0.1. Conjuntos de etiquetas

La sección anterior dió una descripción general de los tipos de clases sintácticas a las que pertenecen las palabras. Esta sección presenta los conjuntos de etiquetas actuales utilizados en la etiquetación gramatical. Es decir, las etiquetas que se corresponden con cada una de estas clases sintácticas. Los conjuntos de etiquetas, también llamados tagsets, ...

Etiqueta	Descripción	Ejemplo
CC	Coordinating conjunction	*and*
CD	Cardinal number	*1, third*
DT	Determiner	*the*
EX	Existential	*there there is*
FW	Foreign word	*d'hoevre*
IN	Preposition/subordinating conjunction	*in, of, like*
JJ	Adjective	*green*
JJR	Adjective, comparative	*greener*
JJS	Adjective, superlative	*greenest*
LS	List marker	*1)*
MD	Modal	*could, will*
NN	Noun, singular or mass	*table*
NNS	Noun plural	*tables*
NNP	Proper noun, singular	*John*
NNPS	Proper noun, plural	*Vikings*
PDT	Predeterminer both	*the boys*
POS	Possessive ending	*friend's*
PRP	Personal pronoun	*I, he, it*
PRP$	Possessive pronoun	*my, his*
RB	Adverb	*however, usually, naturally, here, good*
RBR	Adverb, comparative	*better*
RBS	Adverb, superlative	*best*
RP	Particle	*give up*
SYM	Symbol	*+, %,&*
TO	To	*to go, to him*
UH	Interjection	*uhhuhhuhh*
VB	Verb, base form	*take*
VBD	Verb, past tense	*took*
VBG	Verb, gerund/present participle	*taking*
VBN	Verb, past participle	*taken*
VBP	Verb, sing. present, non-3d	*take*
VBZ	Verb, 3rd person sing. present	*takes*
WDT	Wh-determiner	*which*
WP	Wh-pronoun	*who, what*
WP$	Possessive wh-pronoun	*whose*
WRB	Wh-abverb	*where, when*
$	Dollar sign	*$*
#	Pound sign	*#*
”	Left quote	(' or ”)
”	Right quote	(' or ”)
(	Left parenthesis	([, (, {, ¡)
)	Right parenthesis	( ],), }, ¿)
,	Comma	,
.	Sentence-final punc	(. ! ?)
:	Mid-sentence punc	(: ; ... -)

Hay un pequeño número de conjuntos de etiquetas o tagsets populares para el idioma inglés, muchos de los cuales evolucionaron a partir del conjunto de eti-

quetas utilizado para el corpus Brown. Este conjunto de etiquetas se conoció como el Brown Corpus Tag-set, un conjunto de 87 etiquetas que se utilizó para etiquetar el corpus Brown, un corpus de 1 millón de palabras de ejemplos obtenidos desde 500 textos escritos de diferentes géneros (diarios, novelas, no ficción, académico, étc.) que fué ensamblado en la Universidad Brown entre 1963 y 1964. Este corpus fué etiquetado gramaticalmente aplicando en primera instancia un etiquetador automático, el programa TAGGIT, y luego corregido manualmente.

Al lado del conjunto de etiquetas original Brown se encuentran dos de los conjuntos de etiquetas más utilizados: el conjunto de etiquetas reducido Pen Treebank (de 45 etiquetas) y el conjunto de etiquetas CLAWS C5 de tamaño medio (62 etiquetas) utilizado por Lancaster UCREL en el proyecto CLAWS (Constitutent Likelihood Automatic Word-tagging System) para etiquetar el corpus British National Corpus (BNC).

El conjunto de etiquetas Penn Treebank mostrado anteriormente fué utilizado para etiquetar el corpus Brown, en el corpus Wall Street Journal y el corpus Switchboard entre otros. En realidad, quizás en parte por su pequeño tamaño es uno de los conjuntos de etiquetas más utilizado. A continuación se exhiben algunos ejemplos de oraciones del corpus Brown etiquetadas con el conjunto de etiquetas Penn Treebank. Representaremos una palabra etiquetada mediante la colocación de una barra oblicua seguida de su etiqueta:

1. The/DT grand/JJ jury/NN commented/VBD on/IN a/DT number/NN of/IN other/JJ topics/NNS ./.

2. **There/EX** are/VBP 70/CD children/NNS **there/RB**

3. Although/IN preliminary/JJ findings/NNS were/VBD **reported/VBN** more/RBR than/IN a/DT year/NN aho/IN ,/, the/DT latest/JJS results/NNS appear/VBP in/IN today/NN **'s/POS** New/NNP England/NNP Journal/NNP of/IN Medicine/NNP ./.

