

LOS DICCIONARIOS ELECTRÓNICOS FRASEOLÓGICOS, PERSPECTIVAS PARA LA LENGUA Y LA TRADUCCIÓN

PEDRO MOGORRÓN HUERTA
Universidad de Alicante
Pedro.mogorron@ua.es

Resumen

This paper shows the poor treatment of phraseological units provided by traditional monolingual and bilingual dictionaries. It then goes on to focus on the recent launching of electronic dictionaries, their applications and the automated treatment of language, lexicon and its possible classifications.

Taking verbal locutions as a starting point, the author suggests a new way to compile an electronic dictionary, attempting to make up for the usual lack of information about phraseological units in traditional dictionaries. The paper concludes showing all the possible applications of these new dictionaries in lexicology and in contrastive semantics, as well as in translation.

1. INTRODUCCIÓN

La finalidad de los diccionarios generales en soporte papel ha sido y es aún hoy en día, en el caso de los diccionarios monolingües, la de servir de apoyo a los usuarios de una lengua para mejorar sus competencias lingüísticas y la de ayudar a los estudiantes y a los traductores realizando un análisis contrastivo (siempre parcial) en el caso de los diccionarios bilingües.

Tradicionalmente, los diccionarios generales se han organizado alfabéticamente en torno a las entradas del léxico que recogían, reagrupando todas aquellas unidades que presentan la misma forma bajo una misma entrada. Éste ha sido un grave error de conceptualización, pues no pueden realmente existir unidades léxicas con varios significados ya que entonces en el fondo se están describiendo tantas unidades léxicas como significados o subentradas existan con una determinada palabra en ese diccionario.

En efecto, para todas y cada una de las unidades léxicas existe un triplete constituido por una forma (x), un significado (x') y una combinatoria propia $\Sigma=x''$.

Si consideramos por ejemplo los ejemplos siguientes:

- (1) Max pone el libre en el estante = (deja, guarda).
- (2) María le pone el vestido rojo a su hijo = (viste).
- (3) El sol pone moreno = (tuesta la piel)
- (4) Juan se pone enfermo = (enfermar)
- (5) Max pone hoy todo su dinero en la cuenta del banco = (depositar).
- (6) Max pone un telegrama a Mar = (enviar).
- (7) Mar pone una tienda = (montar)

Cada una de estas unidades léxicas representadas aunque comparta una forma común con los demás ejemplos (el verbo poner o cualquiera de sus formas lematizada), se caracteriza de éstas por tener una combinatoria y un significado propio totalmente independiente de los demás.

Existen asimismo otras unidades léxicas que presentan también esa misma forma, y que tradicionalmente han sido reunidas bajo la misma entrada en los diccionarios clásicos generales pero que sin embargo no se pueden acoger al triplete $\Sigma = x+y$ ya que tienen un significado que no es composicional. En efecto, en determinadas unidades léxicas, la suma de los significados de los componentes (a, b, c) no da como sería de suponer (a + b + c), sino que da un significado (a + b + c = d) que no se parece en nada a la suma del significado de cada uno de los componentes tomado individualmente. Ej:

- (8) Poner(le) alguien el cascabel al gato.
- (9) Poner el grito en el cielo.
- (10) Poner a alguien como un trapo
- (11) Poner las cartas boca arriba
- (12) Poner verde a alguien

Esta muestra de agrupación de significados bajo una misma forma nos introduce de lleno en uno de los grandes problemas de elaboración de los diccionarios:

Los problemas fundamentales con los que debe enfrentarse tanto los redactores como los usuarios de un diccionario son la polisemia y la fijación léxica. «escribir» o «leer» un diccionario consiste en gran medida, en saber delimitar unidades léxicas acerca de las cuales se da o se recibe una información, más o menos formalizada según la microestructura de cada diccionario. (Blanco, 1999: 1).

Si ahora a la entrada en la que figuran los ejemplos (1, 2, 3, 4, 5, 6, 7) le añadimos los ejemplos (8, 9, 10, 11, 12, etc.)¹ que presentan fijación léxica, entonces nos podremos encontrar con una entrada de dimensiones realmente difíciles de manejar, pues algunas entradas tienen una enorme cantidad de Unidades léxicas: por ej: agua, boca, casa, cuenta, etc.

Las unidades con fijación léxica (8, 9, 10, 11, 12) que presentan un significado componen parte de lo que hoy en día se ha venido a llamar fraseología. Nos proponemos a continuación mostrar en este artículo, el tratamiento tan deficiente que han recibido en los diccionarios clásicos generales a pesar de la importancia que tienen en el uso y comprensión de la lengua, y exponer el tratamiento que se les debería otorgar. Para ello, hablaremos de la reciente aparición en el marco de la lingüística informatizada, de los diccionarios electrónicos, y de los logros que se pueden y esperan conseguir gracias a ellos. También trataremos de sus principales objetivos: así como de sus posibles aplicaciones al mundo de la fraseología y de la traducción.

2. LAS UNIDADES FRASEOLÓGICAS

Una lengua es un sistema de comunicación lingüístico propio de una nación o de una comunidad lingüística que se caracteriza además por ser vehículo de una cultura propia. Para

¹ Tenemos censados por ahora 262 ejemplos de locuciones verbales con poner(le).

hablarla y comprenderla es preciso conocer sus reglas gramaticales así como gran parte del léxico que la compone. Además: Mogorrón (2002:15).

El conocimiento de una lengua se mide generalmente por el dominio y la soltura con los que un hablante utiliza las unidades fraseológicas propias y características de esa lengua. En los usuarios nativos estas expresiones representan un *saber lingüístico, cultural y social* común depositado por la vía de la experiencia. En los usuarios no nativos que aprenden un *segundo idioma*, las unidades fraseológicas suponen, en general, un reto, pues representan la mayor dificultad con la que se van a encontrar, debido a la gran cantidad de información lingüística, social y cultural que han de aprender a utilizar y a diferenciar de las características propias de su idioma materno.

Estas unidades fraseológicas, que podremos en lo sucesivo abreviar bajo la forma UFS, se utilizan con un objetivo discursivo muy claro. Representan el deseo del locutor de utilizar en el acto de comunicación una determinada fórmula refrendada por la mayor parte de la comunidad lingüística, sumándole de esa forma un matiz cultural, social, generacional, etc., en vez de utilizar un término neutro o una perífrasis verbal.

Aunque ya a principios del siglo XX algunos autores² habían iniciado una tímida aproximación al fenómeno de la fijación en la lengua y en el discurso, en la segunda mitad del siglo la lingüística ha descubierto en las fijaciones y en las unidades fraseológicas³ un material no solamente de estudio, sino también de enorme interés y trascendental importancia en la lengua y en el desarrollo, adquisición y utilización de la competencia comunicativa de los hablantes⁴.

2.1. Tratamiento de las unidades fraseológicas en los diccionarios generales

Si bien es verdad que en un diccionario de corte general, (ya sea monolingüe o bilingüe), se debería poder encontrar cualquier vocablo, cualquier unidad léxica simple o compleja, de significado composicional o fijado que se buscara en él, también de sobra es sabido que numerosas razones teóricas y técnicas: motivos de adecuación, de espacio físico, amplitud de la tarea a realizar⁵, han limitado esa posibilidad.

2 C Bally (1905), Oscar Jespersen (1926).

3 Queremos aclarar que no es nuestra intención aquí presentar un estado de la cuestión acerca de los estudios fraseológicos y / o de sus posibles clasificaciones. Ver para ello Blanco, (1999), Corpas (1996b 2003), Martins (1996), Mogorrón (2002), Ruiz Gurillo (1997).

