

Léxico generativo y mecanismos de control en el proceso de interpretación

Pablo GAMALLO OTERO y Michel CHAMBREUIL

Laboratoire de Recherche sur le Langage, Université Blaise Pascal - Clermont 2

34 avenue Carnot. 63000 Clermont-Ferrand. France.

Tel : 33 4 73 40 63 37

Telecopia : 33 4 73 40 64 43

Correo electrónico : {gamallo,chambreuil}@lrl.univ-bpclermont.fr

Resumen

Los enfoques generativos del léxico se enfrentan a problemas ligados a la "sobre-generación" y "sub-generación" de significados en la interpretación de expresiones compuestas. La raíz de estos problemas hay que buscarla tanto en una concepción restrictiva, y por tanto cerrada, de la información léxica, como en la falta de control sobre el proceso de generación. A fin de regular la sobre y sub-generación, esta comunicación propone un modelo generativo basado, por un lado, en una concepción enciclopédica y abierta de la información léxica, y por otro, en un mecanismo de control que guíe y oriente el proceso de generación del significado en la interpretación de expresiones compuestas.

Palabras clave

procesamiento semántico del lenguaje natural, léxico generativo, composicionalidad, información enciclopédica

1 Introducción : los límites del léxico generativo en un marco modular

1.1 Modularidad y generación

En los últimos años, numerosos trabajos relativos al léxico, y realizados desde una perspectiva computacional, se apoyan en dos hipótesis fundamentales que conciernen a la naturaleza de la información léxica y al papel de ésta en el mecanismo composicional.

La primera hipótesis postula la separación entre dos tipos de información : la información directamente ligada a las entradas léxicas y la información (de naturaleza enciclopédica) asociada a las expresiones lingüísticas indirectamente a través de fuentes no léxicas. Estos dos tipos de información se organizan y

estructuran en dos módulos diferentes : un módulo léxico y un módulo enciclopédico. Según esta hipótesis "modular", las expresiones lingüísticas conllevan no solamente un contenido informativo de naturaleza léxica, sino también un contenido enciclopédico.

La segunda hipótesis postula que el significado de una expresión compuesta se genera a partir de las informaciones léxicas que conllevan las expresiones constituyentes. Para construir o generar un significado lingüístico es, por tanto, necesario combinar e integrar en el mecanismo composicional "down-top" las informaciones de naturaleza léxica asociadas a las expresiones constituyentes. Esta segunda hipótesis nos sitúa en un enfoque "generativo" del léxico.

Entre los múltiples trabajos basados en la doble hipótesis "modular-generativa", se podrían citar: Bierwich M. 1989, Lang E. et alii 1991, Sowa J. 1993, Franks B. 1995, Pustejovsky J. 1995.

1.2 Sobre-generación y sub-generación

La gran dificultad del enfoque "modular-generativo" consiste en buscar un equilibrio entre la información léxica asociada a una entrada y la información que realmente se explota en el proceso de generación del significado de una expresión compuesta. Tal equilibrio debería permitir generar el contenido informativo apropiado que conlleva la expresión compuesta. Sin embargo, los hechos muestran que, ya sea por exceso, se generan significados que no son aceptables (es decir se "sobre-genera"), ya sea por defecto, no se genera más que una parte de los significados posibles de la expresión compuesta (es decir se "sub-genera"). Como botón de muestra, veamos cómo podrían ser analizadas las expresiones compuestas 'coche rojo' y 'empezar la habitación' en el marco teórico del léxico generativo de Pustejovsky.

