

Theories of Meaning¹ (SEP)

Jeff Speaks

First published Tue Jan 26, 2010; substantive revision Tue Jun 1, 2010

URL = <http://plato.stanford.edu/entries/meaning/>

La expresión ‘teoría del significado’ ha estado presente, de una u otra manera, en un gran número de disputas filosóficas a lo largo del último medio siglo. Desafortunadamente, la misma expresión ha sido utilizada para significar muchas cosas diferentes.

En lo que sigue nos ocuparemos de dos tipos de ‘teorías del significado’. El primer tipo –una *teoría semántica*– es la clase de teorías que intentan asignar *contenidos semánticos* a las expresiones de un lenguaje. Este tipo de enfoque puede ser subdividido de acuerdo a si asigna o no *proposiciones* como significado de las oraciones y, si ese es el caso, de acuerdo al enfoque que cada teoría proponga en relación a la *naturaleza* de las proposiciones.

El segundo tipo de teorías –*teorías fundacionales* del significado– está formado por aquellas que explican los hechos *en virtud de los cuales* las expresiones poseen [o llegan a poseer] los contenidos semánticos que de hecho poseen. Las teorías fundacionales pueden ser divididas, a su vez, entre aquellas que explican los significados de las expresiones de un lenguaje –tal como son usados por determinado un grupo— en función de los contenidos de los *estados mentales* de los miembros de ese grupo [‘de hablantes’]; y aquellas otras que intentan explicarlos sin hacer referencia a tales contenidos.

- 1. Two kinds of theory of meaning (Dos tipos de teorías del significado)
- 2. Semantic theories (Teorías Semánticas)
 - 2.1 Propositional semantic theories (Teorías semánticas proposicionales)
 - 2.1.1 The theory of reference (Teoría de la referencia)

¹ La expresión ‘meaning’ puede traducirse (y de hecho se traduce) tanto por ‘significado’ como por ‘significación’, por lo que el título del presente artículo bien podría ser ‘Teorías de la significación’. En diferentes traducciones, y dependiendo del tema tratado, se traduce por una palabra o por otra (por ejemplo, la distinción de Quine entre ‘theory of reference’ y ‘theory of meaning’ se traduce como la distinción entre ‘teoría de la referencia’ y ‘teoría de la significación’. [nota mía: esto creo que es importante aclararlo, porque se presta a todo tipo de confusiones, en inglés ‘meaning’ y ‘meaningful’ tienen que ver tanto con el hecho *de que* (por ejemplo) alguna expresión posea o no significado, como con la cuestión de *cuál* es el significado de determinada expresión, además de otros usos relacionados. Además ‘to mean’ utilizado como verbo transitivo puede traducirse por ‘significar (algo, el hablante)’ o por ‘querer decir (algo, el hablante)’. En nuestro caso, traduciremos siempre ‘meaning’ = ‘significado’.

- 2.1.2 Theories of reference vs. semantic theories (Teorías de la referencia vs. Teorías semánticas)
- 2.1.3 The relationship between content and reference (Relación entre contenido y referencia)
- 2.1.4 Character and content, context and circumstance (Carácter y contenido, contexto y circunstancia)
- 2.1.5 Possible worlds semantics (Semántica de mundos posibles)
- 2.1.6 Russellian propositions (Proposiciones russellianas)
- 2.1.7 Fregean propositions (Proposiciones fregeanas)
- 2.2 Non-propositional theories (Teorías no proposicionales)
 - 2.2.1 The Davidsonian program (El programa de Davidson)
 - 2.2.2 Chomskyan internalist semantics (Semántica internalista chomskyana)
- 3. Foundational theories of meaning (Teorías fundacionales del significado)
 - 3.1 Mentalist theories (Teorías mentalistas)
 - 3.1.1 The Gricean program (El programa de Grice)
 - 3.1.2 Meaning, belief, and convention (Significado, creencia y convención)
 - 3.1.3 Mental representation-based theories (Teorías basadas en la representación mental)
 - 3.2 Non-mentalist theories (Teorías no mentalistas)
 - 3.2.1 Causal origin (Origen causal)
 - 3.2.2 Truth-maximization and the principle of charity (Maximización de la verdad y el principio de caridad)
 - 3.2.3 Regularities in use (Regularidades en el uso)
 - 3.2.4 Social norms (Normas sociales)
- Bibliography (Bibliografía)

1. Dos tipos de teorías del significado

En “Semántica General”, David Lewis escribió:

Distingo [entre] dos tópicos: primero, la descripción de lenguajes o gramáticas posibles en tanto sistemas semánticos abstractos de acuerdo a los cuales ciertos símbolos son asociados con aspectos del mundo; y segundo, la descripción de los hechos psicológicos y sociológicos de acuerdo a los cuales uno de estos sistemas abstractos es utilizado por cierta persona o población. Mezclar ambos tópicos sólo daría lugar a confusión. (Lewis, 1970, 19)

Lewis tenía razón. Aunque los filósofos no siempre se las hayan ingeniado para mantener ambas cuestiones separadas, existe una clara distinción entre las preguntas ‘¿Cuál es el significado de este o aquel símbolo (para una determinada persona o grupo)?’ por un lado; y ‘¿En virtud de qué hechos relativos a esa persona o grupo tiene ese símbolo el significado que tiene?’², por otro.

A estas dos preguntas diferentes corresponden dos tipos de teorías del significado diferentes. Un tipo de teoría del significado –una *teoría semántica*– consiste en una especificación de los significados de las palabras y oraciones en un cierto sistema de símbolos. Por lo que las teorías semánticas responderán a la pregunta ‘¿cuál es el significado de tal o cual expresión?’. Un tipo diferente de teoría –una *teoría fundacional* del significado – intentará explicar *qué hechos relativos a cierta persona o grupo de personas otorgan a los símbolos que componen su lenguaje los significados que tienen*. Ciertamente, la forma que deba adoptar una teoría semántica correcta impondrá ciertas limitaciones sobre cuáles deban ser las características de una teoría fundacional del significado correcta, y viceversa; pero ello no cambia el hecho de que las teorías semánticas y las teorías fundacionales del significado son sencillamente *dos clases diferentes* de teorías, cuya función es responder diferentes tipos de preguntas.

Para aclarar la distinción entre unas y otras, puede ser de ayuda considerar una distinción análoga. Imaginemos a un antropólogo especializado en estudiar lo que los grupos humanos consideran buenos modales o ‘etiqueta’, enviado a observar una tribu distante. Una de las tareas que el antropólogo puede acometer es la de describir los buenos modales de esa tribu –describir las diferentes categorías según las cuales los miembros de la tribu clasifican las acciones, y decir qué

² [nota mía] Una expresión menos exacta, pero quizás más clara de esta preocupación sería: *¿por qué tienen los significados que tienen?*

tipo de acciones corresponden a cada categoría. Esta sería una empresa análoga a la del filósofo del lenguaje interesado en la semántica; su trabajo será decir cuáles son los significados de los diferentes tipos de expresiones de un lenguaje, y qué significados corresponden a determinados tipos de expresiones.

Pero nuestro antropólogo también puede estar interesado en la *naturaleza* de esos ‘buenos modales’; puede preguntarse, por ejemplo, cómo un cierto conjunto de reglas de comportamiento llega a convertirse en el sistema que de hecho gobierna el comportamiento de ese grupo. Dado que, presumiblemente, el hecho de que un grupo obedezca un sistema de etiqueta en lugar de otro cualquiera debe poder explicarse por algún hecho característico de ese grupo; el antropólogo podría formular esta nueva pregunta de la siguiente manera: ‘¿En virtud de qué hechos relativos a una determinada persona o grupo llega esa persona o grupo a adoptar un sistema de etiqueta [‘buenos modales’] por sobre otro?’. En ese caso, nuestro antropólogo se verá embarcado en una tarea similar a la de la construcción de una teoría fundacional del significado. No se interesará por las propiedades particulares de ciertos tipos de acciones en un determinado grupo, sino en cómo ciertos tipos de acciones pueden, en cualquier grupo, llegar a adquirir tales propiedades [1]. Claro que nuestro investigador bien puede estar interesado en ambas preguntas al mismo tiempo; pero se trata de preguntas claramente diferentes. Del mismo modo, las teorías semánticas y las teorías fundacionales del significado son, claramente, dos tipos diferentes de teorías.

En los últimos tiempos de la historia de la filosofía, la expresión ‘teoría del significado’ ha sido utilizada para referirse a ambas clases de teorías (semánticas y fundacionales). Como esto puede llevar a confusión, en lo que sigue evitaré utilizar la expresión que da título a este artículo, y atenerme a las más específicas ‘teorías semánticas’ y ‘teorías fundacionales del significado’. ‘Teoría del significado’, *simpliciter*, ha de ser entendido de manera ambigua, correspondiendo a ambas interpretaciones.

Antes de adentrarnos en el tratamiento de estos dos tipos de teorías, es importante hacer notar que una prominente tradición en la filosofía del lenguaje *niega* que existan hechos [objetivos] sobre el significado de las expresiones lingüísticas (Ver, por ejemplo, Quine 1960 y Kripke 1982; para una discusión crítica, ver Soames 1999.) Si este tipo de escepticismo relativo al significado es correcto, entonces *ninguna* teoría semántica y *ninguna* una teoría fundacional podrá ser verdadera, dado que sencillamente *no hay* ningún hecho relevante del tipo de los que estas teorías pretenden describir o analizar. La discusión de tales argumentos escépticos me llevaría más allá del alcance del presente

artículo, por lo que en adelante sencillamente asumiré que este tipo de escepticismo sobre el significado es falso.

2. Teorías Semánticas

Cualquier intento de explicar los diferentes enfoques de las teorías semánticas en la filosofía del lenguaje contemporánea parece afrontar, desde el comienzo, un obstáculo insuperable. Dado que no podemos decir que haya dos lenguajes que compartan la misma semántica –es decir, que se compongan de las mismas palabras, con los mismos significados, etc.– parece difícil ver cómo es posible decir algo sobre diferentes enfoques semánticos en general (en lugar de hablar de las diferentes semánticas de cada lenguaje particular). Pero hay una solución relativamente sencilla para este problema: si bien es evidente que una cosa es la semántica del idioma español y otra la semántica del francés, la mayoría tiende a asumir que *cualquier* lenguaje natural contará con una teoría semántica cuya *forma* (en un sentido que luego explicaremos) será básicamente la misma. Por tanto, nuestro objetivo será presentar al lector los principales enfoques en la semántica de los lenguajes naturales –esto es, las formas básicas que puede tomar una semántica de los lenguajes naturales—y no un examen detallado de las múltiples posiciones sobre la semántica de algunas expresiones particulares. (Para algunas de esas discusiones, ver artículos, [names](#), [descriptions](#), [propositional attitude reports](#), y [natural kinds](#).)

Antes de pasar al análisis de las teorías del significado, una última aclaración. Para que un teórico semántico pueda emprender la explicación de los significados de las expresiones de un lenguaje, primero necesita contar con una idea clara sobre cuáles son las cosas cuyos significados intentará explicar [(es decir, cuáles son las cosas que *tienen* significado)]. Así planteado, puede no parecer un gran problema: después de todo, ¿qué otras cosas *tienen* significado más que las *oraciones* y las *partes de oraciones*? Esto es correcto en un principio; pero la tarea de explicar cuáles son las partes semánticamente significativas de una oración, así como de qué manera se combinan entre ellas para dar lugar a una oración, está lejos de ser una tarea trivial y reviste importantes consecuencias para cualquier clase de teoría semántica. El tratamiento de este tipo de teorías, que tratan de la *forma lógica*, o la *sintaxis* del lenguaje, se encuentra más allá del alcance del presente artículo; por lo que, desafortunadamente, no podremos tratar de figuras como la de Richard Montague, cuyo trabajo sobre la sintaxis y su conexión con la semántica ha sido central para el desarrollo de la teoría semántica en las últimas décadas. (Los ensayos de Montague están recopilados en Montague 1974; para una discusión de la importancia de su trabajo, ver §3.3 de Soames 2010.)

La mayoría de los filósofos del lenguaje actuales consideran que el significado de una expresión es un cierto tipo de entidad, y que el trabajo de la semántica será el de encontrar la correspondencia entre ciertas expresiones y con ciertas entidades. Para estos filósofos, la pregunta central sobre la forma correcta que deberá adoptar una teoría semántica se relaciona con la naturaleza de estas entidades. Como las entidades que corresponden a las oraciones se denominan *proposiciones*, las denominaremos *teorías semánticas proposicionales*. Sin embargo, no todos los filósofos del lenguaje creen que los significados de las oraciones consistan en proposiciones, y algunos ni siquiera aceptan que haya tales entidades. En atención a ello, dividiré los enfoques semánticos en teorías semánticas *proposicionales* y *no-proposicionales*.

2.1 Teorías semánticas proposicionales

La manera más sencilla de comprender los diferentes tipos de teorías semánticas proposicionales será comenzando con otro tipo de teoría: una *teoría de la referencia*.

2.1.1 Teoría de la referencia

Una teoría de la referencia es una teoría que, al igual que una teoría semántica proposicional, busca encontrar la correspondencia que se da entre ciertas expresiones y ciertos valores. Sin embargo, a diferencia de ésta, una teoría de la referencia no busca establecer la correspondencia entre las expresiones y sus significados; sino que intenta establecer la correspondencia entre las expresiones y la manera en que cada una de esas expresiones contribuye la determinación del *valor de verdad* de las oraciones en las que aparecen. (Aunque luego veremos que esta perspectiva sobre la referencia de una expresión debe ser acotada en cierta forma).

Esta concepción de la teoría de la referencia tiene sus raíces en el intento de Gottlob Frege de formular una lógica capaz de formalizar las inferencias matemáticas (Ver, sobre todo, Frege 1879, y 1892). Para ilustrar cómo se puede construir una teoría de la referencia de esta clase, conviene empezar con el ejemplo de los nombres propios. Consideremos las siguientes oraciones:

- (1) Barack Obama es el cuadragésimo cuarto [44°] presidente de los Estados Unidos.
- (2) John McCain es el cuadragésimo cuarto [44°] presidente de los Estados Unidos.

(1) es verdadera, mientras que (2) es falsa. Obviamente, esa diferencia en los valores de verdad debe estar dada por alguna diferencia entre las expresiones ‘Barack Obama’ y ‘John McCain’. Pero, ¿qué característica de esas expresiones explica la diferencia en valor de verdad entre una y otra oración?. Es plausible que se trate del hecho de que la expresión ‘Barack Obama’ *está por* [‘refiere’, ‘denota’ o ‘representa’] el hombre que de hecho *es* el 44º presidente de los Estados Unidos, mientras que la expresión ‘John McCain’ *está por* [‘refiere’, ‘denota’ o ‘representa’] otro hombre, que *no es* el 44º presidente de los Estados Unidos. Esto indica que la *referencia* de un nombre propio —es decir, su contribución a la determinación de las condiciones de verdad de las oraciones en que aparece— es el objeto *por el cual* está. [(el objeto al que ese nombre ‘refiere’)].

Desde este punto de partida hay sólo un paso para que arribemos a ciertas conclusiones sobre la referencia de otros tipos de expresiones. Consideremos el siguiente par de oraciones:

(3) Barack Obama es demócrata.

(4) Barack Obama es republicano.

