

Sistema Diccional Multilingüe: aproximación funcional

Agirre E.(*) , Arregi X., Artola X., Díaz de Ilarraza A., Sarasola K.

Informatika Fakultatea
P.K. 649
20080 DONOSTIA (Basque Country - Spain)
xabier@si.ehu.es

Evrard F.

ENSEEIH
2, rue Charles Camichel
31071 TOULOUSE (France)
evrard@enseeih.fr

Resumen

El Sistema Diccional Multilingüe (SDM) que aquí presentamos se ha concebido como un sistema de ayuda al traductor humano. SDM es un sistema computacional on-line ideado para su uso en entornos de trabajo para la traducción. Su base de conocimiento se construye a partir de diccionarios de distintas lenguas (en la implementación actual de diccionarios de francés y de euskara). El desarrollo del sistema monolingüe SIAD, Sistema Inteligente de Ayuda Diccional [Artola, 93], constituye la base y el punto de partida en el diseño de la Base de Conocimiento Diccional Multilingüe.

La descripción del entorno funcional constituye la primera fase en el diseño de SDM. La metodología seguida en la definición de las funciones de SDM se ha basado en la observación directa de la labor de los traductores y en entrevistas personales a traductores profesionales. Mediante este análisis se ha caracterizado el uso típico de los diccionarios en la traducción, incidiendo sobre todo en los aspectos cualitativos. Actualmente se está trabajando en la representación del conocimiento diccional multilingüe, análisis de las relaciones existentes y características que debe tener dicha representación para abordar la puesta en marcha de la funcionalidad descrita.

Tema(s): Lexicografía computacional

(*): En el transcurso de este trabajo disfrutaba de una beca del Plan de Formación de Investigadores del Gobierno Vasco.

1.- Introducción

Es indudable que el diccionario es una herramienta de trabajo necesaria en el proceso de traducción, no ya entre traductores que no dominen a la perfección las lenguas objeto de traducción, sino incluso entre traductores expertos o profesionales. Este es un hecho evidente. Sin embargo, lo que ya no es tan evidente es cómo ha de utilizarse el diccionario, cuáles son las formas de acceso más eficientes y qué planes o estrategias han de seguirse en la búsqueda de respuestas satisfactorias a demandas concretas. Esta dificultad intrínseca al uso del diccionario origina que en muchas -demasiadas- ocasiones la sensación de fracaso se apodere del traductor no-experto, que no es capaz de encontrar lo que buscaba en las páginas del diccionario.

El Sistema Diccional Multilingüe (SDM) que aquí presentamos se ha concebido como un sistema de ayuda al traductor humano, para posibilitar que el uso de los diccionarios sea más ágil y eficiente. SDM es un sistema computacional on-line ideado para el uso en entornos de trabajo para la traducción. Su base de conocimiento se construye a partir de diccionarios de distintas lenguas (en la implementación actual de diccionarios de francés y de euskara). La naturaleza de los diccionarios, asimismo, es heterogénea: se ha considerado la interconexión de diccionarios monolingües explicativos con diccionarios bilingües, diccionarios de sinónimos y diccionarios especializados, éstos últimos fundamentalmente técnicos.

El desarrollo del sistema monolingüe SIAD, Sistema Inteligente de Ayuda Diccional [Artola, 93], constituye la base y el punto de partida en el diseño de la Base de Conocimiento Diccional Multilingüe. Las técnicas de adquisición, representación y explotación del conocimiento diccional desarrolladas en SIAD han de ser enriquecidas en SDM, adecuándolas a las necesidades del traductor humano.

La interconexión de diferentes diccionarios posibilita la explotación no sólo de las relaciones intradicionales sino también de las interdicionales. El sistema de inferencia del que se dotará a SDM permitirá extraer información diccional que, en principio, no es explícita en los diccionarios de base. Para ello el traductor humano podrá hacer uso de ciertas operaciones que han sido definidas a partir de la observación de procesos humanos de traducción.

En el siguiente apartado se describe brevemente el modelo de representación del conocimiento diccional. La metodología que se ha seguido para la determinación y especificación de las funciones de SDM se explica en el apartado 3. En el punto 4 se

recogen diversas consideraciones generales sobre el proceso de traducción y el uso de los diccionarios, consideraciones que son relevantes a la hora de establecer el aspecto funcional de SDM. En el apartado 4, se describe la funcionalidad que ofrece SDM. Finalmente, en el apartado 5 se recogen las conclusiones más relevantes.

2.- La base de conocimientos de SDM. El Sistema Diccional Monolingüe SIAD

La base de conocimientos de SDM se obtiene a partir de la información contenida en diccionarios de distintas lenguas (francés y euskera). El sistema, en la versión que se está diseñando, se compone de tres módulos de conocimiento (SDMOL1, SDMOL2 y SBL1/2), contruídos a partir de la información extraída de dos diccionarios monolingües, correspondientes a las lenguas fuente y objeto de traducción, y uno bilingüe, que relaciona ambas. Los módulos monolingües se estructuran según la propuesta de representación desarrollada en SIAD (sistema monolingüe concebido para asistir al usuario humano en las tareas de comprensión y producción del lenguaje) [Agirre et al., 93]. La estructura de SDM queda reflejada en la figura 1.