- 'coche rojo': el adjetivo de color 'rojo' se caracteriza por aplicarse a la información relativa a los constituyentes materiales de un objeto. Por otra parte, la entrada léxica 'coche' contiene toda la información relativa a las partes constituyentes de un coche. La interpretación de 'coche rojo' genera por tanto el contenido informativo *coche cuyos constituyentes son de color rojo*. Es decir, un coche rojo es un objeto cuyos elementos internos —la carrocería, las ruedas, el volante, las paredes internas, los asientos...— son todos de color rojo. Sin embargo, en realidad la condición necesaria para que un coche sea rojo es que tenga únicamente la superficie externa de la carrocería de color rojo. Nos encontramos, por consiguiente, ante un caso particular de sobre-generación.
- 'empezar la habitación': el verbo 'empezar' se caracteriza por aplicarse a una información de naturaleza temporal que designa un evento

perfectivo (lo que Pustejovsky llama "transición"). Por otra parte, el tipo de transición accesible a partir de la entrada 'habitación' se reduce al contenido causal-agente de la información *construir la habitación*. La interpretación de 'empezar la habitación' genera por tanto únicamente el contenido informativo *empezar a construir la habitación*. Los otros contenidos posibles — *empezar a limpiar la habitación, empezar a pintar la habitación...*— no son generados a partir del módulo léxico. Los eventos relacionados con la limpieza, la pintura... no forman parte del contenido léxico de la entrada 'habitación'. Nos encontramos por tanto ante un caso particular de sub-generación.

La sobre-generación y la sub-generación muestran los límites de la noción de generación léxica en el marco de una teoría modular de la información. El objetivo de este artículo será el de restituir la noción de generación en el marco de un enfoque no modular.

2 La generación léxica en un marco no modular

2.1 Léxico e información enciclopédica

Frente a los enfoques modulares, existen numerosos trabajos que abogan por una teoría no modular de la información asociada al léxico. Estos enfoques postulan que el contenido informativo asociado a una entrada léxica es una red abierta de conocimientos enciclopédicos. El procesamiento de la información no se separa por tanto en dos módulos diferentes (un módulo directamente ligado al léxico y un módulo enciclopédico). Se parte de la idea de que no hay más que un único módulo informativo —la base de conocimientos enciclopédicos— y que toda información es de naturaleza enciclopédica. Por consiguiente, el contenido informativo de una entrada léxica representa el conjunto de conocimientos enciclopédicos que ésta puede transmitir en un discurso lingüístico.

Entre los diferentes trabajos que postulan la hipótesis de la no modularidad, se podrían citar los enfoques lingüístico-cognitivos de Langacker R.W. 1987, 1991, Lakoff 1987, Taylor J. 1995.

Evidentemente, desde el punto de vista de la generación del significado de una expresión compuesta, la no modularidad implica un problema de sobre-generación. Nuestro objetivo será precisamente el de controlar la sobre-generación a partir de un enfoque enciclopédico de la información léxica. Para ello, vamos a proponer, por un lado, la caracterización de un tipo de relación fundamental para la estructuración y organización de la red de informaciones enciclopédicas : la relación "acceso funcional a". Por otro lado, definiremos un tipo de restricción o condición composicional que, a partir de la relación anterior, permite filtrar el contenido informativo adecuado en el proceso de generación del significado de la expresión compuesta. Este tipo de condición composicional se basa en el principio de "metonimia integrada" (Kleiber G. 1994).

2.2 La relación "acceso funcional a"

En diferentes enfoques psico-lingüísticos de la semántica léxica (Scheuder R. & Flores d'Arcais G. 1989), las informaciones de naturaleza perceptiva desempeñan un papel relevante en la organización de la red conceptual asociada a las entradas léxicas. Toda información perceptiva se presenta como una puerta de acceso a un espacio de informaciones funcionales. Por ejemplo, las dimensiones de un objeto físico determinan la manera en la que el objeto puede ser colocado en un lugar. Es decir, las dimensiones de un objeto mantienen una relación específica con el contenido funcional que concierne a su colocación. En otros términos, la información relativa a las dimensiones del objeto abre el *acceso funcional* al espacio informativo relativo a su colocación.

La relación "acceso funcional a" no sólo se revela pertinente en el caso de las propiedades perceptivas, sino que también puede generalizarse a todo tipo de contenido informativo :

Definición 1

un contenido de información es una llave que abre un espacio funcional de naturaleza enciclopédica.

Esta definición caracteriza la relación "acceso funcional a", por medio de la cual se estructura la red de informaciones enciclopédicas asociada al léxico del lenguaje natural.

