

EL ANÁLISIS SEMÁNTICO-DISTRIBUCIONAL. UNA APORTACIÓN A LA LEXICOGRAFÍA

Humberto Hernández
Universidad de La Laguna

Abstract

Our purpose is to offer a valid method for delimiting the real meanings of a polysemic word in order to avoid the chaotic consequences of the use of *acceptation* as a lexicographic notion.

Revelar la existencia de un número fijo de componentes semánticos en todas las lenguas y construir un metalenguaje semántico universal, cuyo vocabulario estaría formado por componentes de sentido atómicos, han sido los objetivos en que han centrado sus investigaciones la mayoría de los semantistas, aunque los métodos para conseguirlo han diferido notablemente.

Podrían agruparse en torno a dos grandes bloques todas las corrientes semánticas actuales. Por una parte, los mentalistas que, como Ogden y Richards, defienden la no relación directa entre las palabras y las cosas (recuérdese su conocido triángulo), y, por tanto, entienden el significado como una «relación recíproca y reversible entre el sonido (nombre) y el sentido»¹. Los antimentalistas, por otra parte, reacios a operar con entidades mentales y a no considerar fiable el método de la introspección, prefieren suprimir el vértice superior del triángulo («pensamiento» o «referencia». «Sentido» para Ullmann), y establecer una relación directa entre el «nombre» y la «cosa».

La polémica sostenida entre los defensores de ambas concepciones ha sido grande, y, así, los estructuralistas americanos han atacado a los teóricos mentalistas con objeciones tan extremas que parecen querer identifi-

car el significado con el referente². Otros, como Haas, han entendido que «el signo lingüístico se supone que surge de una correspondencia, de una especie de paralelismo psico-físico, entre una forma y un significado»; crítica que, además, ataca al intento de definir el significado en términos de distribución³.

No vamos a detenernos ahora en consideraciones sobre la definición de las unidades semánticas, si bien estamos de acuerdo con R. Trujillo en entender que el principal objeto de la semántica es «reducir lo inabarcable de la información semántica del hablar a una serie de contrastes formales básicos», porque en caso contrario, «sólo quedaría una modesta disciplina escolar, una serie de normas prácticas para el comentario semántico de los textos, esencial, sin duda, desde el punto de vista pedagógico, pero incapaz, como cualquier otro tipo de comentario, de ser objeto de una verdadera ciencia»⁴. Si vamos, por el contrario, a procurar aportar datos que ayuden a delimitar dichas unidades, pues es éste el primer paso que habría que dar en toda teoría lingüística, ya que si no se pudiesen determinar las entidades invariantes de una lengua, no sólo negaríamos la existencia de una semántica científica, sino también la de una auténtica lingüística⁵. Gran parte de los trabajos realizados hasta hoy en este sentido han sido realizados por investigadores de tendencia mentalista, los métodos que han predominado han sido los intuitivos y los resultados obtenidos podrían ser, en consecuencia, de dudosa validez.

En *Philosophical Investigations*⁶, L. Wittgenstein formula por primera vez de una manera clara lo que será una concepción operacional del significado: «el significado de una palabra es su uso en el lenguaje»; de este modo se intenta suprimir en la definición y delimitación de las unidades semánticas cualquier proceso mental vago y hacerlo en términos contextuales. Opina S. Ullmann que la doctrina operacional no ha tenido oportunidad de probar sus métodos, y nos remite al intento de R. Wells de estudiar el uso de algunos adjetivos ingleses, intento que, aunque interesante, le parece demasiado limitado para dar una idea de las posibilidades de este método. Pero más adelante —y aquí creemos que se encuentra una de sus principales aportaciones— afirma que la definición operacional debe ser estimada, no como una alternativa, sino como un complemento válido a la teoría referencial, pues «contiene la saludable advertencia que tanto los semánticos como los lexicógrafos harían bien en atender, de que el significado de una palabra *solamente* puede averiguarse estudiando su uso (...). El investigador debe comenzar por reunir una muestra adecuada de contextos y abordarlos luego con un espíritu abierto, permitiendo que el significado o los significados emerjan de los contextos mismos. Una vez que se ha concluido esta fase, puede pasar con seguridad a la fase referencial y procurar formular el significado o los significados así identifica-

dos»⁷. Punto de vista, sin duda, interesante, pues lo que hasta el momento parece haber ocurrido es que la mayoría de los semantistas han utilizado en sus investigaciones sólo uno de los métodos —aferrados, quizás, a dogmatismos doctrinarios que no conducen a ninguna parte—, sin percatarse de que ambos modos de acceso, lejos de excluirse, se complementan.