4 Basta para ello ver la enorme cantidad de grupos de investigación que han trabajado, siguen trabajando o están iniciando sus investigaciones entorno a los fenómenos de fijación lingüística en numerosas lenguas con lo que eso supone en cuanto a proyectos, producción bibliográfica, congresos en los que se habla de ello o en los que únicamente se tratan estos fenómenos. Para tener una pequeña idea acerca de la importancia de estos fenómenos, basta consultar «diez años de investigación en fraseología» de G. Corpas (2003). Además de los grupos mencionados en este libro, se podrían añadir muchísimos más, aunque solamente vamos a reseñar por considerarlos relevantes y con el riesgo de dejar algunos sin señalar, al Laboratoire de Linguistique informatique (LLI) dirigido por G. Gross en Paris 13, al Laboratoire d'Automatique Documentaire et Linguistique (LADL) creado por el recientemente fallecido M. Gross, al prolífico grupo de investigación de la Universidad de Túnez dirigido por el profesor Salah Meiri y que está desarrollando el atlas lingüístico de Túnez, al grupo de Lexicología del departamento de Filología románica de la AUB, dirigido por Xavier Blanco.

5 No conviene olvidar que este tipo de obras es marcadamente editorialista y que por lo tanto se debe a un *factor de rentabilidad no lingüística sino económica con todo lo que esta observación presupone*.

Por lo tanto, queremos dejar patente que aunque el conjunto de unidades fraseológicas o fijaciones seleccionados y presentes tanto en los diccionarios monolingües como bilingües, constituye una parte importante del sistema léxico, sin embargo, esa presencia se ha plasmado generalmente en la incorporación de «cierto número de unidades» a menudo muy reducido, siguiendo un proceso de selección y descripción únicamente alfabético fruto de la propia intuición y competencia lingüística del o de los lexicógrafos encargados de la elaboración de la obra y además sin ningún criterio de clasificación exhaustivo. Esas limitaciones han acotado el uso de la lengua analizada y utilizada en los diccionarios, reduciendo ésta a una lengua llamada estándar que recoge las unidades léxicas y fraseológicas más frecuentes y compartidas por la mayor parte de los usuarios. Este proceso y criterios de selección han dejado de lado numerosas otras expresiones que poseen, dentro del macromodelo diasistemático⁶ una marca características de grupos sociales, usos regionales, generacionales, frecuencias de uso, expresiones en desuso, etc. Ej:

Modelo de marcaje diasistemático:

diatrático	estándar, culto, familiar, popular, coloquial, vulgar, argótico.
diatópico	americanismos, dialectología, regionalismo
diacrónico	arcaico, neologismo
diaintegrativo	latín, argot
dianormativo	correcto, incorrecto
diaconnotativo	infantil, peyorativo, apreciativo
diamediático	oral, escrito
diafásico	formal, informal
diatextual	periodístico, administrativo, literario
diatécnico	lenguaje de especialidad
diafrecuencia	raro, frecuente

Cuadro 1.

Existe pues frecuentemente, un desfase considerable entre el continuo e imparable proceso de remodelación de los actos de comunicación constantemente obligados a readaptarse y el material que los lexicógrafos insertan en los diccionarios. En efecto, por un lado están las UFS de la lengua clásica, a menudo ya en desuso por pertenecer a un uso no ya de otras generaciones sino también a una concepción diacrónica y que por lo tanto ya no son representativas de los usos y estados lingüísticos que se van sucediendo. Por otro lado, se halla el uso presente, marcadamente innovador de la lengua, que se debe ir transformando y adaptando a los constantes cambios que se producen en la lengua por los usos generacionales y las diferentes necesidades y realidades sociales que se plasma en nuevos vocablos y expresiones. Este desfase, que es una de las deficiencias que nos proponemos paliar con los diccionarios electrónicos, es muy patente en los diccionarios clásicos.

6 Una de las descripciones del modelo diasistemático mas completa ha sido realizada por F.J. Haussman, ver para ello Blanco (2001: 52,53).

En efecto, numerosas UFS que aparecen en los diccionarios son las huellas de épocas pasadas, ya no son representativas del estado vigente de la lengua activa pues pertenecen, a un uso generacional caduco a menudo diacrónico, a una concepción social y laboral diferente⁷. Esa falta de adecuación conceptual o composicional supone que el hablante utiliza unidades fraseológicas que le suponen los siguientes problemas de comprensión e interpretación.

- Se usan expresiones con palabras que solamente se utilizan ya en la lengua en esas expresiones. Ej:
 - Salir de **naja**.
 - Prendre la poudre **d'escampette**.
- Se usan expresiones que pertenecen a otras realidades: vocabulario del campo, refranes meteorológicos, mundo laboral ya caduco como es el caso con todo lo referido al campo del caballo, y que ya no son sentidas como propias por los usuarios de la lengua al no saber a que se refieren exactamente, al no conocer el posible hecho histórico en el que se basan. Ej:
 - Cumplir con la parroquia⁸. (DUE).= Comulgar por Pascua.
 - Costar un Potosí. (DUE). = Costar algo mucho dinero, referencia a la colonización española de Hispanoamérica.
 - Colgarle a alguien el / un Sambenito / San-Benito (DUE) = Acusar injustamente a alguien
 - De casta le viene al galgo el ser rabilargo. (DUE). = La cualidad de una persona le viene de familia
 - Poner a alguien a los pies de los caballos. (DUE). = desairar, humillar, maltratar.
 - Pasar más hambre que un maestro de escuela⁹.

Este fenómeno se puede apreciar aún con mayor claridad en los grupos de expresiones parasinónimas. (Ver Cuadro 2 y 3). En el primer cuadro:

7 Según Rey-Debove (1971: 98) «En Synchronie pure, c'est à dire dans l'instant des contemporains de tous les âges compris entre 15 et 75 ans ont des échanges réels. Mais une personne de 75 ans garde les habitudes de langage de sa jeunesse, correspondant à un autre système lexical, et qui persiste grâce à la mémoire et aux échanges avec les personnes du même âge. Une grande partie du vocabulaire actif de ses 15 ans (communication avec les personnes âgées) lui ne fonctionne plus dans aucun échange, faute d'interlocuteurs. C'est de cette manière que la langue perd d'un côté ce qu'elle acquiert de l'autre.»

8 Las siglas de las obras referenciadas corresponden a l Diccionario de la Real Academia Española de la lengua para RAE, al diccionario de uso del español para DUE, Gran diccionario Larousse Español-Francés / Francés-español para LBI, Dictionnaire des Expressions et Locutions de Alain Rey para AR, Thesaurus Larousse para TL, PR para Petit Robert. También hay que señalar que las locuciones verbales que figuran por debajo de la línea horizontal y que no tienen referencia no han sido encontradas en diccionarios, pero sin embargo son bastante frecuentes en la lengua activa española o francesa.

9 No viene referenciada esta locución muy conocida y muy usada en los diccionarios que hemos consultado (DUE, DRAE, En Larousse).

Estar en Babia. (DUE). Estar en las Batuecas. (DUE). Estar en Belén. (DUE). Estar en la higuera. (DUE). Estar en el <i>limbo</i> . (DUE). Estar en la <i>luna</i> . (DUE). Estar en las <i>nubes</i> . (DUE). Mirar a las musarañas. (DUE). Mirar las telarañas. (LBI).	Bayer aux corneilles.(AR). Être dans les <i>limbes</i> . (AR). Être dans la <i>lune</i> . (AR). Être dans les <i>nuages</i> . (AR). N'avoir pas les pieds sur terre. (TL).
<hr/> Estar en la parra.	

Cuadro nº 2

se pueden observar varias expresiones que significan prácticamente lo mismo si nos atenemos a las definiciones que de ellas dan los diccionarios (DUE, DRAE): *estar distraído*, *no enterarse de algo*. Sin embargo estas locuciones no son utilizadas indistintamente por las distintas generaciones de usuarios que componen la población total activa que utiliza una determinada lengua. Así, sorprendentemente, según una encuesta que hemos realizado entre nuestros alumnos en la U de Alicante, las expresiones *estar en Babia*, *en las Batuecas*, incluso *estar en Belén*, así como *Mirar a las musarañas* o *a las telarañas* ya no son comprendidas por la mayoría de los jóvenes estudiantes de 20 años.

