

Propuesta de anotación semántica para una base de datos paremiológica*

Elena Alonso Pérez-Ávila
Università di Roma “La Sapienza” / Università di Torino

An electronic tool, such as an on line multilingual paremiological database, that would enable researchers or translators to search paremiological units of many languages and manipulate the information stored about them more easily, would greatly benefit the field of paremiology.

This paper deals with how the information provided by SpanishWordNet and MultiWordnet may be used to tag semantically paremiological units in Spanish and Italian within the database. In other words, tags related to the WordNet ontology are attached to each term in the proverb in order to provide more information about the domain that the proverb belongs to. We propose this annotation as a methodology to classify paremiological units that can be shared by different linguistic communities since it is based on an already widely used lexical resource developed in many languages: WordNet.

Unfortunately, it is not possible to tag the proverb as a whole unit due to its particular features: the meaning of the whole proverb cannot be easily derived from the meanings of its separate components. At the moment we are trying to supply the database with as much information as possible on the semantics of the components of the proverbs and the relations, such as hyponymy, that those parts present with the rest of the lexicon.

1. Introducción: una base de datos paremiológica

Las unidades paremiológicas se han visto relegadas a un segundo plano en las investigaciones lingüísticas. Este hecho puede deberse al descenso que se ha manifestado en su uso y a las peculiares características que sitúan a estas unidades a medio camino entre la lingüística y los estudios antropológicos y de folklore.¹ En el ámbito del proyecto *Gnome*² se pretende ofrecer un instrumento que favorezca la incorporación de estos elementos de las lenguas, las paremias, a la lingüística actual.

El proyecto *Gnome* se contextualiza en una línea de investigación que tiene como meta resolver los problemas lexicográficos y lexicológicos relativos a las unidades paremiológicas. El trabajo consiste en recuperar las colecciones paremiográficas de distintas lenguas, digitalizarlas y organizarlas en una base de datos para que puedan tener acceso a la materia paremiológica tanto profesionales de las lenguas como personas interesadas en la paremiología. Los integrantes de este proyecto creemos que una base de datos digitalizada y consultable en Internet es el medio idóneo para poder insertar todas las etiquetas necesarias en las que almacenar los datos pertinentes de cada paremia: las fuentes de las que procede; la información léxica, gramatical, semántica y cultural; el tipo de textos en que ocurre, etc., y, asimismo, se podrán desarrollar programas informáticos —o utilizar los ya existentes— con el fin de posibilitar la recuperación

* Este trabajo se ha llevado a cabo durante la realización de un período de formación organizado por el ENEA, *Ente per le Nuove Tecnologie, l'Energia e l'Ambiente*, con sede en Roma.

¹ Así lo señalaba Casares (1950: 203): “(...) no parecerá injustificado sostener que el estudio de los proverbios como tales y no como textos de lengua (...) debe dejarse para la paremiología, ciencia que guarda más relación con el folklore y con la psicología colectiva o étnica que con las disciplinas lingüísticas”.

² Proyecto en el que participan diversas universidades españolas e italianas e instituciones públicas de ambos países. Más información sobre el proyecto en <http://www.gr-gnome.eu/>.

de los datos que le interesen a cada usuario de forma rápida y eficiente. Si bien es cierto que no se podrán resolver las dificultades de la traducción automática de paremias, este sistema de anotación múltiple podrá constituir una óptima herramienta a disposición de los traductores que encuentren dificultades a la hora de encontrar las correspondencias paremiológicas en otras lenguas, especialmente si se utilizan programas adecuados de recuperación de información.

El objetivo de este trabajo es el de proponer una metodología de anotación semántica de unidades paremiológicas que favorezca la comprensión de las mismas, contribuya a su clasificación en el marco de una base de datos y que, al mismo tiempo, se rija por unos criterios que puedan ser compartidos por distintos grupos de investigación y aplicables en distintas lenguas.

2. Heterogeneidad en las clasificaciones paremiográficas

Durante la fase de elaboración del corpus de paremias que formará la base de datos multilingüe hemos detectado una falta de sistematicidad en las clasificaciones de la materia paremiológica existentes. Esta heterogeneidad puede dificultar que un usuario localice la correspondencia de una paremia en otra lengua o, incluso, una misma paremia en dos repertorios distintos de una lengua.