Una vez más, la primera de ellas (3) es verdadera, mientras que la segunda (4) es falsa. Ya sabíamos que la referencia de ‘Barack Obama’ es aquel hombre *por el cual está* esa expresión [‘denotado’ por ella, o al que ella ‘refiere’]. O sea que, dado que la referencia consiste en esa capacidad (de una expresión) de afectar el valor de verdad de las oraciones en que aparece, podemos decir que la referencia de los *predicados*, --tales como ‘es demócrata’ o ‘es republicano’— deberá ser algo que, combinado con un objeto, dará como resultado un determinado valor de verdad. De acuerdo a esto, es natural asumir que la referencia de este tipo de predicados será una función de *objetos* a [en] *valores de verdad*. La referencia de ‘es demócrata’ será aquella función que devuelva el valor ‘verdadero’ cuando su valor de entrada sea un objeto que es miembro del partido Demócrata (y el valor ‘falso’ en caso contrario); mientras que la referencia de ‘es republicano’ será una función que devuelva el valor ‘verdadero’ cuando tome como entrada un objeto que es miembro del partido Republicano (y el valor ‘falso’ en caso contrario). Así se explica el hecho de que (3) sea verdadera, y (4) falsa: Obama *es* un miembro del partido Demócrata, y *no es* un miembro del partido Republicano.

A medida que intentamos extender este tipo de teorías de la referencia para abarcar más y más tipos de expresiones en un lenguaje natural, el asunto se vuelve más complicado, y más controvertido.

(Para una introducción, ver Heim and Kratzer 1998.) De todos modos, lo que hemos expuesto hasta aquí es suficiente para dar una idea general de cómo podríamos proceder. Algunos predicados, como por ejemplo ‘ama a’, combinan *dos* nombres para formar una oración. Por lo que la referencia de este tipo de predicados (de ‘*dos lugares*’) deberá ser algo capaz de combinar un *par* de objetos de manera tal que esa combinación determine un valor de verdad – quizás una función que vaya de pares ordenados de objetos a [*en*] valores de verdad, y que devuelva el valor ‘verdadero’ cuando tome como entrada un *par* de objetos cuyo primer miembro ama al segundo, y ‘falso’ en caso contrario.

2.1.2 Teorías de la referencia vs. Teorías semánticas

Supongamos entonces que contamos con una teoría de la referencia para un lenguaje (en el sentido aclarado más arriba). ¿Tendríamos ya una teoría semántica satisfactoria de ese lenguaje?

Hay argumentos plausibles que indican lo contrario. Adoptando un ejemplo de Quine (1970), consideremos el siguiente par de oraciones

(5) Todas las criaturas con corazón son criaturas con corazón.

(6) Todas las criaturas con corazón son criaturas con riñones.

Supongamos que ambas son verdaderas. Si es así, entonces parece que todas las criaturas con corazón serán también criaturas con riñones. Por lo que, desde el punto de vista de una teoría de la referencia, (5) y (6) son equivalentes: difieren únicamente en la sustitución de ‘criaturas con corazón’ por ‘criaturas con riñones’, expresiones que tienen la misma referencia (que están por la misma función de objetos a [*en*] valores de verdad) .

Aun así, hay una clara diferencia intuitiva entre el significado de (5) y el de (6); parece que las oraciones, en cierto sentido, *dicen* cosas diferentes. (5) parece expresar la noción (trivial y aburrida) de que cada criatura con corazón es una criatura con corazón, mientras que (6) expresa la afirmación nada trivial y potencialmente informativa de que cada criatura que tiene corazón tiene, además, riñones. Lo que sugiere que hay una importante diferencia entre (5) y (6), diferencia que no podemos capturar con nuestra teoría de la referencia.

Podemos generar ejemplos de esta clase utilizando pares de expresiones de otros tipos, que también comparten una misma referencia, pero que intuitivamente también difieren en significado; por

ejemplo, 'Clark Kent' y 'Superman', o (un ejemplo famoso de Frege 1892/1960) 'la estrella de la mañana' y 'la estrella de la tarde' [i.e. 'el lucero del alba' y 'el lucero de la tarde'].

Puede parecer que se trata de un argumento más bien débil sobre lo incompleto de la teoría de la referencia, ya que parece descansar en intuiciones relativas a qué tan "informativas" son algunas oraciones como (5) y (6). Pero el argumento puede ser reforzado si incluimos oraciones del tipo de (5) y (6) dentro de oraciones más complejas, por ejemplo:

(7) Juan cree que todas las criaturas con corazón son criaturas con corazón.

(8) Juan cree que todas las criaturas con corazón son criaturas con riñones.

(7) y (8) difieren solo con respecto a las expresiones subrayadas y, como hemos notado anteriormente, esas expresiones tienen la misma referencia [son *correferenciales*, o comparten referencia]. Pero a pesar de ello, parece claro que (7) y (8) podrían tener distintos valores de verdad: alguien puede creer que todas las criaturas con corazón tienen corazón sin tener ninguna opinión sobre la cuestión de si esas criaturas poseen, además, riñones. Pero esto quiere decir que la referencia de las expresiones no está cumpliendo con la función explicativa que nos hizo ocuparnos de ellas en un primer momento: no explica cuál es la contribución de estas expresiones en la determinación del valor de verdad de las oraciones en que aparecen. (Por supuesto, aún podríamos mantener que la referencia de una expresión explica su contribución a la determinación del valor de verdad de una clase *limitada* de oraciones en las que aparece.) Pero para que podamos explicar cómo es que (7) y (8) difieren en valor de verdad en términos de las propiedades de las expresiones que las componen, debe haber alguna otra clase de valor en las expresiones, alguna clase de significado que vaya más allá de la referencia.

Las oraciones como (7) y (8) se denominan *atribuciones de creencias* [belief ascriptions], por la obvia razón de que atribuyen a un sujeto una determinada creencia. Las atribuciones de creencias son un tipo de *atribuciones de actitudes proposicionales* –entre las que se incluyen otros tipos de atribuciones, como las atribuciones de conocimiento, de deseo, o de juicio-. Como quedará claro en lo que sigue, las atribuciones de actitudes proposicionales han jugado un papel importante en los debates semánticos más recientes. Una de las razones por las que han sido tan importantes es la ejemplificada por (7) y (8). Dado que estas oraciones pueden diferir en su valor de verdad a pesar que sólo se diferencian por las expresiones subrayadas, y dado que ambas expresiones comparten una misma referencia y ocupan el mismo lugar en la estructura de ambas oraciones, se dice que (7)

y (8) contienen un *contexto no extensional*: esto es, a grandes rasgos, un “lugar” de la oración en el cual, si sustituimos el término que ocupa ese lugar por otro con la misma referencia, la oración cambia su valor de verdad. (Se denominan ‘contextos no extensionales’ porque ‘extensión’ es otra forma de decir ‘referencia’.)

Podemos construir un argumento similar acerca de la incompletud de la teoría de la referencia basándonos en la sustitución de oraciones completas. Una teoría de la referencia le asigna valores a las expresiones suboracionales³, valores que explican su contribución a la determinación del valor de verdad de la oración en que aparecen; pero a cada oración sólo se le asignan los valores ‘verdadero’ o ‘falso’. Consideremos el siguiente par de oraciones:

(9) María cree que Barack Obama es el presidente de los Estados Unidos.

(10) María cree que John Key es el primer ministro de Nueva Zelanda.

Como ambas de las oraciones subrayadas son verdaderas, (9) y (10) difieren únicamente con respecto a la sustitución de expresiones que comparten una misma referencia (las oraciones subrayadas). Sin embargo, claramente (9) y (10) podrían no coincidir en sus respectivos valores de verdad.

Esto parece sugerir que, además de su valor de verdad, una teoría semántica debería asignar algún otro valor a las oraciones. Otro camino por el que puede llegarse a la misma conclusión es considerando que afirmaciones como las que siguen pueden ser verdaderas:

Juan cree tres cosas sobre Indiana, y las tres son falsas.

Hay muchas verdades necesarias que no son a priori, y mi oración favorita es la que expresa una de ellas.

Para sacarte un diez debes que creer todo lo que yo diga.

Este tipo de oraciones parece mostrar que los portadores de verdad, así como de las *propiedades modales* (necesidad/posibilidad) o de las *propiedades epistémicas* (por ej. ‘cognoscible a priori’/‘cognoscible a posteriori’) son ciertos *objetos de estados mentales* (como por ejemplo las

³ En el original, ‘*subsential expressions*’ (es decir, las expresiones que aparecen “dentro” de las oraciones, o de las que se “compone” la oración).

creencias)⁴. Pero ¿qué son estas cosas? La teoría de la referencia no nos da ninguna respuesta sobre ello.

Los simpatizantes del enfoque proposicional intentarán ofrecer una teoría sobre esta clase de entidades; y de esta manera tratarán de ofrecer respuestas a los dos problemas de la teoría de la referencia mencionados más arriba: (i) la falta de una explicación para el hecho de que la verdad de (5) resulte trivial y a priori, mientras que no ocurre lo mismo con (6), y (ii) el hecho (ejemplificado por (7)/(8) y (9)/(10)) de que algunas oraciones que sólo difieren por la sustitución de expresiones correferenciales puedan tener diferentes valores de verdad.

Una teoría proposicional, por lo tanto, no abandonará por completo la teoría de la referencia, sino que argumenta que una teoría semántica debe prestar atención a algo más que sólo la teoría de la referencia. Las expresiones suboracionales, además de una *referencia*, poseen un *contenido*. Los contenidos de las oraciones (lo que las oraciones *expresan*) son lo que llamamos *proposiciones*.

2.1.3 Relación entre contenido y referencia

La pregunta que surge naturalmente es: ¿Qué clase de cosas son los *contenidos*? A continuación trataremos algunas de las respuestas más relevantes. Pero, antes de intentar cualquier teoría sobre ellos, podemos mencionar algunas características generales sobre el rol que los contenidos deberán jugar en este tipo de teorías.

Primero, ¿cuál es la relación entre el contenido y la referencia? Examinemos la cuestión teniendo en cuenta las oraciones; en este caso se trata de la pregunta sobre la relación entre la *proposición* expresada por una oración y el *valor de verdad* de esa oración. El ejemplo de (9) y (10) muestra que dos oraciones pueden expresar diferentes proposiciones pero coincidir en su valor de verdad.

Después de todo, cada una de estas oraciones atribuye creencias diferentes a María; y si el objeto de las creencias son las proposiciones, a cada una de las expresiones subrayadas le corresponderá a una proposición diferente. Sin embargo, ambas oraciones son verdaderas.

¿Es posible la relación inversa? ¿Pueden dos oraciones expresar la misma proposición, pero diferir en valor de verdad? No parece posible, como lo muestra una vez más el papel que juegan las proposiciones en tanto objeto de creencias. Supongamos que ustedes y yo creemos exactamente lo mismo –creemos, por ejemplo, que el mundo es de una determinada manera--. ¿Puede mi creencia

⁴ [nota mía] Es decir, parece que lo verdadero o falso (necesario o posible, etc.) es, por ejemplo, *lo que* María cree (en el caso de (9) y (10)).

ser verdadera, y la de ustedes falsa? Pareciera que no; es incoherente decir que tanto ustedes como yo creemos que el mundo es de una determinada manera (de *la misma* manera, tanto “para mí” como “para ustedes”) y que sin embargo de algún modo mi creencia es verdadera mientras que la suya es falsa. Lo que sugiere que, siempre que dos oraciones expresen la misma proposición, deberán tener el mismo valor de verdad.

En general, parece plausible sostener que dos oraciones con el mismo contenido —es decir, que *expresan* la misma proposición— deben tener siempre la misma referencia; mientras que dos expresiones con la misma referencia pueden diferir en contenido. Esto es lo que expresa el *slogan* Fregeano según el cual “el sentido determina la referencia” (‘sentido’ es la traducción convencional de lo que Frege llamó ‘*Sinn*’, y que aquí llamamos ‘contenido’)⁵.

Pero, si esto es así para las oraciones, ¿se da también para el caso de las expresiones suboracionales? Todo indica que debería ser así. Supongamos, por absurdo, que dos expresiones suboracionales, *e* y *e**, expresan el mismo contenido pero tienen distinta referencia. Es plausible que dos oraciones que sólo difieren en la sustitución de expresiones [suboracionales] con el mismo contenido tengan a su vez el mismo contenido (principio que, sin embargo, no es de ninguna manera incontrovertido; ver la entrada ‘composicionalidad’). Pero, si esto es así, entonces las oraciones que sólo difieran en la sustitución de *e* por *e** tendrán el mismo contenido. Y si estas oraciones pueden tener diferentes valores de verdad, entonces, para cada par de expresiones que difieran en su referencia habrá un par de oraciones que difieren sólo por la sustitución de esas expresiones pero tienen diferentes valores de verdad. De modo que, si pudiera haber un par de expresiones como *e* y *e** (que difieren en su referencia pero no en su contenido), podría haber un par de oraciones con el mismo contenido (esto es, que expresen la misma proposición) pero que tengan diferentes valores de verdad. Esto es imposible [ver párrafo anterior], por lo que no puede haber tal par de expresiones, y el contenido determinará la referencia tanto para las oraciones completas como para las expresiones que forman parte de ellas.

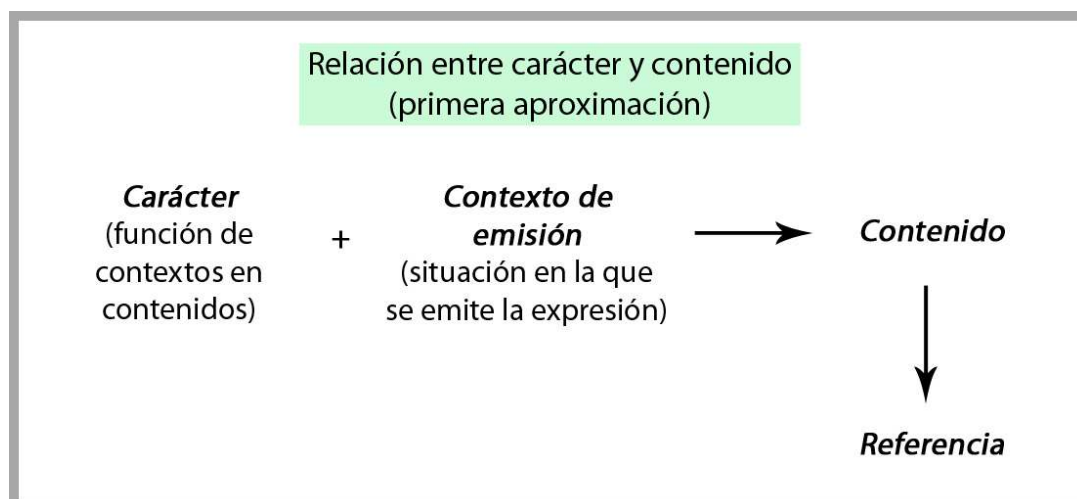
Este resultado —que el contenido determina la referencia— da cuenta de algo que parece razonable exigirle a cualquier teoría semántica: le asigna a cada expresión un valor —un contenido— que determina la referencia de esa expresión.

2.1.4 Carácter y contenido, contexto y circunstancia

⁵ La aclaración entre paréntesis pertenece al original, nosotros, a su vez, traducimos ‘*sense*’ por ‘sentido’ y ‘*content*’ por ‘contenido’.