Por otro lado, existe justamente el fenómeno inverso. Es decir que la lengua activa de una determinada fase lingüística crea otras unidades que no aparecen en los diccionarios que suelen tardar bastante tiempo en incorporar las nuevas creaciones o adaptaciones de la lengua. Éstos las excluyen debido a su inestabilidad o a su caducidad en la lengua, pues numerosas UFS aparecen, son utilizadas cierto tiempo: Ej.

— estar de, ir de guateque

y posteriormente son reemplazadas por otras más modernas,

— estar de, ir de guerra, hacer la ruta (del bacalao).

— estar hasta el punto com¹⁰.

También es verdad que en otras ocasiones, los diccionarios, que como hemos señalado anteriormente suelen reflejar una lengua estándar, no incorporan expresiones que se mantienen en la lengua colectiva durante mucho tiempo debido a diferentes niveles de uso. Pondremos como ejemplo la expresiones

— ser más agarrado que un tango y ser más agarrado que un catalán

de uso corriente que no aparecen mientras que:

— ser más agarrado que un chotis

si que aparece en el DRAE. Otro ejemplo representativo podría ser la serie de expresiones parasinónimas de estar harto de algo o alguien, (Ver cuadro 3) que tienen escasa repercusión en los diccionarios clásicos que no incluyen las expresiones siguientes que figuran por deba-

¹⁰ Esta expresión sería un ejemplo de muy reciente aparición en la lengua activa. Sólo con el paso del tiempo podremos saber su éxito real en la lengua.

jo de la línea horizontal (estar hasta los huevos, hasta el coño, etc.) que sin embargo, tienen gran representatividad en la lengua activa de los jóvenes y de los no tan jóvenes.

Estar hasta el coco de. (DUE).	En avoir assez.(AR).
Estar hasta los cojones de. (DUE).	En avoir marre. (fam AR).
Estar hasta la coronilla de. (DUE).	En avoir par dessus la tête. (fam AR).
Estar hasta el culo de. (DUE).	En avoir par dessus les épaules. (fam,AR).
Estar hasta el gollete de. (RAE).	En avoir plein les bottes. (fam,PR).
Estar hasta el gorro de. (DUE).	En avoir plein le cul. (fam, AR).
Estar hasta el moño de. (DUE).	En avoir plein le dos. (fam,AR).
Estar hasta /las narices de. (DUE).	En avoir ras-le-bol. (fam,AR).
Estar hasta la punta de los pelos. (DUE).	En avoir son bol. (fam,AR).
Estar hasta los pelos de. (DUE).	En avoir (son) pied. (Arg, AR).
Estar hasta por encima de la cabeza de. (DUE).	En avoir soupé (fam,TL),
<hr/>	
Estar hasta el coño de.	En avoir ras le baba.
Estar hasta el punto com de	En avoir ras la casquette.
Estar hasta los huevos de.	En avoir ras les couilles.
Estar hasta los mismísmos de.	En avoir ras le jonc.
Estar hasta el nabo de.	En avoir ras les pâquerettes.
Estar hasta los ovarios de.	
Estar hasta el pirri de.	
Estar hasta el pito de.	

Cuadro nº 3

Convendría pues, que los diccionarios tratasen debidamente estas UFS ya que al no estar debidamente recogidos por los diccionarios, estos desfases entre el uso de la lengua activa, estándar, clásica, popular, etc., su escasa representatividad y clasificación en los diccionarios, dificultan el conocimiento de numerosas fórmulas a:

- usuarios de un idioma.
- estudiantes que descan aprender ese idioma.
- traductores que no saben realmente cual de las combinaciones es el equivalente real de traducción.

Además de estos desfases de incorporación de las unidades fraseológicas, también cabría referirse al tratamiento que se les da, con la escasa información de carácter gramatical y la mala gestión o presentación de esta misma, en efecto, se suelen ordenar las UFS alfabéticamente sin aplicar una estructuración sistemática para no mezclar los distintos grupos de unidades fraseológicas. Estas obras deberían cuanto menos tratarlas y dar la misma cantidad de información que para el resto de las unidades simples no fijadas de la lengua, para permitirle a los usuarios disponer de toda la información morfosintáctica necesaria para utilizarlas correctamente. Así por ejemplo la gran variedad estructural de UFS no viene tratada ya que los diccionarios las incorporan de forma desordenada. Por otro lado existen varias deficiencias en la presentación de estas formas. Así I Santamaría (2003: 48) aprecia:

- indecisión en la forma de la expresión aportando una única de las variantes gráficas sin que esta sea siempre la más usual (a boca de jarro en lugar de abocajarro).
- Indecisión a la hora de lematizar la expresión que una veces aparece con un verbo con el que se combina frecuentemente, otras aparece sin forma verbal (de estampía / salir de estampía).
- Inclusión de las expresiones con o sin preposiciones: en ayunas / ayunas.

Recientemente, se están elaborando versiones comerciales de estos diccionarios en soporte informático: DUE, DRAE, Enciclopedia Larousse, Petit Robert, Enciclopedia Británica, etc. Sin embargo, ese cambio de soporte no ha supuesto ningún cambio apreciable en la estructura o en el contenido de estas obras, pues las únicas diferencias que se han podido observar han sido la rapidez de las consultas a realizar, la posibilidad de navegar por los artículos así como la de seleccionar y buscar estructuras con pequeños localizadores, ya que tanto la estructuración de la información como la de los contenidos siguen siendo prácticamente similares.

3. APARICIÓN DE LOS DICCIONARIOS ELECTRÓNICOS

Por otro lado y también recientemente, ha hecho su aparición en el mundo de la lingüística otro tipo de diccionarios cuya elaboración está relacionada con posibles aplicaciones al tratamiento automático de textos en el campo de la lingüística aplicada. Se trata de los diccionarios electrónicos. Al contrario de lo que había ocurrido con la elaboración y el contenido parcial de los diccionarios generales clásicos, ya desde sus inicios, la concepción de este nuevo tipo de diccionarios se está realizando *asumiendo como objetivo a alcanzar*, el de la máxima exhaustividad posible. En efecto, un diccionario electrónico no es una sencilla lista alfabética muy completa de palabras simples o compuestas que se podrá posteriormente aplicar al análisis automático de textos. No. Un diccionario electrónico es mucho más que eso. Se trata de una base de datos que contiene una enorme cantidad de datos recogidos por los lexicógrafos que deberá ser gestionada por la informática en función de las necesidades que se hallan programado y que se deseen alcanzar.

Para poder entender las posibilidades de los diccionarios electrónicos hay que saber que se basan en torno a dos piedras angulares. Por un lado tendremos una base lingüística y por otro lado la parte informática que se va a encargar de gestionar lo más eficazmente posible toda la cantidad de información lingüística recogida por los especialistas en lengua. Ésta podrá ser por lo tanto: morfológica, sintáctica, fonética, semántica, etc. Sin embargo, la parte informática solamente podrá analizar la información que la base de datos lingüística haya recopilado y estructurado y realizar procesos selectivos de búsqueda sobre ésta. Es decir, que el mejor programa informático no podrá hacer maravillas por mucho que se quiera a la hora de gestionar la información para buscar resultados y posibles aplicaciones. Son pues dos procesos que deben ir a la par. El lingüista debe saber lo que necesita el informático, y el informático a su vez lo que desea conseguir el lingüista. Una vez estén definidos entre los dos los objetivos a alcanzar, se deben crear una serie de ficheros informáticos que contendrán toda la información en forma de entradas léxicas (simples o complejas), en forma de códigos morfológicos identificables por los programas.