1. Chi han ducati, signori son chiamati.

Esta unidad paremiológica se localiza bajo el epígrafe “avere” en la clasificación por medio de una palabra clave efectuada por Strafforello (1883). En cambio, Gella Iturriaga (1983) clasifica la correspondencia española del ejemplo 1 en el capítulo denominado “ducado”:

2. Los que han ducados, señores son llamados.

Comprobamos que el mayor inconveniente que presenta este sistema de clasificación es que la elección de la palabra clave depende del juicio subjetivo del paremiógrafo que, como vemos, puede variar en cada caso. Tampoco facilitan el proceso de localización de paremias las clasificaciones onomasiológicas —también denominadas ideológico-temáticas—, en las que, de nuevo, la organización se somete a la subjetividad del autor. Así, Martínez Kleiser (1993) incluye el ejemplo 2 entre las paremias clasificadas como “Dinero. Adulación al dinero”, mientras que Giusti-Capponi (1853 [2001]) clasifica bajo “Povertà, ricchezza” una variante del ejemplo 1:

3. Quei ch’han ducati, signori son chiamati.

Resulta difícil encontrar una justificación que privilegie una propuesta sobre las demás dado que en cada una se sigue la lógica organizativa que tenía en mente su autor, cuyos criterios no siempre se rigen por principios estrictamente científicos. Esta circunstancia se ve agravada por el hecho de que los criterios taxonómicos usados en los diccionarios tradicionales determinan generalmente una elección a la hora de asignar las paremias a un término o a un campo conceptual, en detrimento de otras opciones válidas. A este respecto habría que añadir además la tendencia general en lexicografía a reproducir en gran medida las fuentes preexistentes, lo que a veces implica el solapamiento de criterios híbridos. Consideramos, por tanto, oportuno aprovechar el ambiente informático en la medida en que favorece la labor de etiquetado múltiple para proponer un sistema de clasificación de paremias basado en un método de anotación formal para que pueda ser procesado por sistemas informáticos.

3. Uso de bases de datos léxicas como sistema de anotación semántica

Hemos optado por el uso de dos bases de datos léxicas, SpanishWordNet y MultiWordNet, para proponer una metodología de anotación semántica de las paremias.

Estas bases de datos léxicas se insertan en el marco de tres proyectos: EuroWordNet, MultiWordNet y Meaning. EuroWordNet (EWN) (Vossen 1998) se sustenta en los principios psicolingüísticos de ordenación del léxico de WordNet (Miller *et al.* 1990), en el que los conceptos están representados por grupos de sinónimos, *synsets*, que a su vez están relacionados entre sí a través de relaciones semánticas y léxicas formando una gran red de palabras. EWN tiene como finalidad la creación de bases de datos léxicas independientes en distintas lenguas para después agrupar todos esos wordnets en una base de datos única y multilingüe. En nuestro trabajo estamos utilizando uno de los wordnets nacidos en el seno de este proyecto:

SpanishWordNet (SWN). Además, hemos aprovechado las nuevas aportaciones del proyecto MultiWordNet (MWN) (Pianta *et al.* 2002), en su versión para la lengua italiana e inglesa, que mantiene las líneas generales de estructuración semántica pero con ciertas divergencias con respecto a los principios establecidos en EWN. Esos dos proyectos han tenido continuidad en el proyecto Meaning (Atserias *et al.* 2003) que prosigue la línea de EWN de crear una base de datos léxica multilingüe; en este caso han trabajado con cinco lenguas: el vasco, el catalán, el inglés, el italiano y el español. En la construcción del Meaning, *multilingual Central Repository* se han aprovechado los recursos de EWN, como el índice interlingüístico y las ontologías; y de MWN, como los *WordNet Domains*, pero también se ha añadido mucha más información léxica y más relaciones entre los *synsets*.

WordNet se ha convertido, en los últimos años, en un recurso léxico imprescindible para múltiples aplicaciones, especialmente en las aproximaciones al procesamiento semántico, llevadas a cabo en todo el mundo en el ámbito del procesamiento del lenguaje natural. Es por este motivo por el que consideramos que constituye una herramienta de gran ayuda para establecer unos criterios de anotación más objetivos que las intuiciones del estudioso que, aunque puedan ser más acertadas en algunos casos particulares, dificultan la aplicación de un mismo sistema de clasificación en distintas lenguas y no favorecen que distintos grupos de trabajo puedan compartir la información.