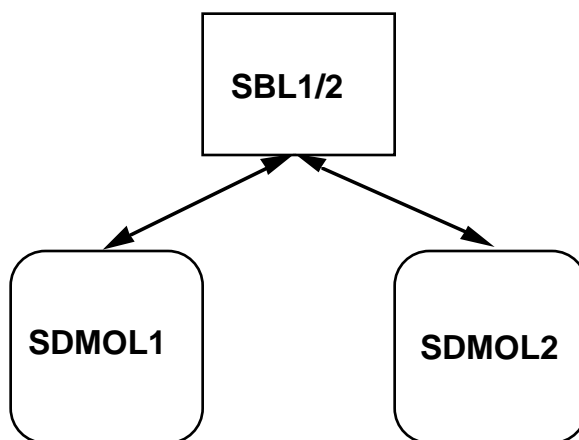


Fig. 1.- Estructura de SDM.

El esquema de representación utilizado en SIAD para la Base de Conocimiento Diccional (BCD) se compone de tres sub-bases de conocimiento:

- KB-THESAURUS refleja la representación del diccionario como una red semántica de frames, donde cada frame representa un concepto (acepción de palabra) o una configuración (representación de estructuras sintagmáticas asociadas a las ocurrencias de los conceptos en las definiciones). Los frames están relacionados mediante slots que representan relaciones léxico-semánticas tales como la sinonimia, relaciones taxonómicas (hiperonimia, hiponimia, la propia taxonomía), relaciones meronímicas (parte-de,

elemento de, conjunto-de, miembro-de), relaciones específicas obtenidas a partir de relatores metalingüísticos, etc.

- KB-DICTIONARY permite acceder desde el nivel de palabra de diccionario al correspondiente nivel de concepto de la BCD. Cada unidad de esta base de conocimientos representa una entrada (palabra) del diccionario y está directamente ligada a sus distintas acepciones en KB-THESAURUS.

- KB-STRUCTURES contiene metaconocimiento sobre los conceptos y relaciones de KB-DICTIONARY y KB-THESAURUS. En esta base de conocimientos se especifican todas las estructuras y se definen las características de los slots. Las unidades de KB-THESAURUS y KB-DICTIONARY son subclases o instancias de clases definidas en KB-STRUCTURES.

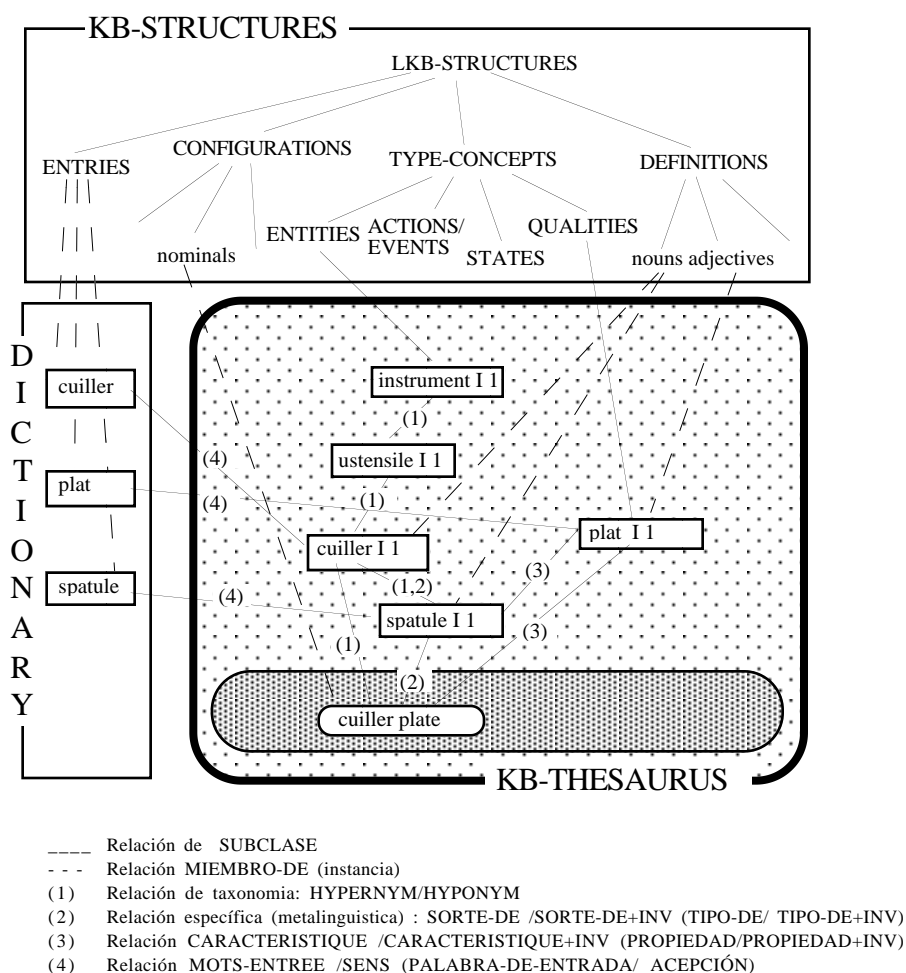


Fig. 2.- Estructura de la Base de Conocimientos Diccional de SIAD.

En la figura 2 se muestra la representación obtenida para el concepto *spatule I 1*: *sorte de cuiller plate* (un tipo de cuchara plana). A pesar de que no está explícito en la definición, *spatule* aparece como un tipo de utensilio en la taxonomía de conceptos y por

tanto heredará algunas de sus características (según el tipo de herencia definido para cada una de ellas). También se muestran los tipos de conceptos usados: *spatule I I* y *cuiller I I* son definiciones de nombres y considerados subclases de ENTITIES mientras que *plat I I* (un adjetivo) es subclase de QUALITIES. La unidad-configuración que representa al sintagma nominal *cuiller plate* se trata como un hipónimo de su concepto núcleo (*cuiller I I*).

