

SENSEVAL, una aproximación computacional al significado



Mariona Taulé

Profesora del Departamento de Lingüística General (Universidad de Barcelona) Consultora de los Estudios de Humanidades y Filología (UOC)
mtaule@pcb.ub.es



M. Antònia Martí

Profesora del Departamento de Lingüística General (Universidad de Barcelona)
amarti@lingua.fil.ub.es

Resumen:

El propósito de este artículo es presentar de manera general el certamen internacional SENSEVAL cuyo objetivo básico es la evaluación objetiva de técnicas, métodos y sistemas de desambiguación semántica automática. Con este fin, primero se da una presentación global de SENSEVAL, sus objetivos y organización, a continuación se expone la metodología de evaluación desarrollada, consistente en la definición de tareas y la preparación de los recursos lingüísticos necesarios y, finalmente, se analizan las repercusiones en el ámbito lingüístico y lexicográfico que hay detrás de esta metodología. Para ejemplificar la metodología nos basaremos en los recursos desarrollados para el SENSEVAL-2 del español.

1. Introducción

Uno de los problemas centrales y más difíciles de tratar en las distintas aplicaciones de procesamiento del lenguaje natural (en adelante, PLN) es, sin duda, la ambigüedad, una propiedad intrínseca y característica del lenguaje. Desde un punto de vista lingüístico, se entiende por ambigüedad, aquel fenómeno que se da cuando una misma forma lingüística se puede interpretar de más de una manera distinta. Cuando la ambigüedad está motivada por los múltiples significados de una misma palabra hablamos entonces de ambigüedad semántica y, en concreto, de ambigüedad léxica. Este es el caso, por ejemplo, de la palabra *partido* que puede significar básicamente tres cosas distintas: a) un nombre que designa una organización política ("Militaba en un partido político"), b) un nombre que designa una prueba deportiva ("El partido de fútbol terminó con un empate"), y, por último, c) el participio del verbo *partir* que indica que algo está dividido o separado ("El buque ha quedado partido en dos"). Como se puede observar, según la categoría y, en especial, según el contexto en que aparece la palabra adquiere uno u otro significado. En este sentido, no se puede tratar el significado de las palabras de una manera aislada, porque habitualmente es el contexto el que perfila, concreta o modula su significado y es a partir de éste que se puede obtener el sentido pleno o más apropiado. Cuando hablamos de contexto nos referimos tanto al contexto sintáctico inmediato, a la oración en la que ocurre la palabra, a las palabras con las que se combina, como al

contexto más amplio, por ejemplo el párrafo o texto en el que se encuentra la palabra.

La ambigüedad léxica no supone, en principio, y exceptuando casos muy concretos, ningún problema para los humanos, pero sí, en cambio, resulta de una extraordinaria complejidad para los sistemas de tratamiento automático del lenguaje y, en especial, cuando lo que se debe determinar es el significado de una expresión lingüística, es decir de una determinada palabra o secuencia de palabras. La desambiguación semántica no es una tarea en absoluto sencilla. Para un programa, cualquier forma (o palabra) a la que se le pueda asignar más de una interpretación, ya sea morfosintáctica o semántica, es una entidad en principio ambigua, porque el programa deberá decidir cual de las posibles interpretaciones es la más apropiada, eso sí, siempre teniendo en cuenta el contexto en el cual aparece. Si no partimos del contexto, el programa o módulo de desambiguación no tiene ningún sentido. Pues bien, los desambiguadores morfosintácticos (*taggers*) y los sistemas de desambiguación semántica (*Word Sense Disambiguation*) son los que se encargan de decidir y asignar la correcta interpretación morfosintáctica o semántica, respectivamente, de una palabra concreta. Cuando la desambiguación se realiza a nivel de categoría léxica se utiliza un *tagger* y cuando la desambiguación implica determinar el sentido o significado más probable de una determinada palabra entonces se utiliza un desambiguador semántico (o léxico). En general, el proceso de desambiguación semántica se realiza una vez analizado y desambiguado morfosintácticamente el texto, ya que de este modo se reduce la ambigüedad semántica al plano de categoría. Por ejemplo, en el caso de la palabra *partido*, el desambiguador morfosintáctico distingue entre las formas nominales y las formas de participio y, en consecuencia, reduce la ambigüedad de dicha palabra a aquellas ocurrencias con categoría nombre. Siguiendo con el ejemplo, el desambiguador semántico sólo tendrá que decidir, en el caso de las formas nominales, entre dos sentidos posibles ("partido político" o "prueba deportiva") y no tres. Por lo tanto, la aplicación previa del desambiguador morfosintáctico reduce considerablemente las posibilidades de asignación semántica de sentidos.

Otro aspecto a tener en cuenta en el proceso de desambiguación semántica es que para interpretar el significado de una oración es necesario, previamente, determinar el significado de cada una de las palabras que la conforman. Por lo tanto, para resolver la semántica de la oración hay que resolver primero la ambigüedad léxica.