2.3 El principio de "metonimia integrada"

Como ya hemos dicho, el enfoque enciclopédico de la información léxica da lugar al problema de la sobre-generación de significados en el proceso de interpretación de expresiones compuestas. Este tipo de enfoque debe recurrir, por tanto, a una noción de filtro que controle el mecanismo de composición.

El filtro que proponemos es el llamado principio de "metonimia integrada". Según este principio, una información parcial relativa a una entidad es una propiedad central y representativa de tal entidad si, en determinadas circunstancias, esta información parcial *caracteriza* completamente a la entidad. En otros términos, una parte constituyente se concibe como una propiedad central del todo si existe un contexto en el que la parte caracterice o represente al todo.

Técnicamente, la noción de contexto caracterizador puede establecerse recurriendo a la relación "acceso funcional a" : un contenido informativo caracteriza o representa una entidad cuando abre el *acceso funcional* a un espacio enciclopédico relativo a la entidad.

El principio de "metonimia integrada" podría definirse de la siguiente manera :

Definición 2

una información parcial sobre una entidad es una propiedad central y representativa de tal entidad si, a partir de esta información, se da "acceso funcional a" un espacio informativo asociado a la entidad.

Este principio regula el mecanismo de composición. La interpretación de una expresión compuesta genera únicamente el contenido

informativo filtrado por el principio de "metonimia integrada". En otros términos, en el proceso de interpretación se generan únicamente las propiedades consideradas centrales con respecto al objeto designado por la expresión compuesta.

Veamos como funciona este filtro con ejemplos muy simples de tipo "nombre-adjetivo".

3 El control del proceso de generación del significado en la composición "nombre-adjetivo"

Son pocos los trabajos en semántica léxica que se han enfrentado directamente al problema de la sobre-generación (Jayez J. & Godard D. 1995, Gamallo P. 1998). Mientras que esta problemática en el plano puramente lingüístico se trivializa y se subestima, en el plano computacional del procesamiento del lenguaje natural, aparece como un fenómeno complejo que se revela fundamental para comprender el proceso mismo de interpretación semántica. A fin de mostrar la sutileza que subyace a este fenómeno, vamos a centrar el análisis en la interpretación de expresiones relativamente sencillas de tipo "nombre-adjetivo": 'coche rojo', 'coche rápido' y 'coche grande'. Por motivos de espacio, el análisis de estos ejemplos se sitúa en un nivel de representación intuitivo (en Gamallo P. 1998, se expone una representación en un lenguaje formal inspirado en la teoría de situaciones).

El mecanismo que genera la significación de las expresiones "nombre-adjetivo" será controlado por las condiciones que previamente hemos definido (definición 1 y definición 2), definiciones relativas a la relación "acceso funcional a" y al principio de "metonimia integrada".

3.1 'coche rojo'

En un enfoque no modular, a la entrada léxica 'coche' se le asocia la red de informaciones enciclopédicas que conciernen directa o

indirectamente a la noción de coche. Entre estas informaciones, se encuentran los constituyentes materiales que pueden poseer un color particular. El adjetivo 'rojo', por su parte, se aplica precisamente a las partes materiales del objeto modificado.

La interpretación de la expresión 'coche rojo' sobre-genera, por tanto, un contenido informativo donde el color rojo se manifiesta en todos los constituyentes materiales del coche: la carrocería, las ruedas, el volante, las paredes internas, los asientos...

A fin de filtrar la significación apropiada, se debe recurrir al principio de metonimia integrada. Entre los diferentes contenidos informativos a los que se aplica el adjetivo 'rojo' —el color de la carrocería externa, el color de las ruedas, el color de las paredes internas...—, el único que abre un *acceso funcional* a un espacio enciclopédico representativo del coche es el que se refiere al color de la carrocería externa. El color de la carrocería es el acceso al espacio funcional que concierne a la identificación, visualización y diferenciación del coche con respecto a otros coches. En otros términos, dado que el color de la carrocería permite a un agente conceptualizador identificar y distinguir el coche, tal información se concibe como una propiedad central y representativa de éste, y por tanto directamente generable en el proceso de composición. Las informaciones relativas al color de las ruedas, del volante, de las paredes internas..., al no ser propiedades centrales del coche, no se generan en el proceso de interpretación de 'coche rojo'.

