

Etiquetado de metáforas lingüísticas en un conjunto de documentos en español

Linguistic metaphor labelling for a Spanish document dataset

Fernando Martínez Santiago, Miguel A. García Cumbreiras,
Manuel C. Díaz Galiano, Arturo Montejo Ráez

Departamento de Informática, Escuela Politécnica Superior
Universidad de Jaén, E-23071 - Jaén
{dofer,magc,amontejo,mcdiaz}@ujaen.es

Resumen: En este trabajo se han etiquetado manualmente las metáforas lingüísticas presentes en la colección de documentos utilizados en SemEval 2013 en la tarea correspondiente a desambiguación léxica en español. El objetivo del trabajo es doble: por un lado realizar una primera aproximación de las dificultades específicas que presenta la identificación de metáforas en el idioma español, y por otro crear un nuevo recurso lingüístico conformado por una colección documental en español que tiene etiquetadas ciertas palabras tanto con su sentido literal como metafórico.

Palabras clave: metáfora lingüística, metáfora cognitiva, recursos lingüísticos, ambigüedad léxica

Abstract: This paper introduces the work performed to manually label linguistic metaphors in the document collection of SemEval 2013, in the Spanish lexical disambiguation task. The objectives of this work are two: first, to make a prior identification of the difficulties inherent in metaphor detection in Spanish, and second, to generate a new linguistic resource as a collection of Spanish documents with certain terms label with both, the literal and the metaphoric sense.

Keywords: linguistic metaphor, cognitive metaphor, linguistic resource, lexical ambiguity

1 Introducción

La metáfora tradicionalmente es entendida como un recurso lingüístico con una finalidad principalmente artística o retórica. Sin embargo, la metáfora también es una herramienta de uso cotidiano que permite explicar un dominio conceptual en términos de otro dominio (Lakoff y Johnson, 2003). Por dominio conceptual entendemos cualquier organización coherente de experiencia. Por ejemplo, en la frase “Alicia no ve el problema” se utiliza una experiencia física (ver) para explicar una experiencia más abstracta (conocimiento). Se establece así un dominio origen y uno destino.

Una metáfora *lingüística* es la expresión lingüística de una metáfora conceptual. En el ejemplo anterior la metáfora lingüística es el verbo “ver”.

Otros ejemplos de metáforas son:

- (i) “Resolveremos el problema más adelante” (*tiempo es movimiento*)

En una primera clasificación podemos hablar de dos tipos de metáforas: conceptual o cognitiva y lingüística.

El proceso de aproximarnos a un dominio conceptual a partir de otro dominio mejor conocido es lo que denominamos metáfora conceptual o cognitiva, como en el ejemplo anterior. Una metáfora conceptual o cognitiva se refiere a la comprensión de una idea, o dominio conceptual, en términos de otro, denominados dominio origen y destino. Siguiendo con el ejemplo anterior, se está asimilando el proceso de ver, que es el dominio destino, con la adquisición de conocimiento, que es el dominio fuente u origen.

- (ii) “Él va sin *dirección* por la vida “ (*vida es viaje*)
- (iii) “Tus declaraciones son *indefendibles*” (argumento es guerra)
- (iv) “Tengo la *mente* puesta en otro lugar” (*mente es objeto físico*)
- (v) “¿Cómo puedo *matar* un proceso?” (*proceso es ser vivo*)

(vi) “ *Parad el mundo, que me bajo*”
(*mundo es transporte*)

En (Kövecses, 2010) se proponen otros criterios de clasificación. Quizás el más usual atiende al grado de convencionalidad de la metáfora, bien en su expresión lingüística (se expresa una metáfora conceptual ya conocida de un modo novedoso) o en los dominios involucrados (se involucran dos dominios conceptuales que es novedoso que participen en la misma metáfora). La frase (vi) es un caso de metáfora no convencional a nivel lingüístico: es una metáfora conceptual frecuente, *vida* es un *viaje*, pero utilizada de un modo original.