Conseguir delimitar con precisión los diferentes significados de una palabra acabaría con una de las más graves deficiencias de la lexicografía, ya que las palabras polisémicas en el diccionario aparecen, como es bien sabido, con una serie de acepciones; pero lo que no es tan sabido, es que bajo el término *acepción* se engloban todos los sentidos que se registran en el habla para una determinada palabra, sentidos que, por otra parte, aparecen totalmente desordenados, y, lo que es peor, sin distinguir cuáles de ellos son signos (invariantes) y cuáles simples variantes. Sobre este hecho señala R. Trujillo lo siguiente: «Lo importante, dentro de la descripción lexicográfica, es que se separen antes que nada los distintos signos unidos a un significante y que luego se describan las variantes más normales de cada invariante registrada, empezando por las combinatorias y terminando por las libres. Con esto se evitan las consecuencias caóticas de la aplicación del concepto de *acepción* sin más, ya que éste no tiene cabida en una descripción lingüística, al mezclar indiscriminadamente los signos distintos con sus variantes»⁸. Pero, ¿cuál sería el método válido para llevar a cabo tal delimitación? Por medio del análisis sémico podemos establecer la composición semántica de una unidad y determinar que es diferente de otra al existir una diferencia de expresión. Ahora bien, el problema se plantea cuando a una diferencia semántica corresponde una misma expresión fonológica, que es precisamente el caso que nos ocupa. La semántica más reciente ha utilizado el análisis distribucional ya que supone que el sentido de las palabras está determinado por sus propiedades gramaticales. El punto de partida lo encontramos en Z. Harris, quien formula el principio de que dos morfemas que tienen significado diferente difieren también en alguna parte de su distribución, entendiéndose como tal «la suma de todos los contextos en los cuales este elemento se encuentra; es decir, la suma de todas las posiciones diferentes de un elemento con relación a los otros»⁹. J. Apresjan acotará, posteriormente, el concepto de distribución, pues entiende que «si se transfiere ese concepto al léxico, resultará, en la práctica, imposible de utilizar, puesto que la distribución de la mayoría de los elementos parece casi ilimitada. Es más racional, entonces, representar la distribución de cada elemento bajo cierta forma generalizada y no bajo forma de enumeración de todos los contextos, sin excepción en los que puede encontrarse»¹⁰. Fue este mismo investigador quien a través del análisis de veinte verbos polisémicos tomados del *The Advanced Learner's Dictionary of the English Language*¹¹, elegidos por decenas en orden

continuo a partir del comienzo de las letras A y D, y por el análisis de los verbos *to go* y *to come* en mil doscientas treinta y nueve páginas de texto, llega a distinguir unos sesenta sentidos de los veinte verbos analizados, además de los sentidos del verbo *to come* y una descripción del verbo *to go*; esto le permitió formular que «puede establecerse una correspondencia biunívoca entre ciertos significados y ciertas distribuciones». Su método de trabajo consistió en atribuir fórmulas de distribución a cada uno de los significados de una palabra polisémica que en el diccionario aparecían como unidades independientes; comparaba luego las fórmulas de distribución de los diferentes sentidos y unía en uno solo a aquellos que tenían la misma fórmula. De este modo, si en el diccionario aparecían separados cuatro sentidos de determinada palabra y la distribución difería para dos de ellos coincidiendo para los otros dos, consideraba que para la lengua lo esencial era la diferencia entre esos dos sentidos; así, por ejemplo, el *The Advanced Learner's Dictionary of the English Language* distingue para el verbo *to dare* los siguientes sentidos: 1) 'arriesgarse a, tener coraje', y 2) 'tener el coraje de, osar'; al observar que la distribución coincide para ambos sentidos, deduce que la lengua no hace distinción entre ellos y los une en uno solo¹².

