

McRae K., Ferreti T.R. & Amoyte L. (1997). "Thematic Roles as Verb-specific Concepts". En Lexical Representations and Sentence Processing. M. C. MacDonald (ed), (pp. 137-176).

Pustejovsky J. (1995). *The Generative Lexicon*. MIT Press : Cambridge.