Un tercer criterio de clasificación atiende a la riqueza del trasvase de información del dominio fuente al dominio destino. En los casos (i) (ii) (iii) y (v) un dominio fuente presta no solo el concepto si no su estructura al dominio destino. Por ejemplo, al asimilar *tiempo* con *movimiento* inferimos que el tiempo es un ente con una posición y movimientos determinados. Este tipo de metáforas son denominadas *estructurales*, en contraposición a las denominadas metáforas *ontológicas*, más vagas e imprecisas. Por ejemplo, en la frase (iv) nos referimos a la *mente* como un *objeto físico*, es un modo de aproximarnos a la idea de mente. En el presente trabajo:

- se realiza un estudio de metáforas lingüísticas en español,
- se propone un método de etiquetado manual, derivado del propuesto en (Steen, y otros, 2010) y adaptado al caso del español y
- se genera un recurso a partir de la colección de documentos utilizados en SemEval 2013 (Lefever y Hoste, 2013) en la tarea correspondiente a desambiguación léxica en español. Estos documentos han sido etiquetados manualmente con los sentidos metafóricos de ciertas palabras.

Las palabras etiquetadas son los sintagmas nominales desambiguados en la colección, verbos, adjetivos y adverbios. Se trata de un recurso que cuenta con nombres desambiguados y además etiquetados según sean metafóricos o no. Los verbos, que no están desambiguados en la colección original, se han clasificado según sean metafóricos o no debido a la elevada frecuencia con que estos suelen presentar un sentido metafórico: (Cameron, 2003) estimó que el 50% de las metáforas presentes en libros de texto se corresponde con verbos.

El resto del artículo está organizado como sigue: la sección 2 presenta una breve motivación sobre el interés de la metáfora desde el punto de vista del procesamiento del lenguaje natural. En la sección 3 se describe con detalle el procedimiento de etiquetado que se ha seguido. En la sección 4 se presenta la colección documental y el grado de acuerdo entre los anotadores. En la sección 5 indicamos algunas consideraciones sobre el corpus anotado: metáforas encontradas y se discuten algunos casos particulares. El artículo concluye esbozando las líneas de trabajo futuras.

2 *La metáfora en el procesamiento del lenguaje natural*

La metáfora es un concepto ubicuo, cotidiano, y no siempre es un proceso creativo o novedoso. Así pues, ¿es necesario interpretarlas de un modo diferente al de expresiones no metafóricas?

En lo relativo a las metáforas convencionales, existen estudios que demuestran que incluso la comprensión de las metáforas convencionales conlleva la activación en el cerebro de las zonas que se asocian tanto con el dominio fuente como con el dominio destino (Boroditsky, 2001) (Gibbs, Bogdanovich, Sykes y Barr, 1997) (Shutova E. , 2010). Por lo tanto se puede esperar que un sistema que requiera interpretar un texto debe tener acceso a los mismos dominios conceptuales que una persona utilizaría para interpretar ese mismo texto.

Son muchas las tareas donde esta detección e interpretación de las metáforas pueden ayudar.

Por ejemplo, (Shutova, Teufel y Korhonen, 2012) encontraron que un 44% de las traducciones realizadas por Google Translate de inglés al ruso contenían algún error debido a la mala traducción de al menos un término usado metafóricamente, convencional o no. Esto mismo es fácil encontrarlo en traducciones entre inglés y español. Así en la frase “Juan soltó todo lo que sabía”, la interpretación del verbo *soltar* es metafórica, pero convencional. Sin embargo se traduce erróneamente por *give* (dar).

La implicación textual (TE, del inglés *textual entailment*) o los sistemas de búsquedas de respuestas son otras tareas que podrían beneficiarse de un sistema capaz de manejar metáforas. Por ejemplo, en TE la frase “Juan va

a *explotar* si no se calma” implica a “Juan está *enfadado*”. Este ejemplo parte de la metáfora conceptual “*enfado es un líquido caliente en un contenedor cerrado*”.