El primer ejemplo exhibe los determinantes *the* y a, los adjetivos *grand* y *other*, los sustantivos comunes *jury*, *number* y *topics* y el verbo en tiempo pasado *commented*. El segundo ejemplo muestra el uso de la etiqueta ET para marcar la construcción existencial *there* y otro uso de *there* que es etiquetado como un adverbio (RB). El tercer ejemplo muestra la segmentación del morfema posesivo *'s* y un ejemplo de la construcción pasiva 'were reported', en la cual el verbo *reported* está marcado como un pasado participio (VBN) en vez de como un pasado simple (VBD). También es interesante notar que el sustantivo propio *New England* está etiquetado como NNP. Finalmente, se puede observar que como *New England Journal of Medicine* es un sustantivo propio, el etiquetado de Treebank elige marcar cada sustantivo separado como NNP, incluyendo *journal* y *medicine*, que en otros casos pueden ser etiquetados como sustantivos comunes (NN).

Algunas distinciones de etiquetado son muy difíciles de realizar tanto para humanos como para máquinas. Por ejemplo las preposiciones (IN), participios (RP) y adverbios (RB) pueden tener un gran solapamiento. Las palabras como *around* pueden ser de los 3 tipos:

1. Mrs./NNP Shaefer/NNP never/RB got/VBD **around/RP** to/TO joining/VBG

2. All/DT we/PRP gotta/VBN do/VB is/VBZ go/VB **around/IN** the/DT corner/NN

3. Chateau/NNP Petrus/NNP costs/VBZ **around/RB** 250/CD

Tomar estas decisiones requiere un conocimiento sofisticado de la sintaxtis; los manuales de etiquetado ofrecen varias heurísticas que pueden ayudar a los codificadores humanos a tomar estas decisiones, y también puede proveer características útiles para los etiquetadores automáticos. Por ejemplo 2 heurísticas de Santorini son que las preposiciones generalmente son asociadas con la frase nominal siguiente (aunque también pueden ser seguidas por frases preposicionales), y que la palabra *around* es etiquetada como un adverbio cuando significa aproximadamente. Además, mientras que los participios pueden preceder o seguir a un objeto de frase sustantiva, como en los siguientes ejemplos:

1. She told off/RP her friends

2. She told her friends off/RP.

las preposiciones no puede seguir sus frases nominales (* se utiliza para marcar oraciones no gramaticales)

1. She stepped off/IN the train

2. *She stepped the train off/IN.

Otra dificultad es etiquetar las palabras que pueden modificar sustantivos. Algunas veces los modificadores que preceden sustantivos son sustantivos comunes como *cotton below*, otras veces el manual de etiquetado del TreeBank especifica que los modificadores serán etiquetados como adjetivos (*income-tax*) y otras veces como sustantivos propios (*Gramm-Rudman*):

1. cotton/NN sweater/NN

2. income-tax/JJ return/NN

3. the/DT Gramm-Rudman/NP Act/NP

Algunas palabras que pueden ser adjetivos, sustantivos comunes o sustantivos propios son etiquetados en el Treebank como sustantivos comunes cuando actúan como modificadores:

1. Chinese/NN cooking/NN

2. Pacific/NN waters/NNS

Una tercera dificultad conocida en el etiquetado es distinguir pasado participio (VBN) de adjetivos (JJ). Una palabra como *married* es un pasado participio cuando está siendo utilizada en una forma verbal como se muestra a continuación en el primer ejemplo. Y es un adjetivo cuando está siendo utilizada para expresar una propiedad como en el segundo ejemplo.

1. They were married/VBN by the Justice of the Peace yesterday at 5:00

2. At the time, she was already married/JJ.

Los manuales de etiquetado dan varias criterios útiles para decidir como clasificar entre verbo y evento una palabra particular en un contexto específico.

El conjunto de etiquetas del Penn Treebank fué seleccionado a partir del conjunto de etiquetas original del corpus Brown (de 87 etiquetas). Este conjunto reducido deja afuera información que puede ser recuperada desde la identidad del item léxico. Por ejemplo los conjuntos de etiquetas Brown original y C5 incluyen una etiqueta separada para cada una de las diferentes formas de los verbos *do* (C5 propone la etiqueta VDD para *did* y VDG para *doing*), *be* y *have*. Estas etiquetas fueron omitidas en el conjunto de Treebank.