3.- Analisis del uso de los diccionarios

Antecedentes

El uso del diccionario, aún siendo considerado como una actividad común y cotidiana, plantea importantes interrogantes que hoy día siguen vigentes. En los últimos años se ha investigado el tema desde diferentes perspectivas, tratando de incidir en la tipología de los diccionarios, en la de los propios usuarios, en las necesidades o bien en las formas de acceso. Presentamos a continuación algunos estudios de carácter práctico que se basan en datos empíricos conseguidos a partir de los propios usuarios.

En 1962 Clarence Barnhart publicó un estudio sobre el uso de diccionarios monolingües comerciales. Basó su análisis en 108 cuestionarios, que fueron enviados a profesores de inglés y en los que se trataba de ponderar el uso de seis tipos de información, a saber: significado, ortografía, pronunciación, sinonimia, etiquetas de uso y etimología.

Randolph Quirk dió a conocer su análisis en 1973, y en él trataba de determinar las actitudes de los usuarios de diccionarios, así como sus expectativas y prejuicios. Para ello se sirvió de un cuestionario de 30 preguntas dirigido a 220 estudiantes de distintas áreas.

Zelko Bujas (1975) trató de testear el rendimiento de un diccionario inglés-croata en la lectura de revistas contemporáneas inglesas por parte de 18 estudiantes. Las conclusiones de este análisis resultaron válidas para mejorar los diccionarios bilingües existentes. Este tipo de test de utilidad también fue seguido por Kurt Opitz (1979) y Karl Müller (1983).

Por otra parte, el interés por determinar el tipo de necesidades a las que suele responder el diccionario ha dado lugar a varios e interesantes estudios. Así, cabe reseñar el trabajo de Tomaszczyk (1979) cuya conclusión más relevante fue que el uso del diccionario depende de la naturaleza de la actividad que se lleva a cabo. En esta línea, Béjoint (1981) trató de clarificar, vía cuestionarios dirigidos a estudiantes franceses de

inglés, la interacción entre las actividades de producción activa y recepción pasiva, y concluyó que los estudiantes no hacen uso de la riqueza de información ofrecida por los diccionarios. Hartmann, por su parte, se centró en la traducción, considerando a ésta como la actividad más popular, y dedujo que la necesidad más importante del traductor radica en el significado y en la gramática. Baxter (1980) concluyó por medio de un cuestionario dirigido a estudiantes japoneses de inglés que éstos tienden a hacer uso de diccionarios bilingües en detrimento de los monolingües, aunque en su opinión se debería estimular el uso de diccionarios monolingües. Y, por su parte, Kürhn y Püschel (1982) establecieron que algunos diccionarios especializados (especialmente de sinónimos) se utilizan por lo menos tanto como los de propósito general.

Uno de los aspectos en los que menos se ha profundizado es el relacionado con las estrategias y caminos seguidos por los usuarios en la búsqueda de respuestas a sus requerimientos. Cabe citar, el trabajo de Marsha Bensoussan y sus colegas (1983) en dos universidades de Israel, en el que se obtuvo como conclusión que el uso de los diccionarios no cambia significativamente el rendimiento en la comprensión de textos por parte de estudiantes de inglés, dado que los de alto nivel no lo necesitan demasiado y los de bajo nivel no saben sacar provecho de él. Josh Ard (1982) centró su atención en el uso de diccionarios bilingües por parte de los estudiantes de inglés (de nivel intermedio-alto) en las tareas de escritura o redacción. Una de las técnicas de recogida de datos que se siguió en este interesante trabajo consistió en filmar la actividad de dos estudiantes y en la obtención de datos a partir de protocolos orales. Evelyn Mitchell (1983) estudió las estrategias de lectura entre estudiantes de escuelas secundarias y dedujo que éstos tienen serias dificultades en la localización de palabras y sentidos en los diccionarios. Se han de considerar de especial interés los trabajos de Hatherall (1984) y Wiegand (1985), centrados en el uso del diccionario en la traducción alemán-inglés e inglés-alemán, respectivamente. Wiegand reseñó mediante un estudio basado en protocolos escritos que los diccionarios monolingües de la lengua destino no satisfacen las cuestiones gramaticales o de uso que plantea la traducción. Starren y Thelen publicaron en 1990 los primeros resultados de un estudio sobre el uso de los diccionarios en el proceso de traducción, estudio que se llevó a cabo entre estudiantes de la escuela holandesa de traducción.

En el marco del proyecto EURALEX/AILA (1988) se ha desarrollado un amplio estudio sobre el uso y la efectividad de los diccionarios entre estudiantes franceses, alemanes, italianos y españoles de inglés. En el análisis se han manejado 1100 respuestas recibidas desde siete países y clasificadas en cuatro grupos según el nivel de conocimiento. Cabe reseñar, entre los numerosos datos cuantificados y porcentajes obtenidos, que los diccionarios bilingües son más utilizados en la comprensión y

traducción de textos, mientras que para la información sobre modos de uso se recurre preferentemente a los monolingües. Son reseñables, asimismo, los estudios sobre las técnicas de búsqueda de unidades multipalabra y el manejo de información gramatical y de metalenguaje.