Por lo tanto se requiere la aplicación del conocimiento sobre el dominio fuente para inferir que el grado de enfado es muy elevado. Este tipo de razonamientos donde se requiere metáforas estructurales con un rico traspaso de conocimiento de un dominio a otro es conocido como *implicación metafórica* (Kövecses, 2010), lo cual es posible cuando el nivel de paralelismo estructural entre el dominio fuente y destino es muy elevado.

Tal como se ha descrito previamente, en este trabajo se ha procedido a etiquetar un corpus utilizado en tareas de desambiguación del sentido de las palabras (*Word Sense Disambiguation*, WSD). La relación entre WSD y metáforas es clara en determinados sentidos de algunas palabras, los cuales tienen un origen metafórico. Tomemos como ejemplo las dos siguientes frases:

(i) “El reclamo podría obligar a los jueces de apelación a *lidiar* con la pregunta”

(ii) “Pero los abogados del señor Hayes han presentado escritos en la corte que *delinean* los argumentos...”

El primer ejemplo *lidiar* se interpreta a partir de un dominio fuente más específico para explicar un dominio más genérico, y ambos dominios están presentes en el diccionario¹: 1. “Luchar con el toro incitándolo y esquivando sus acometidas hasta darle muerte” y 2. “Batallar, pelear”. Claramente el sentido 2 es metafórico a partir del sentido 1, y es posible que un sistema WSD pueda detectar correctamente ambos sentidos. Sin embargo, en otros casos el sentido metafórico de la palabra no se corresponde con un sentido reconocido en un diccionario, por lo que deberían generarse dinámicamente nuevos sentidos acorde al contexto de la palabra. Tal es el caso del verbo “*delinear*” en la frase (ii) donde se está realizando la metáfora conceptual “*argumento es construcción*”². Sin embargo, el único

significado de “*delinear*” es que está recogido en el diccionario de la RAE “trazar las líneas de una figura” por lo que no sería posible establecer tal metáfora conceptual con tan solo la ayuda de diccionarios.

Finalmente, existen estudios que cuantifican que en promedio se encuentra al menos un término usado metafóricamente cada tres frases (Shutova y Teufel, 2010). Se trata pues de un fenómeno que además de requerir su identificación e interpretación en multitud de escenarios, es muy frecuente.

En resumen, la detección e interpretación de la metáfora permite establecer un paralelismo entre el dominio origen y destino. En algunos casos, esto permite inferir nuevo conocimiento en el dominio destino al aplicar en este el conocimiento que se tiene del dominio origen, incluso en el caso de metáforas convencionales. En el caso de las metáforas creativas, éstas implican significados novedosos de palabras, lo que imposibilita asignar correctamente un sentido a tales palabras con la sola ayuda de un diccionario. Finalmente, no se trata de un fenómeno lingüístico poco frecuente, sino que se utilizan en el lenguaje cotidiano, y aun en tales casos es sabido que accedemos tanto al dominio fuente como destino (Kövecses, 2010). Todo ello justifica el interés de la metáfora en el ámbito del procesamiento del lenguaje natural.

En el caso del español podemos encontrar al menos dos corpus desambiguados que han estudiado e incorporado sentidos metafóricos. SenseM (Castellón, 2012) es un corpus constituido por 100 oraciones para cada uno de los 250 verbos más frecuentes del español, etiquetado sintácticamente y semánticamente. AnCora (Taulé, Martí y Recasens, 2008) es un corpus del catalán y del español con distintos niveles de anotación, desde el lema y la categoría morfológica hasta las clases semánticas de los verbos

3 Descripción del procedimiento de etiquetado de metáforas

El procedimiento de etiquetado de metáforas que proponemos está basado en MIP (*Metaphor Identification Procedure*), que fue desarrollado por (Pragglejaz Group, 2007) para el etiquetado de metáforas lingüísticas formadas por una única unidad léxica, generalmente una palabra. En el resto de esta sección se describe este procedimiento, que se ha seguido para etiquetar

¹ Sentidos extraídos del diccionario de la Real Academia Española

² En este ejemplo, hemos considerado construcción como dominio conceptual fuente ya que el dibujo del plano es una de las primeras fases de la construcción, de modo análogo a como la documentación presentada se entiende que constituye una de fase preliminar en elaboración del argumento de la defensa.

un corpus y se destacan algunas diferencias significativas respecto de MIP.