3.1. Dicionarios electrónicos desarrollados para sistemas de TAL (Tratamiento Automático del Lenguaje)

Las aplicaciones contempladas por los diccionarios electrónicos son muy numerosas, pero van a depender en gran parte de la información y del tratamiento que se le haya dado a ésta. En efecto, la utilización de la información en tratamiento automático de textos exige que las palabras pertenecientes a los textos que se han introducido, estén etiquetadas de manera que el diccionario pueda catalogarlas. A partir de técnicas basadas en el procesamiento del lenguaje natural, se han desarrollado sistemas para la lematización, es decir para el etiquetado automático morfológico y sintáctico de los textos de un corpus, que consisten en la lectura y en la división del texto en unidades relevantes que serán más tarde utilizadas para trabajar el análisis de la palabra. Las características que normalmente se indican durante este etiquetado hacen referencia a:

- Aspectos de estructuras de texto: marcas tipográficas, divisiones textuales, párrafos, citas, títulos.
- las propiedades morfosintácticas de la palabra y puede ir desde los datos necesarios a la conjugación de un verbo, hasta todos los datos necesarios para poder realizar un análisis textual.
- Caracterización de las unidades léxicas, léxico en general, extranjerismos, nombres propios, siglas, abreviaciones.
- Funciones sintácticas de cada constituyente y representación por medio de árboles sintácticos. Etc.

De esta forma, dependiendo de la aplicación planteada se podrá por ejemplo realizar: una comprobación ortográfica, un análisis sintáctico de una frase o de un texto, la búsqueda de un corpus para poder contrastar y verificar los elementos analizados así como la frecuencia de aparición de éstos mismos proporcionando datos acerca de la frecuencia de aparición de las palabras en términos cuantitativos y del tipo de textos en que aparece. Finalmente, se podrán como veremos posteriormente aplicar toda la información a procesos de aprendizaje y traducción automáticos¹¹.

3.2. Dicionarios electrónicos concebidos para realizar una descripción gramatical completa del léxico y de la sintaxis

Se han elaborado numerosos prototipos de analizadores textuales, pero sin embargo ninguno de ellos incorpora un tratamiento completo del léxico y de la gramática. Desde principios del siglo XX, los lingüistas parecían haber renunciado a constituir inventarios, pues por ejemplo los estudios transformacionalistas se basaban en pequeños grupos de ejemplos, estos últimos sin embargo serán fundamentales para Maurice Gross (1975:20):

L'étude systématique des propriétés syntaxiques s'impose aujourd'hui pour des raisons diverses. L'existence de théories transformationnelles fait qu'une étude systématique peut être délimitée avec beaucoup plus de précisions qu'à l'intérieur des cadres traditionnels. D'autre part, la situation présente de la syntaxe transformationnelle est particulière. Après

¹¹ Los dos sistemas de TAL más representativos son: SYSTRAN y SADAW.

un période où des succès ont pu laisser croire que l'emploi des transformations dans les descriptions allait régulariser considérablement ces dernières, il est devenu clair que les nouvelles règles continuaient à comporter des exceptions en nombre sensible.

Este gran número de excepciones que dificultan el estudio y el conocimiento de las lenguas va a ser pues el que va a provocar un cambio radical en el objetivo y por lo tanto en la estructuración y en la cantidad de información necesaria. Para ello, se va a tratar toda la información fijándose como meta una descripción ortográfica, fonológica morfológica, sintáctica y textual utilizando como método de trabajo un sistema muy lineal, esquemático y riguroso de tal forma que en esta descripción de las unidades léxicas de la lengua prime la transparencia de las unidades de significación para evitar posibles confusiones en torno a la polisemia y a la fijación léxica.

Este marco teórico y metodológico elaborado por Maurice Gross y su equipo del LADL¹² ha sido llamado Lexique-Grammaire y su objetivo principal es la construcción conjunta de la gramática y del léxico de una lengua dada, describiendo las propiedades sintácticas, distribucionales y transformacionales de las unidades léxicas (verbos, nombres, adjetivos, adverbios, etc.) para proporcionar datos lingüísticos sobre el funcionamiento en el marco de las frases de cada elemento susceptible de ser integrado en sistemas informáticos para poder ser analizados automáticamente en aplicaciones para el tratamiento automático de textos. Para conseguir toda la información necesaria, se organiza el léxico a analizar integrándolo en tablas en las que se averigua si éste acepta una serie de propiedades distribucionales y transformacionales: M.Gross (1975,150)

Nous avons représenté un segment de la grammaire du français au moyen de matrices binaires [... chaque matrice, c'est-à-dire chaque table correspond à une classe de structures [... une ligne correspond à une entrée verbale, une colonne à une propriété syntaxique. Le signe + à l'intersection d'une ligne V et d'une colonne P indique que le verbe V a la propriété P, le signe - indique que v n'a pas la propriété P.

Siguiendo estos procedimientos, se han elaborado varios diccionarios electrónicos:

- DELAS (diccionario de las formas simples, con más de 90.000 entradas).
- DELAF (diccionario de las formas flexionadas).
- DELAC (diccionario de las formas complejas, con más de 100.000 nombres compuestos y más de 7000 adverbios).
- DELAP (diccionario de transcripciones fonéticas).

Hay que señalar que numerosos grupos de investigadores Integrados dentro de la red Relex: <http://infolingu.univ-mlv.fr/Relex/dictionaries.html> han desarrollado o están desarrollando en la actualidad diccionarios electrónicos utilizando el sistema DELA del LADL. En Portugal dirigido por E.Ranchood en la Universidad de Lisboa, En Italia dirigido por A. Elia en la Universidad de Salerno, en España el grupo de lingüística aplicada a las lenguas

12 Laboratoire d'Automatique Documentaire et Linguistique. Este grupo de investigación constituye la UR n° 819 del CNRS dirido por Maurice Gross se ubicó primero en Paris VII, y se ha integrado a la muerte del mismo en el IGM de la Universidad de Mame La Vallée. HYPERLINK <<http://www-igm.univ-mlv.fr>> <http://www-igm.univ-mlv.fr>

románicas de la UAB dirigido por Xavier Blanco: <http://oid.uab.es/lexsen>, en alemán, el sistema CISLEX, diccionario electrónico coordinado por el CIS Centrum für Informations und Sprachverarbeitung (LMU-München), <http://cis.uni-muenchen.de>, ha seguido el modelo DELA, además podríamos señalar también la construcción de sistemas similares para el griego, el coreano, el malgache, etc.

3.3. Las unidades fraseológicas en los diccionarios electrónicos

Si examinamos los nombres de los diccionarios electrónicos mencionados hasta ahora y las unidades léxicas tratadas, observamos que aún no hemos mencionado el tratamiento dado a aquellas ULS cuyo significado no se derivaba de la suma de sus componentes. Faltaría pues un diccionario electrónico fraseológico que analizase estas UFS, aunque quizás sería más adecuado y realista debido a la gran diversidad de UFS existentes y a la enorme cantidad de éstas, elaborar varios diccionarios dependiendo de su estructura y denominación.

Esa parece ser la línea de trabajo e investigación decidida y elegida. Así M. Gross elaboró un diccionario electrónico de más de 45000 «expressions figées»¹³ debidamente clasificadas según su tipo y número de complementos y analizadas sintácticamente, siguiendo los mismos criterios de los DELA, que está siendo terminado por los miembros de su equipo. Ej: Cuadro 4 tabla CPP¹⁴.

