

WEB SEMÁNTICA, MARCADORES DISCURSIVOS Y METARREPRESENTACIÓN

Sebastián Bonilla
Universidad Pompeu Fabra

RESUMEN: El proyecto de implementación de inteligencia en la futura Web Semántica dibuja un nuevo escenario tecnológico ideal(izado) en el que la información, organizada ontológicamente y recorrida por agentes inteligentes, comunicativos, ejecutivos, autónomos y adaptativos, deviene conocimiento.

El problema esencial de este planteamiento radica en que el acceso al significado no es un fenómeno representacional (lingüístico-gramatical), sino metarrepresentacional (cognitivo-pragmático).

En este trabajo, se propone el diseño de un agente metarrepresentacional que localice la información relevante guiado por los marcadores discursivos que acotan zonas metadiscursivas de reformulación y de conclusión en los textos. La ventaja que ofrece esta línea de investigación es que permite operar no con la información descontextualizada inscrita en el código del recurso (como se plantea en la propuesta semántica), sino con la información contextualizada y, por ello enriquecida, contenida en el propio discurso.

PALABRAS CLAVE: Web Semántica. Metarrepresentación. Marcadores Discursivos.

ABSTRACT: The project of intelligence implementation in the future Semantic Web draws a new ideal(ized) technological scene in which information happens knowledge, through the ontological organization of information, and by means of the design of communicative, executive, autonomous, and adaptive intelligent agents.

The essential problem of this exposition is that access to meaning is not a representational phenomenon (grammar-linguistic), but metarrepresentational one (pragmatic-cognitive).

In this work, we propose the design of a metarrepresentational agent to locate relevant information guided by discourse markers that limits metadiscursive zones of reformulation and conclusion in texts. The advantage that offers this way of investigation is that it allows to operate not with the descontextualized information insert in the resource code (as considers in the semantic proposal), but the contextualized information and, for that reason enriched, contained in the own discourse.

KEYWORDS: Semantic Web. Metarepresentation. Discourse Markers.

1. WEB SEMÁNTICA, ONTOLOGÍAS Y AGENTES INTELIGENTES.

Como cualquier usuario de Internet habrá advertido por su propia experiencia personal diaria, la web actual está construida con una cantidad ingente de recursos poco estructurados y pobremente definidos. El precio que se paga por ello es la emergencia de una cantidad desproporcionada de información irrelevante.

Una de las causas formales infraestructurales que explican esta situación es que la web actual se basa en el lenguaje HTML. La insatisfacción creciente con respecto a este lenguaje de programación de usuario, que permite codificar hipertextualmente textos, imágenes, sonidos y multimedia, podría sintetizarse en la opinión generalizada entre los especialistas de que el HTML es, simplemente, un lenguaje estructural de maquetación.

En 1999, Tim Berners-Lee, el creador de la World Wide Web, escribió por primera vez acerca de los requisitos hipotéticos que debería de cumplir una futura Web Semántica que facilitase “la implementación del significado, de la inteligencia en la red” (W3C, 1999). En el discurso académico generado recientemente en torno a Internet, se retoma este planteamiento y empieza a detectarse un cierto consenso en torno a la idea de que el futuro de la web se está gestando en la investigación en formas de intelección artificiales cualitativas similares a la humana (Berners-Lee, 2001).

El primer paso importante hacia esa web de naturaleza semántica que conoceremos en el futuro se basa en la adopción generalizada del lenguaje XML (Cover, 1998), incluido desde hace varios años ya en los programas de usuario de creación de páginas web, y que añade a las propiedades acreditadas del HTML la posibilidad de inclusión en el nivel de código de una infraestructura de metadatos que aporte una descripción ontológica explícita de la información contenida en el recurso.