Es cierto que muchas veces a una diferencia de matriz semántica corresponden diferentes distribuciones; por ejemplo, $S_{(\text{inanimado})} + \text{comprender} + \text{CD} = \text{'contener'}$ («La finca comprende un monte de caza y una dehesa»), pero $S_{(\text{animado})} + \text{comprender} + \text{CD} = \text{'entender'}$ («No comprendo este escrito»). También puede ocurrir que dos signos diferentes tengan idéntico esquema distribucional, y, así, $S\emptyset + \text{haber}_{(3^{\text{a}} \text{ pers. sing.})} + \text{CD}$, vale tanto para el significado 'tener lugar algo' («Esta mañana hubo junta»), como para 'existir' («En la reunión había quince personas») que, evidentemente, son signos distintos. Es más, variantes semánticas de un mismo signo pueden realizar fórmulas distribucionales diferentes: $S_{(\text{persona-libro})} + \text{decir} + \text{CD} = \text{'narrar, contar'}$ («Juan dice cosas interesantes»), y $S\emptyset + \text{decir}_{(\text{prml.})} + \text{CD/Modal} = \text{'debe expresarse'}$ («Se dice anduve y no andé»), que son variantes de la invariante 'manifestar con palabras...'. Todo esto demuestra que es insostenible la teoría que defiende la relación isomórfica entre distribución sintáctica y significado, que, dicho sea de paso, ha sido aceptada hasta hoy por muchos lingüistas; el mismo Gerd Wotjak no se decide a contradecir a Apresjan: «Aunque en muchos casos se puede demostrar, efectivamente, la existencia de una relación inmediata entre distribución y significado, queremos preservarnos de admitir, siguiendo enteramente a G. Helbig, una correspondencia directa general entre la estructura sintáctica y el contenido»¹³. Creemos que ha sido R. Trujillo quien ha concedido a la distribución sintáctica su justo valor: «a una diferencia de matriz semántica puede corresponder una misma expresión fonológica y

actuar entonces como mecanismo significante diferencial una fórmula distribucional determinada»¹⁴.

En consecuencia, y dado que la distribución sintáctica por sí sola es insuficiente para delimitar los diferentes significados de las palabras, proponemos el siguiente método: en primer lugar realizar el análisis semántico¹⁵ de la unidad en cuestión, que nos daría el núcleo semántico irreductible, entendiendo que habría tantos signos como núcleos, y a continuación utilizar el análisis distribucional como método de comprobación empírico, que viene a coincidir, en cierto modo, con el planteamiento metodológico defendido por Ullmann.

En un trabajo que realizamos con los sesenta verbos más frecuentes del español¹⁶ según el *Frequency Dictionary of Spanish Words*¹⁷ aplicamos el método semántico-distribucional con el fin de determinar el número de significados que poseían dichas palabras, y de las 872 acepciones registradas por el *Diccionario* académico para esos sesenta verbos sólo resultaron 189 invariantes. Apresjan había distinguido 60 sentidos para los veinte verbos polisémicos analizados. Obsérvese que la relación entre el número de invariantes resultantes y el número de palabras es muy próxima: 3 en el estudio de Apresjan, 3,15 en el nuestro. Ciertamente es que el criterio de selección del corpus analizado no fue el mismo, sin embargo, los datos dicen mucho en favor de este tipo de análisis.

Notas

1. Vid. S. Ullmann, *Semántica*, Madrid, Aguilar, 1965, pp. 65-66.
2. *Ibíd.*, p. 67.
3. *Ibíd.*, p. 70.
4. Vid. «Sobre la naturaleza de los rasgos semánticos distintivos», en *Logos Semantikos. In Honorem E. Coseriu*, Vol. III, Madrid, Gredos, 1981, p. 162.
5. Vid. R. Trujillo, *Elementos de semántica lingüística*, Madrid, Cátedra, 1976, p. 54.
6. Oxford, 1953. Citamos por Ullmann, op. cit., pp. 73-74.
7. Vid. S. Ullmann, op. cit., pp. 76-77.
8. Vid. *Elementos...*, cit., p. 249.
9. *Methods in Structural Linguistics*, The Un. of Chicago Press, pp. 15-16. Citado por J. Apresjan, «Análisis distribucional de los significados y campos semánticos estructurados», en T. Todorov (comp.), *Investigaciones semánticas*, Buenos Aires, Nueva Visión, 1978, p. 53.
10. Vid. J. Apresjan, art. cit., p. 54.
11. De A. Hornby.
12. Vid. J. Apresjan, art. cit., p. 63.
13. *Investigaciones sobre la estructura del significado*, Madrid, Gredos, 1979, p. 154.
14. *Elementos...*, p. 124.
15. Hay que tener en cuenta que el análisis semántico en lexicografía debe estar basado en la situación cultural del entorno, pues exigir al lexicógrafo que para definir una unidad tenga que situarla previamente en su campo semántico y delimitar la parcela significativa que en él ocupa sería identificar la semántica teórica con la lexicografía. Vid., por ejemplo, L.F. Lara, «Del análisis semántico en lexicografía», en V.V.A.A., *Investigaciones Lingüísticas en lexicografía*, México, El Colegio de México, 1979, pp. 157-266.
16. Vid. nuestra Memoria de Licenciatura inédita, *Análisis semántico-distribucional de verbos españoles*, La Laguna, 1982.
17. A. Juilland y E. Chang-Rodríguez, London, The Hague, Paris, Mouton & Co., 1964.