Hacia la funcionalidad de SDM

Más de la mitad de los trabajos mencionados se podrían calificar como "estudios vía cuestionario" y sólo unos pocos se basan en técnicas de observación directa. Los estudios a base de cuestionarios se han puesto en tela de juicio y actualmente se tiende a estudios basados en la técnica de observación directa complementados con protocolos, registro de entrevistas personales o incluso tests computerizados.

En nuestro caso se ha optado por la observación directa: dados varios textos para traducir y varios diccionarios (de características diferentes), se han registrado todos y cada uno de los problemas, resoluciones y tareas llevadas a cabo por los traductores.

Se puede considerar, por una parte, que el desarrollo de estos protocolos es dirigido por el observador, ya que el texto que se traduce no lo elige el traductor. Además, no siempre se traducen párrafos enteros, sino palabras, expresiones o sintagmas contextualizados, mediante los cuales se pretende analizar cómo resuelven los traductores situaciones bien concretas y determinadas. Se han seleccionado textos de naturaleza diversa, tanto literarios como técnicos, y las traducciones han sido bidireccionales, francés-euskara y euskara-francés. Por cada consulta llevada a cabo se ha registrado el término que se pretende consultar, el diccionario elegido, la entrada del diccionario a la que se accede, el tipo de consulta, el resultado obtenido y, si resultara de interés, alguna observación.

Pero, por otra parte, habría que considerar que el método es no-dirigido, en la medida en que el traductor que es observado no tiene por qué limitarse al marco preestablecido por el observador. De hecho, la iniciativa de la traducción le corresponde exclusivamente al traductor, incluso a la hora de elegir y utilizar los diccionarios que considere adecuados. Este es un aspecto reseñable de los protocolos que ha permitido detectar algunas utilidades al margen de las previsibles a priori.

Se ha contado con la colaboración de siete personas, tres de ellas son traductores profesionales y las cuatro restantes, a pesar de que traducen textos con relativa frecuencia, se pueden considerar como no expertos. Este grupo también es heterogéneo en lo que concierne al nivel de conocimiento del euskara y francés, lo cual ha permitido contrastar diferentes usos del diccionario en función de los niveles de conocimiento o del

tipo de usuario. Con los traductores profesionales se han mantenido entrevistas personales adicionales con el fin de profundizar en aspectos tales como las características que deberían tener los diccionarios dirigidos a la traducción, los problemas que plantea la elección del léxico adecuado, la necesidad de los sistemas computacionales de ayuda al traductor y la funcionalidad que deberían ofrecer estos sistemas.

4.- Aspectos observados en el proceso de traducción y la utilización de los diccionarios

La dificultad para modelizar el proceso de traducción humana ha quedado evidenciada tras las diferentes sesiones que se han llevado a cabo. El traductor, sobre todo aquel que no es profesional, no sigue un método concreto ni se guía por directrices establecidas. Aún en el caso de los profesionales sería difícil establecer pautas de actuación que fueran más allá de principios básicos; aquello de que "cada maestrillo tiene su librillo" es perfectamente aplicable a los traductores. A lo sumo se podrían discernir tres tareas muy generales que, de alguna forma, se dan en todo proceso de traducción: comprensión del texto original, búsqueda de equivalentes y producción del texto objetivo. Estas tareas no se desarrollan, necesariamente, de forma lineal en el tiempo, como en el caso de la traducción automática, sino que interaccionan y se solapan en función del texto y del traductor.

El traductor profesional centra su atención en la comprensión para, a partir del texto comprendido, producir un texto equivalente en la lengua destino. La búsqueda de equivalentes para los términos del texto original, aún siendo importante, no es la piedra angular de la traducción. No trata de buscar la equivalencia uno a uno entre los términos del texto original y objetivo. Por esta razón es característico entre estos traductores el uso de diccionarios monolingües. La frecuencia de acceso al diccionario no es alta y normalmente busca información muy puntual y precisa: definición de una acepción concreta, modos y ejemplos de uso, información gramatical,...

Para el traductor no experto, sin embargo, la búsqueda de equivalentes adecuados adquiere mucha más importancia, sobre todo cuando el conocimiento de las lenguas no es tan profundo. Se ha podido constatar que, a veces, incluso sin comprender perfectamente el contexto, el traductor tiende sin más a buscar equivalentes. Tampoco es extraño en este contexto que el traductor recurra a una tercera lengua (de la que tiene un conocimiento más preciso) que sirve de "pivote" o "puente" en la traducción. En estas circunstancias la tendencia a hacer uso de diccionarios bilingües se manifiesta claramente. La principal causa de esta elección es que el traductor a menudo tiene dificultades para comprender las definiciones de los diccionarios monolingües, bien porque no entiende alguna(s)

palabra(s) de las definiciones, bien porque se le escapa la estructura de las mismas. Ciertamente se trata de un problema de difícil solución, aunque sea patente el esfuerzo de los lexicógrafos por simplificar las definiciones.

En nuestro análisis se ha constatado que el uso de diccionarios bilingües especializados es generalizado cuando se trata de traducir textos técnicos. En este contexto los traductores, tanto expertos como no-expertos, asumen que el riesgo de cometer algún error de tipo léxico debido al uso de diccionarios bilingües se reduce notablemente.

En los protocolos se ha evidenciado que el diccionario debe ayudar tanto en la comprensión y en la búsqueda de equivalentes como en la producción del texto objetivo. Ciertamente, es en la producción dónde se manifiestan con mayor claridad las carencias de los diccionarios.