En el contexto de este nuevo planteamiento de una web en la que se haya codificado semánticamente el significado de la información, será posible diseñar *robots* de búsqueda, gestores de contenidos y agentes autónomos que “entiendan” los documentos y realicen procesos “inteligentes” de búsqueda, extracción y tratamiento de la información relevante para el usuario. En palabras de Berners-Lee, “no se está hablando de una inteligencia artificial mágica que permita a los ordenadores entender las palabras de los usuarios, sino únicamente

de la habilidad de una máquina para resolver problemas bien definidos, a través de operaciones bien definidas que se llevarán a cabo sobre datos bien definidos” (W3C, 1999).

La Web Semántica aspira a mejorar sustancialmente la interacción entre los sistemas informáticos y los seres humanos, dotando a los primeros de una mayor inteligencia y autonomía, y a los segundos de un nuevo infouniverso conceptual ergonomizado intelectivamente que satisfaga la necesidad de convertir la información en conocimiento (para una profundización en este planteamiento, véase Codina y Rovira, 2006).

Para entender el cambio que supondrá la creación de la Web Semántica, puede servir como ejemplo de qué manera ayudará a resolver uno de los problemas fundamentales de los actuales buscadores de Internet. En este momento, estos ingenios basan su operatividad casi exclusivamente en las palabras clave contenidas en el código del recurso. Los futuros buscadores semánticos, en cambio, centrarán su eficacia en la información ontológica. Una ontología es una definición formal explícita, y estructurada en taxonomías, de un cuerpo de conocimiento (Gruber, 1993). La hipótesis que subyace en esta nueva estrategia de investigación es que una ontología diseñada para ser procesada automáticamente permitirá, en teoría, simular que un ordenador manipula la información de manera funcionalmente parecida a como un humano entiende el lenguaje. Bien es verdad que una máquina no puede comprender la información que procesa en ningún sentido profundo de la expresión; no obstante, la Inteligencia Artificial podría acercarse un paso más a la humana si los datos que procesa están estipulados semánticamente mediante ontologías. En una Web Semántica estructurada ontológicamente, los buscadores de Internet dejarán de arrojar millones de resultados indiscriminados, la mayor parte de ellos irrelevantes, y ofrecerán algo similar a la información cualitativa que podría haber seleccionado un especialista humano en la materia.

En esta línea de trabajo, se inscribe el proyecto *OntoTag* (Aguado et alii, 2002), un modelo de anotación semántica de textos que (a) cumple con las recomendaciones EAGLES

(1999), fijadas por la Unión Europea, en su iniciativa “Corpus Encoding Standard”, (b) aplica los principios descriptivos del modelo abstracto del mundo (u ontología) de la Semántica Ontológica (Niremburg & Raskin, 2001) (para un estado de la cuestión sobre esta disciplina, véase OntoWeb, 2002), y (c) implementa los lenguajes de marcado de última generación RDF(S)/XML.

OntoTag, al igual que otros proyectos de investigación similares (Benjamins et alii (1999), se plantea como objetivo la anotación semántica de documentos web que pueda ayudar a las computadoras a comprender la información textual en el contexto del proyecto de creación de la Web Semántica. La peculiaridad de *OntoTag* es que aplica la metodología de la Lingüística de Corpus, entendido “corpus” en el sentido de Leech (1997) como “un conjunto de material lingüístico que existe en forma electrónica y que puede ser procesado por un ordenador con distintos fines, como la investigación lingüística y la ingeniería del lenguaje”. En el proyecto, también se aplican las dos modalidades de anotación semántica previstas por McEney & Wilson (2001): (a) las relaciones semánticas entre elementos del texto (agentes, pacientes y participantes en la acción), y (b) las características semánticas de cada una de las palabras que componen el texto.

No obstante, los responsables del modelo *OntoTag* reconocen que “no hay un acuerdo universal en el ámbito de la semántica sobre qué características de las palabras se deben anotar” y que “todavía falta por determinar de manera exhaustiva el conjunto de categorías semántico-cognitivas básicas”, y reconocen, como principal inconveniente para su proyecto, las limitaciones impuestas por el estadio actual de la tecnología: “el proceso de obtener automáticamente páginas compactas, legibles y verificables es una tarea de muy difícil delimitación y especificación en toda su magnitud; por otro lado, la inclusión del nivel de anotación semántica en un documento web conlleva un evidente (aunque no exagerado) aumento del tiempo de descarga del documento desde la red”.