El problema de la resolución de la ambigüedad semántica (o léxica) es un problema subyacente en muchas de las aplicaciones de PLN, especialmente en los sistemas de traducción automática y de recuperación de la información, un problema que no es nuevo, que se planteó desde el origen del tratamiento automático del lenguaje (Ide y Véronis, 1998). En general, todos estos sistemas incorporan un módulo, más o menos efectivo, de desambiguación semántica para poder obtener resultados de análisis óptimos. Pero recientemente el interés por este tipo de técnicas o de sistemas se ha visto considerablemente incrementado, y posiblemente se deba en gran parte a la disponibilidad cada vez mayor de corpus de grandes dimensiones, que permiten aplicar métodos estadísticos para tratar la desambiguación semántica y que, normalmente, resultan muy efectivos. Los corpus, colecciones de texto en formato electrónico, se convierten en extraordinarios repositorios de información a partir de los cuales se pueden encontrar, obtener y, por lo tanto, aprender los múltiples contextos en los que puede aparecer una determinada palabra, convirtiéndose, así, en una fuente de información fundamental para los sistemas de desambiguación.

La tendencia general es el desarrollo de sistemas de desambiguación más genéricos, es decir independientes de la aplicación, cuyo objetivo se centra en resolver la asignación de sentidos. Esta proliferación de sistemas ha propiciado la creación del concurso SENSEVAL (acrónimo de *Sense Evaluation*), que permite comparar y evaluar los resultados obtenidos por los diversos sistemas de desambiguación semántica automática de una manera objetiva. Para poder realizar esta evaluación objetivamente es necesario desarrollar una metodología de evaluación que permita contrastar en igualdad de condiciones los distintos sistemas. Una metodología que implica, en cierto modo, una revisión exhaustiva de la definición de sentido y de cómo discriminar o distinguir los sentidos básicos de las palabras. Es decir, una metodología que también ha tenido una repercusión importante en planteamientos fundamentales de lexicografía computacional y de lingüística en general.

El objetivo de este artículo es presentar la metodología desarrollada en SENSEVAL, explicar por qué se ha convertido en el referente o marco imprescindible para los sistemas de desambiguación semántica y reflexionar sobre las repercusiones en los ámbitos lingüístico y lexicográfico que hay detrás de esta metodología. Para presentarlo hemos organizado el texto en tres partes: en la primera, se da una presentación global de SENSEVAL, sus objetivos y organización, a continuación, se expone la metodología de evaluación desarrollada y, por último, las repercusiones que ha tenido en la lingüística y la lexicografía. Para ejemplificar la metodología nos basaremos en los recursos desarrollados para el SENSEVAL-2 del español.^[1]

2. ¿Qué es SENSEVAL?

SENSEVAL^[2] es un certamen internacional sobre desambiguación semántica automática (Kilgarrif y Palmer, 2000; Edmonds y Cotton, 2001), que se celebra bianualmente desde 1998 y cuyo objetivo principal es la evaluación objetiva de sistemas, métodos y técnicas que determinan automáticamente el significado de una palabra en un contexto. La evaluación se realiza para diferentes palabras, diferentes variedades lingüísticas y diferentes lenguas. En este marco de intercambio científico no sólo se trata de explorar, de contrastar y comparar los aspectos tanto teóricos como técnicos de los distintos programas diseñados para la desambiguación semántica de sentidos (*Word Sense Disambiguation*, WSD), sino que también se ha generado, gracias a esta iniciativa, toda una reflexión sobre el léxico, sobre cómo hay que tratar, analizar y, en definitiva, representar el significado de las palabras, y, por lo tanto, se plantea toda una problemática entorno de la polisemia desde una perspectiva experimental muy interesante y novedosa.

El primer certamen SENSEVAL^[3] tuvo lugar en 1998 y las lenguas participantes fueron el inglés, el francés y el italiano. La metodología desarrollada permitía evaluar los sistemas de desambiguación exclusivamente en función de una muestra léxica, es decir, había que determinar automáticamente el sentido de una única palabra en un contexto determinado. Con SENSEVAL-2,^[4] las lenguas participantes se incrementaron hasta un total de 12 (inglés, francés, italiano, español, vasco, danés, sueco, holandés, estonio, checo, chino y japonés) al igual que también se incrementaron el tipo de tareas posibles en las que podían concursar los distintos participantes. En este sentido, se diseñaron tres tipos de tareas que básicamente se diferenciaban entre sí por el número de palabras que los programas tenían que desambiguar. Con la finalidad de entender mejor la metodología que se sigue en SENSEVAL, presentamos a continuación de manera más detallada en qué consiste cada una de estas tareas y los recursos desarrollados para poder realizarlas.

3. Metodología

Los diferentes algoritmos y sistemas de desambiguación semántica que participan en SENSEVAL tienen por objetivo resolver el problema de determinar automáticamente el sentido de las palabras sometidas a examen, en un contexto determinado, es decir, decidir y asignar a cada palabra el significado más probable en función del contexto sintáctico en el que aparece.

Para poder realizar una evaluación objetiva, primero es necesario desarrollar una metodología de evaluación en la que se definan explícitamente las diversas tareas en las que pueden participar los distintos programas y a partir de las cuales comparar los resultados y evaluarlos. En segundo lugar, hay que proporcionar a los participantes los datos necesarios, corpus y léxicos, para poder realizar de manera objetiva la evaluación. Se entiende que se trata de los

1. Los recursos desarrollados para la realización de SENSEVAL-2 han sido posibles gracias a la colaboración desinteresada de los grupos de investigación del Centro de Lenguaje y Computación de la Universidad de Barcelona (CLIC-UB), del centro de Técnicas y Aplicaciones del Lenguaje y el Habla (TALP) de la Universidad Politécnica de Cataluña (UPC) y del Grupo de Procesamiento de Lenguaje Natural (GPLN) de la Universidad Nacional de Educación a Distancia (UNED). Estos tres grupos se han encargado de la organización del SENSEVAL-2 español.