A pesar de todo, la idea de que podemos asignar un contenido (tal como lo caracterizamos más arriba) a todas las expresiones del lenguaje natural, presenta una evidente dificultad. Muchas expresiones, como por ejemplo 'yo', 'aquí', 'ahora', etc. poseen una referencia diferente al ser emitidas por diferentes hablantes en diferentes situaciones. Claramente, no podemos asignarle a la expresión 'yo' un *único* contenido que determine su referencia, ya que esta expresión tiene diferentes referencias en diferentes situaciones. Estas 'situaciones' son llamadas típicamente *contextos de emisión*, o simplemente *contextos*, y las expresiones cuya referencia depende del contexto son llamadas *indexicales*, o *expresiones dependientes del contexto*.

La existencia de este tipo de expresiones hace evidente que una teoría semántica deberá hacer más que sólo asignar un determinado contenido a cada tipo de expresiones del lenguaje. Expresiones como 'yo' deberán asociarse con alguna regla que determine su contenido en un determinado contexto de emisión. Estas reglas, que consisten en (o determinan) funciones de contextos a [en] contenidos, son denominadas *caracteres* [se hablará así, del *carácter* de una expresión, como antes se habló de su referencia o de su contenido]. (La terminología empleada aquí, así como la concepción de la relación entre contexto, contenido y referencia se debe a Kaplan, 1989). El carácter de 'yo' deberá ser, pues, una función de contextos a [en] contenidos según la cual, en un contexto en que yo sea el hablante, determine un contenido que a su vez me determinará (a mí, i.e. el hablante en ese contexto) como la referencia de esa expresión; en un contexto en que el hablante sea, por ejemplo, Barack Obama, la función determinará un contenido cuya referencia es Barack Obama, y así sucesivamente.



Aquí encontramos otra ambigüedad en la expresión 'significado' que puede dar lugar a confusiones. ¿Qué constituye el verdadero *significado* de una expresión? ¿Su *carácter*, o su *contenido* (en el contexto relevante)? Pero esto no es más que una vana disputa terminológica. Las expresiones

tienen caracteres que, en cierto contexto, determinan un contenido. Podemos hablar tanto de carácter como de contenido, y ambos son importantes. Nada ganamos discutiendo si uno u otro “merecen” el título de *significado*. Lo importante es tener en claro la distinción, y ser conscientes de la importancia de considerar que las expresiones poseen *tanto* un carácter *como* un contenido (relativo a un contexto).

¿Cuántos tipos de expresiones indexicales hay? Los candidatos obvios son ‘yo’, ‘aquí’, ‘ahora’... Pero, fuera de estos candidatos, hay lugar para amplios desacuerdos (para una buena reseña, ver Cappelen y Le Pore, 2005). Ciertos argumentos parecen mostrar que casi todas las expresiones son indexicales. Consideremos una expresión que en apariencia no es “sensible al contexto”, como por ejemplo: ‘la segunda ciudad más grande de los Estados Unidos’. No parece tratarse de una expresión sensible al contexto, porque parece que siempre refiere a la misma ciudad (Los Ángeles) ya sea emitida por mí, por ustedes, o por cualquier otro hablante. Pero consideremos una oración como:

(11) Hace cien años, la segunda ciudad más grande de los Estados Unidos era Chicago.

Esta oración es verdadera. Pero para que sea verdadera, la expresión ‘la segunda ciudad más grande de los Estados Unidos’ tendría que referir –en (11)— a Chicago. Si esto es así, resulta que esta expresión se comporta, después de todo, como un indexical – es decir, su referencia depende en este sentido del contexto de emisión. Así, según este argumento, en la oración (11) la frase ‘hace cien años’ modifica o ‘desplaza’ [*shift*] el contexto. En (11) ‘la segunda ciudad más grande de los Estados Unidos’ refiere a la ciudad que cumplía con esa condición *hace cien años*. [es decir, su referencia es la referencia que esa expresión *habría tenido* hace cien años]

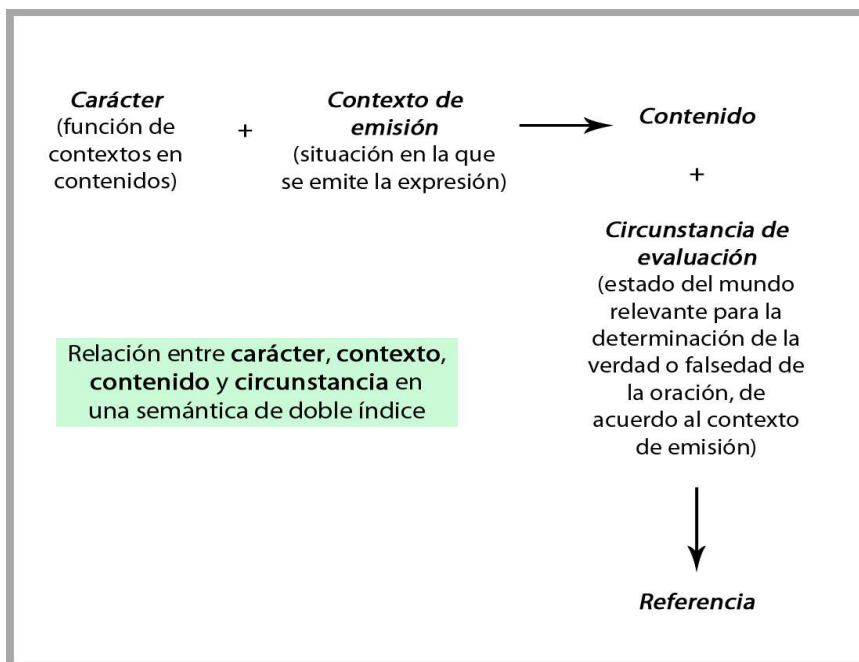
Sin embargo, parece que aquí hay algo que no puede estar del todo bien, como lo demuestra el siguiente ejemplo:

(12) En cien años, yo no existiré.

Supongamos que esta oración, emitida por mí, es verdadera. Entonces, si lo que dijimos sobre (11) es correcto, la expresión ‘yo’ en (12) referirá a lo que referiría si fuera emitida por algún hablante dentro de cien años. Es decir que, si asumimos que (12) es verdadera, entonces la expresión (‘yo’) ya no refiere a mí –después de todo, en cien años yo ya no voy a ser capaz de emitir ninguna

expresión. Sin embargo, parece que ‘yo’ en (12) sí refiere a mí: después de todo, es *sobre* mí (el actual hablante o emisor). ¿Qué está sucediendo aquí?

Lo que muestran ejemplos como (12) es que la referencia de una expresión debe ser relativizada no sólo con respecto a un contexto de emisión, sino también en relación a una *circunstancia de evaluación* –esto es, algún estado posible del mundo que sea relevante para la determinación de la verdad o falsedad de esa oración. En el caso de muchas oraciones sencillas, el contexto (de emisión) coincide con la circunstancia (de evaluación), dado que ambos consisten en un determinado estado del mundo al momento de la emisión, más un lugar y un emisor. Pero en oraciones como (12) el contexto y la circunstancia ya no coinciden. Frases como ‘dentro de cien años’ modifican (o ‘desplazan’ [*shift*]) la *circunstancia de evaluación* relevante para la verdad o falsedad de la oración – es decir, cambian el estado del mundo relevante a la evaluación de su verdad o falsedad— pero *no* modifican su *contexto de emisión*. Es por eso que cuando emito (12), la expresión ‘yo’ refiere a mí, a pesar de que en cien años yo [(i.e. la referencia de esa expresión en el contexto de emisión)] no existiré.



Esto es lo que suele llamarse **semántica de doble índice** –los dos índices corresponden al contexto de emisión y las circunstancias de evaluación, respectivamente.

La explicación clásica de una semántica de doble índice se debe a Kaplan (1989); otra discusión importante la encontramos en Kamp (1971). Para una interpretación diferente de este enfoque, ver Lewis (1980).

La semántica de doble índice explica cómo es posible aceptar que la referencia de ‘la segunda ciudad más grande de los Estados Unidos’ en (11) es Chicago, sin que sea necesario considerar que ‘la segunda ciudad más grande de Estados Unidos’ es una expresión indexical (como ‘yo’, por ejemplo). Según este enfoque, la expresión ‘la segunda ciudad más grande de los Estados Unidos’ no varía de contexto dependiendo del contexto de emisión, sino que su contenido es tal que determina una referencia diferente en diferentes circunstancias de evaluación. En particular, su referencia será la ciudad de Los Ángeles de acuerdo al actual estado del mundo, mientras que su referencia de acuerdo al estado del mundo hace cien años (en 1910) será la ciudad de Chicago [2]. Y como ‘la segunda ciudad más grande de los Estados Unidos’ refiere a cosas diferentes con respecto a diferentes circunstancias, esta expresión *no es un designador rígido* – modo en que se denomina una expresión que refiere al *mismo* objeto de acuerdo a *cualquier* circunstancia de evaluación en la que exista ese objeto, y que nunca refiere a ninguna otra cosa de acuerdo a ninguna otra circunstancia de evaluación (el término ‘designador rígido’ se debe a Kripke, 1972)

2.1.5 Semántica de mundos posibles

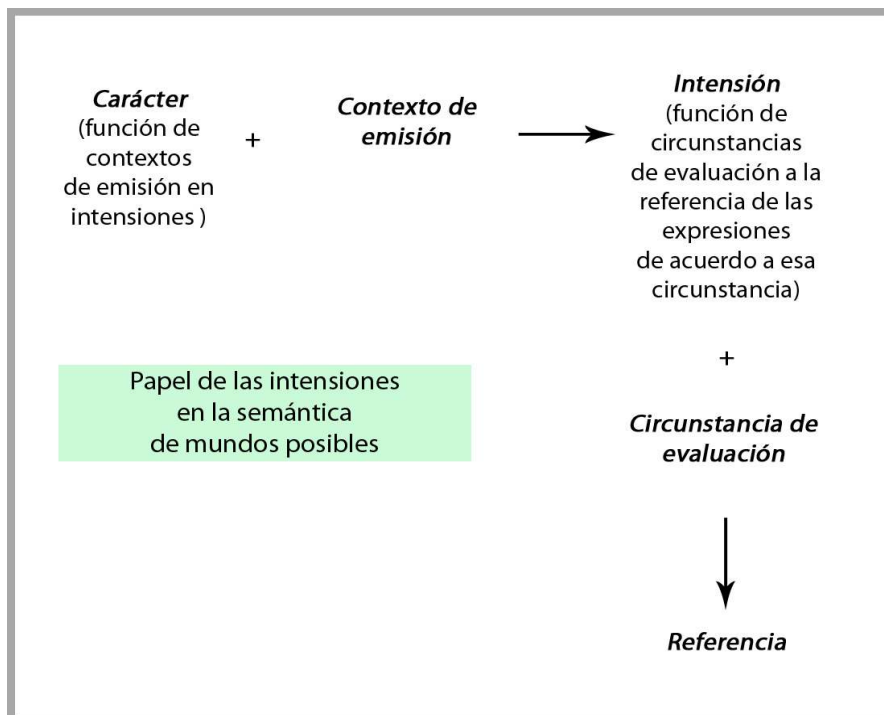
Ahora sabemos que las expresiones se asocian con caracteres, que son funciones de contextos a [en] contenidos; y sabemos que los contenidos son aquellas cosas que, de acuerdo a cada circunstancia de evaluación, determinan una referencia. Ya podemos plantear la pregunta central de la teoría semántica (proposicional): ¿Qué cosas con los contenidos? Lo anterior sugiere una respuesta convenientemente minimalista: quizás, si los contenidos son cosas que, en conjunto con circunstancias de evaluación, determinan una referencia, los contenidos no serán al fin y al cabo otra cosa que funciones que van de circunstancias de evaluación a [en] una referencia.

Este enfoque parece abstracto pero en cierto sentido es bastante intuitivo. La idea es que el significado de una expresión no es aquello *por lo cual está* [*stands for*] esa expresión, en la circunstancia relevante), sino una regla que nos dice por qué cosa *estaría* esa expresión⁶ si el mundo fuera de tal o cual manera. De modo que, según esta concepción, el contenido de la expresión ‘el hombre más alto del mundo’ no será simplemente aquel hombre que resulta ser el más alto según el actual estado del mundo, sino una función que vaya de ciertas maneras en las que el mundo puede

⁶ O ‘a qué cosa referiría esa expresión’.

ser, a [en] hombres –es decir, una función que para cada manera en la que pueda ser el mundo, asigne como referente de esa expresión a aquel hombre que resulte ser el más alto en ese “mundo” (esto, en el caso de que haya tal hombre en el mundo relevante, en caso contrario la expresión carecerá de referente). Lo cual se corresponde con la idea intuitiva de que para entender una expresión no necesitamos saber a qué refiere en realidad esa expresión –después de todo, podemos entender el ‘significado’ de la expresión ‘el hombre más alto’ aunque no conozcamos precisamente *quién* es en realidad ese hombre; sólo tenemos que saber a qué referiría esa expresión, dada cierta información sobre el mundo (en este caso, la altura de los hombres en ese mundo).

Estas funciones, o reglas, son llamadas *intensiones* (siguiendo a Carnap, 1947). La semántica de mundos posibles es la concepción según la cual los contenidos son intensiones (y por tanto que los caracteres son funciones de contextos a [en] intensiones -- es decir, funciones *de* contextos a [en] funciones de circunstancias de evaluación a [en] la referencia.) ‘Intensión’ suele utilizarse de manera más general como un sinónimo de ‘contenido’. Este uso se presta a confusiones, por lo que conviene reservarlo para funciones de contextos a [en] referentes. Pero que los contenidos sean intensiones, como proponen los partidarios de mundos posibles, es materia de controversia.



Para una discusión de la aplicación de este marco de la semántica de mundos posibles al análisis del lenguaje natural, ver Lewis, 1970. La intensión de una oración —esto es, la proposición que esa oración expresa, según este enfoque— será una función de mundos a [en] valores de verdad. En particular, será aquella función que ‘devuelva’ (o determine) el valor ‘verdadero’ para cada mundo respecto al cual esa oración sea verdadera, y ‘falso’ en caso contrario. La intensión de un predicado simple como ‘es rojo’, será una función de mundos a objetos que le asigne, en cada mundo, el valor ‘verdadero’ si el objeto en cuestión es rojo, y ‘falso’ en caso contrario. De hecho, la semántica de mundos posibles considera que el significado de las expresiones consiste en funciones que van de mundos a aquellos valores que una teoría de la referencia asignaría a esas expresiones en el mundo relevante: en ese sentido, las intensiones serían como una ‘capa extra’ que se superpone a la teoría de la referencia.

Esta ‘capa extra’ promete solucionar los problemas que acarrear los contextos no extensionales⁷, tales como los ilustrados por los ejemplos de ‘criatura con corazón’ y ‘criatura con riñones’ en (7) y (8). En ese caso la preocupación era que, si el significado consiste únicamente en la referencia, entonces cualquier par de oraciones que sólo difieran en la sustitución de este tipo de expresiones deberían tener el mismo valor de verdad. Pero vimos que (7) y (8) no necesariamente tienen el mismo valor de verdad. Los partidarios de la semántica de mundos posibles pueden resolver este problema identificando el *significado* de estas expresiones con su *intensión* (no con su referencia), y haciendo notar que, aunque ‘criatura con corazón’ y ‘criatura con riñones’ comparten una misma *referencia*, cada una tiene una *intensión* diferente. Después de todo, incluso si en nuestro mundo cada criatura que posee corazón posee también riñones (y viceversa), parece posible que exista algún otro mundo en el cual no toda criatura con corazón tenga, además, riñones. Y dado que, de acuerdo a esa circunstancia de evaluación los dos términos no tienen la misma referencia, sus intensiones (funciones de circunstancias de evaluación a referentes) serán también diferentes. De esta manera, la semántica de mundos posibles hace lugar al hecho de que (7) y (8) tengan diferentes valores de verdad (como hemos visto que efectivamente sucede).