A estos inconvenientes señalados por los propios responsables del proyecto *OntoTag*, podrían añadirse dos reflexiones críticas de más profundo calado. La primera es el peligro de obsolescencia que sufre la tecnología de anotación, marcaje y etiquetaje de corpus, a causa de la continua evolución de los estándares. La segunda es que la tarea de anotar semánticamente la totalidad de la Web, de convertirla en un corpus marcado y etiquetado ontológicamente, dada su magnitud y mutabilidad, parece un empeño inasequible (Bonilla, en preparación a).

Dejando a un lado las dificultades actuales que plantea la construcción de una web enriquecida ontológicamente, el paso siguiente del proceso que culminaría en la Web Semántica sería hacer factible el diseño de productos informáticos como los *agentes inteligentes*, que gestionarán las necesidades de los usuarios humanos de manera ejecutiva, autónoma y adaptativa.

El agente inteligente ideal (Hendler, 1999), auténtico reto para los investigadores en Ciencias de la Computación, en primer lugar, será *comunicativo*, es decir, interactuará de una manera fluida con los objetivos y las preferencias personales del usuario. En segundo término, será *ejecutivo*, estará capacitado para tomar decisiones y no únicamente para presentarle al usuario varias opciones para que éste elija una. Tercero, será *autónomo*, capaz de actuar sin que el usuario tenga que controlarlo exhaustiva y continuamente. Y en última instancia, será *adaptativo*, podrá aprender de su propia experiencia de funcionamiento y de las preferencias idiosincrásicas de su usuario.

Formulado en otros términos, los agentes inteligentes se harán cargo de una manera competente del trabajo más rutinario y, en ocasiones, físicamente inabarcable, que en la actualidad se ven obligados a realizar manualmente los usuarios que navegan e interactúan con la web. Hendler (1999) enuncia así el optimismo de los especialistas al respecto: “En pocas palabras, si en estos momentos no está usted utilizando tecnología basada en agentes inteligentes, no se preocupe, porque muy pronto lo hará”.

La Web Semántica dibuja un nuevo escenario tecnológico ideal(izado) en el que la información, organizada ontológicamente y recorrida por agentes inteligentes, comunicativos, ejecutivos, autónomos y adaptativos, deviene conocimiento.

2. SEMÁNTICA Y SIGNIFICADO: ALGUNOS PROBLEMAS DE IMPLEMENTACIÓN COMPUTACIONAL.

En el diseño de programas para representar y recuperar información, los expertos en Computación orientada hacia la web están exprimiendo y llevando casi al límite la capacidad de aplicación de los modelos lingüísticos actuales basados en reglas formales en el campo de la Morfología y de la Sintaxis. La tendencia más reciente de investigación, en la que se cifra la esperanza de dar un salto cualitativo en la disciplina, gira en torno a la Semántica.

Desde la perspectiva que ofrece el desarrollo reciente de la Lingüística, el punto de vista semántico, aplicado en el campo interdisciplinario de la web, quizás está creando unas expectativas excesivamente entusiastas, sin duda alentadas por la capacidad de fascinación que ejerce las palabras *semántica* y *significado*, sin duda a causa de que está empleándose la expresión *Web Semántica* no en un sentido estrictamente técnico, sino como una metáfora afortunada, como un hallazgo terminológico, como si fuera un *meme* (Dawkins, 1976).

No parece ocioso, por lo tanto, revisar críticamente de manera breve qué es la Semántica, en qué consiste el significado y en qué medida pueden aplicarse ambos conceptos en el desarrollo futuro de la web.