2. La organización de SENSEVAL^[url1] se realiza con la colaboración desinteresada de diferentes grupos de investigación.

3. El primer ejercicio de SENSEVAL, y de hecho también de su homónimo ROMANSEVAL (Véronis, 1998; Calzolari y Corazzari, 2000), tuvo lugar en 1998, bajo la supervisión del grupo SIGLEX, y culminó en un taller (*workshop*) en Hertsmonceux Castle, Inglaterra. SIGLEX^[url2] (Special Interest Group of the Association for Computational Linguistics) provee un marco de referencia para la investigación en lexicografía, semántica léxica computacional y otras áreas afines.

4. SENSEVAL-2 se organizó en el 2000-2001 con el apoyo de EURALEX, ELSNET, EPSRC y ELRA, y tuvo lugar en Toulouse conjuntamente con el congreso de la ACL-2001.

misimos datos para todos los sistemas que participan en la desambiguación de una misma lengua. Finalmente, hay que elaborar también un programa que contabilice, compare y evalúe los resultados obtenidos por los diferentes participantes (Melamed y Resnik, 2000).

3.1. Tipos de tareas

En SENSEVAL-2^[5] se diseñaron tres tipos de tareas para la evaluación de los sistemas de desambiguación: tarea basada en una muestra léxica (*lexical sample task*), tarea léxica completa (*all-words task*) y tarea de traducción (*translation task*).

La tarea basada en la muestra léxica sólo se fija en una única palabra por frase, mientras que en la tarea léxica completa, los sistemas tienen que desambiguar semánticamente todas las palabras, exceptuando las funcionales (conjunciones, preposiciones, artículos, etc.), que aparecen en las frases de los corpus seleccionados. La tarea de traducción, en la que sólo ha participado la lengua japonesa, es de hecho un subtipo de muestra léxica porque sólo hay que desambiguar una única palabra; la diferencia es que el sentido de la palabra se define de acuerdo a su traducción.

En definitiva, en la tarea léxica completa se evalúa la capacidad del sistema automático de desambiguar cada una de las palabras (u ocurrencias) que aparecen en el corpus que se proporciona. En las tareas basadas en una muestra léxica se evalúa a los sistemas en función de la desambiguación de las diferentes ocurrencias de una muestra de palabras previamente seleccionadas en el contexto de una frase.

En la tabla siguiente se muestra de manera esquemática las tareas en las que han participado las distintas lenguas y el número de equipos participantes (entre paréntesis):

TAREAS	
Tarea basada en una muestra léxica	Español (5), vasco (2), inglés (15), italiano (2), japonés (3), coreano (2), sueco (5)
Tarea de traducción	Japonés (8)
Tarea léxica completa	Checo (1), estonio(2), holandés ⁱ , inglés (12)

ⁱ En el caso de la lengua holandesa se prepararon los datos pero no estuvieron disponibles para el concurso

Tabla 1. Tareas en las que participan las distintas lenguas

Para poder llevar a cabo dichas tareas es necesario disponer de dos tipos básicos de datos: un diccionario (o léxico) y un corpus etiquetado manualmente que servirá como *gold-standard*, y a partir del cual se creará el corpus de entrenamiento y el corpus de evaluación.^[6] Todos los sistemas que concursan tienen que analizar semánticamente el mismo corpus y, en consecuencia, es necesario que cada lengua participante disponga de este corpus y de un diccionario en el que se incluyan los distintos significados de las palabras que se deben desambiguar.

5. A partir de ahora se expone la metodología desarrollada y los datos y resultados obtenidos en SENSEVAL-2 porque abarca más lenguas y más tareas. De hecho, la metodología del primer concurso SENSEVAL queda incluida en el segundo.

6. En Rigau *et al.* (2001) y en Taulé (2002) se da una descripción completa de los recursos lingüísticos elaborados para SENSEVAL-2.

3.1.1. Léxico

En el caso de la tarea basada en una muestra léxica, el léxico que se proporciona tiene como requisito contener el conjunto de palabras seleccionadas para el ejercicio.^[7] Para la tarea léxica completa evidentemente se precisa de un léxico extenso. En cualquier caso se trata de léxicos disponibles en formato electrónico.

A continuación, se dan como ejemplo dos entradas, *grano* y *popular*, del léxico^[8] elaborado expresamente para que la lengua española pudiera participar en la tarea de muestra léxica.

grano#NCMS#1#Pequeño bulto o forúnculo que sale en la piel: la nariz roja y con muchos granos

#SIN:forúnculo

#COL:grano de pus

#SYNSESET:08681778n/08677034n#

grano#NCMS#2#Pequeña partícula o elemento que forma parte de un todo: doce granos de uva; granos de polvo

#SIN:?

#COL:grano de arena, grano de uva, grano de azúcar

#SYNSESET:05734433n#

grano#NCMS#3#Semilla o fruto de los cereales o de otras plantas: granos de cereal; grano de maíz

#SIN:?