Sin embargo, el principal problema al que se enfrenta la semántica de mundos posibles tiene que ver, precisamente, con las oraciones de la forma de (7) y (8). Es decir, oraciones que atribuyen *actitudes proposicionales*, como creencias, a los sujetos. Para considerar este problema, podemos comenzar por preguntarnos ¿de acuerdo a la semántica de mundos posibles, qué características deben tener dos oraciones para que podamos decir que tienen el mismo contenido (es decir, que

⁷ Ver página 10.

expresan la misma proposición)? Como los contenidos son intensiones, y las intensiones son funciones de circunstancias de evaluación a referentes, parece que de acuerdo a la semántica de mundos posibles, dos oraciones tienen el mismo contenido siempre que compartan el mismo valor de verdad de acuerdo a cualquier circunstancia de evaluación. Dicho de otra forma, dos oraciones expresan la misma proposición si y sólo si no es posible que difieran en valor de verdad.

El problema es que hay oraciones que tienen el mismo valor de verdad en toda circunstancia de evaluación, pero aún así parecen tener significados diferentes. Consideremos, por ejemplo:

(13) $2+2=4$.

(14) Hay infinitos números primos.

(13) y (14) nos recuerdan en cierto modo a (5) y (6): (13) parece consistir en una trivialidad cuya verdad es conocida por cualquiera, mientras que (14) parece consistir en una afirmación más sustancial, cuya verdad bien podríamos no conocer. Sin embargo, ambas son verdades necesarias: como cualquier verdad de la matemática, ninguna de ellas depende de ninguna característica del mundo, sino que ambas son verdaderas de acuerdo a *cualquier* circunstancia de evaluación. Por lo que (13) y (14) tienen la misma intensión y, de acuerdo a la semántica de mundos posibles, deberían tener [(o 'expresar')] el mismo contenido.

Esto resulta profundamente contrario a la intuición. El problema (como en el caso de (5) y (6)) puede ser resaltado incrustando estas oraciones dentro de oraciones de atribución de actitudes proposicionales:

(15) Juan cree que $2+2=4$.

(16) Juan cree que hay infinitos números primos.

Como hemos visto, los partidarios de la semántica de mundos posibles están obligados a considerar que las oraciones subrayadas (13) y (14) tienen el mismo contenido. Es decir que el partidario de la semántica de mundos posibles deberá sostener que (15) y (16) son un par de oraciones que sólo difieren por la sustitución de expresiones con el mismo contenido. Por lo que a su vez estará obligado a afirmar que este par de oraciones expresa la misma proposición, y que ambas tienen el mismo valor de verdad; sin embargo, (15) y (16) —como (7) y (8)— evidentemente pueden tener diferente valor de verdad, y por lo tanto *no* expresan la misma proposición.

Efectivamente, el problema es incluso más grave (como se muestra en Soames, 1988).

Consideremos un par de oraciones como las siguientes:

(17) La hierba es verde.

(18) La hierba es verde y hay infinitos números primos.

La segunda –(18)— consiste en la conjunción de (17) y de una verdad necesaria; por lo que (18) es verdadera si y sólo si (17) es verdadera. Pero si esto es así, tienen la misma intensión y, de acuerdo a la semántica de mundos posibles, el mismo contenido, por lo que las dos oraciones siguientes no pueden diferir en su valor de verdad:

(19) Juan cree que la hierba es verde.

(20) Juan cree que la hierba es verde y hay infinitos números primos.

Y este último par de oraciones no puede tener diferentes valores de verdad, porque (19) y (20) difieren sólo por la sustitución de (17) y (18), y éstas, a su vez (de acuerdo a la semántica de mundos posibles) tienen el mismo contenido. Peor aún, parece que la creencia es distributiva respecto a la conjunción, en el sentido de que quien cree la conjunción de un par de proposiciones, cree también cada una de esas proposiciones por separado. Pero entonces, si (20) es verdadera, también lo será (16), y cualquiera que crea que la hierba es verde creará también que hay infinitos números primos. Esta línea de argumentación se puede generalizar para mostrar que cualquier persona que crea cualquier proposición debe al mismo tiempo creer todas y cada una de las verdades necesarias. Este resultado es, como poco, una consecuencia bastante poco intuitiva de la estrategia de los semánticos de mundos posibles, estrategia que se basa precisamente en identificar los *contenidos* con las *intensiones*. Parece que deberíamos tratar de encontrar un enfoque del que no se derive esta incómoda consecuencia.

Para un intento de responder a estos argumentos desde el marco de la semántica de mundos posibles, ver Stalnaker, 1984; para una discusión de un enfoque relacionado, que intenta evitar estos problemas, ver [la entrada] situaciones en la semántica de lenguajes naturales.

2.1.6 Proposiciones russellianas

Lo que necesitamos es una concepción semántica capaz de explicar cómo es posible que oraciones como (13) y (14), y por lo tanto también (15) y (16), expresen diferentes proposiciones. Es decir, necesitamos una concepción de las proposiciones que haga lugar a la posibilidad de que un par de oraciones puedan ser verdaderas de acuerdo a cualquier circunstancia, pero aún así puedan tener contenidos diferentes.

Una suposición natural es que (13) y (14) tienen diferentes contenidos simplemente porque son *sobre [acerca de]* diferentes cosas. Por ejemplo, (14) afirma algo sobre el conjunto de los números primos, mientras que (13) es sobre la relación entre los números 2 y 4. Sería deseable que una teoría semántica fuera sensible a este tipo de diferencias: esto es, que nos permitiera considerar que dos oraciones que poseen diferentes *temas* (que son, en este sentido, *sobre* cosas diferentes) expresan *contenidos* diferentes. Una forma de asegurarnos este resultado es pensar en los contenidos de las expresiones suboracionales como componentes de la proposición expresada por la oración en su conjunto. La diferencia en los contenidos de las expresiones suboracionales resultaría entonces suficientes para explicar las diferencias en el contenido de la oración completa. Así, por ejemplo, si (14) contiene una expresión que refiere a los números primos, mientras que (13) no contiene tal expresión, ambas oraciones expresarán proposiciones diferentes.

Los defensores de este tipo de enfoque consideran que las proposiciones son *estructuradas*: poseen *constituyentes* que incluyen los significados de las expresiones que componen la totalidad de la oración mediante la cual se expresa la proposición relevante (para una mayor discusión, ver [la entrada] proposiciones estructuradas). Una pregunta importante que podemos realizar a este tipo de enfoques es: ¿Qué se quiere decir cuando se dice que un objeto abstracto, como es una proposición, es ‘estructurada’ o ‘posee constituyentes’? Pero esta pregunta nos llevaría demasiado lejos hacia el campo de la metafísica. La principal cuestión *semántica* que deberán solucionar los partidarios de las proposiciones estructuradas será la siguiente: ¿Qué clase de cosas son los constituyentes de una proposición? [o ¿En qué consisten los constituyentes de una proposición?]

La respuesta a esta dificultad está disponible para quienes proponen una concepción que se denomina de las *proposiciones russellianas*. (Este enfoque se denomina ‘russellianismo’ debido a su similitud con la explicación del contenido defendida en el capítulo IV, de Russell, 1903). Visto de esta forma, el russellianismo sólo consiste en una concepción general acerca de qué son los constituyentes de las proposiciones, y por tanto quien sostiene esta posición no se compromete con ningún tipo de concepción particular sobre los contenidos de algún tipo de expresiones. Sin embargo, la mayoría de los russellianos aceptan también cierto enfoque respecto el contenido de los

nombres propios, conocida como *millianismo*. Esta posición consiste en afirmar que el significado de un nombre propio no es otra cosa que el objeto por el cual está ese nombre (si es que hay tal objeto) [el objeto al que ese nombre ‘refiere’, o que es ‘denotado’ por ese nombre].

Hay mucho que decir a favor del russellianismo. No sólo resuelve el problema que notamos anteriormente en la semántica de mundos posibles, sino que además concuerda con la idea intuitiva de que la función de los nombres es identificar o individualizar⁸ objetos, mientras que la función de los predicados consiste (obviamente) en predicar propiedades de esos objetos.

A pesar de ello, las teorías semánticas de la variedad Millian-Russell también tienen sus dificultades. Algunas de ellas son de naturaleza metafísica, y se basan en la premisa de que las proposiciones que poseen objetos entre sus constituyentes no pueden existir en circunstancias en las que esos objetos no existan (para una discusión de este tema, ver [la entrada] [proposiciones singulares, secciones 4 y 5](#)). Dentro de las objeciones propiamente semánticas que se le pueden realizar a este tipo de concepciones, hay dos que son especialmente importantes.

La primera de ellas involucra la existencia de *nombre vacíos*: nombres que carecen de referente. Es un lugar común aceptar que de hecho utilizamos nombres de ese tipo. Un ejemplo que lo demuestra es ‘Vulcano’, nombre que se propuso para el planeta que se encuentra entre Mercurio y el Sol, causante de las perturbaciones en la órbita de Mercurio. Dado que una postura del estilo Mill-Russell declara que el contenido de un nombre es su referente, quien sostenga esta postura estará comprometido con la afirmación de que los nombres ‘vacíos’ carecen de contenido. Pero esto resulta sorprendente: de hecho, parece que podemos utilizar nombres ‘vacíos’ en ciertas oraciones para expresar proposiciones y formar creencias acerca del mundo. La posición Mill-Russell nos debe alguna explicación acerca de cómo es esto posible (dado que tales nombres carecen, según ellos, de contenido). (En Braun, 1993, puede encontrarse una excelente discusión de este problema desde un punto de vista del millianismo.)

Pero quizás el problema más grave al que debe enfrentarse esta posición es llamado ‘rompecabezas [acertijo] de Frege’. Consideremos las siguientes oraciones:

(21) Clark Kent es Clark Kent.

(22) Clark Kent es Superman.

⁸ “single out”

De acuerdo al defensor de la postura de Mill- Russell, las oraciones (21) y (22) sólo se diferencian por la sustitución de expresiones con el mismo contenido: después de todo, ‘Clark Kent’ y ‘Superman’ son nombres propios que refieren al mismo objeto, y de acuerdo a esta postura el contenido de un nombre propio no es otra cosa que el objeto referido por el nombre. Pero este resultado es sorprendente. Ambas oraciones parecen tener un significado diferente, porque (21) parece expresar una afirmación trivial y obvia, mientras que (22) expresa una afirmación informativa y nada trivial.

Una vez más, este tipo de objeción puede reforzarse si incrustamos estas oraciones, intuitivamente diferentes, en oraciones de atribución de actitudes proposicionales:

(23) Lois cree que Clark Kent es Clark Kent.

(24) Lois cree que Clark Kent es Superman.

El problema que el russelliano deberá enfrentar aquí es análogo al problema que (15) y (16) planteaban para la semántica de mundos posibles. En este caso, como en aquél, tenemos un par de atribuciones de creencia que pueden tomar valores de verdad diferentes a pesar de que ambas oraciones sólo se diferencian por la sustitución de expresiones (que la teoría semántica del caso considera) sinónimas.

Los russellianos han propuesto una variedad de soluciones para el ‘rompecabezas [acertijo] de Frege’. Muchos creen que la intuición de acuerdo a la cual consideramos que (23) y (24) pueden diferir en valor de verdad se basa en realidad en una equivocación. Este error puede ser explicado, al menos parcialmente, en términos de una confusión entre la proposición que la oración expresa (semánticamente) en un contexto, y la proposición que los hablantes típicamente intentarían expresar (pragmáticamente) mediante la emisión de esas oraciones (Salmon 1986; Soames 2002). Otra manera de dar cuenta de esta confusión es afirmando que una misma proposición puede ser objeto de creencia [i.e. puede ser creída] bajo diferentes ‘presentaciones proposicionales’ [*‘propositional guises’*] (Salmon, 1986). También se lo puede intentar explicar haciendo referencia a la incapacidad de integrar porciones de información que han sido almacenadas bajo representaciones mentales diferentes (Braun y Saul, 2002) [3]. Una última alternativa para el russelliano sería aceptar el hecho de que (23) y (24) no coincidan en su valor de verdad, renunciando con ello a la idea de que las oraciones que sólo difieren por la sustitución de nombres propios con el mismo contenido expresen la misma proposición (Fine, 2007).

2.1.7 Proposiciones fregeanas

Sin embargo, estas no son las únicas respuestas posibles al rompecabezas fregeano. Así como el russelliano respondía al problema de (15) y (16) sosteniendo que dos oraciones con la misma intensión puedan tener diferentes significados, uno podría responder al problema de (23) y (24) argumentando que dos nombres pueden referir al mismo objeto pero tener significados diferentes, lo que permitiría dar cuenta de la diferencia entre los valores de verdad de (23) y (24). En esto consistiría una respuesta *fregeana* al rompecabezas fregeano, respuesta que a su vez implica abandonar el enfoque russelliano (o al menos la semántica al estilo Mill-Russell).