3.1 Procedimiento de etiquetado propuesto

A continuación se muestran los pasos definidos como procedimiento para el etiquetado de metáforas:

1. Lea todo el texto para establecer una comprensión general del significado
2. Determine las unidades léxicas en el texto
3. Establezca el significado de cada unidad léxica en el texto, teniendo en cuenta su contexto, la situación evocada por el texto (significado contextual). Tome en cuenta lo que viene antes y después de la unidad léxica
4. Para cada unidad léxica, determine si tal unidad puede tener, en algún otro contexto, un significado diferente y más *básico*. Para nuestros propósitos, significados *básicos* tienden a ser:

- a. Más concreto (lo que evocan es fácil de imaginar, ver, oír, sentir, oler y saborear)
- b. Relativo al movimiento, o a acciones corporales
- c. Más preciso (en oposición a vago)
- d. Históricamente más antiguo

Los significados *básicos* no son necesariamente los significados más frecuentes de la unidad léxica

En caso que efectivamente la unidad léxica pueda tener en otro contexto un significado más básico: decida si el significado encontrado en el texto puede ser interpretado a partir del significado básico.

5. En caso afirmativo, marque la unidad léxica como metafórica

Directrices adicionales:

En caso de duda en el paso 3, se recomienda buscar la unidad léxica en el diccionario de la Real Academia Española. Si la duda persiste, consultar en otros recursos como BabelNet e intentar identificar el significado contextual y el significado básico entre los provistos en el diccionario

En caso de duda en el paso 4, trate de identificar el dominio fuente y el dominio destino.

Casos particulares que debe considerar:

- La personalización es metafórica si se asigna a una cualidad/funcionalidad a parte que no le corresponde (“el corazón *rugió* enfurecido” es metáfora, “sus ojos *buscaban* los ojos del niño” es sinécdoque, no metáfora)
- Son frecuentes las expresiones multi-palabras donde una de las palabras es usada metafóricamente. Por ejemplo “efecto *invernadero*”. Se deben considerar las unidades léxicas que la conforman separadamente si tal multi-palabra no se encuentra registrada en el diccionario y además el anotador manual no marca la expresión multi-palabra como unidad léxica. En otro caso, se trata de una única unidad léxica constituida por más de una palabra

Las palabras que aparecen en frases o expresiones hechas (una frase o expresión que tiene forma fija, tiene sentido figurado y es de uso común por la mayoría de hablantes de una comunidad lingüística) tal como “se marchó con las *manos vacías*” no deben, en general, considerarse metafóricas.

3.2 Diferencias entre MIP y el procedimiento de etiquetado propuesto

El método descrito difiere de MIP en algunos aspectos. En primer lugar MIP se basa sistemáticamente en diccionarios para que el anotador decida el significado base y los posibles significados metafóricos. Sin embargo, (Lönneker y Eilts, 2004) encontraron que el modo en que se incluyen significados metafóricos en diccionarios es poco sistemático, algunos sentidos aparecen otros no. Por este motivo (Shutova, Devereux y Korhonen, 2013) propone no utilizar diccionario alguno, sino confiar únicamente en el conocimiento del anotador. Posiblemente esta estrategia posibilita etiquetar significados metafóricos que no se encuentran registrados en el diccionario, pero a costa de una menor cohesión entre los anotadores. Nosotros hemos tomado una posición intermedia: al etiquetador no se le obliga a revisar cada término en un diccionario pero, para aquellos términos que le resulten dudosos, sí se le facilita un procedimiento basado en un diccionario, entre otros recursos.

Recursos utilizados. Si bien MIP no es dependiente de un idioma concreto, en su formulación original utiliza diccionarios

específicos del idioma inglés. En nuestro caso, estos recursos se han sustituido por el diccionario de la Real Academia del Español y BabelNet.