Como señalan Leech (1974) y Lyons (1977) en sus manuales clásicos de la disciplina, la teoría semántica cubre un amplísimo espectro de cuestiones relacionadas con el significado,

la interpretación y la comprensión del lenguaje; pero si ambos especialistas tuvieran que consensuar una definición general minimalista, ésta sería aproximadamente: “la Semántica estudia las múltiples relaciones de significación que establecen los signos con sus referentes”.

En cualquier caso, y aquí está la clave que quizás haya deslumbrado a la vanguardia de la Inteligencia Artificial aplicada en la web, la idea central de la Semántica (que requiere ser revisada críticamente) es que las palabras codifican significados.

En la lógica en que se basa el proyecto de una Web Semántica, si pudiera reconstruirse la ontología de una palabra, entonces se habría conseguido formalizar su significado y, mediante un protocolo de implementación computacional, este significado podría programarse en el código del recurso, para integrarlo luego operativamente en la web. Desde este punto de vista un tanto optimista, la Inteligencia Artificial habría encontrado en las ontologías una posible vía de acceso eficaz hacia la simulación de la comprensión humana del lenguaje.

Desde nuestro punto de vista, el problema esencial de este planteamiento radica en que el significado, como se argumentará más adelante, no está en las palabras, sino en la mente de quien las procesa. Una de las tesis básicas de este artículo es que el significado no es un fenómeno representacional (lingüístico-gramatical), sino que posee una naturaleza metarrepresentacional (cognitivo-pragmática).

Los diccionarios y las enciclopedias, auténticas proezas culturales que sintetizan lingüísticamente buena parte del saber humano, generan la ilusión de que el significado está en las palabras, de que el significado consiste, en última instancia, en un conjunto estructurado (es decir, en una ontología) de unas palabras que sirven para definir a otras palabras.

Adviértase que la implementación de la Semántica en la web crearía un bucle autorreferencial irresoluble (Hofstadter, 1979): palabras que remiten a su vez a otras palabras

que remiten a su vez a otras palabras... Si el significado de una palabra está en otras palabras: ¿dónde está el significado de esas “otras palabras” y así *ad infinitum*?

A la propuesta semántico-computacional, puede oponérsele también una contraargumentación biológica: una lesión cerebral localizada en una determinada región neuronal puede impedir que una persona acceda al significado del lenguaje. Las palabras estarán físicamente disponibles en la voz que las pronuncia o en las letras impresas de los libros, pero el lesionado no podrá comprenderlas por sí mismas, en sí mismas, en modo alguno. El significado no está en las palabras, se genera en la mente y tiene un soporte biológico.

La singularidad de la cognición humana se explica por el funcionamiento del cerebro, capaz de procesar información mediante trillones de sinapsis tridimensionales. Desde una perspectiva estrictamente funcionalista, los expertos más optimistas en Inteligencia Artificial consideran que esta ingente capacidad de procesamiento complejo estaría en el origen de la autoconciencia, de la intencionalidad, de la comprensión, etc. Desde este punto de vista, bastaría con que se fuese cumpliendo la ley de Moore de multiplicación progresiva de la potencia de cómputo de los procesadores sintéticos, y cuando se acerquen o iguallen la capacidad humana de procesamiento, simplemente la Inteligencia Artificial dará el salto a la autoconciencia, a la intencionalidad, a la comprensión: ¿no se dio este salto a la humanidad desde la biología de la animalidad? ¿por qué no podría suceder lo mismo desde el mundo de lo artificial?

En oposición a estas ideas todavía especulativas, parece sólidamente establecido que el significado es un fenómeno mental que remite a una capacidad cognitiva humana, adquirida genéticamente por nuestra especie a lo largo de millones de años de selección evolutiva. Por lo tanto, la semántica y el acceso al significado tienen una base biológica ligada a la actividad intelectual de un organismo vivo.

No puede olvidarse que el proyecto de Web Semántica es telemático, y no puede ser de otra manera, es decir, su funcionamiento requiere la interacción entre mentes asistidas por ordenadores interconectados en red. Por lo tanto, el significado se procesa en la mente, no en la red misma, y, en consecuencia, no hay posibilidad alguna actualmente de que la web desarrolle una inteligencia autónoma, basada en la gestión del significado, independiente de sus gestores y usuarios humanos.