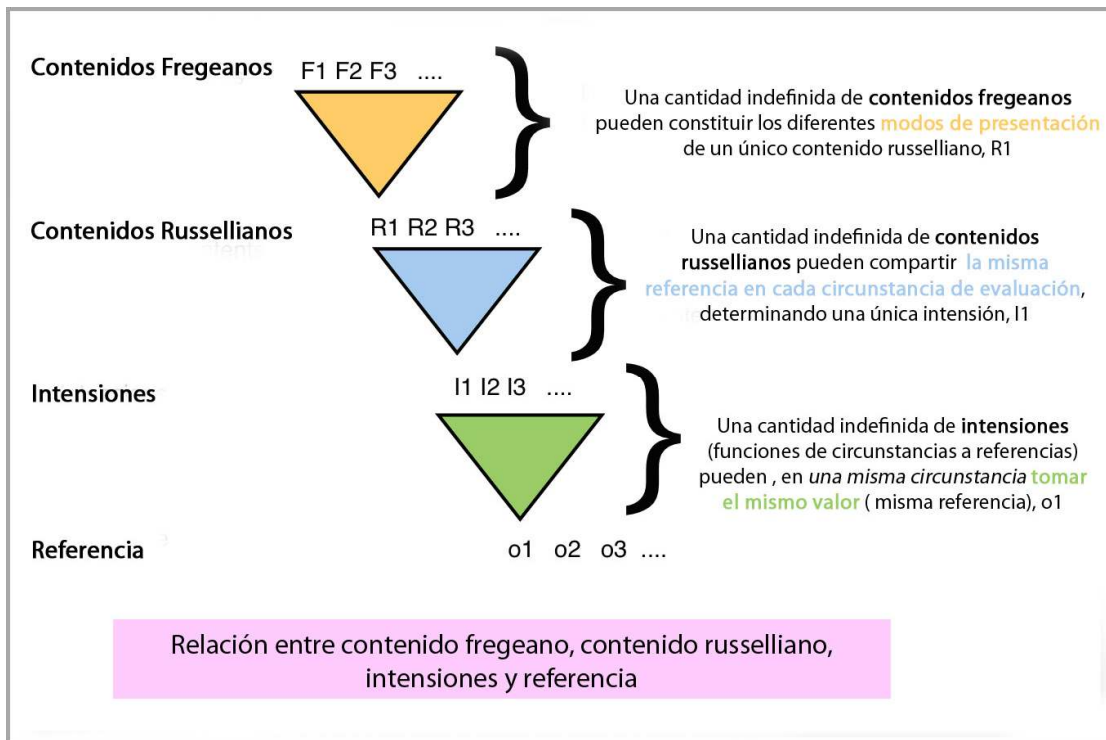
Los fregeanos, al igual que los russellianos, consideran que la proposición que una oración expresa es una cierta entidad estructurada por constituyentes (los contenidos de las expresiones que conforman la oración). Pero, a diferencia de los russellianos, los fregeanos no piensan que los constituyentes de las proposiciones sean los objetos, propiedades y relaciones que esas expresiones representan [*stand for*]; en lugar de eso, los fregeanos consideran que los contenidos son ‘modos de presentación’ o ‘maneras de pensar’ acerca de objetos, propiedades y relaciones. El término estándar con el que se denominan estos modos de presentación es *sentido*. (Como en el caso de ‘intensión’, ‘sentido’ se utiliza a veces como sinónimo de ‘contenido’. Pero, al igual que con ‘intensión’, evitaremos confusiones si restringimos el uso de ‘sentido’ para: ‘contenido, de acuerdo a una semántica fregeana’. Es discutible que exista tal cosa como estos sentidos, también es motivo de debate que estos constituyan los ‘contenidos’ de las expresiones.) El mismo Frege ilustró su posición con respecto a los sentidos mediante una analogía:

“La referencia de un nombre propio es el objeto mismo que designamos por medio de él; la representación que tenemos en este caso es completamente subjetiva; entre ambos está el sentido, que ciertamente ya no es subjetivo como la representación, pero que tampoco es el objeto mismo. El siguiente símil es quizás apropiado para esclarecer estas relaciones. Alguien observa la Luna a través de un telescopio. Comparo la Luna con la referencia; es el objeto de observación, que viene dado por la imagen real que se proyecta en la lente del objetivo del interior del telescopio y por la imagen que se produce en la retina del observador. A la primera imagen [la del interior del telescopio] la comparo con el sentido; a la segunda [la de la retina], con la representación o intuición. La imagen del telescopio es, ciertamente, unilateral, depende del lugar de observación; pero es, con todo, objetiva en la medida en que puede servir a muchos observadores. En cualquier caso, podría disponerse de

tal manera que muchos la usaran al mismo tiempo. Pero, por lo que respecta a las imágenes de la retina, cada uno tendrá la suya propia.” (Frege 1892/1960)⁹

Los sentidos son, entonces, objetivos, en tanto que más de una persona puede expresar pensamientos con un mismo sentido, que se corresponde en relación de muchos a uno con algún objeto. Así como las proposiciones russellianas se corresponden de muchos a una con respecto a las intensiones, las proposiciones fregeanas corresponden de muchas a una con respecto a las proposiciones russellianas. Esto es lo que a veces se expresa diciendo que las proposiciones fregeanas son más ‘exhaustivas’ [*fine-grained*, es decir, ‘hilan más fino’] que los contenidos russellianos (o que las intensiones).

De hecho, podemos pensar que las tres teorías semánticas proposicionales, y la teoría de la referencia, de las que hemos tratado aquí, se relacionan de la manera ilustrada en el gráfico siguiente.



⁹ Transcribimos la traducción de Valdés Villanueva. El artículo citado es *Sobre Sentido y Referencia* (compilado en Valdés Villanueva, Luis M., “*La búsqueda del significado. Lecturas de filosofía del lenguaje*”. Ed. Tecnos, Madrid, 2000.)

El principal argumento a favor de la semántica fregeana (que motivó los esfuerzos del mismo Frege) es que ofrece una solución elegante al ‘rompecabezas[acertijo] de Frege’. Nos dice que en casos como los de (23) y (24), en los que parece haber una diferencia de contenido entre una oración y otra, esa diferencia de hecho existe. Los nombres comparten una misma *referencia*, pero poseen diferentes *sentidos*, debido a que se trata de diferentes ‘modos de presentación’ de esa misma referencia.

La dificultad principal que deberá resolver el fregeanismo es la de proporcionar una explicación no metafórica acerca de la naturaleza del sentido. Se trata de un problema más preocupante para el semántico fregeano que para el defensor de una semántica de mundos posibles o de una semántica russelliana. Esto se debe a que, a diferencia de ellos, el fregeano da cuenta del significado de las expresiones mediante la introducción de una nueva clase de entidades [(los ‘sentidos’)], mientras que los otros dos explicaban el fenómeno en base a algún tipo de entidades previamente aceptadas – funciones, objetos, propiedades, relaciones-. [4]

Un primer paso para hacerle frente a esta dificultad consiste en proveer un criterio para determinar cuándo dos expresiones tienen diferente significado. En su artículo de 1906, ‘Un breve resumen de mis doctrinas lógicas’, Frege parece proponer el siguiente criterio:

Criterio fregeano de diferencia de sentidos

Dos oraciones O y O^* difieren en su sentido si y sólo si cualquier agente racional que comprendiera ambas pudiera, luego de una consideración reflexiva, juzgar que O es verdadera sin juzgar que O^* es verdadera.

Una preocupación acerca de esta formulación tiene que ver con la aparente existencia de pares de oraciones, del tipo ‘Si Obama existe, entonces Obama=Obama’ y ‘Si McCain existe, entonces McCain=McCain’, dos oraciones que cualquier persona racional capaz de comprenderla juzgaría verdaderas al mismo tiempo. Pero, intuitivamente, el contenido de estas oraciones no es el mismo – mientras que el criterio anterior no acepta esta diferencia—. Una forma de solucionar este problema sería reformular el criterio de la diferencia entre sentidos de las expresiones en términos de las diferencias que resulten de la sustitución de unas expresiones por otras:

Dos expresiones e y e^* difieren en sentido si y sólo si, dadas un par de oraciones O y O^* , (i) estas oraciones difieren solamente por la sustitución de e por e^* , y (ii) estas oraciones son tales que algún agente racional que entendiera ambas sería capaz, previa consideración reflexiva, de juzgar que O es verdadera sin juzgar que O^* es verdadera.

Esta reformulación del criterio de Frege es un caso especial, dado que las oraciones son, por supuesto, expresiones; resuelve el problema de las verdades obvias, ya que pareciera que la sustitución de oraciones de este tipo puede cambiar el valor de verdad de una atribución de actitud proposicional. Más aun, este criterio permite el resultado deseado de que nombres correferenciales como ‘Superman’ y ‘Clark Kent’ tienen sentidos diferentes, ya que hay algún agente racional y reflexivo (Lois Lane, en este caso) capaz de pensar que (21) es verdadera, sin por ello aceptar la verdad de (22).

Pero, aunque podamos así determinar cuándo dos nombres difieren por su sentido, aún no sabemos exactamente *en qué consiste* el sentido de los nombres. La siguiente es una manera plausible, en principio, para tratar de explicar qué es el sentido de un nombre. Sabemos que, sea lo que sea, el contenido de un nombre deberá ser algo capaz de determinar como su referencia aquél objeto que el nombre representa [*stands for*, aquel objeto *por el cual está* ese nombre, i.e. su referente]. Y sabemos que, si el fregeanismo es correcto, debe tratarse de alguna otra cosa que el objeto mismo. Aquí es natural pensar que el contenido de un nombre consistirá en alguna condición que sólo es satisfecha (completamente) por el referente de ese nombre. Los nombres correferenciales pueden tener diferentes sentidos porque cada objeto satisface siempre más de una condición (por ejemplo, [el objeto] Superman/Clark Kent satisface tanto la condición de ser “el superhéroe más admirado por Lois”, como la de ser “el reportero que ella menos admira”). Desde este enfoque, es natural sostener que los significados de los nombres son *descripciones definidas* –frases de la forma ‘el tal y tal’—. Después de todo, este tipo de frases parecen diseñadas para individualizar al único objeto que satisfaga esa condición (lo que viene después de ‘el ...’), en caso de que tal objeto exista. (para más discusión de este tema, ver [la entrada] [descripciones](#)). Esta postura fregeana acerca de los nombres se denomina *descriptivismo fregeano*.

Sin embargo, como argumenta Saul Kripke en *Naming and Necessity*, el descriptivismo fregeano enfrenta serios problemas. El siguiente es uno de los argumentos contra él, el llamado *argumento modal*. Consideremos el nombre propio ‘Aristóteles’, y supongamos, por razones expositivas, que el sentido que yo asocio con ese nombre es el que corresponde a la descripción definida ‘el más grande filósofo de la antigüedad’. Ahora consideremos el siguiente par de oraciones:

(25) Necesariamente, si Aristóteles existe, entonces Aristóteles es Aristóteles.

(26) Necesariamente, si Aristóteles existe, entonces Aristóteles es el más grande filósofo de la antigüedad.

Si el descriptivismo fregeano fuera correcto, y si 'el más grande filósofo de la antigüedad' fuera de hecho la descripción que asocio con el nombre 'Aristóteles', entonces (25) y (26) sólo se diferenciarían por la sustitución de expresiones que poseen el mismo contenido (partes subrayadas). Pero si esto es así, entonces (25) y (26) expresan la misma proposición, y por lo tanto tienen el mismo valor de verdad. Y esto parece ser un error: aunque (25) parece ser siempre verdadera (Aristóteles no podría no haber sido Aristóteles), (26) podría ser falsa (Aristóteles podría haber elegido ser zapatero en lugar de filósofo, por ejemplo; o, quizás si Platón hubiera trabajado más duro, el filósofo más grande de la antigüedad habría sido él —y no Aristóteles—).

Los descriptivistas fregeanos han ofrecido varias respuestas a los argumentos de Kripke; ver, especialmente Plantinga (1978), Dummett (1981), y Sosa (2001). Para respuestas a estos argumentos fregeanos, ver Soames (1998, 2002) y Caplan (2005). Para un breve esbozo de los otros argumentos de Kripke contra el descriptivismo fregeano, ver [la entrada] nombres, §2.4.

Los argumentos kripkeanos ofrecen una potente razón para que incluso los fregeanos consideren abandonar el *descriptivismo* fregeano, aceptando que el sentido de los nombres propios no está dado por las descripciones definidas que los hablantes asocian con tales nombres. El principal problema de este *fregeanismo no descriptivo* será el de explicar cómo es que el sentido de los nombres propios es capaz de determinar su referencia [(si no es mediante la especificación de las condiciones que el referente debe cumplir, es decir, mediante una descripción definida)]. McDowell, 1977, y Evans, 1981, defienden un fregeanismo no descriptivo; para una versión que abandona la idea de que el sentido determina la referencia, ver Chalmers, 2004, 2006.

Es necesario mencionar otros dos problemas a los que se enfrenta el semántico fregeano. El primero cuestiona su afirmación de que esa postura proporciona una solución aceptable al 'rompecabezas [acertijo] de Frege'. En efecto, el fregeano sólo resuelve el problema postulando diferentes sentidos para dar cuenta de las aparentes diferencias en los valores de verdad. Pero este tipo de solución, si es aceptada en un nivel general, parece llevarnos al sorprendente resultado de que no podríamos encontrar *ningún* par de expresiones que compartan el mismo sentido.

Consideremos un par de expresiones que de hecho parecen tener el mismo sentido, por ejemplo 'ketchup' y 'catsup'. (Este ejemplo, así como el argumento que desarrollaremos a continuación, pertenecen a Salmon, 1990). Tomemos el caso de Beto, un cocinero bastante confundido, que piensa que la sustancia roja y sabrosa que viene en los envases con la etiqueta 'catsup' es un tipo de condimento diferente que el que normalmente se etiqueta como 'ketchup'. Consideremos las dos oraciones siguientes:

(27) Beto cree que el catsup es catsup.

(28) Beto cree que el catsup es ketchup.

(27) y (28) son bastante similares a (23) y (24): en los dos casos parece tratarse de pares de oraciones que difieren en valor de verdad, pero que sólo se diferencian por la sustitución de las expresiones subrayadas. En atención a la consistencia, un fregeano debería tratar de explicar la aparente diferencia en los valores de verdad de (27) y (28) de la misma manera en que explicaba la diferencia entre (23) y (24) —esto es, postulando una diferencia de significado entre las expresiones subrayadas. Pero, por un lado, es difícil ver cómo las expresiones 'catsup' y 'ketchup' puedan tener diferentes significados; y, por otro, parece que podríamos generar ejemplos análogos para cualquier par de expresiones supuestamente sinónimas. (Una serie de ejemplos profundamente relacionados al presentado aquí se desarrolla en detalle en Kripke, 1979.)

El ejemplo de 'catsup' y 'ketchup' se relaciona con una segunda dificultad para la semántica fregeana. Esta dificultad consiste en la inquietud inversa a la planteada por los fregeanos a los russellianos. Puede sostenerse plausiblemente que el criterio fregeano para la diferencia de sentidos 'hila demasiado fino', permitiendo establecer distinciones entre los contenidos que en realidad no existen. Una de las formas que puede tomar esta línea argumental se basa (una vez más) en las atribuciones de actitudes proposicionales. Parece que si emito la oración 'Hamurabi pensaba que Héspero era visible sólo por la mañana', lo que estoy diciendo es verdadero si y sólo si cierto pensamiento de Hamurabi tiene el *mismo* contenido que la oración 'Héspero es visible sólo por la mañana' (tal y como es usada por mí). De acuerdo a una concepción russelliana, la anterior condición impone una restricción razonable en cuanto a la verdad o falsedad de mi atribución de actitud proposicional; sólo requiere que Hamurabi crea que cierto objeto instancia la propiedad de ser visible por la mañana. Pero, desde un punto de vista fregeano, lo que se necesitaría es que Hamurabi pensara en el planeta Venus de según el mismo 'modo de presentación' bajo el cual *yo pienso* 'Héspero'. Esto no parece plausible, porque parece que puedo reportar correctamente las

creencias de Hamurabi sin necesidad de conocer lo más mínimo acerca del ‘modo de presentación’ bajo el cual él pensaba en tal o cual objeto. (Para un intento reciente de desarrollar una semántica para las atribuciones de actitudes proposicionales que evita este problema, ver Chalmers, 2007.)

2.2 Teorías no proposicionales

De manera que, aunque hay poderosos motivos para adoptar teorías semánticas proposicionales, cada una de estas teorías enfrenta varios desafíos. Estos desafíos han hecho que algunos consideren que la idea que fundamenta las semánticas proposicionales —la noción de que la tarea que una teoría semántica debe llevar a cabo es la de hacer corresponder sistemáticamente ciertas expresiones con ciertas entidades que constituyen sus significados— está fundamentalmente desencaminada. Esta es precisamente la idea que Wittgenstein intentaba parodiar cuando escribió: “*Se dice: no importa la palabra, sino su significado; y con ello se piensa en el significado como en una cosa de la índole de la palabra, pero aun así diferente de la palabra. Aquí la palabra, ahí el significado. El dinero, y la vaca que se puede comprar con él.*” (Wittgenstein, *Investigaciones Filosóficas*, pág. 120)

Aunque Wittgenstein mismo no consideraba que fuera posible construir una teoría semántica sistemática, ha habido filósofos posteriores que no compartieron tamaña postura antiteórica. Muchos de esos filósofos, sin embargo, han compartido su oposición a la idea de “los significados como entidades”. La siguiente sección intentará dar alguna idea de cómo podría funcionar un enfoque semántico que no se comprometa con las proposiciones y sus constituyentes [(en tanto entidades)]. Lo haremos presentando dos teorías representativas de esta tradición.

