

SEMINARIO DE LINGÜÍSTICA MATEMÁTICA

Participantes: J. Catron de García, V. Demonte, Ch. Destroyer, E. García Camarero, P. García Domínguez, J. Gerday de Polonio, I. Gómez de Liaño, C. Piera, A. Ríos, J. Sadaba, F. Saltor, C. Shields, F. Theunissen, R. Velilla.

Coordinador: V. Sánchez de Zavala.

Sesiones: días 5 y 19 de diciembre.

UNA TECNICA PARA LA IDENTIFICACION Y DESCRIPCION DE RELACIONES SEMANTICAS: LA TIPIACION DE PREDICACIONES.

Por Violeta Demonte

1. Supuestos básicos:

- a) Los significados de los elementos del lenguaje son relativos entre sí de una manera no fortuita y pueden ser descritos en términos de una estructura semántica.
- b) Toda estructura se plantea como un inventario de unidades y sus relaciones; las fundamentales son:
 - la relación de inclusión en clases (paradigmática)
 - la de "patterns" de coocurrencia (sintagmática)
 - la covariación de forma - significado.
- c) Esa estructura es susceptible de estudiarse por medio de la observación de sus manifestaciones en la conducta oral o en textos registrados.

1.1. Principios metodológicos:

Se aplica el método inductivo: Desde un conjunto de supuestos generales acerca de la naturaleza de los datos se deriva, en vez de conclusiones sobre detalles de los mismos, un conjunto de procedimientos que sirven para verificar las hipótesis y para investigar esos detalles. En dos partes: a) el desarrollo de una estrategia de investigación (estudio piloto). b) la aplicación de esa estrategia a un área dada del problema.

En nuestro caso interesará el paso a .

1.2. Objetivos y técnicas:

Para la descripción de una estructura lingüística se cuenta con cuatro elementos básicos: Unidades - Relaciones de inclusión en clases - Relaciones de "patterns" de coocurrencia - Covariación forma-significado.

Los dos primeros proveen el marco de referencia para establecer el objetivo de una descripción lingüística:

- a) Identificar las unidades pertinentes
- b) Clasificarlas en términos de miembros en clases y subclases de alcance cada vez más limitado.

Los otros dos sirven para el desarrollo de técnicas analíticas:

- a) De las relaciones de coocurrencia se derivan las técnicas de análisis distribucional; que se han utilizado tradicionalmente para identificar unidades de un orden alto de complejidad (oraciones, cláusulas, etc.) y asignar a ellas, en clases de distribución, las unidades mínimas del tipo del morfema.

Aplicadas en el plano semántico suponen: a) definir el tipo de unidad lingüística a ser examinada (the current forms) b) indicar las unidades del contexto anterior que son relevantes para los propósitos de la investigación: los definidores contextuales.

- b) Las otras se denominarán técnicas de forma-significado: Significado y forma tienen dos manifestaciones que son observables y tratables operacionalmente: Traducción y paráfrasis. La elección de una u otra depende de que el estudio se conduzca sobre bases monolingüales o bilingüales. Este estudio se ocupa de la paráfrasis. Parafraseo es la operación en la cual una declaración original, o parte de ella, se reemplaza por otra declaración, o parte de declaración, sin que se pierda el significado original.

2. Especificación de la técnica de Tipificación de Predicaciones

2.1. La técnica consiste en pedir a varios sujetos que parafraseen las oraciones de un texto de acuerdo con ciertas instrucciones específicas.

2.1.1. Instrucciones dadas a los informantes

Se dan verbalmente e informalmente. Lo esencial es responder con lo primero que se les ocurra. No se exige paralelismo gramatical entre la forma original y su paráfrasis, por eso las instrucciones tampoco se dan en términos gramaticales. La única condición es que se satisfaga la propiedad de comparación: semejanza semántica entre la original y la resultante.

2.1.2. Operaciones del parafraseo

Para poder aprehender sus pasos debemos antes definir:

Predicación: Toda oración que presenta una conexión funcional entre dos términos A y B.