Ciertas distinciones sintácticas no fueron marcadas en el conjunto de etiquetas Penn Treebank porque las oraciones de Treebank fueron parseadas, no meramente etiquetadas, y por lo tanto alguna información sintáctica es representada en la estructura de la frase. Por ejempo, la etiqueta simple IN es utilizada para preposiciones como para conjunciones subordinadas desde que la estructura de arbol de la oración las desambigua (las conjunciones subordinadas siempre preceden cláusulas, las preposiciones preceden frases nominales o frases preposicionales). La mayoría de las situaciones de etiquetado, sin embargo, no involucran corpora parseado; por esta razón el conjunto del Penn Treebank no es suficientemente específico para muchos usos. Los conjuntos de etiquetas de Brown original y C5 , por ejemplo, distinguen preposiciones (IN) de conjunciones subordinadas (CS), como en los siguiente ejemplos:

1. **after/CS** spending/VBG a/AT few/AP days/NNS at/IN the/AT Brown/NP Palace/NN Hotel/NN

2. **after/IN** a/AT wedding/NN trip/NN to/IN Corpus/NP Christi/NP ./.

El conjunto de etiquetas Brown original y C5 también tiene dos etiquetas para la palabra *to*; en Brown el uso del infinitivo es etiquetado como TO, mientras que las preposiciones utilizan IN:

1. **to/TO** give/VB priority/NN **to/IN** teacher/NN pay/NN raises/NNS

Brown también tiene la etiqueta NR para sustantivos adverbiales como *home*, *west*, *Monday* y *tomorrow*. Como el Treebank carece de esta etiqueta, hay una política mucho menos consciente para sustantivos adverbiales; *Monday*, *Tuesday* y otros días de la semana son marcados como NNP, *tomorrow*, *west* y *home* son marcados algunas veces como NN y algunas veces como RB. Esto hace el conjunto de etiquetas del Treebank menos útil para tareas de alto nivel linguístico como la detección del tiempo de frases. Sin embargo, el conjunto de etiquetas de Treebank ha sido el más utilizado en la evaluación de algoritmos de etiquetación automática. Esta es la razón por la cual elegimos este conjunto de etiquetas para utilizar en el desarrollo del presente trabajo.

0.2. Diccionario COBUILD

Como se menciona anteriormente, para suplir la falta de corpus de entrenamiento sin caer en la tediosa y costosa tarea de anotar un nuevo corpus manualmetne, se introduce una fuente de información existente y manualmente anotada. Estamos hablando de un diccionario, que no es ni más ni menos que un conjunto de palabras con su/s posible/s etiqueta/s y uno o más ejemplos en donde cada palabra aparece con cada una de esas etiquetas. El primer

paso es elegir un diccionario y extraer esta información. El diccionario elegido fué Cobuild.

Cobuild es un diccionario basado en la información del corpus Bank of English y el corpus Collins. Su siglas significan: Collins Birmingham University International Language Database.

El corpus Collins es una base de datos con alrededor de 2.5 billones de palabras en Inglés. Contiene material escrito de websites, diarios, revistas y libros publicados en todo el mundo, y material hablado de radio, TV y conversaciones diarias.

A su vez, el Bank of English forma parte del corpus Collins. Contiene 650 millones de palabras de una cuidadosa selección de fuentes, para dar un reflejo preciso y balanceado del Inglés que es usado día a día.

Como el corpus es tan extenso, se pueden apreciar una gran cantidad de ejemplos de como las personas utilizan realmente las palabras. Se puede entender como son utilizadas, que significan, que palabras ocurren juntas y que tan a menudo. Esta información sobre la frecuencia ha ayudado a decidir que palabras incluir en el diccionario Cobuild. Por ejemplo, alrededor del 90 % del inglés hablado y escrito está constituído de aproximadamente 3.500 palabras.

El diccionario Cobuild fué concebido teniendo especial atención en los ejemplos expuestos. El proceso de agregado de palabras al diccionario es muy cuidadoso: cuando un editor quiere agregar una nueva palabra al diccionario, busca en el corpus cada ejemplo que contenga esa palabra. La palabra aparece en una larga lista de oraciones y el editor decide cuál de todos los ejemplos expresa mejor el sentido que está buscando en esa palabra. Todos los ejemplos del diccionario Cobuild muestran patrones gramaticales típicos, vocabulario típico y contextos típicos para cada palabra.