#COL:grano tostado, grano de trigo, grano de arroz, grano de café, grano de maíz

#SYNSESET:07123531n/07123713n#

LOCUCIONES:

ir al grano#

popular#AQCS#1#Relativo al pueblo: música popular; tradiciones populares

#SIN:folclórico

#ANT:?

#COL:?

#SYNSESET:00537687a/00363173a/05171170a/02139261a/00834378a#

popular#AQCS#2#Que gusta, que es muy conocido: un cantante popular; un actor muy popular

#SIN:famoso

#ANT:desconocido, anónimo

#COL:?

#SYNSESET:01386309a/01047611a#

popular#AQCS#3#Relativo al partido popular: dirigente del partido popular; un diputado popular

#SIN:?

7. Este es el caso del SENSEVAL español, que no disponía de un léxico completo y se realizó uno parcial diseñado especialmente para el concurso.

8. El léxico (o minidiccionario) del SENSEVAL español ha sido elaborado íntegramente por el grupo de lingüistas del Centro de Lenguaje y Computación de la Universidad de Barcelona (CLIC-UB).

#ANT:?

#COL:dirigente popular, sede popular, partido popular

#SYSNET:?#

LOCUCIONES:

tribunal popular#

La información que contienen las entradas de este léxico^[9] (en adelante, "Minidir"), especialmente confeccionado para SENSEVAL y, por lo tanto, con una clara orientación para ser utilizado por un sistema de desambiguación semántica, se estructura en siete (en el caso de las entradas nominales y verbales) u ocho (en las entradas adjetivales) campos distintos: el lema que identifica la entrada, la categoría léxica,^[10] el número de acepción, la definición seguida opcionalmente de un ejemplo, los sinónimos, los antónimos en el caso de las entradas adjetivas, una lista de las colocaciones más frecuentes y el *synset* o *synsets*^[11] correspondientes de EuroWordNet.^[12] Esta información la tiene asociada cada una de las acepciones de la entrada y al final también se propone una lista de locuciones.

Uno de los aspectos más problemáticos que se han planteado en la elaboración del "Minidir" es la asignación del *synset* correspondiente a cada uno de los sentidos de las palabras seleccionadas, es decir compatibilizar la información del diccionario con la de la base de conocimiento EuroWordNet. En el "Minidir" se ha intentado incluir sólo los significados básicos, obteniendo así un léxico que presenta poca granularidad de sentidos, un promedio de cuatro sentidos para las entradas nominales y adjetivas, y de seis para los verbos. Sin embargo, en EuroWordNet las palabras representadas presentan una gran granularidad, es decir especificidad, de sentidos. Por este motivo, nos encontramos que en el "Minidir" una misma acepción se puede corresponder con más de un *synset* de la base de conocimiento (por ejemplo: "grano_1: Pequeño bulto o forúnculo que sale en la piel" se corresponde con dos *synsets* distintos –08681778n/08677034n–), y así se hace constar en los sentidos correspondientes.

Además de los lexicones proporcionados también se ha utilizado la versión 1.7 de las bases de conocimiento léxico-semántico WordNet,^[13] para las tareas del inglés, y EuroWordNet, para las tareas del español, italiano y estonio. Por este motivo, se han asociado las definiciones de cada palabra del léxico con el *synset* correspondiente para tener ambas fuentes relacionadas.

9. El diccionario sólo incluye las 40 palabras seleccionadas para la prueba (18 nombres, 9 adjetivos y 13 verbos).

10. La categoría léxica (nombre, adjetivo o verbo) se representa con etiquetas *parole* (Eagles, programa Eureka 1989-1995) abreviadas.

11. En el modelo de WordNet los conceptos se definen por conjuntos de sinonimia, denominados *synsets*, conectados entre ellos por relaciones semánticas (hiponimia, hiperonimia, antonimia, meronimia, etc.).

12. En Vossen (1999) hay una descripción completa de EuroWordNet^[uri3]. Sitio web de EuroWordNet en español, catalán y vasco: <<http://nipadio.lsi.upc.es/wei.html>^[uri4]>.

13. En Miller y Fellbaum (1991) hay una descripción completa de WordNet^[uri5].

A continuación, se listan las palabras seleccionadas para realizar la tarea basada en una muestra léxica del español:

NOMBRES
<i>arte, autoridad, bomba, canal, circuito, corazón, corona, gracia, grano, hermano, masa, naturaleza, operación, órgano, partido, pasaje, programa, tabla</i>
VERBOS
<i>actuar, apoyar, apuntar, clavar, conducir, copiar, coronar, explotar, saltar, tocar, tratar, usar, vencer</i>
ADJETIVOS
<i>brillante, claro, ciego, local, natural, popular, simple, verde, vital</i>

Tabla 2. Palabras seleccionadas para el SENSEVAL español

3.1.2. Corpus

El segundo recurso lingüístico necesario para llevar a cabo la evaluación de los sistemas es la creación del corpus, el denominado "*gold-standard*", a partir del cual se obtendrá el corpus de entrenamiento y el corpus de evaluación. Para poder evaluar y contrastar los resultados de los distintos sistemas es necesario disponer de este corpus desambiguado manualmente de modo que pueda utilizarse como referencia a partir de la cual comparar los resultados obtenidos por los sistemas participantes.