Otro de los aspectos relevantes para determinar la funcionalidad de SDM es el relacionado con la unidad de traducción. Es el texto original en su integridad el objeto de traducción y, por tanto, se ha de preservar su unidad. Sin embargo, la dinámica de la traducción y la imposibilidad manifiesta de traducciones texto-texto obliga al traductor a sucesivas divisiones del texto en unidades más pequeñas y accesibles. Este proceso descendente concluye normalmente cuando las unidades son de nivel léxico, ya sean palabras, ya sean locuciones. Es ésta la fase en que se interacciona con el diccionario. Ahora bien, todo término ha "heredado" características del texto y de la oración a la que pertenece, y el traductor no puede ignorarlas cuando accede al diccionario. De hecho, es el traductor quien debe salvar las distancias entre el término contextualizado del texto y las entradas descontextualizadas del diccionario; pero, en cualquier caso, el diccionario debería aportar información sobre los contextos de uso de los distintos términos.

5.- Funcionalidad de SDM

Como resultado del análisis llevado a cabo en torno a la actividad de los traductores y el uso del diccionario, además de extraer algunas consideraciones sobre el proceso de traducción, se ha determinado y caracterizado el conjunto de operaciones -búsquedas y verificaciones- que se dan sobre los diccionarios. Se podría hablar de la "demanda real" a la que se somete al diccionario por parte del traductor.

Dichas operaciones constituyen el entorno funcional del Sistema Diccional Multilingüe que se va a describir a continuación. Las funciones se agrupan en clases

según la tarea -comprensión del texto original, búsqueda de equivalentes y producción del texto objetivo- en la que se hace uso mayormente de ellas. Las funciones englobadas en la comprensión del texto fuente y la producción del texto objetivo también se dan en entornos monolingües y su especificación se describe ampliamente en [Arregi et al., 91]. Con el fin de adaptarlas al entorno multilingüe y multidiccional, a estas funciones se les han añadido parámetros que especifican cuál es la lengua en la que se ha de presentar el resultado y cuál es el diccionario al que se desea acceder (en el anexo I se recogen la especificación y ejemplos de estas funciones).

Comprensión del texto fuente

Las funciones englobadas en este grupo son:

- a) petición de definición
- b) reformulación de definición
- c) petición de propiedades de un concepto

Producción del texto objetivo

En esta actividad se utilizan básicamente estas funciones:

- d) búsqueda tesaúrica de conceptos
- e) búsqueda de relación entre conceptos
- f) petición de diferencias entre conceptos
- g) verificación de definición de un concepto
- h) verificación de propiedades de un concepto

Búsqueda de equivalentes para la traducción

El concepto de equivalencia en la traducción ha sido objeto de múltiples reflexiones. Existe una premisa que es asumida por la gran mayoría de autores: la equivalencia, entendida en su dimensión formal o matemática, es difícilmente aplicable en la traducción, entre otras razones porque cada lengua es producto y, por tanto, reflejo de determinados condicionantes culturales que, evidentemente, no se dan en todas las demás. Para la equivalencia total podemos definir que: dos unidades de traducción son equivalentes en un determinado contexto si y sólo si son análogos: i) sus significados, ii) sus niveles de registro y modos de uso, iii) sus connotaciones y metafóricidad.

i) PETICION DE EQUIVALENTES

Sería deseable que cuando un traductor busca equivalentes para una unidad de traducción léxica se preserven todos los rasgos de equivalencia. No obstante, en los diccionarios bilingües a partir de los que se extraen las relaciones de equivalencia no es frecuente

encontrar referencias explícitas a los niveles de registro y al uso figurativo de las palabras. Estas limitaciones se reflejan en nuestra base de conocimiento y, en consecuencia, la búsqueda de equivalentes totales no siempre resultará exitosa, bien porque existen diferencias en los rasgos, bien porque éstos no son explícitos. Por tanto, debemos tener en cuenta necesariamente la equivalencia parcial. En la especificación de EQUIV vamos a considerar que dos conceptos son equivalentes (aunque sólo sea parcialmente) si sus significados son análogos. Queda en manos del usuario decidir sobre el resultado de la función valorando los datos ofrecidos sobre uso y sentido figurativo de los conceptos. Por este motivo, definimos una unidad de traducción como una tripleta formada por un concepto, una etiqueta de uso y un indicador de sentido figurativo.

En nuestro análisis se ha demostrado que se da un porcentaje bastante alto de fracasos en la búsqueda de equivalentes para locuciones o, en general, elementos léxicos multipalabra. La función EQUIV de SDM facilita al traductor el acceso a este tipo de unidades complejas.

Otro de los problemas a tener en cuenta en los entornos multilingües está relacionado con los 'vacíos' léxicos, que se dan cuando a una unidad léxica de la lengua fuente no le corresponde ninguna unidad de la lengua destino. Este fenómeno, que no es en absoluto extraño, se debe tratar explotando las relaciones de sinonimia, hiperonimia/hiponimia, sus respectivos equivalentes y definiciones, de forma que el traductor pueda captar el significado de la unidad fuente y producir el equivalente (que no debe ser necesariamente una unidad léxica).

Especificación:

Datos: $U \in \text{unidad traducción}, L \in \text{lengua}, T \in \text{área}$

Resultados: $E \in \text{lista-unidades-traducción-equivalentes}$.

El parámetro área hace referencia al ámbito en el que se hace uso de la unidad léxica (es particularmente relevante en el tratamiento de textos técnicos).

Excepción: Cuando se dan 'vacíos' léxicos se utilizan los operadores 'C' y 'D' para expresar que el equivalente es una generalización o una especialización respectivamente del original.