Cobuild fué el primer diccionario en presentar una cantidad exhaustiva del vocabulario inglés derivado de observaciones directas del lenguaje. Desde entonces el equipo de Cobuild ha continuado recolectando textos desde todas las fuentes para crear el corpus The Bank of English. The Bank of English es una colección o corpus de alrededor de 650 millones de palabras de inglés escrito y hablado ubicado en computadoras para el estudio del uso del lenguaje.

The Bank of English contiene un amplio rango de tipos diferentes de lenguaje escrito y hablado proveniente de cientos de fuentes diferentes. Aunque la mayoría de las fuentes son británicas, aproximadamente el 25 % de la información proviene de fuentes de inglés americano y alrededor del 5 % de otras variedades nativas del inglés como Australia y Singapur. Los textos escritos provienen de diarios, revistas, libros de ficción y de no ficción, folletos, informes y cartas. Dos tercios del corpus están confeccionados a partir del lenguaje de los medios: diarios, revistas, radio y televisión. Esta es una categoría significativa en vista de que millones de personas leen y escuchan el lenguaje presente en los medios. Publicaciones internacionales, nacionales y locales también fueron incluídas para capturar un rango general de temas importantes y estilos. Hay otros cientos de libros y revistas de especial interés, que abordan temas desde aeróbicos a zoología. Sin embargo los libros de texto técnicos, científicos, manuales, étc., no fueron incluídos en el corpus.

El lenguaje hablado informal es representado por grabaciones de conversaciones diarias casuales, reuniones, entrevistas y discusiones. Alrededor de 15 millones de palabras de The Bank of English son transcripciones de lenguaje hablado de esta clase. Luego son seleccionados para incluir un amplio rango de

temas y situaciones de habla.

El propósito de recolectar todo esta valiosa información en computadoras fué para permitir a los linguistas (escritores de diccionarios) tener acceso a la mayor cantidad de información posible sobre cada una de las palabras que está definiendo. Desde luego, los linguistas son elegidos a partir de su habilidad con el lenguaje, pero incluso el linguista mas experimentado no puede deducir solo por su intuición todos los hechos relevantes sobre todas las palabras en un lenguaje. El corpus, y el software que se utiliza para analizarlo ayudó al equipo de Cobuild a ordenar la información y ganar valiosa percepción sobre la manera en que las palabras son utilizadas: sus significados, sus patrones gramaticales típicos y las maneras en que están relacionadas con otras palabras.

Muchas palabras tienen más de una clase de palabra gramatical y a menudo es de mucha ayuda para los linguistas mirar solo a una clase de palabra por vez. Para ayudarlos a hacer esto, se ha desarrollado un software que muestra las clases de palabras en cada linea del corpus. De esta manera los linguistas pueden mirar la información completa con el código de clase de palabra o pueden preguntar solo por verbos, sustantivos, étc.

Este tipo de software les permite a los linguistas a tomar decisiones sobre los diferentes sentidos de las palabras, el lenguaje de las definiciones, la selección de ejemplos, y la información gramatical dada. El corpus permite realizar esta tarea con confianza y exactitud. Y cuanto más grande es el corpus mayor es la confianza y la exactitud.

0.2.1. Método de construcción

En 1987 se publicó el diccionario Cobuild basado en un corpus de 20 millones de palabras. A continuación se construyó un nuevo corpus, el Bank ok English con alrededor de 200 millones de palabras. La nueva edición del diccionario Cobuild se basa en este nuevo corpus.

La construcción de este nuevo diccionario, es un proceso en donde se decide que palabras y frases presentes en el corpus incluir. Luego se examina el lenguaje palabra por palabra y frase por frase con el objetivo de dar clara cuenta de cada significado y uso. Entonces se escribe una definición, se eligen ejemplos típicos, y se agrega información sobre la pronunciación, la gramática, semántica, pragmatismos y frecuencia para completar cada entrada.