Evidentemente, la relación específica entre el color de la carrocería y la función de identificación del coche es un caso particular de una relación de alcance general: el color de la superficie externa de un objeto es la puerta de *acceso funcional* al proceso de identificación y distinción del objeto. Esta relación general es el filtro que permite controlar la interpretación de expresiones como 'frasco rojo', 'libreta azul', etc.

3.2 'coche rápido'

Entre las informaciones enciclopédicas asociadas a la entrada 'coche' se encuentra la idea de velocidad, ya sea relativa a la actividad de

desplazamiento, o al proceso de fabricación, o a la acción de aparcar, etc.

El adjetivo 'rápido', por su parte, se aplica precisamente a las acciones o procesos susceptibles de contener la idea de velocidad.

La interpretación de la expresión 'coche rápido' sobre-genera, por tanto, un contenido informativo donde la rapidez se manifiesta en todas las acciones o procesos relativos al coche : el desplazamiento, la fabricación, la acción de aparcar, etc.

A fin de filtrar la significación apropiada, se debe recurrir al principio de metonimia integrada. Entre los contenidos informativos a los cuales se aplica el adjetivo 'rápido', el único que abre un *acceso funcional* a un espacio enciclopédico que concierne directamente al coche es la información relativa a la velocidad de desplazamiento. La velocidad de desplazamiento es la puerta de acceso a diferentes tipos de espacios informativos ligados al coche : la potencia del motor, la relación calidad-precio... En cambio, la velocidad en la fabricación y la velocidad en la acción de aparcar abren el acceso a espacios informativos que no caracterizan directamente al coche, sino al agente-fabricante o al agente-conductor. Tales informaciones, al no ser por tanto propiedades centrales del coche, no se pueden generar en el proceso de interpretación de 'coche rápido'.

La relación específica entre la velocidad de desplazamiento y la potencia interna del coche es un caso particular de una relación de alcance general : la velocidad de una actividad es la puerta de *acceso funcional* al espacio informativo relativo a las capacidades internas del agente que realiza tal actividad.

Esta relación general es el filtro que impide generar un significado aceptable en la interpretación de una expresión como 'cuchillo rápido' : tanto la velocidad de fabricación como la velocidad en la acción de cortar no caracterizan directamente al cuchillo, sino al agente-fabricante o al agente-cortador. Tales informaciones no son por tanto propiedades generables a partir de la entrada 'cuchillo'.

3.3 'coche grande'

Entre las informaciones enciclopédicas asociadas a la entrada 'coche' se encuentra la idea de dimensión, ya sea relativa a la superficie externa o al volumen interno.

El adjetivo 'grande', por su parte, se aplica a los constituyentes dimensionales del coche, es decir a su superficie externa y a su volumen.

La interpretación de la expresión 'coche grande' genera por tanto un doble contenido : objeto físico de grandes dimensiones y espacio con un volumen amplio.

El principio de metonimia integrada permite la generación de estos dos contenidos informativos. Tanto las dimensiones externas como el volumen interno son puertas de *acceso funcional* a espacios enciclopédicos que caracterizan directamente al coche. Las dimensiones externas dan acceso a la función de colocación en un espacio (es decir a la acción de aparcar el coche). El volumen interno, por su parte, abre el acceso al espacio funcional relativo a las acciones de llenar-vaciar el coche.

Estos dos contenidos informativos son, por tanto, propiedades centrales del coche, y por ello pueden generarse en el proceso de interpretación de 'coche grande'.

4 *Perspectivas de investigación : el control del proceso de generación en el discurso*

La interpretación de este último ejemplo — 'coche grande'— nos sitúa frente a un problema fundamental en el procesamiento semántico del lenguaje natural :

¿ cómo y dónde parar el proceso inferencial cuando se manipulan informaciones de naturaleza enciclopédica ?

Por ejemplo, el procesamiento de la expresión 'coche grande' desencadena un proceso inferencial que recupera todas las informaciones ligadas tanto a la superficie externa como al volumen interno del coche. Entre las

informaciones recuperadas se encuentran diferentes tipos de funciones : la colocación en un espacio, la posibilidad de contener objetos que ocupen todo o parte del volumen, etc.