Forma señal: Toda forma de habla particular que señala la elección de un tipo determinado de oración.

Forma reemplazante o esquema de parafraseo: Toda conexión contenida en el esquema parafrástico, indicativa de la oposición de los dos términos. Así, si la predicción se asigna al tipo: "A se basa en B" la forma reemplazante es "se basa en"

La tarea concreta del informante consiste en: Ante un corpus de Predicaciones debe:

- a) Asignarlas a un tipo de predicación determinado (éstas están presentes en una lista que el sujeto maneja)
- b) Indicar sus formas señal
- c) Señalar para cada una un esquema de parafraseo (también se listan)
- d) Desarrollar la paráfrasis.

2.1.3. Organización de un archivo de datos semánticos

Con los siguientes componentes: a) Registro de documentos. b) Registro de tipos de predicaciones. c) Registro de predicaciones modificadas. d) Registro de formas-señal: Cuando éstas están formadas por más de una palabra se clasifican por la principal. e) Registro de familias de formas-señal: Las formas-señal morfológicamente relativas se integran en una familia: ej.: aplicar, aplicación, aplicable, aplica, etc. Cuando se trata de morfemas irregulares se agrupan bajo la forma canónica: ej.: "fui" bajo "ir".

2.1.4. Concepto de "Rasgos semánticos" y de "espectro semántico"

Las unidades pertinentes son las formas señal que se caracterizan en términos del conjunto de tipos de predicación a los cuales señalan. El supuesto operacional es que a cada tipo de predicación corresponde un rasgo semántico de la forma señal. Por rasgo semántico se entiende las invariantes características de significado de una unidad lingüística dada: las constantes que subyacen las variaciones exhibidas por la unidad.

A una forma dada que se haya identificado como señal se le asignarán tantos rasgos semánticos como sea el número de tipos de predicación que hayan sido sugeridos por ella en el curso de procesar una muestra considerable de lenguaje

Tipos de predic. (rasgos semánticos)

| | |
|-------------------------|---|
| Ej.: | Indicación de origen (A nace de B) |
| Forma señal: Establecer | Indicación de sucesión (A es seguido por B) Indicación de relación básica (A se basa en B) |

Ejemplo de parafraseo para: Indicación de sucesión

Declaración Original: Las especies domesticadas establecen importantes variaciones.

Forma - Señal: establecen

Forma reemplazante: "es seguida por"

Declaración resultante: La domesticación de las especies es seguida por modificaciones.

Rasgo semántico (inferido del tipo de Predicac.) : Sucesión

Por lo tanto pueden asignarse a "establecer" por lo menos tres rasgos semánticos. Del examen de una muestra mayor quizá podrían emerger otros rasgos adicionales.

Considerada una muestra suficientemente larga podría analizarse la frecuencia relativa con que cada rasgo semántico se asigna a una forma señal, y esto nos daría el factor de relieve de cada rasgo.

Todos los rasgos semánticos de una forma señal, modificados por sus respectivos factores de relieve, constituyen el espectro semántico de esa forma-señal.

3. Análisis de datos

Como se ha señalado en 1.2 el objeto último de este tipo de investigación es el desarrollo de un método para una clasificación semántica. Se propone en consecuencia que dicho método puede basarse en el análisis de espectros semánticos. Si el paso primero fue la constitución de espectros semánticos para cada forma señal y si las formas señal son susceptibles de agruparse en familias, conviene ahora plantear la conformación de espectros de familias de formas-señal. En este sentido un espectro semántico consiste de todos los rasgos semánticos asignados a una familia dada de formas-señal, modificados por sus respectivos factores de relieve. Se supone pues que el espectro constituye una caracterización semántica de la familia de formas señal por medio de cuyos miembros se derivó. A esta segunda representación operacional (los espectros de familias de formas-señal) sigue una serie de pasos estadísticos muy precisos, y que detallaremos sólo sumariamente, que contribuyen, por un lado, a la interpretación de relaciones semánticas, y por otro (causa de lo antedicho) significan la inclusión de los rasgos semánticos analizados en clases cada vez más amplias e internamente más heterogéneas.