Sintaxis:

petición-equivalentes=

```
"EQUIV" "(" unidad-traducción "," lengua-resultado "," área ","  
  lista-unidades-traducción-equivalentes ")".
```

Ejemplos:

1. T.- **EQUIV** ((|accusatif I 1|, ,), euskara, gram, ?LP)
S.- LP = ((|akusatibo I 1|, ,))
2. T.- **EQUIV** ((|coup_de_bec I 1|, ,), euskara, común, ?LP)
S.- LP = ((|mokokada I 1|, ,))
3. T.- **EQUIV** ((|cervelle I 2|, ,fig), euskara, común, ?LP)
S.- LP = ((|adimen I 1|, ,) , (|buru I 1|, ,fig))
4. T.- **EQUIV** ((|pattar I 1|, ,), francés, común, ?LP)
S.- LP = ((D,|eau-de-vie I 1|, ,))
5. T.- **EQUIV** ((|txakolin I 1|, ,), francés, común, ?LP)
S.- LP = ((C,|vin I 1|, ,))

Otras funciones

j) COMPATIBILIDAD SEMANTICA

Por medio de esta función se trata de determinar si dos unidades léxicas se pueden ligar según una relación dada atendiendo a sus restricciones de selección.

Especificación: COMPATIBILIDAD SEMANTICA

Datos: $U_1, U_2 \in \text{concepto}; R \in \text{relación}$

Resultados: *true* si U_1 y U_2 son relacionables mediante R . Dos unidades serán relacionables cuando sus rasgos semánticos no imposibilitan esa relación.

Sintaxis:

compatibilidad-semántica=

```
"COMP-SEM" "(" concepto "," concepto ","  
  relación "," resultado-compatibilidad ")".
```

Ejemplos:

1. T.- **COMP-SEM** (|poisson I 1|, |ruminer I 1|, AGENT)
S.- False
Establece que el concepto **pez** no puede ser **agente** del verbo **rumiar**.

k) FUNCIONES GRAMATICALES

Al traductor en el ejercicio de su labor se le plantean numerosos problemas de tipo gramatical. En las protocolos se ha evidenciado que, incluso ante estas cuestiones, el traductor tiende a consultar el diccionario. Bien es cierto que no es el diccionario la herramienta más idónea para aclarar las cuestiones de índole gramatical, pero en cualquier caso, es deseable que un sistema computacional como SDM soporte en alguna medida las peticiones gramaticales de los usuarios. A continuación se especifican las funciones que se han recogido en los protocolos:

Funciones morfológicas

Las funciones de esta subsección tienden a salvar las distancias entre la unidad léxica flexionada del texto y la entrada del diccionario. La implementación y puesta en marcha de estas funciones exige la incorporación e integración en SDM de analizadores/productores morfológicos.

Especificación: ANALISIS MORFOLOGICO

Datos: $U \in$ *unidad léxica*

Resultados: $L \in$ *lista-análisis*, análisis es un par de la forma:
(lema, información-morfológica).

Sintaxis:

análisis-morfológico=

"ANALI" "(" *unidad-léxica* ", " *lista-análisis* ")"

Ejemplos:

- T.-** ANALI ('liztorrari', ?LL)
S.- LL = (('liztor', (DATIVO, def, sg)))

Especificación: PRODUCCION-FORMA-LEXICA

Datos: $C \in$ *concepto*; $K \in$ *categoría-gramatical*; $M \in$ *morfología-resultado*

Resultados: $U \in$ *unidad léxica*, según la información aportada por M.

Sintaxis:

producción-forma=

"PROD" "(" *concepto* ", " *categoría* ", " *morfología* ", "
unidad-flexionada ")"

Ejemplos:

- T.-** PROD (|liztor I 1|, nom., (DATIVO, def, sg), ?U)
S.- U = 'liztorrari'

Especificación: TRADUCCION-LEXICA

Datos: $U_1 \in$ *unidad léxica*, $L \in$ *lengua*

Resultados: $U_2 \in$ *unidad léxica*, los lemas de U_1 y U_2 son equivalentes y la morfología de ambas formas es idéntica.

Sintaxis:

producción-forma=
"TRAD-LEX" "(" unidad léxica "," lengua "," unidad-flexionada ")"

Ejemplos:

1. **T.-** **TRAD-LEX** ('goaz', francés, ?U)
S.- U = 'allons'

Funciones sintácticas

Especificación: PATRON-SINTACTICO

Datos: $P \in$ *identificador-patrón-sintáctico*; $L \in$ *lengua*

Resultados: $S \in$ *lista-construcciones-sintácticas*

Sintaxis:

patrón-sintáctico=
"PAT-SINT" "(" *identificador-patrón-sintáctico* "," lengua ","
lista-construcciones-sintácticas ")"

Ejemplos:

1. **T.-** **PAT-SINT** ('CONCESIVA', euskara, ?LS)
S.- LS = ('nahiz' 'nahiz eta' 'ba... ere' 'arren')

Especificación: REGIMEN-VERBAL

Datos: $C \in$ *concepto verbal*

Resultados: $G \in$ *lista-casos-gramaticales-subcategorizados*

Sintaxis:

régimen-verbal=
"REG-VERB" "(" *concepto-verbal* "," lista-casos-subcategorizados ")"

Ejemplos:

1. **T.-** **REG-VERB** (|mintza(tu) I 1|, ?LG)
S.- LG = (nor [nori/norekin] [zertaz])
El verbo *mintza(tu)* (hablar) siempre debe llevar el caso *nor* (sujeto),
opcionalmente el caso *nori* o *norekin*, y también opcionalmente *zertaz*.