En el caso de las tareas basadas en una muestra léxica, la elaboración del *gold-standard* se realiza a partir de las palabras preseleccionadas para la evaluación, obteniendo como contexto mínimo la oración en la que aparecen dichas palabras. Siguiendo las indicaciones propuestas por la organización de SENSEVAL (Edmonds y Cotton, 2001), se intenta proporcionar como mínimo 15 ocurrencias distintas de cada sentido o acepción de dichas palabras.^[14]

Una vez seleccionado el corpus se procede a la anotación semántica, es decir a su desambiguación, que consiste en asignar a cada palabra el sentido más apropiado del diccionario. Este proceso se realiza manualmente, y por más de una persona, intentando que se produzca un 90% de acuerdo final entre los distintos anotadores. Una parte del *gold-standard* se entrega a los participantes un mes antes del concurso para que sus sistemas de desambiguación lo utilicen como corpus de entrenamiento y el resto se utiliza como corpus de evaluación para el concurso, borrando previamente las anotaciones semánticas. Los sistemas disponen de una semana para entregar los resultados obtenidos por sus sistemas a la organización, que los contrastará con las anotaciones propuestas en el *gold-standard* por los anotadores humanos.

Hay que señalar que también se distingue entre los sistemas supervisados, es decir aquellos que utilizan el corpus de entrenamiento previamente anotado, y los sistemas no supervisados que no lo utilizan, porque usan otro tipo de fuentes de conocimiento como son diccionarios u otros corpus disponibles. En el caso del español, todos los sistemas que se presentaron eran supervisados y requerían, por lo tanto, del corpus de entrenamiento.

A continuación, se muestra la entrada verbal *explotar* y un trozo del *gold-standard* español en la que aparece dicho verbo desambiguado semánticamente.

- Entrada léxica de *explotar*:

14. Hay casos en los que no es posible proporcionar 15 ocurrencias por sentido por no encontrarse más ejemplos en los corpus a partir de los cuales se obtiene el *gold-standard*.

explotar#VM#1#Obtener beneficio de algo mediante el trabajo: explotar una mina

#SIN:?

#COL:?

#SYNSET:00660933v/00661466v#

explotar#VM#2#Hacer trabajar a alguien de forma abusiva en beneficio propio: explotaba a sus empleados

#SIN:abusar, aprovecharse

#COL:?

#SYNSET:01363660v

#

explotar#VM#3#Valerse de una situación o de una relación en beneficio propio: explotaba sus encantos para obtener trabajo

#SIN:aprovecharse

e

#COL:?

#SYNSET:00659593v/01303135v#

explotar#VM#4#Hacer explosión una cosa, en especial un artefacto: una bomba muy potente explotó cerca del mercado

#SIN:estallar

#COL:?

#SYNSET:00185045v/00185462v/00643046v#

explotar#VM#5#Manifestar o emerger de repente un sentimiento o una situación latente: el presidente explotó indignado por las declaraciones

#SIN:estallar

#COL:?

#SYNSET:00185244v/00404923v#

- Muestra del *gold-standard* del español:

La empresa Boliden, que **explota** la mina de Aznalcóllar, dice ahora que aplaza la reapertura de la instalación. #1

El suceso más grave tuvo lugar en Bilbao, donde un artefacto casero **explotó** ante la puerta del domicilio de Dimas_Sañudo, candidato del PSE a la alcaldía. #4

Se **explota** el filón de su etapa en Ercros y de lo que él mismo ha definido como optimización fiscal: el uso de una sociedad patrimonial para pagar menos al fisco. #3

Tamames **explotó**: "¡Poner a Clinton de modelo de virtud!" #5

Un pequeño artefacto **explotó** ayer en el buzón del domicilio de la concejal del PP en Zarauz_Lucía_Peralta, sin que se registraran heridos. #4

Según esas fuentes, la mujer, identificada por la policía como Esmá_Yurdakul, hizo **explotar** los artefactos que llevaba encima cuando pasó al_lado_de dos vehículos policiales estacionados en el centro de la plaza Taksim, una de las más frecuentadas

de la capital. #4

Dave_Morris y Helen_Steel fueron condenados en 1997 a pagar 15 millones de

pesetas por repartir panfletos en los que afirmaban que la empresa produce alimentos perniciosos para la salud, que maltrata a los animales y que *explota* a sus empleados. #2

El primer ministro italiano, Massimo_D'Alema , aseguró ayer que "Albania corre el riesgo de *explotar*", como consecuencia de la avalancha de refugiados procedentes de Kosovo. #5

La firma que *explota* la ruta a Barcelona acumula quejas de los usuarios y sanciones. #1

Los inmigrantes eran *explotados* por el grupo como peones en unas condiciones de absoluta esclavitud. #2

Constructoras más que dudosas se han valido de obreros sin papeles, a los que se ha *explotado* de manera vergonzosa. #2

El fuego hizo que *explotara* una de las lunas de cristal del cajero y los vidrios hirieron al trabajador Carlos_Elizalde, produciéndole heridas en la cara, los brazos y las piernas. #4