0.2.2. Evidencia

Un diccionario debe comenzar por la evidencia, los hechos. Los hablantes de un lenguaje conocen mucho sobre éste, a partir de que cada día leen y hablan sin esfuerzo durante horas. Sin embargo no son capaces de explicar que es lo que hacen. Utilizar un lenguaje es una habilidad para la cual la mayoría de las personas no son completamente conscientes; no pueden examinarlo en detalle, simplemente lo utilizan para comunicarse. Aquellos que aprenden a observar el lenguaje cuidadosamente pueden expresar y organizar algunos de los hechos sobre éste basados en la experiencia. De todas maneras hay muchos hechos sobre el lenguaje que no pueden ser descubiertos simplemente pensando y reflexionando sobre él, incluso leyendo y escuchando muy atentamente. Es por eso que Cobuild estableció la utilización de computadoras para identificar estos hechos.

0.2.3. Un corpus

El resultado de este fué que Cobuild estableció un nuevo tipo de evidencia, una colección de textos en inglés llamado corpus ubicado en una computadora de manera que pueda consultarse instantáneamente. Los creadores de Cobuild sabían que necesitaban millones de palabras de inglés, hablado y escrito, americano y británico, formal e informal, sobre hechos y sobre ficción, étc. Esta evidencia reunida durante varios años, permitió encontrar que palabras y expresiones son más utilizadas. Cuando una palabra tiene varios significados existe la capacidad de ver cuales son los significados importantes, y cuales frases se deben incluir. Tomaron como filosofía y fueron conscientes de que todos los detalles de un uso natural de una palabra son escenciales y no pueden ser falsificados. Se dieron cuenta de que debían utilizar ejemplos reales siguiendo la tradición de los grandes linguistas, en lugar de crearlos.

0.2.4. The Bank of English

Hace varios años que se ha hecho mucho más fácil reunir grandes cantidades de lenguaje hablado y escrito. Los publicadores de libros, revistas y diarios se hicieron conscientes de la gran cantidad de lenguaje que pasaba a través de sus manos y de que debe haber muchas buenas razones para conservarlo en formatos electrónicos. Un negocio apareció para el lenguaje electrónico entre la gente que quería encontrar o verificar sentencias, particularmente en las noticias, revistas y lenguaje legal. Gradualmente millones de palabras comenzaron a estar disponibles para los estudiosos del lenguaje. Hoy en día el problema no es encontrar el lenguaje sino manejarlo y realizar sensibles y balanceadas selecciones para las tareas analíticas.

Diseñando el corpus The Bank of English se balancearon un número de factores(inglés hablado y escrito, americano y británico y otras características: hablantes de comunidades nativas, libros y revistas y más clasificaciones dentro de ésas)

Dentro del componente hablado, el tipo de lenguaje más dificil de recolectar fué como siempre la conversación informal grabada en la vida diaria de la gente común, sin pensar de que su lenguaje está siendo preservado en un corpus. Cada conversación tiene que ser grabada y transcripta por expertos para luego ser ingresada en una computdora. Esta clase de lenguaje improvisado es de un interés particular para los constructores de diccionarios. El Bank of English cuenta con un total de 15 millones de palabras de este tipo de grabaciones de lenguaje hablado.

0.2.5. La lista de palabras principales

Es mucho más fácil decidir que palabras y frases incluir y cuales omitir, cuando se tienen cifras exactas a partir de una cantidad tan grande de lenguaje. Las computadoras pueden verificar instantáneamente la actividad del lenguaje de miles de hablantes y escritores. Un diccionario (incluso un gran diccionario) es capaz de elegir solo los hechos más importantes del lenguaje para presentar y los compiladores necesitan buena evidencia para sus selecciones. Cobuild se especializa en presentar las palabras y frases que son frecuentes en el uso diario. Lejos de ser un registro histórico del lenguaje es más bien una muestra del lenguaje contamporáneo.

0.2.6. Frecuencia

Cobuild brinda información sobre la frecuencia de las palabras principales. Se establecieron 5 bandas de frecuencias. Comenzando con las palabras muy comunes (las de mayor frecuencia), oscila entre un vocabulario básico a uno intermedio y luego hasta cubrir el núcleo del vocabulario. Las palabras principales sin marca de frecuencia son las menos comunes, sin embargo vale la pena incluirlas en el diccionario. El punto es que el idioma inglés utiliza un número bastante pequeño de palabras para la mayoría de los propósitos pero también tiene disponible un rico y amplio vocabulario. Por ejemplo *be* y *because* son naturalmente pertenecientes a la banda de mayor frecuencia, por el otro lado, palabras como *barracuda*, *basalt* y *basrelief* no son tan frecuentes. Estas últimas son claramente utilizadas en ocasiones particulares. Cabe aclarar que incluso las palabras infrecuentes incluídas en el diccionario fueron seleccionadas por su utilidad relativa entre miles de palabras posibles.