Sintetizaremos los pasos de este análisis, que se aplican a las familias de formas-señal en tanto éstas se interpretan como unidades analíticas:

- a) Teniendo en cuenta los porcentajes máximos de tipos de predicación en el espectro semántico de cada familia de formas-señal, se realiza un listado decreciente de las mismas. Aquellas familias que comparten un mismo tipo de predicación como "cima" (peak) pueden formar una clase de rasgos prominentes (class peak-derived).

- b) Con cada familia, es posible establecer una comparación, por pares, entre su espectro y los de todas las demás familias, sumando en cada par de espectros los porcentajes de solapamiento respecto de cada tipo de predicación. Los items así obtenidos deberían ordenarse en orden de similaridad decreciente para formar las series de semejanza semántica de esa familia. Obviamente, habrá tantas series de semejanza como familias se consideren .
- c) Para cada familia pueden listarse todos los miembros de sus series de semejanza que tienen un índice de similaridad de un 50% o más, y repetir el listado con las familias de cada uno de los miembros de esa serie primera (teniendo en cuenta en todos los casos el umbral mínimo de semejanza: 50%). La agrupación resultante puede identificarse como un "Grupo de semejanza semántica". Debe inferirse que si las "clases de rasgos prominentes" reflejan semejanzas semánticas en términos de un solo rasgo, estos grupos nos dan semejanzas en términos de compuestos semánticos y pueden postularse pues como clases semánticas tentativas. v. 1.2. .
- d) Para cada grupo de semejanza deberían sumarse los porcentajes de calificación (score) de cada tipo de predicación y dividir la suma por el número de familias del grupo hasta obtener un Espectro semántico de grupo.
- e) El paso b puede repetirse con cada familia, pero comparando esta vez la semejanza de su espectro semántico con cada uno de los grupos de espectros semánticos. Clasificados en orden decreciente formarían las Serie de Grupos de semejanza.

-

El procedimiento puede aplicarse repetidamente para procesar extensiones de la muestra inicial. Con cada aplicación de esta serie de pasos se espera que tanto las familias como sus clasificaciones se modifiquen a la luz del aumento de los datos de base. Es obvio también que el número de familias aumentará tanto como el texto procesado, aunque no necesariamente en forma lineal.

La ventaja de este proceso consiste en que las familias y sus variaciones pueden redefinirse sin que haya necesidad de cambiar la técnica.

Por otra parte estos pasos analíticos pueden aplicarse a otros tipos de unidades semánticas, supuesto que tales unidades se traten en términos de un conjunto finito de representaciones análogo al de los tipos de predicaciones.

Bibliografía

Detalles de la aplicación de esta técnica a partir de un corpus de textos técnicos:

- Predication-Typing: A pilot study in semantic Analysis.
By: Paul Garvin, Jocelyn Brewer and Madeleine Mathiot.
Supplement to Language. Volume 43. Number 2. Part II.
June 1967.

DIVERSOS ESTUDIOS SOBRE EL ANALISIS DE COMPONENTES SEMANTICOS

Por. V. Sánchez de Zavala

Conviene completar el estudio del día pasado sobre la tipificación de predicaciones con otro análisis experimental de los componentes semánticos de las formas léxicas verbales: el de E. H. Bendix Componential Analysis of General Vocabulary: The semantic structure of a set of verbs in English, Hindi and Japanese (Bloomington de Indiana / La Haya, I. Univ., Mouton, 1966).

En esta obra lo que se pretende es poner a punto un método de análisis del significado de elementos léxicos en su funcionamiento efectivo: si, por ejemplo, se trata de pretender esclarecer lo que significa el verbo dar, lo que se analizará será el significado de expresiones en las que una "forma finita" se emplee de modo simple; por ejemplo, en oraciones simples con el esquema

.... da a

o sea,

A da B a C ;