El corpus para el SENSEVAL español se ha elaborado a partir de dos fuentes distintas: a) *El Periódico*,^[15] un corpus que recoge las noticias del año 2000 extraídas de dicho periódico; se trata por lo tanto de una recopilación de textos en lenguaje estándar y de temática general. b) *Lexesp*,^[16] un corpus equilibrado de 5,5 millones de palabras, que incluye textos de diversos temas (científicos, económicos, jurídicos, literarios, periodísticos, etc.) y estilos (ensayo, novela, etc.) que recogen diversos registros de la lengua (estándar, técnico, culto, etc.). Al tratarse de dos corpus analizados y desambiguados morfológicamente,^[17] es posible realizar de forma rápida la búsqueda automática de las frases que contienen las palabras seleccionadas para el ejercicio (una búsqueda combinada de lema y categoría con el fin de extraer todas las formas en las que aparece dicha palabra en los corpus). La elaboración del *gold standard* se realizó en dos fases: primero, se procedió a la etiquetación semántica del corpus y, en una segunda fase, se compararon los resultados aportados por los distintos equipos de anotadores;^[18] en aquellos casos en los que no se producía acuerdo en la asignación de sentidos (en un 46% de las ocurrencias) un único anotador intervenía como árbitro para asignar la etiqueta definitiva.

En cuanto a los sistemas participantes, en la tarea léxica del español se presentaron un total de 12 sistemas de 5 equipos distintos: Universidad de Alicante (UA), Universidad John Hopkins (JHU), Universidad de Standford (SU), Universidad de Maryland (UMD) y Universidad de Manitoba (que presentó 8 algoritmos distintos de desambiguación, D6-D10, dX, dY y dZ). El sistema que obtuvo mejores resultados, es decir, que se aproximaba más a los resultados obtenidos manualmente, fue el sistema de aprendizaje supervisado de la Universidad John Hopkins. Este sistema obtuvo mejores resultados en 12 de las 39 palabras seleccionadas, así como también fue el mejor en la desambiguación de nombres y verbos, pero no en el caso de los adjetivos, para los cuales la Universidad de Standford obtuvo mejores resultados. En general, en todos los sistemas la desambiguación de nombres y adjetivos se aproximaba más a los resultados etiquetados del *gold-standard* que la de los verbos. En la tabla 3, se muestran los porcentajes de acuerdo obtenidos por los distintos sistemas participantes.

Sistemas	UA	SU	JHU	UMD	d6	d7	d8	d9	d10	dX	dY	dZ
Acuerdo	0,51	0,63	0,65	0,61	0,55	0,57	0,59	0,53	0,59	0,55	0,51	0,57
Kappa	0,20	0,34	0,47	0,20	0,13	0,19	0,23	0,06	0,24	0,15	-0,03	0,15

Tabla 3. Porcentajes de acuerdo y medida Kappa

15. El corpus de *El Periódico* ha sido cedido por el proyecto HERMES (CICYT TIC2000-0335-C03-02).

16. El corpus *Lexesp* ha sido proporcionado por el proyecto Lexesp-III (DGICYT APC 99-0105).

17. Para el análisis morfológico se ha utilizado el analizador morfológico MACO+ (Carmona *et al.*, 1998) y para la desambiguación morfológica el tagger RELAX (Padró, 1998).

18. Los equipos de desambiguación estaban constituidos por: el equipo de lingüistas de CLiC (UB) y los equipos, integrados básicamente por informáticos, de TALP (UPC) y GPLN (UNED).

Como se puede observar el sistema presentado por la JHU es el que tiene una media de acierto superior: 0,65 ó 0,47 si aplicamos la medida Kappa.

Para ver los resultados obtenidos por los diferentes participantes más detalladamente, así como los recursos desarrollados para el ejercicio consúltese la página oficial de SENSEVAL^[url6].

4. Repercusiones de SENSEVAL en los estudios de semántica léxica

La metodología desarrollada en SENSEVAL, orientada principalmente a la evaluación y comparación de sistemas automáticos de desambiguación semántica, se ha centrado fundamentalmente en la valoración de los algoritmos de dichos sistemas. Estos algoritmos se han evaluado independientemente de la calidad lingüística y lexicográfica de los recursos lingüísticos, tanto de las fuentes léxicas como de los corpus anotados a mano (o *gold-standard*) utilizados. En este contexto, se ha generado una discusión paralela en torno a cuestiones básicas de lexicografía y semántica, que se encuentran en la base de todo el proceso de desambiguación. Los puntos en que se centra actualmente el debate son:

- La validez del proceso de etiquetado, tanto del etiquetado manual para los corpus de aprendizaje como del etiquetado automático, en los procesos de desambiguación semántica automática.
- La objetividad en los procesos de evaluación de los sistemas de desambiguación semántica automática.
- La calidad de las fuentes de información y su adecuación para las tareas de desambiguación semántica de sentidos.
- La misma existencia de los sentidos, su descripción y representación.

Una de las primeras cuestiones que cabe plantearse es si es posible o no la desambiguación semántica, tanto manual como automática. Si los humanos actuamos de manera poco sistemática en esta tarea, entonces el valor de los corpus etiquetados semánticamente es cuestionable y también lo será el desarrollo de sistemas que realicen la desambiguación semántica de manera automática. Si bien es cierto que las máquinas realizan determinadas tareas mejor que los humanos, ello no es así cuando se halla implicado el significado, el contexto y, en general, la inteligencia. En estos casos, se puede considerar que la manera en que lo realiza un humano es el nivel máximo de calidad al que se puede aspirar. Así, la calidad de los sistemas de desambiguación semántica automática se evalúa comparando sus resultados con corpus etiquetados manualmente (*gold-standard*), lo cual implica analizar cómo los humanos realizamos esta tarea y la validez de los resultados obtenidos.