Uso de dominios cognitivos. Aunque en el presente trabajo sólo se etiquetan metáforas lingüísticas, detectar un significado base y un significado metafórico conlleva, implícitamente, reconocer que existen dos dominios cognitivos bien diferenciados aun sin llegar a identificarlos. En casos dudosos hemos encontrado útil tratar de identificar explícitamente los dominios fuente y destino. Para ello se facilita a los anotadores dos listas con los dominios fuente y destino más usuales, como las mostradas previamente. Estas listas se han adaptado de las propuestas en (Kövecses, 2010).

Finalmente, el procedimiento propuesto va acompañado de un conjunto de reglas que deben aplicarse en algunos casos particulares, pero relativamente frecuentes. Se ha seguido el criterio que se describe para cada uno de esos casos en (Steen, y otros, 2010). Solo en el caso de las expresiones multi-palabra hemos tenido que seguir un criterio ligeramente diferente. Mientras que para MIP una expresión multi-palabra es una única unidad léxica solo si esta se encuentra en el diccionario, nosotros hemos ampliado este supuesto al caso que el analizador sintáctico que han utilizado marque la expresión multi-palabra como tal. En definitiva, el criterio que se sigue con las expresiones multi-palabra y las expresiones hechas es el mismo: solo se marca metafóricamente sus constituyentes si la expresión completa es utilizada metafóricamente.

4 Descripción y etiquetado de la colección documental

4.1 La colección documental: metáfora y desambiguación

La colección documental etiquetada se corresponde con la colección de textos suministrada para la evaluación de sistemas en la tarea *Multilingual Word Sense Disambiguation* correspondiente al *SemEval 2013* (Lefever y Hoste, 2013). Es una colección de 13 textos relativos a noticias con un total de 306 oraciones conformados por 8779 palabras, de las cuales 1198 son verbos y 1481 están desambiguadas léxicamente. Estos textos provienen del corpus paralelo Europarl, y

fueron desambiguados y lematizados manualmente por los organizadores de la tarea.

Se ha seleccionado esta colección porque existe una clara relación entre ambigüedad léxica y metáfora. Por ejemplo, MIP utiliza los sentidos registrados en diccionarios para diferenciar el sentido metafórico y sentido básico. En consecuencia MIP se puede interpretar como un procedimiento de desambiguación manual con especial énfasis en metáfora (Shutova y Teufel, 2010). La diferencia clave entre ambas tareas es que, mientras en el caso de la ambigüedad léxica se parte de un conjunto predefinido de sentidos, en el caso de la metáfora estos significados deben ser inducidos automáticamente lo que impide utilizar un enfoque tradicional de desambiguación léxica en el ámbito de la metáfora lingüística (Shutova, Teufel y Korhonen, 2012), (Pustejovsky, 1995). Sin embargo, a pesar de la estrecha relación entre ambigüedad léxica y metáfora, hasta donde nuestro conocimiento llega no existe ningún recurso, en español u otro idioma, que etiquete los términos en ambas vertientes.

4.2 Esquema de anotación

El corpus ha sido anotado por tres personas que tienen el español como idioma nativo, pero sin conocimientos específicos relativos a metáforas conceptuales. Por ello previamente se les entrenó a través de algunos ejemplos. Luego se les facilitaron las instrucciones para el procedimiento de etiquetado descrito en la sección 3.1. Con la finalidad de comprobar que se había asimilado correctamente la tarea, se les solicitó que etiquetaran cinco frases de ejemplo siguiendo tal procedimiento y se revisó el resultado de forma conjunta. Finalmente se facilitó la colección documental con las palabras resaltadas a etiquetar como metafóricas o no. Del total de 8779 palabras del corpus, sólo se han anotado nombres y verbos, de tal modo que cada anotador ha etiquetado 2679 unidades léxicas.