En el ámbito de trabajo de WordNet se han formalizado unidades léxicas simples y también algunas unidades fraseológicas entre las que se incluyen algunas colocaciones; por ejemplo, hemos localizado la colocación “propiedad privada” en el *synset* formado por “bien 4, patrimonio 2, propiedad privada 1, pertenencia 2”. No se han codificado, ni en Princeton WordNet (Miller *et al.* 1990) ni en ninguno de los wordnets de otras lenguas, ni expresiones idiomáticas ni paremias, pues su propia naturaleza lingüística obstaculiza su procesamiento (Fellbaum, 1998).

Por tanto, al utilizar WordNet como sistema de anotación nos hemos preguntado si es posible etiquetar una paremia completa a través de un sólo concepto, un sólo *synset*, y la respuesta ha sido negativa. El hecho de que constituyan textos autónomos, con peculiaridades tanto sintácticas como semánticas, implica necesariamente una descripción semántica compleja: no se trata de unidades sintagmáticas simples potencialmente sustituibles por un solo concepto. Muchos especialistas han advertido este problema, y así Casares (1969: 198) señala que “(...) el refrán no es un hecho expresivo reducible a una equivalencia conceptual unitaria”; idea que comparte Fernández-Sevilla (1985: 93-94): “[los refranes] sólo pueden ser explicados por perífrasis complejas (...). El refrán, por ser un texto complejo, tiene sentido complejo que se percibe como resultante de una o más asociaciones”.

En conclusión, no procederemos a la anotación del significado global de cada paremia, sino que se llevará a cabo una anotación múltiple a través del etiquetado de cada término de las paremias. No hemos tomado en consideración aún diferentes aspectos pragmáticos, implícitos en la misma naturaleza de los diferentes tipos de paremias, que por obvias razones de tendencia a la universalidad —entendida como validez en todos los mundos posibles, en virtud del carácter tautológico con el que algunas se utilizan—, de valor alegórico, analógico o ejemplar, hacen de las paremias un recurso expresivo con una referencialidad muy variable. Como es lógico, toda aproximación a la elaboración de instrumentos informáticos para el tratamiento automático o automático asistido del lenguaje se ha de mover en un nivel heurístico, buscando criterios viables que sucesivamente han de trabajarse para alcanzar los objetivos propuestos. Será necesario, por tanto, añadir fases sucesivas de anotación para poder dar cuenta de la complejidad semántica de las paremias.

4. Metodología

Como se indica en el apartado precedente, nuestra propuesta de anotación no pretende ser unívoca ni inequívoca —dada la especial naturaleza de las paremias, consistiría una empresa imposible—, sino que se trata de llevar a cabo un sistema de anotación múltiple que enriquezca la información de la que se dispone para cada unidad, de manera que se mejore la comprensión

de las mismas y se incrementen las posibilidades de localización de correspondencias paremiológicas en otras lenguas.

Para la elaboración de este estudio hemos trabajado con un reducido corpus de paremias relativas al dominio conceptual del dinero y procedentes de diversas recopilaciones paremiográficas en lengua española y en lengua italiana.

En la primera fase de anotación semántica, hemos utilizado para procesar el corpus de paremias españolas Freeling 1.5³ (Atserias *et al.* 2006), que efectúa una anotación tanto gramatical como semántica, esta última basada precisamente en los datos de SWN. En un primer momento, consideramos que Freeling constituía una herramienta útil ya que proporciona rápida y automáticamente todos los *synsets* asociados a un *word form*, es decir, a un significante. A continuación, el investigador deberá eliminar los *synsets* que no se corresponden con cada contexto y corregir los problemas que se detecten en las etiquetas establecidas automáticamente por Freeling 1.5. No obstante, en las fases sucesivas fuimos conscientes de que los errores y lagunas en la anotación semántica automática resultaban demasiado numerosos, posiblemente debido al hecho de que únicamente un subgrupo de los *synsets* de SWN ha sido añadido a Freeling 1.5⁴, de manera que se hizo necesario completar siempre el análisis mediante una búsqueda manual en la base de datos de SWN.