A raíz de las investigaciones llevadas a cabo en el área de la desambiguación semántica automática y de los experimentos orientados a evaluar la calidad de las fuentes y la validez del proceso de anotación (Véronis, 2002; Krishnamurthy y Nicholls, 1998), se ha puesto de manifiesto el escaso acuerdo entre los anotadores humanos y la poca adecuación de las fuentes de conocimiento léxico para la realización de la desambiguación semántica automática, que ha conducido no sólo al replanteamiento del concepto mismo de sentido, de cómo se debe describir y representar, sino que también se cuestiona la consistencia de los métodos de discriminación de sentidos. Experimentos anteriores se habían centrado en analizar cómo los humanos realizamos esta tarea (Ahlsvede, 1995; Ahlsvede y Lorand, 1993; Fellbaum, Grabowsky y Landes, 1998), aunque estos estudios o bien son poco sistemáticos o tratan un número muy reducido de palabras. Véronis (2002) propone un experimento con el objetivo de mostrar la falta de acuerdo en la anotación manual de corpus que concluye con un diagnóstico sobre las razones del problema: la no adecuación de los diccionarios de uso común ni de las ontologías existentes para la realización de la desambiguación semántica automática.

Otra de las consecuencias de este tipo de investigaciones es el replanteamiento de algunas cuestiones que atañen a los fundamentos de la polisemia. Se ha evidenciado la necesidad de definir unas bases teóricas y metodológicas para el tratamiento de los sentidos, tanto en lo que se refiere a su discriminación o identificación como a los criterios de granularidad (especificidad) con que deben tratarse para tareas de desambiguación semántica automática. En consecuencia, se han cuestionado algunos de los procedimientos del trabajo lexicográfico, en especial el recurso de la intuición para la distinción de sentidos, para abrirlo a nuevas orientaciones y tipos de conocimiento (Kilgarrif, 1998; Miller y Leacock, 2000) como la información sintáctica, colocativa y ejemplos extraídos de corpus. En esta línea merecen destacarse los trabajos de Hanks (2000), Palmer (1998), Mana y Corazzari (2000) y Véronis (2002).

Otra de las líneas de trabajo reciente se funda en la idea de que la comparación entre las lenguas es útil para la desambiguación semántica automática. Se parte de la hipótesis de que la correspondencia entre las palabras y los sentidos varía de manera significativa de una lengua a otra (Ide, 2000). Resnik y Yarowsky (1997) sugieren que, para los propósitos de la desambiguación semántica automática, los múltiples sentidos de una palabra podrían determinarse considerándose sólo las distinciones de sentido lexicalizadas entre lenguas (*cross-linguistically*).

Son muchas las críticas que se han levantado en contra de la exclusiva representación de los sentidos en forma de definiciones de un diccionario, tanto por la falta de información sobre las preferencias léxicas y las estructuras sintácticas que aceptan como por lo impreciso del lenguaje utilizado y el propio método de análisis, la introspección. Se trata de ver en qué medida las técnicas y métodos que se proponen desde el procesamiento del lenguaje natural pueden aplicarse para mejorar las fuentes de conocimiento existentes, proponer nuevas fuentes de información y aportar criterios para la discriminación y desambiguación de sentidos.

En este sentido, SENSEVAL se ha convertido en un nuevo foro de discusión, de análisis y de intercambio científico en el que tienen cabida todos aquellos aspectos vinculados con la resolución de la ambigüedad semántica, ya sea desde un punto de vista más técnico como es el caso de las estrategias y métodos propuestos para su tratamiento automático, ya sea desde una perspectiva más lingüística que pone en evidencia la dificultad que implica la delimitación de sentidos y la necesidad de un estudio en profundidad sobre su discriminación, es decir sobre la polisemia y sobre los criterios lingüísticos que permitan su delimitación de una manera objetiva, rigurosa y adecuada para los sistemas de desambiguación semántica automática.

Lista de URL:

- [url1]:<http://www.senseval.org/>
- [url2]:<http://www.siglex.org/>
- [url3]:<http://www.illc.uva.nl/EuroWordNet/>
- [url4]:<http://nipadio.lsi.upc.es/wei.html>
- [url5]:<http://www.cogsci.princeton.edu/~wn/>
- [url6]:<http://www.senseval.org>

Bibliografía:

- AHLWEDE, T.E. (1995). "Word sense disambiguation by humans informants". En: *Proceedings of the 6th Midwest Artificial Intelligence and Cognitive Society Conference* (abril 1995: Carbondale, Illinois), págs. 73-78.
- AHLWEDE, T.E.; LORAND, D. (1993). "The ambiguity questionnaire: a study of lexical disambiguation by human informants". En: *Proceedings of the 6th Midwest Artificial Intelligence and Cognitive Society Conference* (Chesterton, Indiana), págs. 21-25.