Dado que la tarea propuesta es inherentemente subjetiva, es esencial cuantificar el acuerdo existente entre los anotadores, esto es, la similitud de las anotaciones producidas por diferentes anotadores. Se evaluó la confiabilidad del sistema de anotación propuesto en términos del valor estadístico κ (Siegel y Castellan, 1988) obteniendo un valor $\kappa = 0,62$ ($n = 2$; $N = 2679$;

$k = 3$), donde n representa el número de categorías (metafórico/no metafórico), N para el número de casos anotados y k el número de anotadores. Este nivel de acuerdo es considerado sustancial (Landis y Koch, 1977).

5 *Análisis de los resultados*

En el presente trabajo se ha seguido el criterio de que para que un término sea considerado metafórico debe haber consenso de al menos dos anotadores. De este modo, se ha encontrado que 113 de las 306 oraciones contienen al menos un término metafórico, lo que supone que se encuentra al menos una metáfora cada tres oraciones aproximadamente. Este resultado es consistente con (Shutova y Teufel, 2010).

	Metafóricos	Literales	Total
Verbos	110	1088	1198
Nombres	176	1305	1481
Total	286	1393	2679

Tabla 1. Metáforas anotadas

En cuanto al número de verbos y nombres etiquetados metafóricamente (Tabla 1), se han contado 110 verbos utilizados metafóricamente, lo que supone un 10,9% del total. En el caso de los nombres, en 176 casos han sido utilizados metafóricamente (8,4%). Esto supone que un 39% del total de las metáforas tiene origen verbal. Este porcentaje es sensiblemente inferior al 50% que apunta (Cameron, 2003), si bien el motivo puede ser que mientras el trabajo de Cameron está centrado en textos educativos, el presente trabajo parte de un colección de textos cuyo origen son noticias publicadas en diversos medios de comunicación. Además, la restricción impuesta de consenso entre dos anotadores resta necesariamente el número total de metáforas anotadas finalmente.

5.1 Consideraciones adicionales

5.1.1 Metáforas novedosas o creativas

A los anotadores también se les solicitó que indicaran aquellas metáforas que son nóveles como tales esto es, metáforas cuyo significado no es posible encontrar en un diccionario. Se etiquetaron tan sólo 13 metáforas de este tipo, menos del 5% del total. Aun siendo un número bajo, una cuestión que surge con las metáforas novedosas es cómo actualmente se están marcando tales palabras en las colecciones

textuales desambiguadas manualmente. Esta cuestión es relevante puesto que guía el modo en que los sistemas de desambiguación son evaluados. Por ejemplo, en la colección utilizada en el presente trabajo, encontramos el siguiente ejemplo:

(i) “Al principio de su mandato, Nicolas_Sarkozy quería ir a buscar los puntos de crecimiento que le faltaban con los *dientes*”

En la frase precedente, “*dientes*” es marcado como metáfora creativa, ya que no es posible encontrar en un diccionario el significado que se le atribuye en esta frase. Además, *dientes* no es metonimia ya que aunque estén representando a la persona, no se usan del modo que en la frase se recoge. Es por lo tanto un significado novedoso el que se le está atribuyendo a la palabra y sin embargo tal palabra está etiquetada manualmente con el sentido más usual de “*dientes*”, lo cual desde nuestro punto de vista es erróneo: un sistema que interpretara la palabra con el significado literal nunca podría interpretar correctamente el significado real de la frase.

Una situación similar se encuentra en la metonimia y la sinécdoque, ya que las utilizamos para evocar un concepto a partir de otro concepto, si bien se diferencian en el mecanismo que activa el concepto destino. Es frecuente encontrar ejemplos de metonimias en textos informativos. De nuevo es posible encontrar en el diccionario significados que tienen su origen en la metonimia, pero otras son novedosas. Por ejemplo:

(ii) “...manifestó el *portavoz* del Ministerio de Petróleo Assam Jihad”

El significado de “*portavoz*” tiene origen metonímico, pero tal significado está reflejado en el diccionario por lo que cabe la desambiguación correcta del término. Sin embargo, dada la frase:

(iii) “...el asesinato deja cuatro *bocas* que alimentar “

El término “*bocas*” es una sinécdoque de los hijos, y tal significado no se encuentra en el diccionario. Entonces ¿cómo debería desambiguarse tal término?