+	-	-	<E>	changer	+	-	-	de	le	tout	-	à	le	tout	-	-
+	-	-	<E>	choisir	+	-	-	entre	la	peste	-	et	le	choléra	-	-
-	+	-	<E>	circuler	-	-	-	de	<E>	bouche	-	en	<E>	bouche	-	-
-	+	-	<E>	circuler	-	-	-	de	<E>	bouche	-	à	<E>	oreille	-	-
-	+	-	<E>	circuler	-	-	-	de	<E>	main	-	en	<E>	main	-	-

Cuadro 4.

También están elaborando la descripción de un proyecto llamado BFQS¹⁵ en el que se analizan todas aquellas variaciones diatópicas de las expresiones fijas que aparecen en los cuatro países mencionados, intentando unificar bajo un mismo diccionario la descripción de todas las formas. Ej: Ver cuadro 5.

13 Quizás el único punto flaco de este inventario es que reúne bajo una misma catalogación a estructuras con valores muy diferentes, pero también es verdad que todas ellas responden al concepto común de fijación.

14 Las siglas CPP significan en Lexique-Grammaire que las expresiones tienen todas la estructura siguiente: N0 V Preposición 1 Constante 1 Preposición 2 Constante 2.

15 B de Bélgica, F de Francia, Q de Québec, S de Suiza.

Table CPID-F	B	F	Q	S	Traduction	classe
jouer des coudes	+	+	+	+	se frayer un chemin	BFQS
cracher en l'air	+	+	+	-	faire qqch d'inutile	BFQ
baigner dans l'huile	+	+	-	+	marcher bien	BFS
rouler au super	-	+	+	+	rouler à l'essence «super»	FQS
aller à la selle	-	+	+	-	aller déféquer	FQ
aller aux asperges	-	+	-	-	aller se prostituer	F

Cuadro 5.

Tal y como habíamos señalado para los diccionarios DELA, en italiano, portugués, coreano, alemán se están realizando actualmente clasificaciones de estas estructuras. Las locuciones adverbiales, debido con toda probabilidad a su número más reducido han podido ya ser analizadas en varios idiomas: M. Gross para el francés (1986), J. Baptista para el portugués (2002), M de Gioia para el italiano (2001), D. Cátala en un tesis doctoral en (2003) para el español. También Mirella Connena (U de Bari), está realizando desde hace numerosos años un inventario y su clasificación de los refranes en italiano y francés desde una perspectiva comparada y traductológica.

Vemos pues, como poco a poco se van elaborando diccionarios electrónicos de las léxicas fraseológicas y de su gramática que palian paulatinamente el tradicional vacío que existía en el estudio de estas formas.

3.4. Aparición de la Onomasiología en los diccionarios electrónicos

Otra de las novedades más interesantes que ha aportado el perfeccionamiento de la lingüística informática ha sido la posibilidad de cambiar el tradicional criterio de ordenación alfabética de las unidades seleccionadas. En efecto, aunque ya en el pasado algunos clásicos (Comenius, Francis Bacon) defendieron la necesidad de una clasificación onomasiológica, hasta ahora, las clasificaciones semiasiológicas que parten del significante para llegar al significado, habían ganado siempre la partida y se utilizaban por su indiscutible objetividad e imparcialidad frente a la siempre discutible subjetividad de las clasificaciones extralingüísticas¹⁶. Una base de datos con una clasificación temática exhaustiva en campos léxicos¹⁷ nos permitirá encontrar en décimas de segundo las expresiones de los cuadros 2 y 3 que son parasinónimas, e incluso de expresiones que sean antónimas, e incluso hipónimas. Ahora bien esta otra posibilidad de clasificación supone un esfuerzo de análisis del léxico muy riguroso para no caer en posibles errores, contradicciones y lagunas como muy bien explica Antonio Pamiés¹⁸ (1998, 208):

16 Además hay que señalar que ese tipo de clasificación que suponía una catalogación muy costosa por parte del lexicógrafo también suponía una labor de consulta lenta e incierta para el usuario que debía a menudo someterse a un aprendizaje previo para consultar y utilizar correctamente esas clasificaciones

17 La noción de campo léxico tomada de la lingüística alemana que habla de campo semántico desde Jost Trier ha sido seguida por números lexicólogos como Wartburg, G Matoré, P. Cuiraud, E. Coseriu bien en una lengua o en varias lenguas consiguiendo de esa forma estructuras diferenciadas, diferencias morfosemánticas, socioculturales, históricas, etimológicas, etc.

18 Criterios para una fraseografía onomasiológica automatizable. (1998).

El primer peligro procede de la sinonimia entre descriptores: p.ej., el lexicógrafo puede asignar en un momento dado el descriptor DINERO a un refrán como *lo barato sale caro*, ¿pero qué garantía tenemos de que el usuario no pregunte por COMERCIO? Tampoco hay garantías de que el mismo lexicógrafo asigne unos meses más tarde el descriptor ECONOMÍA o incluso AVARICIA a un duro es un duro, con lo cual ambas expresiones quedarían desconectadas para siempre en el pozo sin fondo de la base de datos.

4. PROPUESTA DE CONTENIDOS DE UN DICCIONARIO ELECTRÓNICO DE UFS

Acabamos de pasar revista a los diferentes modelos de diccionarios electrónicos utilizados en lingüística computacional, que nos van servir, a continuación para elaborar un diccionario electrónico de locuciones verbales¹⁹ multilingüe que sea por un lado una obra de referencia y de consulta para cada idioma debido a su exhaustividad y que facilite por otro lado la traducción, la comprensión y la correcta utilización gramatical y contextual de estas formas en las lenguas analizadas.

La primera fase de este diccionario sería la de elaborar individualmente en cada una de las lenguas de trabajo un repertorio lo más completo posible que supla las deficiencias de contenido en materia fraseológica de los diccionarios clásicos:

- Sumando los contenidos de varios de ellos, (por ej: DUE, DRAE, LBI, Enciclopedia Larousse, etc). Que a veces presentan notables diferencias. Ver I. Santamaría (2003, 105-107) en las que presenta una propuesta sobre las unidades fraseológicas de la entrada agua elaborada a partir de las subentradas de varios diccionarios manejados en su obra.
- Incorporando las creaciones de la lengua activa, realizando filtros a partir de nuestra competencia léxica y haciendo encuestas a personas de diferentes regiones de cada país, para estar seguros de que no se trata de localismos.

Conviene insistir en que en este tipo de trabajos, aunque su finalidad sea un diccionario fraseológico multilingüe, esta primera fase debe realizarse primero individualmente para cada uno de los idiomas, pues si se parte de una única lengua de trabajo, se corre el riesgo de privilegiar uno de los idiomas que sería tomado como referente, dejando de lado un gran número de locuciones verbales, que podremos a continuación denominar LV(S), propias y características a la idiosincrasia histórica y cultural de la comunidad lingüística de cada una de las otras lenguas²⁰. Posteriormente, en una segunda etapa, se realizará una puesta en común en la que sí se deberán realizar cruces entre los diferentes inventarios lingüísticos y en la que alternativamente cada idioma será considerado como lengua de trabajo para intentar incorporar Locuciones verbales de uso frecuente y conocidas que se hubiesen podido olvidar.

A la par que se elabora el repertorio, se puede iniciar también otra parte de este diccionario electrónico que consistirá en clasificar cada LV en clases sintácticas en función de sus componentes para poder averiguar las propiedades distribucionales y transforma-

19 Se podría aplicar exactamente el mismo proceso de análisis y trabajo a las demás UFS. Que tradicionalmente se han dividido las locuciones según su función oracional: (nominal, adverbial, adjetival, verbal) Ver Corpas 1996: 93-94.