Incluimos a continuación un ejemplo del análisis proporcionado por Freeling 1.5 de la siguiente unidad paremiológica:

4. A fuerza de duros caen los más fuertes muros.

A_fuerza_de_a_fuerza_de SPS00

duros duro AQ0MP0 00710342:01106903:01102688:02328596:02211458:01247049:

00797350:00711482:02032175:02327618:02326527:00713158:01731319:01198741:01016585:

01080240:02326025:00582699:00973806:01109786:01110850:01107950

caen caer VMIP3P0 01342394:01345838:00100153:01353782:00301476:01251646:01781656:

01883959:01061625:01787446:00237141:01575833:01624789:01749892:00300889:01343793

los el DA0MP0

más más RG

fuertes fuerte AQ0CP0

02210262:01393833:01448254:02208726:01965724:01141075:01142322:

01451021:01964884:01755935:00673790:01139298:01732107:02211044:00935294:02165307:

01123468:01133353:01109786:01399036:01015627:01112066

muros muro NCMP000 03585958:06796761

.. Fp

Figura 1. Anotación gramatical y semántica realizada por Freeling 1.5

En este caso, al llevar a cabo la comprobación del *output* que ha producido Freeling 1.5 hemos detectado que el término “duro” está etiquetado únicamente como adjetivo y no como sustantivo y, por tanto, no se incluyen los *synsets* que le corresponderían como sustantivo. Hemos procedido entonces a la búsqueda manual en SWN para poder así anotar “duros” con dos etiquetas distintas, de forma que cada uno de los términos se vincule a un dominio semántico distinto. A través de este lenguaje formal, se especifica también la categoría gramatical a la que pertenece cada uno de los significados a los que está asociado este significante:

³ Se trata de un analizador libre para procesar el lenguaje natural basado en las etiquetas propuestas por el grupo EAGLES para las lenguas europeas. Ha sido creado por el TALP Research Centre de la Universitat Politècnica de Catalunya. Más información en: <http://garraf.epsevg.upc.es/freeling/index.php>. [Consulta del 20 de marzo de 2008].

⁴ Dado que se trata de una aplicación *open source* existe la posibilidad de ampliar ese conjunto de *synsets* a todos los existentes en SWN.

a) n 3 [50000419] [Moneda española equivalente a cinco pesetas.].

b) a 3 [01102688].⁵

De esta manera se puede disolver la ambigüedad con la que juega esta paremia para potenciar su capacidad expresiva.

Para poder completar y mejorar la información que se obtiene de Freeling 1.5 en un ámbito concreto, hemos desarrollado una descripción formal del dominio conceptual del dinero, tanto en lengua española como italiana, siguiendo los criterios de organización del léxico que propone WordNet. Dicha descripción formal se ha efectuado interrogando las bases de datos de SWN y MWN por medio de la base de datos *open source* MySQL y de la interfaz de consulta de Meaning⁶. Exponemos a continuación una figura a través de la que queremos mostrar un ejemplo extraído de la red semántica⁷ obtenida como resultado de estas búsquedas y que se basa en la relación semántica fundamental de WordNet: la relación de inclusión, hiponimia o *isa-relation*.