La diferencia entre estas teorías se verá mejor si prestamos atención a nuestro esbozo de teoría de la referencia del punto 2.1.1. Recordemos que las teorías proposicionales suplementan la teoría de la referencia con una ‘capa extra’ —una teoría que asigna cada expresión significativa, no sólo una referencia, sino además un *contenido*—. Una de las alternativas a las teorías proposicionales —las *teorías de condicionales de verdad* davidsonianas— considera que esta ‘capa extra’ es innecesaria, y sostiene que una teoría de la referencia es todo lo que una teoría semántica necesita. Una alternativa más radical —las *teorías internalistas* chomskyanas— no sostiene que una teoría de la referencia sea suficiente, sino que es *excesiva*. De acuerdo a esta posición, los significados de las expresiones de los lenguajes naturales no *consisten en*, ni tampoco *determinan* una referencia.

2.2.1 El programa de Davidson

Una de las fuentes de oposición más importantes a la idea de “significados como entidades” es la obra de Donald Davidson. Él pensaba que la teoría semántica debía tomar la forma de una teoría de la verdad para un lenguaje, a la manera de la teoría que Alfred Tarski nos enseñó a construir (ver Tarski, 1936, y [la entrada] [las definiciones de verdad de Tarski](#))

Para nuestros propósitos, será conveniente pensar en una teoría tarskiana de la verdad como una variante de los tipos de teoría mostrados en 2.1.1 del presente artículo. Recordemos que las teorías de la referencia de este estilo especifican, para cada nombre propio del lenguaje, el objeto al que ese nombre refiere, y para cada predicado simple del lenguaje, el conjunto de objetos que satisfacen ese predicado. Si consideramos una oración que combina un nombre propio con un predicado, como

Amelia canta

la teoría nos dice bajo qué condiciones esa oración sería verdadera: nos dice que tal oración es verdadera si y sólo si el objeto al que refiere el nombre ‘Amelia’ es miembro del conjunto de cosas que satisfacen el predicado ‘canta’ –es decir, el conjunto de cosas que cantan. Puede pensarse que una teoría de la referencia aplicable a todo el lenguaje implicaría que, para cada una de las oraciones de este tipo, hay un *enunciado-T* [*T-sentence*] de la forma:

“Amelia canta” es V [*T*] (en el lenguaje) si y sólo si Amelia canta.

Supongamos que expandimos nuestra teoría de la referencia de modo que implique un *enunciado-T* de este tipo para cada oración del lenguaje (no sólo para las oraciones simples que combinan un nombre y un predicado gonádico). Tendríamos así una teoría tarskiana de la verdad para ese lenguaje. La idea de Tarski era que tal teoría definiría el predicado de verdad (“T”) para ese lenguaje. Davidson, en cambio, creía que en la teoría tarskiana de la verdad podemos encontrar una “sofisticada y poderosa fundamentación para una teoría del significado competente” (Davidson, 1967).

La afirmación es sorprendente. ¿Por qué razón habríamos de contar entre las teorías *semánticas* a una teoría que no consta más que de *enunciados-T*, y que nada afirma sobre el significado o el contenido de las expresiones? La respuesta de Davidson es que conocer esa teoría (para el lenguaje

del caso) sería suficiente para comprender ese lenguaje¹⁰. Si Davidson hubiera tenido razón en esto, entonces contaría con un argumento plausible para sostener que una teoría del significado podría tomar esta forma. Después de todo, si alguien entiende un lenguaje, también conoce los significados de las expresiones en ese lenguaje; y si el conocimiento de la teoría tarskiana de la verdad para ese lenguaje resultara ser suficiente para entender el lenguaje, entonces el conocimiento de dicha teoría sería suficiente para conocer todos los hechos relevantes sobre las expresiones en ese lenguaje. En cuyo caso parecería que la teoría nos ha proporcionado todos los hechos sobre los significados de las expresiones en ese lenguaje que se le podían exigir.

Una ventaja de este tipo de enfoques semánticos es su parsimonia: no echa mano de ninguno de los tipos de entidades –como las *intensiones*, las *proposiciones russellianas*, o los *sentidos fregeanos*—que las semánticas proposicionales asignaban a las expresiones (como vimos anteriormente). Claro que, como también vimos, estas entidades se introdujeron en un intento por proporcionar un tratamiento semántico adecuado para varios tipos de construcciones lingüísticas, y podríamos preguntarnos si es posible dar una teoría tarskiana de la verdad para un lenguaje natural *sin utilizar* intensiones, proposiciones russellianas, o sentidos fregeanos. El programa de Davidson evidentemente se basa en ello, pero es controvertido que tal cosa sea posible. La discusión de este tema queda más allá del alcance del presente artículo; una buena manera de adentrarse en el debate es la discusión acerca de si el programa de Davidson es capaz de ofrecer un tratamiento adecuado de las atribuciones de actitudes proposicionales. (Ver discusión sobre el enfoque paratáctico y las formas lógicas interpretadas en los reportes de actitudes proposicionales. Para un tratamiento de las atribuciones de actitudes en Davidson, ver Davidson, 1968; para mayor discusión, ver, entre otros, Burge 1986; Schiffer 1987; LePore and Loewer 1989; Larson and Ludlow 1993; Soames 2002.)

Pero por ahora dejemos de lado esta dificultad. Supongamos que sí es posible construir una teoría de la verdad para un lenguaje natural, y veamos si este tipo de teoría sería capaz de proporcionar una semántica adecuada. Hay dos razones principales para pensar lo contrario, y ambas se deben en última instancia a Foster (1976). En lo que sigue adoptaré la denominación de Larson y Segal (1995), y las llamaré *el problema de la extensión*, y *el problema de la información*.

¹⁰ Cf. “La tesis de Davidson se apoya en una intuición sólida: conocer el significado de una oración equivale a saber todas las (posibles) ocasiones en que dicha oración es verdadera.” p.135 de: Acero, Bustos, Quesada, “Introducción a la filosofía del Lenguaje”, Ed. Teorema, Madrid, 1989

El problema de la extensión deriva del hecho de que no basta con que una teoría semántica de este tipo (cuyos teoremas son enunciados-T) proporcione teoremas verdaderos; por ejemplo, el siguiente enunciado-T:

‘La nieve es blanca’ es V [T] en castellano *syss* la hierba es verde

es verdadero, pero no nos dice demasiado sobre el significado de ‘La nieve es blanca’. Lo que necesitamos es que por cada oración del lenguaje objeto la teoría semántica implique [*entail*] un (y sólo un) enunciado-T *interpretativo*: es decir, un enunciado-T cuya mitad derecha proporcione el *significado* de la oración mencionada en su mitad izquierda. De acuerdo a esto, nuestra teoría deberá implicar *al menos un* enunciado-T para cada oración en el lenguaje objeto (porque la idea es dar el significado de todas las oraciones del lenguaje). Y deberá implicar *no más de uno* porque, si entre los teoremas de la teoría hubiera más de un enunciado-T por cada oración O del lenguaje objeto, aunque conociéramos todos los teoremas de la teoría seríamos incapaces de comprender O (ya que no sabríamos cuál de los enunciados-T que mencionan a O es su *enunciado interpretativo*).

El problema es que, según parece, una teoría que implique [*entail*] *al menos un* enunciado-T por cada oración del lenguaje, implicará también *más de uno*. Dado el par de oraciones p, q , si la teoría implica el siguiente enunciado-T

O es V en L *syss* p

entonces, dado que p es lógicamente equivalente a $(p \ \& \ \sim(q \ \& \ \sim q))$, la teoría implicará también el siguiente enunciado-T:

O es V en L *syss* $(p \ \& \ \sim(q \ \& \ \sim q))$,

Pero si el primero es el enunciado-T interpretativo de O , este segundo enunciado-T no lo será. Por lo tanto, la teoría implica al menos un enunciado-T no interpretativo, y aunque conociéramos toda la teoría, seríamos incapaces de determinar cuál de los enunciados relevantes es interpretativo y cuál no; en ese caso, no podríamos entender el lenguaje.

El problema de la información consiste en que, aunque la teoría semántica implique todos los enunciados-T interpretativos de las oraciones del lenguaje objeto (y sólo éstos), el conocimiento de lo que dicen estos teoremas no basta para comprender el lenguaje objeto. Parece que de hecho

podemos saber *lo que* dice una serie de enunciados-T interpretativos *sin* saber *que son* interpretativos. Por ejemplo, puedo saber lo que dice el siguiente enunciado-T (interpretativo):

“Londres est Jolie” es V en francés *syss* Londres es bella

Pero aunque sepa lo que dice este enunciado-T, puedo ignorar completamente el significado de la oración mencionada en su mitad izquierda. Después de todo, la verdad de este enunciado parece compatible con el hecho de que la oración de la mitad derecha sea *materialmente equivalente* a la oración mencionada en la mitad izquierda, y que sin embargo no tenga el *mismo significado* que ésta. Lo anterior parece indicar que conocer todo lo que nos dice una teoría de la verdad de este tipo no es suficiente para comprender un lenguaje. (Para respuestas a estas críticas, ver Davidson, 1976, Larson y Segal, 1995, y Kobel, 2001; para críticas a esas respuestas, ver Soames, 1992, y Speaks, 2006.)

2.2.2 Semántica internalista chomskyana

Existe otra alternativa a la semántica proposicional, que difiere del programa de Davidson al menos tanto como éste difiere de los enfoques proposicionales. Esa postura es denominada en ocasiones ‘semántica internalista’, debido a su contraste con las teorías que ubican las propiedades semánticas de las expresiones en cierta relación con el mundo externo. Un enfoque internalista es el asociado al trabajo de Noam Chomsky (ver especialmente, Chomsky, 2000).

Es sencillo comprender a qué se opone este enfoque. El internalista niega un presupuesto común a todos los enfoques tratados hasta aquí: la idea de que al dar el contenido de una expresión, lo que hacemos es sobre todo especificar alguna característica acerca de la relación entre esa expresión y ciertas cosas del mundo [sobre las cuales podemos, a su vez, decir cosas utilizando esa expresión], cosas sobre las que hablamos precisamente utilizando esa expresión. Para el internalista, las expresiones en cuanto tales no guardan ninguna relación semánticamente interesante con las cosas del mundo; por ejemplo, los nombres no *refieren* a los objetos con los que se los asocia. Las oraciones no son verdaderas o falsas, y tampoco expresan proposiciones verdaderas o falsas; y la idea de que podemos comprender los lenguajes naturales mediante una teoría de la referencia es errónea. De acuerdo a esta postura, a veces utilizamos oraciones para decir cosas verdaderas o falsas acerca del mundo, y ocasionalmente utilizamos nombres para referir individuos; pero esto es sólo

una de las cosas que podemos hacer con los nombres y las oraciones, y no se relaciona con el significado de esas expresiones.

Lo que ya es más difícil, en un espacio tan reducido, es tratar aclarar en qué consiste el significado de las expresiones lingüísticas para el internalista. Según McGilvray, 1998: “su tesis básica es que los significados son contenidos intrínsecos de las expresiones... y que son definidos y determinados por su sintaxis, en un sentido amplio”. Esta descripción basta para mostrar la diferencia entre este enfoque y los tratados más arriba; no se trata sólo de que la atención no está puesta en la relación entre ciertos elementos sintácticos y la realidad no-lingüística; sino más bien de que, según esta postura, las propiedades sintácticas y semánticas de las expresiones son consideradas inseparables unas de otras. McGilvray añade que “este inusual enfoque al problema del significado tiene hoy en día pocos defensores”, lo cual es probablemente cierto –aunque quizás no tanto ahora como en 1998, momento en que se publicó el texto citado— Sobre las defensas y desarrollos en este programa, ver McGilvray, 1998, Chomsky, 2000, y Pietroski, 2003,2005.

3. Teorías fundacionales del significado

Pasamos ahora a la segunda clase de ‘teorías del significado’ [(de las mencionadas en el punto 1)]: las teorías fundacionales del significado, que intentan especificar los hechos en virtud de los cuales las expresiones del lenguaje natural llegan a adquirir sus propiedades semánticas.

Las preguntas que este tipo de teorías intentarán responder es una pregunta bastante corriente en varios campos de la filosofía. La *filosofía de la acción* se pregunta por cuáles son los hechos en virtud de los cuales una determinada conducta constituye una acción intencional; el problema filosófico de la *identidad personal*, consiste en preguntarse y tratar de determinar cuáles son los hechos en virtud de los cuales x e y son una misma persona; en el campo de la *ética* nos preguntamos qué determinan que una acción moral sea correcta o incorrecta [‘buena’ o ‘mala’]. Sin embargo, aunque se trate de un tipo de preguntas recurrentes en varios campos de la filosofía, no sabemos exactamente qué condiciones deberán satisfacer las respuestas en este caso, o cuándo podemos esperar que este tipo de preguntas den lugar a respuestas interesantes.

Debido a estas dificultades es que una posición posible respecto a este tipo de teorías del significado es precisamente aquella que *niega* la posibilidad de que podamos encontrar una teoría fundacional correcta. Podríamos, por ejemplo, aceptar alguna de las teorías semánticas esbozadas más arriba, pero sosteniendo al mismo tiempo que los hechos relativos a los significados de las expresiones son *primitivos* –en el sentido de que no hay ninguna explicación sistemática posible acerca de cuáles son los hechos en virtud de los cuales esas expresiones llegan a tener los significados que tienen. (ver, por ejemplo, Johnston, 1988.)

Sin embargo, la mayoría de los filósofos han optado por una posición diferente, y han afirmado que debe de haber algún tipo de teoría sistemática, basada en ciertos hechos relativos a los hablantes de un lenguaje, en virtud de los cuales las palabras llegan a adquirir sus actuales propiedades semánticas. Normalmente, estos filósofos han intentado especificar las propiedades *necesarias* y *suficientes* (así como *explicativamente fundamentales*) por las cuales las expresiones llegan a adquirir un determinado significado.

En los últimos cincuenta años, la visión dominante sobre la fundamentación del significado en la filosofía analítica ha sido la posición *mentalista*, según la cual los significados de las expresiones en

los lenguajes naturales pueden ser explicadas en términos de los contenidos de los estados mentales de los hablantes. En lo que sigue trataremos en principio de las teorías fundacionales del significado de carácter mentalista, para luego ocuparnos de las alternativas no-mentalistas.