5.1.2 Significados base inusuales

En ocasiones es posible encontrar significados muy precisos y concretos, pero muy poco conocidos, o bien muy específicos de determinados ámbitos. En tales casos esos

significados no se han considerado como significados base: aquellas cuyo significado base está ya en desuso y por lo tanto no se concluye que la palabra sea utilizada metafóricamente. Ejemplos son:

(iv) “La *expulsión* presuntamente urdida por dos jugadores del Real_Madrid”

(v) “El texto, que podría proporcionar la base para un *acuerdo* político final”

En la frase (iv) el término *expulsión* no es metafórico a pesar de que *expulsión* también significa “Golpe que da el diestro sacudiendo violentamente con la fuerza de su espada la flaqueza de la del contrario, para desarmarlo”, lo cual es un significado más preciso, pero generalmente desconocido. En el caso de la frase (v), *acuerdo* tampoco es metáfora. *Acuerdo* en el sentido de “*templar instrumentos musicales*” es el significado base porque es un significado más concreto y además, posiblemente más antiguo. Pero aun así no se le da sentido metafórico al *acuerdo* del ejemplo (frase v) porque el significado más concreto y/o antiguo de la palabra es prácticamente desconocido. Lo que subyace en estos dos ejemplos es que, para que haya metáfora, deben ser conocidos los dos sentidos de la palabra (en el caso de la metáfora creativa uno de los dos sentidos no se conoce de antemano, pero se infiere a partir del sentido base y del contexto).

5.1.3 Multi-palabras y unidades léxicas

Algunos términos utilizados metafóricamente no han podido etiquetarse como tales porque en la colección documental aparecían como parte de una expresión multi-palabra. Por ejemplo la expresión *economía emergente* es marcada como una expresión multi-palabra en la colección documental, y por lo tanto debe evaluarse como una única unidad léxica. Sin embargo, *economía emergente* no se encuentra en el diccionario de la RAE ni en BabelNet, lo que habría permitido evaluar *emergente* como una unidad léxica y, consecuentemente, como una metáfora. La solución que se ha adoptado es ampliar la definición de unidad léxica en el caso de multi-palabras: una multi-palabra es una única unidad léxica si se encuentra recogida en el diccionario o bien es marcada como tal por el analizador sintáctico. Este mismo criterio se ha seguido en el caso de las expresiones hechas.

Por ejemplo en la frase (vi) la unidad léxica *mano* no es marcada como metafórica, ya que está siendo utilizada del modo usual en una expresión bien conocida (*ir de la mano*)

(vi) “Con esto los institutos pretenden liberarse de las restricciones que habían *ido de la mano* con la aceptación del dinero”

5.1.4 Personalización

Siguiendo el enfoque propuesto en MIP, la personalización no se ha considerado en general metáfora, salvo que la metáfora sea en relación a la persona o personas referidas mediante la personalización. Esto es, primero se interpreta la personalización, y luego se evalúa si hay metáfora. Por ejemplo:

(vii) “La *ONU* *esboza* un plan para reducir las emisiones”

(viii) “La *Unión Europea* también *estimuló* las conversaciones”

En (vii) el término *ONU* no se refiere a la ONU, sino a algunas personas que pertenecen a esa organización, por lo tanto ni “*ONU*” ni *esbozar* constituye metáfora. En (viii) *Unión Europea* es un caso similar, pero una vez asumido que se refiere a las personas y no a la organización, sí cabe interpretar *estimuló* de un modo metafórico ya que este término tiene un significado base más específico (“agujonear, picar, punzar” en comparación con “incitar, excitar con viveza la ejecución de algo”).