Desde un enfoque psicolingüístico, el problema reside en saber si un agente-receptor "accede" al conjunto de informaciones que puede inferir cuando procesa esta expresión en un discurso, o bien no recupera más que las informaciones que son *pertinentes* con respecto a la información que aporta el discurso. El acceso abierto a todas las informaciones inferidas puede saturar la memoria de procesamiento del agente-receptor. Sin embargo, un acceso restringido a la información pertinente le permite manipular con eficacia un conjunto limitado de informaciones que no saturen su memoria de procesamiento.

La noción psicolingüística de "pertinencia" podría relacionarse, en el plano computacional, con la idea de control del proceso de generación en el discurso. Situemos la expresión 'coche grande' en el enunciado discursivo siguiente :

1. No fui capaz de aparcar el coche grande.

a. Tiene el morro muy largo.

b. ??Es demasiado espacioso.

Debido a las propiedades léxicas del verbo 'aparcar', la expresión compuesta 'aparcar el coche grande' genera un contenido informativo relativo a las dimensiones externas del coche, y no a su volumen interno. Quiere esto decir que la información accesible (y por tanto *pertinente*) para el agente que procesa el discurso es la que configura el coche con respecto a su forma externa. El enunciado (a), que trata de la forma externa del coche, es, así pues, perfectamente aceptable. En cambio, el enunciado (b), donde el adjetivo 'espacioso' genera un contenido relativo al volumen interno del coche, se revela inaceptable.

Este ejemplo parece mostrar que, en el discurso, se produce un control fuerte sobre el proceso de generación del contenido informativo accesible al agente-receptor. Si esta constatación se revela apropiada, uno de los objetivos fundamentales del procesamiento semántico del lenguaje natural será el de establecer y caracterizar los principios que guíen y regulen el

proceso de generación. En este artículo, nos hemos limitado a esbozar algunas ideas que apuntan en esta dirección.

Referencias bibliográficas

- Bierwisch M. (1989). *Dimensional adjectives : Grammatical structure and conceptual interpretation*. Berlin-Heidelberg-New York:Springer-Verlag.
- Franks B. (1995). "Sense generation : a "quasi-classical" approach to concepts and concept combination". *Cognitive Science*, vol. 19, nº 4, (pp. 441-505).
- Gamallo Otero P. (1998). *Construction conceptuelle d'expressions complexes : traitement de la combinaison "nom-adjectif"*. Tesis doctoral de la Universidad Blaise Pascal Clermont 2.
- Jayez J. & Godard D. (1995). "Principles as lexical methods". En las actas de *AAAI workshop on representation and acquisition of lexical knowledge, March 27-29, Stanford University*.
- Kleiber G. (1994). *Nominales. Essais de sémantique référentielle*. Armand Colin, Paris.
- Lakoff G. (1987). *Women, Fire, and Dangerous Things : What Categories Reveal About the Mind*. University of Chicago Press.
- Lang E. & Carstensen K. & Simmons G. (1991). *Modelling Spatial Knowledge on a Linguistic Basis*. Berlin-Heidelberg-New York:Springer-Verlag.
- Langacker R. (1987, 1991). *An introduction to Cognitive Grammar*. Vol I y II, Stanford University Press.
- Pustejovsky J. (1995). *The Generative Lexicon*. MIT Press : Cambridge.
- Schreuder R. & Flores d'Arcais G. (1989). "Psycholinguistics issues in the lexical representation of meaning". En *Lexical representation and Process*, William Marslen-Wilson (ed.), Massachusetts, (pp. 409-436).
- Sowa J. (1993). "Lexical structures and conceptual structures". En *Semantics and the Lexicon*. J. Pustejovsky (ed.), Kluwer academic publisher, Printed in the Netherlands, (pp. 223-262).
- Taylor J. (1995) "Approaches to word meaning : The network model (Langacker) and the two-level model (Bierwisch) in comparison". En *Current Approaches to the Lexicon*. R. Dirven et Johan Vanparys (eds.), Peter Lang, (pp. 3-25).