20 Los trabajos de lexicología plurilingües realizados a partir de un idioma padocén de lo que Alain Rey (1992: 102) ha llamado «Terminologie de départ».

cionales que poseen: (sujeto humano o no humano, régimen preposicional, posibilidad de transformación en forma pasiva, etc.). Queremos recordar que el soporte informático en sus múltiples búsquedas, solamente analizará y tratará la información sintáctica que el lexicógrafo / lingüista le hayan clasificado. Por lo tanto esta labor es más importante de lo que pueda parecer pues sin ella, durante el tratamiento de textos no se reconocerían muchas de las formas tratadas que estuviesen utilizando en ese momento alguna de las propiedades transformacionales. Además, permite consultar las posibles construcciones de las unidades léxicas en caso de duda, pues siempre se representan en el diccionario electrónico bajo su forma más extensa.

Un vez terminado el análisis sintáctico de las clases en cada idioma, se dispondrá en cada lengua de una cantidad enorme de información lingüística que el soporte informático podrá gestionar en una labor de síntesis, estableciendo:

- para cada lengua el tipo de estructura más utilizado, las palabras léxicas más utilizadas, la cantidad de LVS que admiten la posibilidad de modificar su estructura parcialmente con variantes paradigmáticas sin cambiar de significado y sin dejar de ser una unidad léxica fijada, etc.
- Estudios y paralelismos contrastivos acerca de numerosísimos aspectos lingüísticos: las preposiciones y su correspondencia, la determinación más utilizadas, como ejemplo pondremos la mayor utilización por parte del francés de determinantes posesivos que en español: ver cuadro 6.

Hablar entre sus dientes	Parler entre ses dents
Echar agua al vino	Mettre de l'eau dans son vin
Comerse las palabras	Manger ses mots
Bajarse los pantalones	Baisser son pantalon
Meter las narices en algo	Mettre son nez dans

Cuadro 6.

Simultáneamente, a la recopilación y a la clasificación sintáctica, se procederá también a un tercer tipo de clasificación para cada entrada. Esta vez, se trata de una clasificación semántica mediante la cual se agruparan las LV en campos léxico-fraseológicos. En estos campos, se realiza la descripción sintáctica y semántica de cada entrada articulándola de forma más precisa en subclases llamadas genéricamente clases de objeto o clases semánticas. Se procederá así a utilizar unos macro campos temáticos, por ejemplo el de las locuciones verbales de los actos del discurso. A su vez se dividirán estos macro campos en clases de objetos especificando sus argumentos conceptuales: orden insulto, afirmación, agradecimiento disculpas, acusación, felicitación, etc. Si se sistematiza la caracterización de los diferentes empleos de las unidades léxicas analizadas en clases semánticas, se podría resolver gran número de ambigüedades debida a la polisemia ya que cada UL se integra dentro del tratamiento automático mediante un marcado de argumento.

La constitución de estos campos temáticos van a permitir establecer con toda claridad relaciones de polisemia, antonimia, sinonimia, hiponimia, etc. que serán muy importantes para poder encontrar la unidad fraseológica más equivalente en traducción.

Los análisis y las agrupaciones conceptuales o temáticas de las UFS resultan muy necesarios y útiles para la fraseología. Tanto para en una determinada lengua como en fraseología comparada ya que se pueden entonces al tomar el campo temático al completo como punto de referencia realizar numerosas observaciones, similitudes, diferencias entre los sistemas lingüísticos analizados. Pero quizás la más interesante de todas ellas, sea la de intentar encontrar entre las diferencias lingüísticas y culturales que presentan las lenguas entre sí, las huellas de que a pesar de esas grandes variaciones debe existir también un núcleo común del entendimiento entre los seres humanos basado en una conceptualización al menos parcialmente común y que podría también tener algunos patrones gramaticales comunes coincidentes. En efecto, no en vano Lakoff (1987:448) destacaba ya la importancia que suponía para la fraseología y la traducción el que numerosas UFS fuesen signos motivados.

4.1. Localización y presentación de las LV en el diccionario electrónico

La inclusión de las LV o de cualquier UF en los diccionarios clásicos le plantea al lexicógrafo los consabidos problemas de lugar, ordenación y bajo qué entradas hacerlas figurar. Sin duda la mejor opción para el usuario de los diccionarios habría consistido en hacer aparecer las locuciones en cada uno de los componentes que tenga un significado léxico²¹, aunque esta solución no ha sido utilizada hasta ahora porque habría supuesto obras poco rentables económicamente debido a sus considerables dimensiones. Por el contrario, los diccionarios electrónicos permiten gracias a la informática superar esas dificultades y hacer figurar cualquier locución bajo todos los componentes de la expresión, con la excepción de los componentes gramaticales como preposiciones, determinantes, etc. Si mientras que las LV:

- cantar(le) las cuarenta a alguien.
- arrimar el ascua a su sardina.
- venirse el ciclo abajo.

figuran en el DRAE y en el DUE bajo las entradas **cuarenta**, **ascua** y **cielo**, en el diccionario electrónico que estamos elaborando se podrán encontrar gracias a todos sus componentes léxicos, es decir: **cantar**, **cuarenta**, **arrimar**, **ascua**, **sardina**, **venirse**, **cielo** y hasta **abajo**. Pero además, no tenemos que olvidarnos de la clasificación semántica que hemos avanzado y que permitirá a su vez buscar las LV por contenido semántico en el caso de no recordar alguna forma, pero sí el concepto.

Ciertas UFS también plantean problemas de localización y de presentación en los diccionarios tradicionales. Se trata de aquellas unidades que presentan variantes paradigmáticas²², es decir que se pueden cambiar elementos sin alterar el significado Mogorrón (2002: 63-64).

- *seguir* (mantenerse, permanecer) en sus trece.
- *Ganarse la vida* (las habichuelas, el cocido).

Deberán figurar todas las posibles variantes que a menudo no aparecen en los diccionarios clásicos, ver cuadros 2 y 3, para facilitar la labor de búsqueda. Además deberemos a

21 Ya en 1971 Zgusta sugería la posibilidad de incluir las Unidades fraseológicas bajo la entrada de cada uno de sus componentes que tuviese significado léxico. Estas unidades según la terminología de L. Tesnière, (1959: 53) se podrían llamar «mots pleins» en oposición a los «mots-vides» non référentiels.

22 Variantes en sentido estricto según Zuluaga (1980).

partir de una búsqueda en corpus (ver apartado siguiente), averiguar cual de las variantes es la más utilizada para indicarlo a los usuarios y aprendices.

4.2. Información diasistemática

Ahora bien, dado que la finalidad del diccionario electrónico de locuciones verbales que estamos realizando está orientada principalmente a la traducción, su conceptualización y elaboración exigirá no solamente, (como ya hemos presentado), unas clasificaciones y unos análisis morfológicos, sintácticos y semánticos lo más completos posibles, sino que además deberá disponer de cierta información igualmente imprescindible para esa finalidad, para poder ajustarse a otras características propias a cada lengua. Se trata de toda la información llamada diasistemática. (Ver cuadro 1, pág. 5). Pues en toda lengua, cada unidad léxica está caracterizada además de por su significado por una serie de informaciones lingüísticas, o de connotaciones contextuales usadas por los miembros de las comunidades lingüísticas a veces conscientemente, a veces inconscientemente que deben ser sin embargo, tenidas en cuentas en el proceso de traducción, pues sino se corre el riesgo de traducir únicamente el significado general de la UL sin reflejar una marcas características de grupos sociales, usos regionales, generacionales, frecuencias de uso, expresiones en desuso, etc.

La mayoría de estas características no vienen reflejadas ni en los diccionarios generales ni en los bilingües por lo que se dificulta lógicamente el acto de traducción. De todas ellas pensamos que nuestro diccionario electrónico debería mencionar al menos las siguientes: diastrático²³, diatópico, diacrónico, diatécnico, diatextual, diafrecuencia.