3.1 Teorías mentalistas

Lo que todas las teorías mentalistas tienen en común es el hecho de que analizan un tipo de representaciones —las *representaciones lingüísticas*— en términos de otro tipo de representaciones —las *representaciones mentales*—. Para aquellos filósofos interesados en dar cuenta del contenido, o de la representación, en términos no representacionales, las teorías mentalistas serán entonces un primer paso en la tarea de proporcionar una explicación exhaustiva de la fundamentación de las representaciones lingüísticas. El segundo paso de este proyecto consistiría en dar una explicación en el nivel de una teoría de los contenidos mentales. (Para un resumen de esta clase de teorías, ver [la entrada] [representación mental](#), y los ensayos compilados en Stich y Warfield, 1994). De hecho, el creciente interés en las teorías de la representación mental se debe, entre otras cosas, a la popularidad de las teorías mentalistas del significado y a la convicción de que el contenido debería ser explicado en términos no representacionales.

Dado que los mentalistas¹¹ intentan explicar la naturaleza del significado en términos de estados mentales de los hablantes, las teorías mentalistas pueden subdividirse de acuerdo a cuál es el tipo de estados mentales que consideran relevantes para la determinación del significado. Las propuestas más desarrolladas en ese sentido son el llamado programa griceano [(por Paul Grice)], que se propone explicar el significado de las expresiones en términos de las *intenciones comunicativas* de los hablantes; el otro es el enfoque según el cual que el significado de las expresiones es establecido por ciertas *convenciones* que determinan la correspondencia entre oraciones y creencias. Trataremos ambas posturas respectivamente, para discutir luego una tercera alternativa disponible al mentalista.

3.1.1 El programa de Grice

Paul Grice desarrolló un análisis del significado que se basa en la conjunción de dos afirmaciones: (1) los hechos acerca del significado de las expresiones han de ser explicados o analizados en

¹¹ Adoptamos el uso del autor, según el cual, por razones de concisión, quienes proponen o defienden una teoría fundacional del significado *de carácter mentalista* son denominados sencillamente ‘mentalistas’ (comparar con las secciones anteriores, en las que se habla de ‘(semánticos) intensionales’, ‘(semánticos) fregeanos’, etc.). [nota mía. PD: acá convendría aclarar que ‘mentalista’ no tiene nada que ver con la parapsicología, la telepatía, el Tarot, etc. Digo, nunca falta algún despistado...]

términos de hechos relativos a qué es lo que los hablantes intentan decir o comunicar [*mean*] al emitirlos y, (2) que los hechos relativos a lo que los hablantes intentan comunicar mediante sus emisiones pueden ser explicados en términos de sus *intenciones*. En estas dos tesis se resume el llamado ‘programa de Grice’, mediante el que se intenta dar cuenta del significado en términos de los contenidos de las intenciones de los hablantes.

Para entender el enfoque griceano del significado, es importante aclarar primero la distinción entre significado, o contenido, de *las expresiones lingüísticas* –lo que las teorías semánticas de la sección 2 intentaban describir—, y aquello que *los hablantes* significan [(o intentan dar a entender)] mediante sus emisiones de esas expresiones. El siguiente ejemplo ayudará a ilustrar la distinción. (ver [la entrada] [pragmática](#) para un tratamiento más extenso del tema). Supongamos que, en respuesta a una pregunta sobre los restaurantes de mi ciudad, digo, “Bueno, Paraná no es exactamente París”. El significado de esta oración es bastante claro: expresa la proposición (verdadera) de que la ciudad de Paraná no es la ciudad de París. Pero lo que yo quiero decir [(o dar a entender) = ‘*mean*’] cuando emito esa oración no es sólo esta trivialidad: lo que yo intento dar a entender no es solamente que Paraná y París son dos ciudades diferentes, sino además que la calidad de la comida de Paraná no es comparable a la de la comida en París. Y este ejemplo es típico en un aspecto importante: normalmente las proposiciones que los hablantes intentan comunicar mediante sus emisiones incluyen no sólo las proposiciones expresadas por las oraciones emitidas, sino además alguna otra proposición. Cuando preguntamos “¿Qué quiso usted decir con eso?”, la respuesta que esperamos no consiste en que se nos proporcione el significado de [la proposición expresada por] la oración que nuestro interlocutor acaba de emitir¹².

La idea detrás de la tesis (1) de la teoría del significado de Grice es que de entre estos dos fenómenos, el fundamental es el de *significado del hablante*; las oraciones y otras expresiones obtienen su significado de acuerdo a lo que los hablantes *intentan* [hacer, comunicar] mediante esas expresiones. (Para mayores detalles sobre la propuesta de Grice acerca de cómo el significado de las oraciones [*sentence-meaning*] puede ser explicado en términos de significados del hablante [*speaker-meaning*], ver [la entrada] [Paul Grice](#).) Una manera de fundamentar la tesis de que el significado del hablante es explicativamente anterior [*explanatorily prior to*] al significado de las expresiones, es mostrar que los hechos relativos al significado del hablante pueden ser explicados sin hacer uso de

¹² Donde utilizamos ‘emisión’, ‘emitir’, etc. en el libro “*Introducción a la filosofía del Lenguaje*” (Acero, Bustos, Quesada, ed. Teorema, Madrid, 1989), se utiliza ‘proferencia’, ‘proferir’, etc. La expresión en inglés es ‘*utterance*’, ‘*utter*’.

hechos relativos a los significados de las expresiones; y en esto consiste precisamente la etapa (2) del análisis griceano.

Grice consideraba que el significado del hablante podía ser analizado en términos de las intenciones comunicativas de los hablantes –en particular, la intención de lograr que su audiencia forme determinadas creencias. Aunque hay muchas versiones diferentes de este tipo de análisis griceano de los significados del hablante, la siguiente es tan buena como cualquier otra [5]:

(G)¹³ la emisión de x por parte de H [el hablante], significa que p si y sólo si por su emisión de x , H intenta que:

1. su audiencia llegue a creer p
2. su audiencia reconozca la intención del hablante, y
3. (2) sea causa de (1).

[3. que (2) sea razón para que se de (1)]

[o, más explícitamente: 3. que en razón del reconocimiento de la intención del hablante (condición 2), su audiencia llegue a creer p (condición 1)]

Una manera de darnos cuentas de la motivación intuitiva de este tipo de análisis del significado (como G) es comenzando por la idea de que ‘quiero decir algo’ mediante la emisión de cierta expresión involucra mi intento de comunicar ciertas creencias (mías) a mi audiencia. Intentar comunicar mis creencias puede ser considerado, a su vez, como mi intento de que alguien *comparta* esas creencias; sin embargo, puedo tratar de que *a causa de* mi acción alguien forme la creencia de que p , sin que por ello mi acción ‘quiera decir’ (o ‘signifique’) que p . Podemos tratar de ilustrar este hecho mediante un ejemplo. Supongamos que te digo: “Me estás pisando el pie.”; lo que intento que creas es *que dije* ‘Me estás pisando el pie.’ Pero *lo que* quiero decir no es:

“(yo) acabo de emitir la oración: ‘Me estás pisando el pie.’”

¹³ Esta formulación del análisis de Grice se basa en presente artículo, y además en la formulación que aparece en (Acero, Bustos, Quesada, ed. Teorema, Madrid, 1989).

Lo que quiero comunicar mediante mi emisión *no es* el hecho de que *estoy emitiendo cierta oración*. Esa *es* precisamente mi emisión. Pero lo que *intento* [lo que 'quiero decir' o 'significo'] mediante mi emisión es precisamente *que me estás pisando el pie* —o más exactamente, que *dejes de pisarme el pie*.

Este tipo de ejemplos indica que el significado del hablante no tiene que ver sólo con su intención de que su audiencia forme cierta creencia —sino que tendrá que ver con el intento de que su audiencia forme cierta creencia *de una determinada manera* [i.e. *por determinadas razones*]. Pero, además del intento de causar ciertas creencias, ¿qué otra cosa se requiere para el significado de *p*? La idea de Grice es que uno no sólo intenta lograr que su audiencia forme cierta creencia, sino que además lo hagan basados en su reconocimiento de la intención del hablante. En el ejemplo anterior, no espero que llegues a creer que emití cierta oración, sino que espero que formes la creencia de que me estás pisando el pie, basado en tu reconocimiento de mi intención de que eso es lo que intento mediante la emisión de la oración.

Sin embargo, aunque podamos dar una motivación plausible del análisis de (G), y aunque dicho análisis se ajuste bastante bien a muchos casos, también se encuentra expuesto a contraejemplos convincentes. Se los puede agrupar en tres tipos de casos: (i) los casos en los que el hablante significa *p* mediante una emisión, a pesar de saber que su audiencia ya cree *p*, como en los casos de confesión, o en el caso de que le estemos recordando algo a alguien; (ii) los casos en los que el hablante significa *p* mediante su emisión, como en la conclusión de un argumento, en el cual el hablante intenta que su audiencia crea [o llegue a creer] *p* en base a la *evidencia*, y no en base a su reconocimiento de las intenciones del hablante; y (iii) los casos en los que no hay ninguna audiencia, como cuando se utiliza el lenguaje para llevar a cabo pensamientos [por ejemplo, en un 'monólogo interior']. Tales casos nos hacen dudar de que exista una conexión tan estable como la propuesta por Grice entre el significado del hablante y los efectos que este intenta producir en su audiencia, conexión que sería indispensable para fundamentar un análisis del tipo proyectado por Grice. Aun es materia de controversia si alguna explicación en la línea de (G) pueda llegar eventualmente a tener éxito. A pesar de esta controversia, el análisis griceano sigue siendo lo más cercano a la ortodoxia dentro de las teorías fundacionales del significado.

For developments of the Gricean program, see—in addition to the classic essays in Grice (1989)—Schiffer (1972), Neale (1992), and Davis (2002). For an extended criticism, see Schiffer (1987).

Para desarrollos en el programa de Grice, ver —además de los ensayos clásicos del mismo Grice (1989)— Schiffer (1972), Neale (1992), y Davis (2002). Para más críticas, ver Schiffer (1987).

3.1.2 Significado, creencia y convención

Una alternativa importante al análisis griceano, que comparte con éste su enfoque mentalista del análisis del significado en términos de los contenidos de los estados mentales es el que, en lugar de analizar el significado en términos de las *intenciones* de los hablantes, lo hace en términos de sus *creencias*.

Parece plausible que este tipo de análisis pueda llevarse a cabo. Después de todos, existen ciertas regularidades que conectan las emisiones con las creencias de los hablantes. Parece que, a grandes rasgos y en la mayoría de los casos, los hablantes sólo emiten [profieren] sinceramente la oración que (en el contexto) significa p si de hecho creen que p [o 'creen p ']. Podríamos tratar de analizar el significado directamente en términos de las creencias de los hablantes, diciendo que una oración O expresa la proposición p si, típicamente, los miembros de esa comunidad de hablantes no emitirían p si no creyeran p . Pero podemos pensar en una acción o tipo de acción tal que nadie (en esa comunidad) la llevaría a cabo si no creyera la proposición p , y que sin embargo nadie (en esa comunidad) esté al tanto de que los otros miembros actúan de acuerdo a esta regla. Parece plausible sostener que en esa comunidad, dicha acción [*'action-type'*; o sea, la clase o tipo de acción] no expresaría la proposición p , y de hecho no tendría ningún significado.

Casos como el anterior parecen mostrar que las regularidades entre significado y creencias no bastan por sí solas para fundamentar un análisis del significado. Es por ello que muchos defensores del análisis mentalista del significado en términos de creencias han buscado llevarlo a cabo mediante un análisis de las *convenciones* que gobiernan tales regularidades. Existen diferentes análisis que intentan dar cuenta de en qué consiste el hecho de que una determinada regularidad se de por convención [*'está determinada'*, o *'haya sido establecida'* por convención]. De acuerdo a una postura importante, una oración O expresa la proposición p si y sólo si se satisfacen las siguientes condiciones: (1) los miembros de la comunidad de hablantes típicamente emiten O si creen p , y típicamente creen p al oír O ; (2) los miembros creen que (1) es verdadero, y; (3) el hecho de que los miembros creen que (1) es verdadero, y que creen que los otros miembros de la comunidad creen que (1) es verdadero les da una buena razón para actuar como si (1) fuera verdadero¹⁴. (Esta es una

¹⁴ [nota mía] Me parece que una versión menos literal, pero más quizás más idiomática, se obtendría reemplazando '*(1) es verdadero*' por '*se da (1)*' (después de todo, (1) es precisamente una *condición*). Este reemplazo daría la siguiente formulación (a partir de los dos puntos):

"(1) los miembros de la comunidad de hablantes típicamente emiten [profieren] O si creen p , y típicamente creen p al oír O ; (2) los miembros creen que se da (1), y; (3) el hecho de que los miembros creen que se da (1), y que creen que los otros miembros de la comunidad creen que se da (1), les proporciona una buena razón para actuar como si como si se diera (1)."

versión simplificada de la desarrollada en Lewis, 1975). Para una discusión crítica de este tipo de análisis del significado, ver Burge 1975, Hawthorne 1990, Laurence 1996, and Schiffer 2006.

3.1.3 Teorías basadas en la representación mental

Las dos clases de teorías mentalistas esbozadas más arriba intentan explicar el significado en términos de las relaciones entre las expresiones lingüísticas y las actitudes proposicionales de los hablantes del lenguaje del caso. Pero no es la única clase de teoría disponible para quien pretenda analizar el significado en términos de representaciones mentales. Una posición común en la filosofía de la mente y la ciencia cognitiva es que las actitudes proposicionales de los sujetos se fundan en un lenguaje interno del pensamiento, constituido por representaciones mentales. (Ver [la entrada] [teoría computacional de la mente](#).) Podríamos tratar de explicar el significado lingüístico directamente en términos de los contenidos de las representaciones mentales, mostrando la correspondencia entre expresiones lingüísticas y representaciones mentales; así, el significado de una determinada expresión para un cierto individuo estaría dado por contenido de la representación mental asociada a esa expresión. Para una discusión de este tipo de teoría, ver Laurence, 1996.

Así como los defensores de las teorías griceanas y de las teorías basadas en convenciones suelen considerar sus propuestas sólo como un primer paso en el análisis del significado —dado que lo analizan en términos de otro tipo de representaciones mentales—, los defensores de las teorías basadas en la representación mental típicamente intentarán proporcionar un análisis independiente acerca de los contenidos de las representaciones mentales. Para una buena reseña de intentos de construir este último tipo de teorías, ver los ensayos en Stich y Warfield, 1994.

3.2 Teorías no mentalistas

Como ya se ha hecho notar, no todas las teorías fundacionales se proponen explicar el significado en términos de representaciones mentales. Una teoría fundacional no mentalista podría ser preferible por varios motivos; por ejemplo, podríamos ser escépticos respecto a las teorías mentalistas disponibles; podríamos pensar que las representaciones mentales deberían ser analizadas en términos de representaciones lingüísticas (y no a la inversa, como proponen los mentalistas); o podríamos sostener que la representación debería ser analizable en términos no-representacionales, y dudar de que sea posible encontrar una explicación adecuada de la representación que sea capaz de continuar con la tarea explicativa de las teorías mentalistas —que,

como vimos, sólo se proponen reducir el significado a la representación mental, sin expedirse sobre esta última.

Todas las *teorías fundacionales [del significado] no-mentalistas* intentan explicar los significados de las expresiones en términos de su *uso*. Lo cual, en principio, no aclara demasiado: lo mismo puede ser dicho de las teorías mentalistas (la teoría griceana, por ejemplo, sostiene que lo que cuenta es la intención comunicativa con la que se utiliza una expresión). Proponer una teoría fundacional [del significado] no-mentalista es decir cuáles aspectos del uso de una expresión son los que determinan su significado –y hacerlo, a su vez, sin recurrir a hechos relativos a las actitudes proposicionales y las representaciones mentales que acompañan a ese uso. En lo que sigue discutiremos algunos de los aspectos de uso de las expresiones mediante los cuales las teorías no-mentalistas intentan dar cuenta de los significados de esas expresiones.