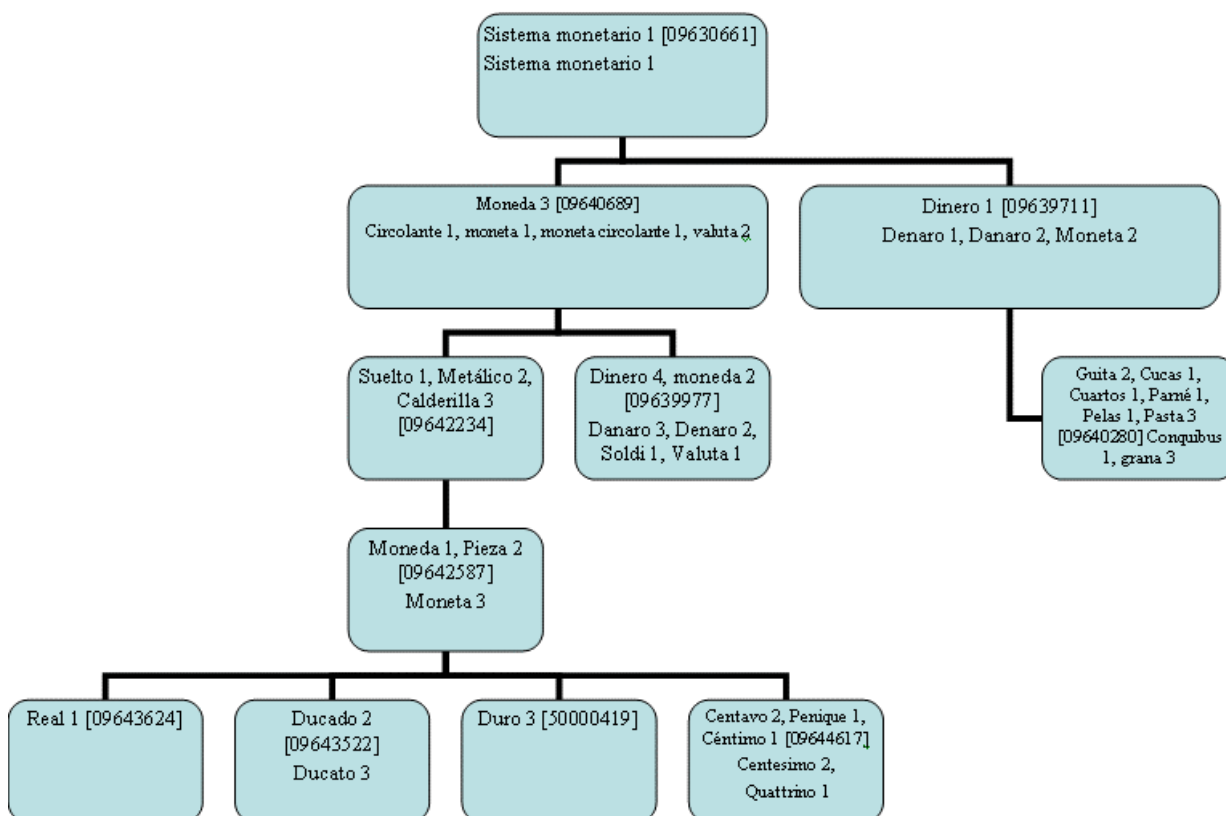


Figura 2. Modelo de referencia de la red semántica relativa al ámbito del dinero⁸

Gracias a la descripción formal de este dominio, las relaciones semánticas de sinonimia, hiperonimia e hiponimia se vuelven detectables por un ordenador. En la siguiente fase hemos procedido a la anotación de las unidades paremiológicas por medio de las etiquetas obtenidas,

⁵ Se trata de un caso de laguna en SWN, pues no se ofrece ni definición ni ejemplo de este concepto, remiten a la explicación de WordNet 1.6 para la lengua inglesa: *not yielding to pressure or easily penetrated*, y señalan que es un atributo de Dureza n 03862786.

⁶ Disponible en <http://www.lsi.upc.es/~nlp/meaning/> [Fecha de consulta: 20 de Marzo de 2008].

⁷ Constituye un mero ejemplo representativo: el árbol de hipónimos de este dominio es demasiado amplio como para ser recogido aquí. Hemos incluido sólo algunos de los hipónimos de cada *synset* para poder proporcionar una idea aproximada del trabajo realizado.

⁸ El *synset* de SWN ha sido situado antes del código que lo representa, el *offset*; mientras que el *synset* italiano (MWN) se sitúa debajo. Como puede observarse, no todos los conceptos se han lexicalizado en ambas lenguas.

los *offset*. Así, en los ejemplos 1, 2 y 3 se anota manualmente los términos “ducados” y “ducati” con el *offset* correspondiente: [09643522] que aporta la siguiente información: i) indica la clase de palabras a la que pertenecen las unidades que forman el *synset*, en este caso, sustantivos; ii) puede ofrecer también una definición; iii) señala su localización en la red semántica: es un cohipónimo de otros muchos tipos de monedas e hipónimo de “moneda 1, pieza 2”, y iv) sirve para desambiguar significados en los casos de palabras polisémicas, en este caso se diferencia “ducado 2”, hipónimo de “moneda 1, pieza 2” de “ducado 1”, hipónimo de “dominio 3, territorio 1” y de “ducado 3”, hipónimo de “categoría 2, rango 2, jerarquía 1, standing 1”.