No obstante, en el horizonte de la investigación futura, uno de los objetivos prioritarios es replicar computacionalmente la comprensión humana del lenguaje; pero quien esto escribe es pesimista con respecto a que las soluciones procedan únicamente de la semántica (véase la propuesta explorativa de naturaleza metarrepresentacional que se presenta en el apartado siguiente).

En estos momentos de la investigación, el problema irresoluble (el acceso computacional a la comprensión) que plantea la implementación del significado en una hipotética Web Semántica puede emplearse como confirmación de las palabras de Searle (2001): “el término Inteligencia Artificial ha sido siempre un error. Es mejor hablar de simulación cognitiva”. Formulado en otros términos, la complejidad casi insalvable de la aplicación de la Semántica en las disciplinas computacionales hace evidente el hecho de que *Inteligencia Artificial* es un meme, una expresión metafórica afortunada que estimula la imaginación proyectiva de los investigadores, pero que dista de ser una realidad empírica. ¿Es la semántica, el significado, lo que realmente necesita la Inteligencia Artificial para dar un salto cualitativo que la acerque a una menor distancia de la humana?

3. METARREPRESENTACIÓN Y MARCADORES DISCURSIVOS. UNA PROPUESTA DE IMPLEMENTACIÓN EN LA WEB SEMÁNTICA.

Pensemos el problema computacional de la implementación del significado en la web desde una perspectiva especulativa diferente de la semántica.

El antropólogo cultural y pragmatista Sperber (2000) propone la hipótesis de que los seres humanos poseemos un mecanismo cognitivo, fruto de nuestra evolución biológica, que nos permite metarrepresentar mentalmente las representaciones lingüísticas. Una metarrepresentación es una representación intelectual de alto nivel del contenido de una representación lingüística. Como ventaja evolutiva crucial, nuestra especie con habilidades metarrepresentacionales puede vivir e interactuar en un mundo donde no sólo hay hechos físicos, sino también hechos mentales.

En una formalización previa de este planteamiento, Wilson y Sperber (1990) sugieren la coexistencia en el lenguaje humano de dos tipos complementarios de información; por un lado, el discurso contiene *información conceptual* (de naturaleza semántica), y por otro, transporta *información computacional* (de naturaleza pragmática y metadiscursiva), es decir, información metarrepresentacional sobre cómo ha de procesarse el contenido semántico-conceptual del discurso; o dicho con otras palabras, instrucciones computacionales de interpretación lingüística (para una profundización en este planteamiento, véase Wilson, 2000 y No, 2000; para una aplicación en el español, véase Escandell, 1989 y Portolés, 2005).

Más en concreto, Wilson y Sperber analizan el funcionamiento de determinados elementos lingüísticos con características computacionales que influyen directamente en la manera en que se procesa y se interpreta la información discursiva, mostrando un interés especial en los *marcadores discursivos*, ya que, en primer lugar, funcionan a modo de estímulo ostensivo (es decir, como un elemento focalizador utilizado de manera intencionada por el comunicador para controlar la atención de su interlocutor), y en segundo lugar, guían el

proceso de interpretación discursiva, influyendo de manera positiva en la minimización del esfuerzo cognitivo que ha de invertir el receptor en la comprensión del significado.

El interés en el funcionamiento de esta clase de elementos en concreto, y, en general, de la información pragmática de tipo computacional que contiene el discurso, desborda el campo especializado de la Lingüística y se extiende a otras áreas de la Inteligencia Artificial. Sirva como ilustración el ejemplo siguiente.

Existe una demanda creciente de programas informáticos que simulen la inteligencia de un ser humano competente en la realización de *resúmenes* y que sean capaces de generar mapas conceptuales jerarquizados de la información relevante contenida en un cuerpo de conocimiento determinado.