3.2.1 Origen causal

En *Naming and Necessity*, Kripke sugiere que la referencia de un nombre podría explicarse en términos de *la historia del uso* de ese nombre, y no por las descripciones definidas que los hablantes asocian a ese nombre. Kripke consideró que, en los casos normales, la explicación de la referencia de un nombre puede darse en dos pasos: explicando en un primer lugar la *introducción* de un nombre, un evento en el cual se asocia por primera vez un término con tal o cual objeto –el acto de ‘bautismo’—; y dando cuenta, posteriormente, de la exitosa *transmisión* – de hablante en hablante – de ese uso [de asociar consistentemente determinado nombre con determinado objeto].

Un enfoque posible para una teoría del significado consistiría en extender las sugerencias de Kripke de dos maneras: primero, considerar que este mecanismo podría servir para explicar no sólo la referencia, sino también el significado [6]; y segundo, extender el análisis a otras partes del discurso (además de los nombres). (Para discusión de este tema, ver Devitt, 1981). De esta manera, podríamos intentar explicar los significados de las expresiones en términos de su *origen causal*.

Vale la pena hacer notar que no está claro que este tipo de proyecto no caiga dentro de la clase de las teorías mentalistas. Podríamos pensar que la introducción de un término [el ‘bautismo’] implica *intentar* que ese nombre sea asociado con tal o cual objeto o propiedad, y asimismo podríamos considerar que la transmisión de un término de hablante en hablante implica que en cada caso se *intente* utilizar el término de la misma manera que el hablante anterior. Si fuera así, entonces quizás

las teorías causales también propongan, a fin de cuentas, un análisis del significado en términos de las intenciones de los hablantes (y no en menor medida que las teorías griceanas).

Existen dos problemas típicos para las teorías causales de esta clase (ya sea que se las elabore en términos mentalistas o no-mentalistas). El primero tiene que ver con que no resulta demasiado natural intentar extender la teoría para abarcar otros términos más allá los nombres. Entre los ejemplos que muchos consideran particularmente problemáticos están los nombres vacíos y los términos carentes de referentes, el vocabulario lógico, y el caso de los predicados (cuyo contenido no parece estar relacionado de manera directa a ninguna experiencia sensible, y para los que parece difícil concebir un acto de ‘bautismo’ inicial). El segundo de los problemas, que en ocasiones se denomina el problema del ‘*qua*’, consiste en la dificultad de explicar cuál de todas las causas de la introducción de un término determina su contenido. Supongamos que el término ‘agua’ fue introducido en la presencia de H₂O. ¿Qué determina que el término correspondiese a *esa* sustancia, y no a cualquier tipo de líquido en general, o a cualquier líquido incoloro presente en la región donde el término fue introducido? Los defensores de este tipo de teorías causales deben proporcionar alguna respuesta a estas preguntas; para una discusión del tema, ver Devitt y Sternly, 1987.

3.2.2 Maximización de la verdad y el principio de caridad

Las teorías causales intentan dar una explicación del significado en términos de las relaciones entre las expresiones y los objetos y propiedades representados por esas expresiones. Un tipo muy diferente de teoría fundacional mantiene el énfasis en la relación entre las expresiones y el mundo, pero otorgando un papel central al *principio de caridad* – principio que sostiene que una asignación correcta de significados a las expresiones del lenguaje de un hablante será aquella que maximice la verdad de las emisiones del hablante.

Un influyente defensor de esta posición fue Donald Davidson, quien justificaba de la siguiente manera la motivación detrás de su propuesta:

“Una causa central de dificultades es la manera en que las creencias y los significados conspiran para dar cuenta de las emisiones. Un hablante que afirma la verdad de una oración lo hace, en parte debido a lo que intenta significar [*‘what he means’*] mediante la emisión de esa oración, y en parte debido a cierta creencia [o ‘debido a lo que cree’]. Si lo único que sabemos es que se trata de una emisión sincera, no podremos inferir la creencia si

ignoramos el significado, y no tenemos posibilidades de inferir el significado si ignoramos la creencia.” (Davidson, 1973^a,314; ver Davidson, 1973b.)

La idea de Davidson era que cualquier intento de determinar los hechos en virtud de los cuales las expresiones poseen cierto significado debe enfrentarse a una suerte de dilema: si tuviéramos una explicación independiente que diera cuenta de en qué consiste el hecho de que un agente posea determinada creencia (con un determinado contenido), podríamos tomar ese dato como base para explicar en qué consiste el hecho de que una oración posea significado; por otro lado, si tuviéramos una explicación independiente de en qué consiste el hecho de que una oración posea significado, podríamos a partir de ella explicar en qué consiste el hecho de que un agente posea cierta creencia (con determinado contenido). Pero en realidad carecemos de una explicación independiente tanto de la primera clase de hechos como de la segunda, y esto se debe a que existen múltiples asignaciones de creencias y significados que son todas igualmente consistentes con una misma conducta lingüística del hablante. La solución que Davidson propone para este dilema es que debemos definir los hechos relativos a las creencias y los significados de manera *simultánea*, por lo cual deberemos definirlos en términos de un *tercer hecho* independiente: el hecho de que las creencias de un agente y los significados de sus palabras son serán dadas por aquella interpretación que maximice la verdad de sus creencias y emisiones.

De esta manera, significados y creencias quedan ligados a la verdad, por lo que este tipo de teoría fundacional implica que es imposible que cualquier hablante competente de un lenguaje tenga creencias radicalmente falsas sobre el mundo; y esto a su vez implica que ciertos niveles de desacuerdo *radical* entre hablantes o comunidades son asimismo imposibles (dados que las creencias de cada comunidad deben ser siempre en su mayor parte verdaderas). Esta es una consecuencia de la posición defendida por Davidson (ver, Davidson, 1974); pero uno podría pensar que tanto el desacuerdo radical como el error radical son perfectamente posibles, y que por lo tanto cualquier teoría que implique lo contrario debe ser incorrecta. (Para discusión de este tema, ver el capítulo 13 de Soames, 2003).

Una preocupación diferente acerca de este tipo de teorías proviene de que su requerimiento de maximizar la verdad de las emisiones de los hablantes no parece suficiente para determinar el significado de las expresiones en ese lenguaje. Parece en principio plausible que puedan ofrecerse *varias* interpretaciones diferentes del mismo lenguaje que cumplan con el requisito de maximizar la verdad. Una forma de ver la fuerza de esta objeción es recordando el caso (mencionado en nuestra discusión del punto 2.1.5) de que un par de oraciones pueden ser verdaderas exactamente en las

mismas circunstancias y sin embargo tener significados diferentes. Una preocupación relacionada sería que una teoría del estilo davidsoniano implicaría una poco plausible *indeterminación del significado*. Para el intento más acabado de Davidson por responder a esta clase de objeciones, ver el capítulo 3 de Davidson, 2005.

3.2.3 Regularidades en el uso

Una forma diferente de desarrollar una teoría fundacional [del significado] no-mentalista se concentra menos en la *relación* entre las expresiones suboracionales o las oraciones y ciertas porciones de la realidad no lingüística, y más en las *regularidades* que gobiernan nuestro uso del lenguaje. Este tipo de propuestas han sido defendidas por varios autores; en lo que sigue esbozaremos una versión de la posición desarrollada en Horwich, 1998, 2005.

La idea central de Horwich es que nuestra aceptación de las oraciones es gobernada por ciertas leyes y que, para toda expresión no ambigua, hay una única ‘regularidad de aceptación’ que explica todos nuestros usos de esa expresión. El tipo de regularidad de aceptación relevante variará de acuerdo al tipo de expresión cuyo significado intentamos explicar. Por ejemplo, nuestro uso de un término relacionado con la percepción, como es ‘rojo’, puede ser explicado por la siguiente regularidad de aceptación:

La disposición a aceptar ‘esto es rojo’ en respuesta al tipo de experiencia visual normalmente provocada por una superficie roja.

Mientras que en el caso de un término lógico como ‘y’ [‘^’ (conjunción)] la regularidad de aceptación involucrará la disposición a aceptar ciertas inferencias que involucren pares de oraciones, y no en la en la disposición a responder a ciertos tipos de experiencias. Por ejemplo:

La disposición a aceptar el esquema “ $p, q // p \text{ y } q.$ ”

Como estos ejemplos ilustran, resulta plausible sostener que una de las ventajas de la propuesta de Horwich reside en su capacidad para tratar expresiones de diferentes categorías.

Al igual que sus competidoras, la teoría de Horwich no es inmune a ciertas objeciones. Podría señalarse con cierta preocupación el hecho de que, si la *actitud* de aceptación de una oración es analizada en términos de creencias, este tipo de explicaciones nos devuelve al campo de las teorías mentalistas. También está la objeción –que afecta a otras teorías del significado basadas en el ‘uso’,

el ‘papel conceptual’ o el ‘papel funcional’ de las expresiones – de que este tipo de explicación implica la existencia de diferencias de significado en casos en que no parece haber tales diferencias. Por ejemplo, parece que diferentes regularidades de aceptación pueden explicar los usos de *un mismo término* por parte de dos personas diferentes. En Schiffer, 2000, se discute el ejemplo de ‘perro’, y de las diferentes regularidades de aceptación que gobiernan el uso de ese término para los casos de personas ciegas, de personas con pocos conocimientos de biología, o de aquellas personas que sólo están familiarizadas con ciertas razas de perros. [7]

3.2.4 Normas sociales

Esta última preocupación concerniente a la teoría de Horwich se debe a que se trata de una explicación de carácter básicamente *individualista*: explica el significado de una expresión para un individuo en términos de las propiedades del uso de ese término por ese individuo. Un tipo considerablemente diferente de teoría del significado se distancia de la preocupación por las leyes que explican el uso de los términos por parte de los *individuos*, e intenta encontrar las normas que gobiernan el uso de los términos *en una sociedad*. Como las otras propuestas aquí reseñadas, la idea de que el significado es producto de este tipo de normas sociales tiene una larga historia; se la asocia particularmente con la obra del último Wittgenstein y sus discípulos filosóficos. (Ver especialmente, Wittgenstein, 1953)

Un defensor reciente de este tipo de enfoque es Robert Brandom. De acuerdo al enfoque de Brandom, el significado de una oración está dado por las condiciones bajo las cuales es apropiado o correcto, en una determinada sociedad, llevar a cabo diferentes actos de habla en los que se utilice esa oración. Para poder desarrollar una teoría de esta clase, debemos lograr dos cosas. Primero, debemos ser capaces de mostrar de qué manera pueden explicarse los significados en términos de estas condiciones normativas – en la formulación (muy poco ortodoxa) de Brandom, debemos poder mostrar de qué manera puede explicarse la semántica en términos de la pragmática. Y segundo, deberemos explicar cómo es que estas condiciones normativas llegan a ser instituidas por ciertas prácticas sociales. Para más detalles, ver Brandom, 1994, en donde se desarrolla extensamente la propuesta; para una discusión crítica sobre el intento de Brandom de llevar a cabo la segunda de las tareas mencionadas más arriba, ver Rosen, 1997.

Ejemplos (oraciones numeradas)

- (1) Barack Obama es el cuadragésimo cuarto [44º] presidente de los Estados Unidos.
- (2) John McCain es el cuadragésimo cuarto [44º] presidente de los Estados Unidos.

- (3) Barack Obama es demócrata.
- (4) Barack Obama es republicano.

- (5) Todas las criaturas con corazón son criaturas con corazón.
- (6) Todas las criaturas con corazón son criaturas con riñones.

- (7) Juan cree que todas las criaturas con corazón son criaturas con corazón.
- (8) Juan cree que todas las criaturas con corazón son criaturas con riñones.

- (9) María cree que Barack Obama es el presidente de los Estados Unidos.
- (10) María cree que John Key es el primer ministro de Nueva Zelanda.

- (13) $2+2=4$.
- (14) Hay infinitos números primos.

- (15) Juan cree que $2+2=4$.
- (16) Juan cree que hay infinitos números primos.

- (17) La hierba es verde.
- (18) La hierba es verde y hay infinitos números primos.

- (19) Juan cree que la hierba es verde.
- (20) Juan cree que la hierba es verde y hay infinitos números primos.

- (21) Clark Kent es Clark Kent.
- (22) Clark Kent es Superman.

- (23) Lois cree que Clark Kent es Clark Kent.
- (24) Lois cree que Clark Kent es Superman.

- (25) Necesariamente, si Aristóteles existe, entonces Aristóteles es Aristóteles.
- (26) Necesariamente, si Aristóteles existe, entonces Aristóteles es el más grande filósofo de la antigüedad.

- (27) Beto cree que el catsup es catsup.
- (28) Beto cree que el catsup es ketchup.

Notas del artículo original

1 Para una discusión de las analogías entre las reglas semánticas y las reglas de etiqueta, ver McKeown-Green, 2002.

2 Esta manera de presentar el argumento presupone que los contenidos pueden tener diferentes referencias en diferentes momentos del tiempo; se trata de una presuposición controvertida, pero cuya discusión nos llevaría demasiado lejos. (Para un argumento contra esa presuposición, ver Richard, 1981). Pero, aunque se tratara de un presupuesto falso, puede desarrollarse un argumento a favor de una semántica de doble índice basándonos en cambios de referencia entre mundos, y no entre momentos del tiempo.

3 De acuerdo al primer tipo de russelliano, no estamos en posición de decidir ninguna disputa sobre la semántica de los nombres hasta que no investiguemos la relación entre la semántica y la pragmática; una tendencia importante en los estudios semánticos de los últimos tiempos ha sido el creciente interés por tratar de dar una explicación del fenómeno lingüístico considerando tanto las explicaciones semánticas como las pragmáticas.

4 Lo que no quiere decir que no haya controversia acerca de la existencia de este tipo de entidades; la hay. (Ver, por ejemplo, el nominalismo en metafísica). El hecho es que muchos creen en la existencia de estos tipos de entidades por razones independientes de la filosofía del lenguaje; es decir, estas entidades no fueron introducidas expresamente para ocupar el lugar de contenido de las expresiones en una teoría semántica.

5 Ver Grice (1957; 1969).

6 Esto es natural para enfoques como el de la postura de Mill-Russell, que ven al significado como algo estrechamente relacionado a la entidad del mundo por la cual está la palabra (o 'a la que la palabra refiere').

7 Horwich, (1998,85-6) sugiere que este tipo de casos pueden resolverse agregando a su análisis una cláusula que determine que para que el hablante signifique [*mean*] (por ejemplo) perro, mediante su uso de 'perro', es suficiente con que esté dispuesto a acudir a un experto, cuyo uso del término se explica por la regularidad canónica (siempre y cuando el uso del hablante no esté 'demasiado alejado' del del experto). Pero es difícil ver cómo la ausencia de tal disposición (quizás producto de un temperamento obstinado, o de simple antipatía a los expertos) pueda hacer que una persona con conocimientos de biología pocos sofisticados signifique [*mean*] mediante su uso de 'perro' algo diferente a lo que significamos cualquiera de nosotros.