Estas informaciones diasistemáticas de las que nos gustaría poder disponer nos sitúan de lleno frente a otro de los grandes problemas de los trabajos lexicográficos. A saber, el de la frecuencia y vigencia de uso de las unidades que componen el diccionario. Tenemos que catalogar cada LV de tal forma que sepamos por un lado, si constituye una muestra representativa de la lengua utilizada por una comunidad lingüística en una determinada época, y además, por otro lado su frecuencia de uso, para averiguar la importancia relativa de esa misma UL en el conjunto del vocabulario de una lengua. No tenemos que olvidar que todo sistema lingüístico está en continua transformación y readaptación y que esa información debe estar disponible para que el usuario sea capaz de interpretarla y de adecuarla a otro sistema lingüístico. Para ello, deberemos hacer uso de los corpus lingüísticos. Se trata de unos conjuntos de textos en formato electrónico, (podrán ser documentos completos o conjuntos de fragmentos de mayor o menor extensión), que son una muestra representativa de la lengua y que permiten al lexicógrafo realizar una investigación con datos objetivos²⁴. Gracias a la rapidez y fiabilidad de la informática, ya se pueden consultar hoy en día diferentes tipos

23 Si bien los niveles estándar y culto no presentan demasiados problemas de identificación, convendría quizás a la vista de los desfases existentes en los diferentes diccionarios consultados, ya citados, limitar el número de los siguientes registros: familiar, popular, coloquial, argótico), (Ver Mogorrón 2002: 67-71) y proceder a una unificación de criterios que permita realizar una valoración y catalogación rigurosa.

24 El procesamiento de todos los datos de los corpus permite en un tiempo record detectar fenómenos y patrones, incluso de baja frecuencia que las muestras no permiten a simple vista detectar, determinar la vigencia de los fenómenos gramaticales, analizar su distribución a lo largo de la evolución de una lengua, almacenar grandes cantidades de textos para después analizarlos, seleccionar los textos más representativos para enfocar la investigación a desarrollar, etc.

de Corpus en función de la idiosincrasia y de las necesidades de cada diccionario. Así el **Corpus diacrónico del español (CORDE)** que refleja la evolución histórica de la lengua española, el **Corpus de referencia del español actual (CREA)** que presenta una enorme cantidad de textos escritos y orales de todos los países de habla hispana, con más de 100 Millones de palabras.

Señalamos además como muestra los siguientes corpus:

FRANTEXT	150 M de palabras para el idioma francés
British National Corpus	100 M de palabras en Inglés
Brown Corpus	100 M de palabras en Inglés americano.
Cobuild	450 m de palabras en inglés

Los medios mediáticos son utilizados actualmente cada vez con más frecuencia para la elaboración de corpus, y que de todos los medios mediáticos, sin ninguna duda el que brilla con luz propia es el macro medio representado por Internet. Hay que señalar los enormes recursos disponibles en Internet y la posibilidad de utilizar buscadores para localizar los documentos útiles. Además se están formando laboratorios de lingüística informatizada especializados en convertir los textos de numerosas versiones electrónicas de periódicos en corpus electrónicos. Señalamos entre ellos a GlossaNet, que utilizando unos programas con amplios recursos lingüísticos envía los resultados de las búsquedas por e-mail.

5. LA TRADUCCIÓN DE LAS LV Y EL DICCIONARIO ELECTRÓNICO?

Traducir significa buscar la mejor forma posible de pasar de un idioma a otro un texto en el que el traductor tiene que conseguir reconstruir el mensaje. Además, tiene que intentar reproducir abundante información de tipo lingüístico, estilístico, cultural, social, etc., que suele estar presente en todo enunciado. Para traducir, los traductores suelen consultar tanto los diccionarios monolingües como los bilingües, en busca de esa tan preciada información. Sin embargo, a pesar de la importancia de la información que debe figurar en los diccionarios, el escaso tratamiento dado a las UFS por estos últimos ha permitido no solamente que éstas siguiesen siendo unas grandes desconocidas para el usuario en general debido a las dificultades de consulta y aprendizaje, sino que además ese desconocimiento llegó a hacer pensar a algunos teóricos que las expresiones idiomáticas eran generalmente intraducibles.

Toda la información, (morfológica, sintáctica, diastemática, semántica), que vamos a recabar durante la elaboración de nuestro diccionario electrónico, para cada LV nos permitirá constituir un banco de datos de incalculables prestaciones. Lo que sí está claro es que un trabajo de esa magnitud es muy necesario, pues va a permitir por un lado a los usuarios de la lengua conocer la gran mayoría de las LV existentes, con sus significados y sus variantes. Por otro lado también va a permitir a los traductores y a los estudiantes extranjeros, al realizar el mismo trabajo individualizadamente para no privilegiar o permitir que un idioma de partida influyese sobre los otros, que éstos encuentren la traducción mas adecuada.

Para ello, se utilizarán dos fascs.

En una primera fase, gracias al diccionario electrónico y a su soporte informático, realizaremos una búsqueda alfabética y semántica de una determinada LV, y de las LVS que

dentro del mismo campo semántico sean sinónimas, antónimas, hipónimas²⁵. Esta es una traducción sistemática que tiene por objeto encontrar, gracias a la elaboración de un paradigma de unidades léxicas relacionadas por el significado en la lengua fuente y a un paradigma similar para la lengua meta, el equivalente más próximo de las LV que queramos traducir. Ejemplo, cuando algo vale mucho dinero, hemos encontrado los paradigmas siguientes en español y en francés:

13. Costar./Valer un ojo de la cara. (DUE)	1. Coûter un prix fou.
14. Costar un riñón. (DUE).	2. Coûter les yeux de la tête. (PR, fam)
15. Costar un Potosí. (DUE).	3. Coûter la peau du dos (th l, fam)
	4. Coûter la peau des fesses. (AR, très fam)
16. Costar un huevo (y la yema del otro).	5. Coûter la peau du cul. (AR, très fam)
17. Costar un cojón.	6. Coûter la peau des miches (AR, très fam)
18. Costar más que una querida.	7. Coûter la peau des couilles (AR, très fam)
19. Costar más que un hijo tonto.	

Podemos observar las siguientes características:

En las expresiones españolas, la nº 15, *costar un Potosí*, pertenece al tipo de expresiones que tienen una palabra diacrítica, o al menos una referencia a su pasado histórico, mientras que en las palabras francesas no hay ninguna de las expresiones con alguna palabra de este tipo.

Las expresiones españolas 13 y 14 y la número 1 francesa pertenecen al mismo registro de lengua estándar.

Las expresiones españolas 16, 17, 18 y 19 no aparecen en los diccionarios españoles, y pertenecen por lo tanto a la lengua oral y familiar, (pero no hay ninguna indicación que nos indique a qué registro de lengua exacto pertenecen).

Las expresiones francesas, 2 y 3 pertenecen al registro familiar.

Las expresiones francesas 4, 5, 6, y 7 pertenecen al registro muy familiar.

En una segunda fase, en la que tendremos en cuenta que no se deben en la medida de lo posible traducir palabras aisladas, aplicaremos ese criterio a la traducción de las LVS. En efecto en la práctica, el traductor no se enfrenta generalmente a estas unidades de forma aislada, sino dentro de un contexto. Por lo tanto realizaremos o buscaremos contextos para estas U léxicas. Esta operación que se ha llegado a realizar manualmente²⁶, en un costoso proceso de lectura, supondría para un corpus de unas 20.000 locuciones verbales como el que estamos elaborando una investigación imposible de realizar debido al coste en tiempo y personal necesario para localizar esos textos. Sin embargo, hoy en día, esas operaciones se pueden realizar como hemos visto en el apartado anterior con la elaboración de los corpus, mecánicamente, en un rápido proceso gracias a las posibilidades de gestión y localización de datos de los programas informáticos.