Entonces Cobuild cuenta con un sistema de frecuencia que marca las palabras principales: una marca significa que la palabra tiene una alta frecuencia y por lo tanto es una palabra común dentro del lenguaje inglés. Dos o más marcas significan que la palabra es parte esencial del vocabulario, cuantas más marcas posee, menos frecuente es.

0.2.7. Ejemplos

Todos los ejemplos fueron seleccionados del corpus The Bank of English. Como antes dijimos, los ejemplos son seleccionados cuidadosamente para mostrar los patrones que frecuentemente son hallados junto a una palabra o frase. El compilador tiene docenas, centenas o miles de ejemplos disponibles y rapidamente escoge los *colocados* (palabras particulares ubicadas cerca de la palabra principal) y las estructuras típicas en donde la palabra o frase es encontrada más a menudo.

Esto significa que los ejemplos cumplen varias funciones. Desde luego ayudan a mostrar el significado de la palabra exhibiendo su uso. Las investigaciones sugieren que un gran número de usuarios comienza con los ejemplos antes que con el significado. Las definiciones de Cobuild son bastante claras por sí mismas y los ejemplos muestran el fraseo característico alrededor de la palabra. Como los ejemplos son piezas de texto genuinas y han sido elegidas cuidadosamente en base al uso de la palabra, pueden ser de confianza para exhibir la palabra en un contexto natural.

0.2.8. Información gramatical

Casi cada sentido de cada entrada en el diccionario Cobuild tiene junto a esta una clasificación gramatical, usualmente una clase de palabra y a menudo también una nota estructural. Esta es la información sobre la que se sustenta este trabajo, ya que en base a ella se construirá el nuevo corpus de entrenamiento.

0.2.9. Pragmatismo

Muchos usos de una palabra necesitan más de una frase para explicar apropiadamente su significado. La gente utiliza palabras para realizar muchas cosas:

hacer invitaciones, expresar sus sentimientos, enfatizar que es lo que está diciendo, étc. El corpus nos brinda evidencia para tales usos que son difíciles de tomar desde cualquier otra fuente, porque nosotros solo los advertimos cuando vemos reunidos muchos ejemplos de ellos.

El estudio y descripción de las formas en que la gente utiliza el lenguaje para realizar cosas es llamado pragmatismo. Este aspecto del lenguaje es muy importante y fácil de omitir. Esto sucede cuando el lenguaje está dando significado adicional. Cobuild posee mucha información sobre pragmatismo y la expone mediante un símbolo especial en cada entrada. Por ejemplo *and things like that* es definido como una expresión utilizada para ensanchar el rango de una lista.

0.2.10. Definición del estilo

La característica más distintiva de Cobuild en su primera versión fué el uso de frases completas en las definiciones. El significado fué establecido de la forma en que una persona ordinaria podría explicárselo a otra.

Generalmente los diccionarios ofrecen definiciones breves y tradicionales, mientras que Cobuild expone definiciones realmente amplias y ricas. Si se observan detenidamente las definiciones particulares se puede apreciar que cada palabra es elegida para ilustrar ciertos aspectos del significado. Y en la medida en que es posible, las palabras utilizadas en una definición son más frecuentes que la palabra que está siendo definida.

Las definiciones cortas no pueden decir demasiado. Por ejemplo, el primer sentido de verbo de *mean* podría ser definido como solo *signify*, que es cierto, pero no es todo lo que se puede decir. Cobuild expone esto: *If you want to know...* es decir que ese sentido surge cuando alguien está en la búsqueda de información. La palabra *if* indica que esta es una opción, pero una perfectamente normal, y *you* nos dice que no es una característica de ningún grupo particular de gente (compararlo con *if a policeman arrests you...*). Entonces la definición dice lo que alguien puede querer saber sobre el significado de una *palabra, código, señal o gesto*, indicando que esas son las típicas clases de temas que serán encontradas con este sentido de *mean*. Solo después de toda esta información viene el equivalente de *signify*: *lo que se refiere a o a que mensaje transmite*. Entonces hay 12 palabras antes de la palabra principal en este sentido, pero cada una de ellas transmite información vital que sería muy difícil de incluir en una definición corta.