Los programas disponibles en el mercado que sumarizan textos de manera automática suelen operar aplicando índices de frecuencias estadísticas de aparición de palabras (que, por otra parte, es la tarea que, hoy por hoy, los ordenadores realizan más rápidamente y mejor). Para que la computadora arroje datos significativos en este nivel de funcionamiento, los programas excluyen del recuento las “palabras vacías”, invariables y sin contenido proposicional inventariable léxico-semánticamente (artículos, pronombres, conjunciones, preposiciones, marcadores, conectores, etc.), ya que, se supone (en nuestro planteamiento, creemos que erróneamente) que no aportan información demasiado interesante al texto. Cuando se pone en marcha, el programa así diseñado identifica las cadenas de caracteres que más se repiten como las palabras claves, como las ideas principales del texto.

Si bien se trata de un planteamiento en apariencia sensato y basado en el sentido común, semejante diseño condena al programa a un funcionamiento en muchas ocasiones extravagante: sabido es que los textos bien contruidos poseen un elevado grado de variación estilística, sinonimia y elisión, y un bajo nivel de repetición. Estas características lingüísticas

elusivas del tratamiento estadístico provocan que el programa desatine continuamente sus decisiones de sumarización.

Ahora bien, si siguiese el planteamiento metarrepresentacional, bastaría con que el programa de resumen automático tuviese en cuenta los marcadores discursivos reformulativos o conclusivos (del tipo *en resumen*, *por lo tanto*, etc.), inscritos en la superficie lingüística por el propio productor del texto para facilitar y controlar la trayectoria interpretativa del receptor, para que, sin duda alguna, su nivel de eficacia traspasase con creces el umbral de lo aceptable.

La idea de que el lenguaje humano posee no únicamente información de naturaleza conceptual, sino también computacional y metarrepresentacional puede ser muy fructífera en la creación de programas que simulen artificialmente la inteligencia.

La mente procesa la información lingüística en tiempo real, por lo tanto, no puede detenerse en el procesamiento morosamente detallado de todos y cada uno de los niveles (fonético, fonológico, morfológico, sintáctico, léxico y semántico) que propone la Lingüística Computacional clásica en sus exhaustivos análisis. Las últimas teorías sobre la inteligencia humana (Hawkins & Blakeslee, 2004) sugieren que la mente no funciona mediante complejos sistemas formales de reglas, sino a través de patrones de memoria y de estrategias eficaces de procesamiento. Guiada por su capacidad innata metarrepresentativa, la mente coge el atajo pragmático-computacional hacia la comprensión (Bonilla, en preparación b).

Ha llegado el momento de realizar nuestro esbozo de propuesta explorativa de naturaleza metarrepresentacional. Imaginemos un ejemplo de aplicación –con cierta exigencia- en la futura Web. Un usuario lanza contra la red una pregunta de investigación, verbigracia, “El estado actual de los estudios en Teoría de Supercuerdas”. Lo que esperaría obtener un usuario exigente cuando realiza su consulta, más que una lista inacabable de páginas web cuyo único nexo en común, en ocasiones, es contener la palabra clave “supercuerda”, es un texto con una estructura informativa clara y que responda de una manera

competente a la cuestión planteada; es decir, lo mismo que esperaría si realizase la consulta a un especialista humano en la materia.

Un agente metarrepresentacional, no basado en la información semántica ontológica, que busque estratégicamente la información relevante para la resolución de la pregunta de investigación, desestimaré todos los recursos de información publicados en una fecha anterior a cinco años. No obstante, exploraré, en la zona de referencias de la bibliografía desestimada, aquellas obras y autores más citados, para establecer cuáles son los trabajos clásicos en los que se asienta la materia, así como las citas literales que más se repiten en la bibliografía más reciente. De esta manera, se obtendría un estado de la cuestión sintético, que podría encabezar la respuesta a la pregunta inicial de investigación.