6 Conclusiones y trabajo futuro

La metáfora es un concepto resbaladizo y ubicuo; quizás por ello es difícil encontrar trabajos relativos al etiquetado, identificación e interpretación de la metáfora en el caso del idioma español. Sin embargo, es un fenómeno lingüístico que debe ser abordado en aquellas tareas que requieran acceder al significado del texto. En este trabajo se presenta un método de etiquetado manual de metáforas para el caso del español, basado en MIP. Tal método ha sido validado sobre una colección documental con verbos y nombres etiquetados según sean usados con un sentido metafórico o no. El acuerdo alcanzado entre los anotadores es considerado sustancial y demuestra la validez del método propuesto. Además, los nombres etiquetados también están desambiguados, lo cual permite en un futuro evaluar el desempeño de los sistemas de desambiguación automática en aquellos términos que son usados con un

sentido metafórico. Como trabajo futuro también se propone aplicar el método de etiquetado manual propuesto en otras colecciones documentales, con la finalidad de estudiar la incidencia de la metáfora en tareas donde el acceso al significado sea relevante, tales como la implicación textual o el análisis de sentimientos.

Agradecimientos

Esta investigación ha sido subvencionada por el proyecto ATTOS (TIN2012-38536-C03-01) financiado por el Gobierno de España y el proyecto europeo FIRST (FP7-287607).

Bibliografía

- Boroditsky, L. 2001. Does Language Shape Thought?: Mandarin and English Speakers' Conceptions of Time. *Cognitive Psychology*, 43, 1-22.
- Cameron, L. 2003. *Metaphor in Educational Discourse*. London: Continuum.
- Castellón, I. 2012. Constitución de un corpus de semántica verbal del español: Metodología de anotación de núcleos argumentales. *RLA*, 13-38.
- Gibbs, R. W., Bogdanovich, J. M., Sykes, J. R. y Barr, D. J. 1997. Metaphor in Idiom Comprehension. *Journal of memory and language*, 37, 141-154.
- Kövecses, Z. 2010. *Metaphor*. Oxford University Press; 2nd edition.
- Lakoff, G. y Johnson, M. 2003. *Metaphors We Live By*. University Of Chicago Press; 2nd edition.
- Landis, J. y Koch, G. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Lefever, E. y Hoste, V. 2013. SemEval-2013 Task 10: Cross-lingual Word Sense Disambiguation. *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)* (págs. 158-166). Atlanta, Georgia, USA: Association for Computational Linguistics.
- Lönneker, B. y Eilts, C. 2004. A current resource and future perspectives for enriching wordnets with. *Proceedings of the second international wordnet conference (GWC 2004)*, (págs. 157-162). Brno.
- Martin, J. 1990. *A Computational Model of Metaphor Interpretation*. San Diego, CA, USA: Academic Press.
- Pragglejaz Group. 2007. MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22, 1-39.
- Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge, MA: MIT Press.
- Shutova, E. 2010. Models of Metaphor in NLP. *Proceedings of ACL 2010*.
- Shutova, E. y Teufel, S. 2010. Metaphor corpus annotated for source-target domain mappings. *Proceedings of LREC 2010*, (págs. 3255-3261). Malta.
- Shutova, E., Devereux, B. y Korhonen, A. 2013. Conceptual metaphor theory meets the data: a corpus-based human annotation study. *Language Resources & Evaluation*, 47(4), 1261-1284.
- Shutova, E., Teufel, S. y Korhonen, A. 2012. Statistical Metaphor Processing. *Computational Linguistics*, 39(2).
- Siegel, S. y Castellan, N. J. 1988. *Nonparametric statistics for the behavioral sciences*. New York, USA: McGraw-Hill Book Company.
- Steen, G., Dorst, A., Hermann, J., Kaal, A., Krennmayr, T. y Pasma, T. 2010. *A Method for Linguistic Metaphor Identification*. John Benjamins Publishing Company.
- Taulé, M., Martí, M. y Recasens, M. 2008. AnCora: Multilevel Annotated Corpora for Catalan and Spanish. *Proceedings of the 6th conference on International Language Resources and Evaluation*, 96-101.