Especificación: COLOCACION LEXICA

Datos: $U \in$ *concepto* ; $F \in$ *función léxica*

Resultados: $L \in$ *lista-unidades-léxicas*

Sintaxis:

colocación=
"COLOC" "(" *concepto* "," *función-léxica* ","
lista-unidades-léxicas ")"
dónde el rango de valores del parámetro *función-léxica* es el
establecido por Mel'cuk [Mel'cuk, 82]

Ejemplos:

1. **T.-** **COLOC** (|beldur I ?|, Magn, ?LC)
S.- LC = ('izugarria' 'hezurretarainoko' '['beldurrak] aidean')

1) PETICION DE EJEMPLOS

La utilidad e importancia de los ejemplos ilustrativos se ha manifestado claramente en las sesiones de traducción efectuadas. Los ejemplos del diccionario sirven para:

- Probar que una palabra o una acepción existe en un lengua, es decir, mostrar esa palabra en un contexto real y en un entorno cultural concreto.
- Ilustrar el significado de la palabra definida.
- Ilustrar el comportamiento gramatical.
- Mostrar aspectos estilísticos.
- Ilustrar casos típicos de colocación.

Especificación:

Datos: $C \in \text{concepto}$, $L \in \text{lengua}$

Resultados: $E \in \text{lista-ejemplos}$.

Sintaxis:

petición-ejemplos=

"EJEM" "(" concepto "," lengua-resultado "," lista-ejemplos ")".

Ejemplos:

1. T.- **EJEM** (|battre I 1|, francés, ?LE)
- S.- LE = ('battre des mains' 'battre les cartes')

m) PARAFRASE

El parafraseo de unidades multipalabra más o menos complejas ha sido una de las funciones sugeridas por los encuestados. La acción de parafrasear implica el manejo de conocimiento no sólo léxico, sino también morfológico y sintáctico. Por ello, su descripción y diseño queda fuera del ámbito de aplicación de SDM en su primera fase, que como ya se ha mencionado, sirve de soporte léxico en la tarea de traducción.

6.- Conclusiones

La descripción del entorno funcional constituye la primera fase en el diseño del Sistema Diccional Multilingüe, sistema computacional de ayuda a un traductor humano. La metodología seguida en la definición de las funciones de SDM se ha basado en la observación directa de la labor de los traductores y en entrevistas personales a traductores profesionales. Mediante este análisis se ha caracterizado el uso típico de los diccionarios en la traducción, incidiendo sobre todo en los aspectos cualitativos.

La interconexión de módulos de conocimiento diccionarial posibilita la explotación de los nexos de equivalencia entre conceptos de distintas lenguas. Esta posibilidad permite enriquecer y adecuar a un entorno multilingüe las funciones definidas para el sistema monolingüe SIAD.

La puesta en marcha de la funcionalidad descrita plantea problemas de diversa naturaleza. En particular, la desambiguación de las palabras (unidades de los textos de traducción y de las definiciones diccionariales) en conceptos (unidades de la red

semántica) no es en absoluto trivial y exige un procesamiento previo, a veces interactivo, de las demandas del usuario. Por otra parte, las limitaciones de un sistema diccionario demandan la integración de analizadores/productores morfológicos para la implementación de funciones tales como ANALI, PROD, TRAD y PARAFRASE. Asimismo, en las funciones PAT-SINT, REG-VERB, COLOC y COMP-SEM se plantea la necesidad de información extradiccional y de formalizar desde el punto de vista lingüístico algunas de las estructuras de datos utilizadas (*identificador-patrón-sintáctico, construcción-sintáctica,...*).

Actualmente se está trabajando en la representación del conocimiento diccionario multilingüe, análisis de las relaciones existentes y características que debe tener dicha representación para abordar la puesta en marcha de la funcionalidad descrita. Se está desarrollando un entorno bilingüe (francés-euskara), a partir de dos diccionarios monolingües y otro bilingüe. La BCD del prototipo implementado de SIAD contiene 6.003 conceptos, extraídos de las definiciones de un diccionario explicativo en francés. En lo concerniente al euskara, se han representado hasta el momento las definiciones de 50 entradas diccionariales.

7.- Agradecimientos

Nuestro más sincero agradecimiento a Edorta Agirre, Olatz Arregi, Itziar Baragaña, Bittori Fernández, Antxon Gurrutxaga, Xabier Mendiguren y Miriam Urkia, las y los traductores que han colaborado en este estudio.

Este trabajo se encuadra dentro de un proyecto subvencionado por la Diputación Foral de Gipuzkoa.

Bibliografía

- Al-Kasimi A.. *Linguistics and Bilingual Dictionaries*. Leiden: E.I.Brill, 1977.
- Agirre E., Arregi X., Artola X.,Díaz de Ilarraza A., Evrard F., Sarasola K.. Construction of Phrasal Pattern Hierarchies to Produce a Semantic Representation of Dictionary Definitions. Enviado a ALKT'93 (Ohio). 1993.
- Ard J.. The use of bilingual dictionaries by ESL student while writing, *ITL Review of Applied Linguistics*, 58, 1-27.1982.
- Arregi X., Artola X.,Díaz de Ilarraza A., Evrard F., Sarasola K.. Aproximación funcional a DIAC: Diccionario inteligente de ayuda a la comprensión, *Boletín SEPLN*, 11, 127-138. 1991.
- Artola X.. HIZTSUA: Hiztegi-sistema urgazle adimendunaren sorkuntza eta eraikuntza./ Conception d'un système intelligent d'aide dictionnaire (SIAD). Tesis. UPV-EHU. 1993.