25 Toda la información del diccionario electrónico que estamos elaborando en varios idiomas, nos va a permitir realizar operaciones de fraseología contrastiva. Ésta persigue como objetivo general la determinación de las semejanzas y diferencias existentes entre los diferentes sistemas fraseológicos analizados, estudiando las correspondencias sintácticas y las relaciones semánticas de cada lengua analizada

26 Hoy en día en aquellos corpus basados en obras literarias con derecho de autor, aún se tiene que hacer manualmente.

Los diccionarios electrónicos permiten, además, gracias a su gran capacidad de memoria, incorporar esos corpus con referencias a fuentes textuales en las que se han encontrado esas expresiones y almacenarlas, almacenarlas y tratar solamente una parte de esa información o tratarla en su totalidad sin que ello afecte en nada el volumen de la obra cosa que un diccionario en soporte papel no permite debido a la importancia que representa en su concepción el tamaño del soporte papel. Hasta la aparición y el perfeccionamiento de la informática no se podía soñar con poder realizar todos esos procesos tan necesarios para poder sacar conclusiones apoyadas en ejemplos.

6. RECAPITULATIVO Y PERSPECTIVAS

La traducción de las unidades fraseológicas supone para cualquier traductor que se precie de serlo un verdadero reto, debido al escaso tratamiento que han tenido en los diccionarios y a la escasa información que se puede conseguir acerca de las propiedades de estas unidades léxicas. Después de mostrar esas deficiencias en la primera parte de este trabajo, hemos expuesto los diferentes tipos de diccionarios electrónicos surgidos recientemente, que combinan a la par la lengua y la informática, para el tratamiento automático del lenguaje (TAL), para elaborar gramáticas electrónicas (Lexique-grammaire) y diccionarios semasiológicos en los que la semántica hace su aparición. Ahora bien, la elaboración de un diccionario representa una labor metódica y compleja que se inserta en un proyecto ambicioso que debe desarrollarse a largo plazo, máxime cuando como el que nosotros proponemos y estamos ya realizando reúne características de los tres simultáneamente.

Con vistas al futuro, se están desarrollando proyectos relacionados con el aprendizaje automático, con el diseño de sistemas capaces de procesar y etiquetar información multimedia y también capaces de tratar esa información para aplicarla en traducción automática, reconocimiento de palabras y construcción de corpus. Por lo tanto o mucho nos equivocamos o los diccionarios electrónicos están destinados a ser utilizados en un futuro cada vez más por estudiantes de un idioma extranjero, usuarios de ese idioma, traductores que deseen servirse de una herramienta de trabajo potente y rápida que les facilite mucho la labor manual que en numerosos casos venían desarrollando hasta la fecha.

REFERENCIAS BIBLIOGRÁFICAS

- Alvar Ezquerro, M. (1981): «Los diccionarios bilingües: su contenido», *Lingüística Española Actual*, II, 1 175-196. (Reeditado en el capítulo 8 de *Lexicografía descriptiva 1993*. Barcelona: bibliograf).
- Alvar Ezquerro, M. (2002): *Los diccionarios de ayer y hoy*, Granada, Comares.
- Bally, C. (1909=1951): *Traité de stylistique française*. Vol.1. Paris. Librairie Klincksieck.
- Baptista, J.; Català, D. (2002). «Compound temporal Adverbs in Portuguese and in Spanish», *Advanced in Natural Language Processing*, Berlin: Springer, págs. 173-176.
- Blanco, Escoda, X (1999). «¿Cómo leer un diccionario? Algunos aportes de la lingüística informática». *Cenários Interactivos. Arte e tecnologia*. Escola Superior de Educação: Santarem, págs. 117-127.
- Blanco, X. (2002): «Les déterminants figés», *Langages*, 145, págs. 61-81.
- Casas Gómez, M. (1999): *Las relaciones léxicas*. Tübingen. Max Niemeyer Verlag.
- Català Guitart, D. (2003): *Les adverbos compostos. Approches contrastives en linguistique appliquée*. Thèse de doctorat. UAB.

- Corpas Pastor, G. (1997): *Manual de fraseología española*. Madrid. Gredos.
- Corpas Pastor, G. (2003): *Diez años de investigación en fraseología*. Vervuert. Lingüística Iberoamericana
- De Gioia, M. (2001): *Avverbi idiomatici dell'italiano. Analisi lessico-grammaticale*. Torino: L'Harmattan Italia.
- Escoda Cabré, M. (2004): *Preclasificación automática de las secuencias NA para la elaboración de diccionarios electrónicos multilingües*. Trabajo de investigación de 3er ciclo y máster Europeo TALN. UAB.
- García Page, M. (1998): «Expresión fija y sinonimia». G.Wotjak ed., *Estudios de fraseología y fraseografía del español actual*, Frankfurt am Main-Madrid: Vervuert-Iberoamericana, págs. 195-201.
- Gross, M. (1982): Une classification des phrases «figées» du français. *Revue Québécoise de linguistique*, vol. 11, n° 2, págs. 151-185.
- Gross, M. (1986): *Grammaire transformationnelle du français. Syntaxe de l'adverbe*, Paris: ASS-TRIL.
- González Rey, I. (2002): *La phraséologie du français*. Toulouse Presses Universitaires du Mirail.
- Guiraud, P. (1955=1969): *La sémantique*. Paris. PUF.
- Jespersen, O. (1926): *The philosophy of grammar*. London: Allen & Unwin Ltd.
- Lakoff, G. (1987): *Women, Fire and Dangerous Things. What categories Reveal about the Mind*. University of Chicago.
- Lamiroy, B.; Leclère, C.; Klein, J.-R.; Labelle, J. (2003): «Les expressions verbales figées dans quatre variétés du français. Le projet «BFQS»». *Cahiers de Lexicologie* 83, Paris: Champion, págs. 153-172.
- Luque Durán, J. de D.; A. Pamies Bertrán (eds.). (1998): *Léxico y fraseología*, serie Collectae, Granada.
- Martin, R. (1996): «Sur les facteurs du figement lexical», *La locution: entre langue et usage*, Paris. Éditions Fontenay / saint-Cloud.
- Mejri, S. (2003): «Traduction linguistique. Problème terminologique ou construction conceptuelle?». En Salah Mejri, Taïeb Baccouche, André Class, Gaston Gross (dirs): *Traduire la langue. Traduire la culture*. Tunis, Maisonneuves & Larose, págs. 177, 191.
- Mel'čuk, I. (1984): *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques*, vol. 1, Les presses de l'université de Montréal.
- Mogorrón Huerta, P. (1996): «Les expressions figées des verbes ser et estar suivies de Prép x. *Linguisticae Investigationes* XX:1, Amsterdam/Philadelphia: John Benjamins, 3-32.
- Mogorrón Huerta P. (2002): *La expresividad en las locuciones verbales en francés y en español*. Publicaciones de la Universidad de Alicante.
- Pamies, A., Bálmacz, M. y Ifiesta, E. (1998): *Criterios para una fraseografía onomasiológica automatizable*.
- Penadés Martínez, I. (2002): *Diccionario de locuciones verbales*, Madrid, Arco/Libros.
- Rey, A. (1979=992): *La terminologie. Noms et notions*. Presses Universitaires de France.
- Rey-Debove, J. (1971): *Étude linguistique et sémiotique des dictionnaires français contemporains*, Paris-La Haya, Mouton.
- Ruiz Gurillo, L. (1997): *Aspectos de fraseología teórica española*. Valencia: Universidad.
- Santamaría Pérez, I. (2003): *La fraseología española en el diccionario bilingüe español-catalán: aplicaciones y contrastes*. Publicaciones Universidad de Alicante.
- Zgusta, (1971): *Manual of lexicography*. The Hague – Paris: Mouton.
- Zuluaga, A. (1980): *Introducción al estudio de las expresiones fijas*. Frankfurt, n° 10 de Studia Romaniaica, Verlag Pcter Lang.