- CARMONA, J.; CERVELL, S.; MÀRQUEZ, L.; MARTÍ, M.A.; PADRÓ, L.; PLACER, R.; RODRÍGUEZ, H.; TAULÉ, M.; TURMO, J. (1998). "An environment for morphosyntactic processing of unrestricted spanish text". En: *Proceedings of the First International Conference on Language Resources and Evaluation, LREC* (Granada, España).
- EDMONDS, P.H.; COTTON, S. (2001). "SENSEVAL-2 overview". En: *Proceedings of SENSEVAL-2, Second International Workshop on Evaluating Word Sense Disambiguation Systems, 39th Annual Meeting of the ACL* (Francia).
- FELLBAUM, C.; GRABOWSKY, L. (1998). "Performance and confidence in a semantic annotation task". En: FELLBAUM, C. (ed.). *WordNet: An electronic database*. Cambridge (Massachussets): MIT Press, págs. 217-237.
- GALE, W.A.; CHURCH, K.W.; YAROWSKY, D. (1993). "A method for disambiguating word senses in a large corpus". *Computer and the Humanities*. Núm. 26, págs. 415-439.
- HANKS, P. (2000). "Do word meanings exist?". *Computers and the Humanities* (número especial: *Evaluating Word Sense Disambiguation Programs*). Vol. 1-2, núm. 34, págs. 205-215.
- IDE, N. (2000). "Cross-lingual sense determination: Can it work?". *Computers and the Humanities* (número especial: *Evaluating Word Sense Disambiguation Programs*). Vol. 1-2, núm. 34, págs. 223-234.
- IDE, N.; VÉRONIS, J. (1998). "Introduction to the special issue on word sense disambiguation: the state of the art". *Computational Linguistics*. Vol. 1, núm. 24, págs. 1-40.
- KILGARRIFF, A. (1998). "SENSEVAL: An exercise in Evaluating Word Sense Disambiguation Programs". LREC-1998.
- KILGARRIF, A.; PALMER, M. (2000). (editores invitados). "Special issue on SENSEVAL: Evaluating Word Sense Disambiguation Programs". *Computers and the Humanities*. Vol. 1-2, núm. 34.
- KILGARRIF, A.; ROSENZWIG, J. (2000). "Framework and results for english SENSEVAL". *Computers and the Humanities*. Vol. 1-2, núm. 34, págs. 15-48.
- KRISHNAMURTHY, R.; NICHOLLS, D. (1998). "Peeling an onion: the lexicographer's experience of manual sense tagging". En: *SENSEVAL Workshop* (2-4 de septiembre de 1998: Herstmonceux Castle, Sussex, Inglaterra).
- MANA, N.; CORAZZARI, O. (2001). "The lexico-semantic annotation of an italian treebank" [en línea]. <<http://www.tcc.itc.it/publications/2001/>>.
- MELAMED, D.; RESNIK, P.H. (2000). "Tagger evaluation given hierarchical tag sets". *Computers and the Humanities*. Vol. 1-2, núm. 34.
- MILLER, G.A.; FELLBAUM, C. (1991, diciembre). "Semantic networks of English". *Cognition*. Núm. 41, págs. 197-229.
- MILLER, G.A.; LEACOCK, C. (2000). "Lexical representations for sentence processing". En: *Polysemy, theoretical and computational approaches*. Oxford University Press.
- PADRÓ, L. (1998). *An hybrid environment for syntax-semantic tagging*. Tesis doctoral. Dep. LSI, UPC.
- PALMER, M. (1998). "Consistent criteria for sense distinctions". *Computers and the Humanities* (número especial: *Evaluating Word Sense Disambiguation Programs*). Vol. 1-2, núm. 34, págs. 217-222.
- RESNIK, P.H.; YAROWSKY, D. (1997). "A perspective on word sense disambiguation methods and their evaluation". En: *Workshop: Tagging text with lexical semantics: Why, what and how?* (ACL SIGLEX, Washington, D.C.).
- RIGAU, G.; TAULÉ, M.; GONZALO, J.; FERNÁNDEZ, A. (2001). "Framework and results for the spanish SENSEVAL". En: *Proceedings of the SENSEVAL-2: Second International Workshop on Evaluating Word Sense Disambiguation Systems* (ACL SIGLEX, Toulouse).
- TAULÉ, M. (2002). *Especificación de los criterios y la metodología seguida en la organización del SENSEVAL-II español*. X-Tract WP-08/02. Barcelona.

VÉRONIS, J. (1998). "A study of polisemy judgements and inter-annotator agreement". En: *Programme and advanced papers of the SENSEVAL workshop* (Herstmonceux Castle, Inglaterra).

VÉRONIS, J. (2002). "Sense tagging: does it make sense?". En: *ACL-Workshop on Word Sense Disambiguation* (2002: Toulouse).

VOSSSEN, P. (ed.). (1999). *EuroEordNet general document* [en línea]. <<http://www.hum.uva.nl/~ewn>>.

Enlaces relacionados:

⇒ SENSEVAL:

<http://www.senseval.org/>

⇒ Grupo de Procesamiento del Lenguaje Natural (Universidad de Sheffield):

http://nlp.shef.ac.uk/research/areas/word_sense_disambiguation.html

⇒ Artículo sobre desambiguación semántica:

<http://www.ilc.cnr.it/EAGLES96/rep2/node39.html>

⇒ Demostración de desambiguación semántica (en alemán):

<http://pi0897.kub.nl/~iris/>

Cita recomendada:

TAULÉ, Mariona; MARTÍ, M. Antònia (2003). "El software libre y las lenguas minoritarias: una oportunidad impagable". *Digitium*, n.º 5 [artículo en línea]. DOI: <http://dx.doi.org/10.7238/d.v0i5.559>

[Fecha de publicación: abril de 2003]