De esta manera, aunque no se haya logrado una explicación global de la paremia a anotar —el poder del dinero que supera los títulos de la nobleza—, sí se puede saber de forma rápida y automática que tiene una relación con el ámbito del dinero y que se está hablando de un tipo de moneda. Se obtiene así mucha más información sobre el léxico de la que aparece en las entradas de un diccionario. Además, tanto la paremia española, el ejemplo 2, como las italianas, 1 y 3, se anotarán por medio del mismo *synset*. Así, en este caso, no resultará difícil localizar las correspondencias paremiológicas en ambas lenguas, ya que aparecerán anotadas con la misma etiqueta. Como hemos señalado al analizar el ejemplo 4, se podrá añadir más de una etiqueta semántica a un mismo término y, por supuesto, añadiremos varias a cada paremia.

5. La anotación semántica como sistema de organización de las unidades paremiológicas en la red semántica

Acabamos de indicar cómo la aplicación de este sistema de anotación puede favorecer la localización⁹ de correspondencias paremiológicas literales —coincidencia en forma y sentido— en distintas lenguas,¹⁰ pero también puede constituir un instrumento útil para identificar tanto variantes como correspondencias conceptuales —coincidencia de sentido, pero no de forma—, que son las más frecuentes en las paremias.

5. Hombre sin parné, como muerto *jié*.

Esta unidad paremiológica (Gella Iturriaga, 1982) puede causar perplejidad tanto a un hablante no nativo de español como a un nativo poco familiarizado con el léxico caló. Mediante la búsqueda de “parné” en SWN conoceremos una serie de sinónimos que pueden presentar el mismo significado en este contexto y que conforman el *synset* al que pertenece: “guita, cucas, cuartos, pelás y pasta” [09640280]. Esta información favorece un primer acercamiento al significado de este término. A continuación, efectuamos la búsqueda del mismo *offset* en MWN y así obtenemos el siguiente *synset*: “conquibus, grana”. No hemos localizado la paremia italiana correspondiente en la que aparezca alguno de los términos que forman parte del *synset* italiano, pero gracias a nuestra metodología de anotación sabemos la situación del *synset* de “parné” en nuestra red semántica del dinero, de manera que podremos buscar paremias anotadas con *synsets* conceptualmente cercanos:

6. Uomo senza soldi è un morto che cammina.

En el ejemplo 6 (Franceschi *et al.* 2000), se ha anotado “*soldi*” con la etiqueta ofrecida por MWN [09639977] “danaro, denaro, soldi, valuta”. Si bien este *synset* no pertenece al mismo hiperónimo directo que “parné”, sí se encuentra en una cercanía conceptual bastante considerable, como se puede comprobar en la figura 2. Asimismo, hemos localizado en Martínez Kleiser (1993) una variante del ejemplo 5:

7. Hombre sin dinero, por muerto lo tengo.

Se puede anotar “dinero” tanto con la misma etiqueta que en el caso de 6, como con la etiqueta del hiperónimo directo de “parné”, en cualquier caso, nuestro sistema de anotación permite

⁹ Esta localización de paremias tendrá que ser necesariamente manual: las etiquetas semánticas propuestas constituyen únicamente una ayuda.

¹⁰ Utilizamos la terminología empleada por Sevilla (2000).

explicitar las relaciones semánticas entre los términos de estas paremias y puede ser una guía a la hora de localizar correspondencias.

6. Conclusiones y nuevas perspectivas

El objetivo de este artículo ha sido el de proponer unos criterios de anotación de paremias españolas e italianas que aporten información semántica a cada unidad paremiológica basada en los principios de WordNet, es decir, se definen las unidades léxicas no de manera absoluta, sino en relación con otras unidades. Dicha información constituye una gran ayuda para mejorar la comprensión del significado de cada paremia y para localizar correspondencias paremiológicas en otras lenguas. Asimismo, estas etiquetas semánticas pueden servir como sistema de clasificación de las paremias presentes en la base de datos del proyecto *Gnome*, en cuyo marco se contextualiza este trabajo.

La ventaja de usar como fuente de información léxica los wordnets disponibles para varias lenguas radica en que son recursos léxicos extensamente utilizados en el ámbito de la inteligencia artificial y su uso es compartido por distintas comunidades lingüísticas. Así, hemos considerado una metodología que parte de una base homogénea para distintas lenguas, la que proponen los distintos wordnets, y que admite el desarrollo autónomo de cada grupo integrante del proyecto.