En paralelo, el agente metarrepresentacional localizará los marcadores discursivos reformulativos, para documentar en los textos seleccionados cuáles son los conceptos fundamentales que reciben una atención privilegiada. De esta suerte, habrá encontrado las zonas de máxima relevancia en las que el texto habla de sí mismo, se hace explícito, se metarrepresenta. La ventaja que tiene este procedimiento, frente a la consideración de las palabras clave del lenguaje HTML, o de los metadatos en el lenguaje XML que aparecen en el código del recurso, es que los conceptos aparecen en su contexto discursivo, con lo que se enriquece enormemente su aportación a la resolución del problema. En este sentido, se habrían identificado y contextualizado los conceptos fundamentales que protagonizan la respuesta temática a la pregunta inicial de investigación.

También en paralelo, el agente metarrepresentacional localizará en los textos seleccionados los marcadores conclusivos que contienen la síntesis de la información más relevante, zonas discursivas, no se olvide, etiquetadas por el propio productor textual. La ventaja que tiene este procedimiento con respecto a la consideración de los sumarios inscritos en el código del recurso (prototípicos del lenguaje XML), de nuevo es su contextualización

discursiva, que sin duda aumentará la sensación de que estamos operando con discurso inteligente y formulado lingüísticamente de una manera impecable.

Con todo este material, el agente metarrepresentacional construirá un texto coherente que entregará en pantalla al usuario. Este documento tendrá la estructura siguiente:

1. Título: La pregunta inicial de investigación.
2. Síntesis del estado de la cuestión: Qué discurso y quiénes lo han enunciado, asentando las bases de la materia. Procedimiento operativo: Búsqueda de coincidencias en la zona de relevancia discursiva etiquetada como “referencias bibliográficas” y sus variantes estilísticas.
3. Propuesta temática: Identificación de los conceptos clave contextualizados discursivamente. Procedimiento operativo: Búsqueda de los marcadores reformulativos que maximizan la relevancia de las zonas metadiscursivas en que el discurso se explica a sí mismo.
4. Conclusiones: Localización de las zonas discursivas en que se establecen las conclusiones de investigación. Procedimiento operativo: Búsqueda de los marcadores conclusivos que acotan las zonas de síntesis en que el autor ha metarrepresentado su propio texto.

4. CONCLUSIONES

El proyecto de implementación de inteligencia en la futura Web Semántica dibuja un nuevo escenario tecnológico ideal(izado) en el que la información, organizada ontológicamente y recorrida por agentes inteligentes, comunicativos, ejecutivos, autónomos y adaptativos, deviene conocimiento.

Desde nuestro punto de vista, el problema esencial de este planteamiento radica en que el significado no es un fenómeno representacional (lingüístico-gramatical), sino metarrepresentacional (cognitivo-pragmático): el significado es un fenómeno mental que remite a una capacidad cognitiva humana, adquirida genéticamente por nuestra especie a lo largo de millones de años de evolución selectiva darwiniana. Por lo tanto, la semántica y el acceso al significado tienen una base biológica ligada a la actividad intelectual de un organismo vivo.

Si se piensa en la resolución del problema computacional de la implementación del significado (de la inteligencia) en la web desde una perspectiva especulativa diferente de la semántica, la idea de que el lenguaje humano posee información de naturaleza metarrepresentacional puede ser muy fructífera en la creación de programas que simulen artificialmente la inteligencia.

En este sentido, podría diseñarse un agente metarrepresentacional que localizara estratégicamente la información relevante guiado por los marcadores discursivos que acotan zonas metadiscursivas de reformulación y de conclusión. La ventaja que ofrece esta línea de investigación es que permite operar no con la información descontextualizada inscrita en el código del recurso (como se plantea en la propuesta semántica), sino la información contextualizada y, por ello enriquecida, contenida en el propio discurso.

NOTA

Este trabajo se enmarca en el proyecto de investigación “Web Semántica y Sistemas de Información Documental” (HUM2004-03162/FILO), dirigido por el Dr. Lluís Codina, y financiado por el Ministerio de Ciencia y Tecnología (Programa nacional de Tecnologías de la Información y de las Comunicaciones).

REFERENCIAS

- Aguado, G. et alii. 2002. "A Semantic Web Page Linguistic Annotation Model. Semantic Web Meets Language Resources". Technical Report WS-02-16. American Association for Artificial Intelligence. California: AAAI Press.
- Benamins, V.R., et alii. 1999. "Building Ontologies for the Internet: A Mid Term Report". *International Journal of Human Computer Studies* 51: 687–712.
- Berners-Lee, T. 2001. "The Semantic Web", *Scientific American*.
- Bonilla, S. (en preparación a). "Marcaje de corpus y Metarrepresentación".
- Bonilla, S. (en preparación b). "Metarrepresentación".
- Codina, Ll. y C. Rovira. 2006. "La Web Semántica". En Tramullas, J. coord. *Tendencias en documentación digital*. 9-54. Gijón: Trea.
- Cover, R. 1998. "XML and Semantic Transparency"
<http://xml.coverpages.org/xmlAndSemantics.html>
- Dawkins, R. 1976. *El gen egoísta*. Barcelona: Salvat.
- EAGLES. 1999. *EAGLES LE3-4244: Preliminary Recommendations on Semantic Encoding, Final Report*.
<http://www.ilc.pi.cnr.it/EAGLES/EAGLESLE.PDF>
- Escandell, V. 1998. "Metapropositions as metarepresentations". Paper delivered to the *Relevance Theory Workshop*. Luton.
- Gruber, T.R. 1993. "A translation Approach to Portable Ontology Specifications". *Knowledge Acquisition* 5.2: 199-220.
- Hawkins, J. & S. Blakeslee. 2004. *On Intelligence*. New York: Times Book.
- Hendler, J. 1999. "Is there an Intelligent Agent in Your Future?"
<http://www.nature.com/webmatters/agents/agents.html>].

- Hofstadter, D. 1987 (1979). *Gödel, Escher, bach. Un Eterno y Grácil Bucle*. Barcelona: Tusquets.
- Leech, G. 1974. *Semantics*. London: Penguin.
- Leech, G. 1997. "Introducing corpus annotation". In Garside R., Leech, G. y McEnery, A. M. eds. *Corpus Annotation: Linguistic Information from Computer Text Corpora*. London: Longman.
- Lyons, J. 1977. *Semantics*. Cambridge: Cambridge University Press.
- McEnery, A. M. y Wilson, A. 2001. *Corpus Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Nirenburg, S. & Raskin, V. 2001. *Ontological Semantics*.
<http://crl.nmsu.edu/Staff.pages/Technical/sergei/book/index-book.html>
- No, E.J. 2000. *Metarepresentation. A Relevance-Theory Approach*. Amsterdam: John Benjamins.
- OntoWeb. 2002.
http://ontoweb.aifb.uni-karlsruhe.de/About/Deliverables/D13_v1-0.zip
- Portolés, J. 2005. "Marcadores del discurso y metarrepresentación". En Casado, M. et alii. eds. *Estudios sobre lo metalingüístico (en español)*. 25-46. Berlin: Peter Lang.
- Searle, J. 2001. *Rationality in Action*. Massachusetts: MIT Press.
- Sperber, D. 2000. "Metarepresentations in an evolutionary perspective". In D. Sperber. ed. *Metarepresentations: A Multidisciplinary Perspective*. 117-137. Oxford: Oxford University Press.
<http://perso.club-internet.fr/sperber/metarep.htm>
- W3C. 1999
<http://www.w3c.org/2001/sw>

Wilson, D. 2000. "Metarepresentation in linguistic communication". In D. Sperber. ed.
Metarepresentations. Oxford: Oxford University Press.

<http://www.phon.ucl.ac.uk/home/robyn/workshop/papers/wilson.htm>

Wilson, D. & D. Sperber. 1990. "Linguistic Form and Relevance". *UCL Working Papers in
Linguistics* 2: 95-112.