Esta metodología no ofrece el significado global de la paremia, ni una asociación de la misma a un sólo concepto en el que se pueda condensar la idea que la define, ni podrá constituir un sistema de traducción automática de proverbios —ideal para la base de datos multilingüe—, pero sí que supone un paso adelante en los intentos de sistematización de las unidades paremiológicas y de asignación de las mismas a un ámbito de conocimiento.

Entre las perspectivas futuras, se prevé el desarrollo de una ontología de dominio cuya implementación consienta la recuperación automática de la información semántica de las paremias relativas a dicho dominio. Este trabajo se enmarca en el sector de investigación del web semántico (Oltromari *et al.*, 2003), cuya potencialidad podrá ser explotada en la base de datos paremiológica multilingüe y consultable en Internet.

Referencias bibliográficas

Repertorios paremiográficos

- Boggione, V.; Massobrio, L. (2004). *Dizionario dei proverbi*. Turín: UTET.
- Giusti, G.; Capponi, G. (2001). *Proverbi toscani*. Roma: Newton & Compton Editori.
- Franceschi, T.; Cervini, C.; Guarino Amato, M. et al. (2000). *Atlante paremiologico italiano. Questionario*. Alessandria: Edizioni dell'Orso.
- Gella Iturriaga, J. (1982). *Las monedas en el refranero*. Madrid: Juan R. Cayón.
- Martínez Kleiser, L. (1993 [1953]). *Refranero general ideológico español*. Madrid: Hernando.
- Rodríguez Marín, F. (1925). *Más de 21000 refranes castellanos no contenidos en la copiosa colección de G. Correas*. Madrid: Tip. de la Revista de archivos, bibliotecas y museos.
- Straffforello, G. (1883). *La sapienza del mondo ovvero dizionario universale dei proverbi di tutti i popoli*. Turín: Augusto Federico Negro.

Otras referencias

- Atserias, J. et al. (2003). "The MEANING Multilingual Central Repository". En Sojka, P. et al. (eds.). *Proceedings of the Second International WordNet Conference, GWC 2004*. Masaryk University: Brno. 23-30.
- Atserias, J. et al. (2006). "FreeLing 1.3: Syntactic and semantic services in an open-source NLP library". En *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*. Génova.
- Casares, J. (1950). "Introducción a la lexicografía". *Revista de Filología Española*, Anejo LII. Madrid. 185-204.
- Fellbaum, C. (1998). "Towards a representation of idioms in WordNet". En Harabagin, S. (ed.). *Proceedings of the Workshop on Usage of WordNet in Natural Language Processing Systems, COLING/ACL 1998*. Montreal. 52-57.
- Fernández-Sevilla, J. (1985). "Consideraciones lexicológicas y lexicográficas sobre el Refranero". En *Estudios románicos dedicados al Profesor Andrés Soria Ortega*. Granada: Departamento de Filología Románica. 89-99.
- Miller, G. et al. (1990). "Introduction to WordNet: An On-line Lexical Database". *International Journal of Lexicography* 3 (4). 235-244.
- Oltromari, A. et al. (2003). "Il ruolo dell'ontologia nella disambiguazione del significato". *Networks* 2003 (2). 14-24.
- Pianta, E.; Bentivogli, L.; Girardi, C. (2002). "MultiWordNet: Developing and Aligned Multilingua Database". En *Proceedings of the First International Conference on Global WordNet*. Central Institute of Indian Languages: Mysore, India. 293-302.
- Sevilla Muñoz, J. (2000). "Técnicas de la "traducción paremiológica" (Francés-Español)". *Proverbium* 17. Vermont: Universidad de Vermont. 369-386.
- Vossen, P. (1998). "Introduction to EuroWordNet". En Ide, N.; Greenstein, D.; Vossen, P. (eds.). *Special Issue on Eurowordnet. Computers and the Humanities* 32 (2, 3). 73-89.
- Meaning, Multilingual Central Repository Interface: <http://www.lsi.upc.es/~nlp/meaning/demo/demo.html> [Fecha de consulta 20 de Marzo de 2008].
- <http://garraf.epsevg.upc.es/freeling/index.php>. [Fecha de consulta: 20 de marzo de 2008].