- Atkins B.T. and Knowles F.E.. Interim report on the EURALEX/AILA Research Project into Dictionary use, in T. Magay, J. Zigány eds., 381-392, *Proceedings BudaLEX'88*. Budapest: akadémiai kiadó, 1990.
- Barnhart C.L.. Problems in editing commercial monolingual dictionaries, in F. Householder, S. Saporta eds., 161-181, *Problems in Lexicography*. Bloomington: Indiana University, 1967.
- Hartman R.R.K.. Four perspectives on dictionary use: a critical review of research methods. The Dictionary and the Language Learner, in A.Cowie ed., 11-28, *Papers from the EURALEX Seminar (Leeds)*, *Lexicographica, Series maior*, 17, Tübingen; Niemeyer. 1985.
- Johnson P. and Johnson H.. Knowledge analysis of tasks: task analysis and specification for human-computer systems, in A.Downton ed., 119-144, *Engineering the human-computer interface*. London: McGraw-Hill, 1991.
- Martin W. and Al B.P.F.. User-orientation in dictionaries: 9 propositions, in T. Magay, J. Zigány eds., 393-399, *Proceedings BudaLEX'88*. Budapest: akadémiai kiadó, 1990.
- Mel'cuk I.A.. Lexical functions in lexicographic description, *Proceedings of the VIII-th Annual meeting of the Berkeley Linguistic Society (Berkeley)*, 427-444. 1982.
- Neubach A. and Cohen A.D.. Processing Strategies and Problems Encountered in the use of Dictionaries, *Dictionaries. Journal of the Dictionary Society of North America*, 10, 1-20. 1988.
- Neubert A.. Fact and Fiction of the Bilingual Dictionary, *Actas del IV congreso Internacional.EURALEX'90 (Benalmádena)*,29-42, Bibliograph, 1992.
- Starren P. and Thelen M.. General dictionaries and students of translation: A report on the use of dictionaries in the translation process, in T. Magay, J. Zigány eds., 447-458, *Proceedings BudaLEX'88*. Budapest: akadémiai kiadó, 1990.
- Zgusta L.. Translational Equivalence in a Bilingual Dictionary.*Dictionaries. Journal of the Dictionary Society of North America*, 9, 1-47. 1987.

ANEXO I

ESPECIFICACION Y EJEMPLOS DE LAS FUNCIONES PARA LA COMPRESION Y PRODUCCION DE TEXTOS (Arregi et al. 91)

Comprensión del texto fuente

a) PETICION DE DEFINICION

Especificación:

Datos: $C \in$ concepto; $N \in$ nivel-explicación; $T \in$ diccionario; $L \in$ lengua

Resultados: $D \in$ definición

Sintaxis:

petición-definición=

"DDEF" "(" concepto-thesaurus "," nivel-explicación "," diccionario
"," lengua-resultado "," definición ")"

Ejemplos:

1. **Traductor.- DDEF** (|guêpe I 1|, textual, LPPL, francés, ?D)
Definición de **avispa** en francés con nivel de explicación "textual"
Sistema.- D= 'insecte hyménoptère à aiguillon'
2. **T.- DDEF** (|guêpe I 1|, textual, AEH, euskara, ?D)
S.- D= 'erlearen antzeko intsektu himenoptero miztoduna'
3. **T.- DDEF** (|guêpe I 1|, heredado, LPPL, francés, ?D)
Definición de **avispa** en francés con nivel de explicación "heredado"
S.- D= (and (|guêpe I 1| HYPERONYME |insecte I 1|)
(|guêpe I 1| CARACTERISTIQUE |hyménoptère I 1|)
(|guêpe I 1| CARACTERISTIQUE |articuler I 1#m|)
(|guêpe I 1| POSSESSION |aiguillon I 1|)
(|guêpe I 1| POSSESSION |patte I 1#n|)
(|guêpe I 1| HYPONYME |frelon I 1|)
(|guêpe I 1| POSSESSEUR |guêpier I 1|))
Avispa es un **insecto himenóptero** y **articulado** que posee un **agijón** y **seis patas**. Un **avispon** es una **avispa** y un **avispero** posee **avispas**

b) REFORMULACION DE DEFINICION

Se trata de una función secundaria que se utiliza después de una petición de definición.

Especificación:

Datos: $D \in$ definición; $N \in$ nivel-explicación; $T \in$ diccionario; $L \in$ lengua

Resultados: $D' \in$ expresión-definición, donde $D' \cong D$

Sintaxis:

reformulación-definición=

"RDEF" "(" definición "," nivel-explicación "," diccionario ","
lengua-resultado "," definición-reformulada ")"

Ejemplos:

1. T.- DDEF (|email I 1|, local, LPPL, francés, ?D)
Definición de **email** en francés (nivel de explicación "local")
- S.- D= (and (|email I 1| HYPERONYME |enduit I 1|)
(|email I 1| CARACTERISTIQUE |vitri fiable I 1|)
(|email I 1| POSSESSEUR |faïence I 1|))
Esmalte es un baño **vitri fiable** de la **procelana**
- T.- RDEF ((and (|email I 1| HYPERONYME |enduit I 1|)
(|email I 1| CARACTERISTIQUE |vitri fiable I 1|)
(|email I 1| POSSESSEUR |faïence I 1|)),
local, LPPL, francés, ?DR)
Reformulación de la definición precedente en francés (nivel de explicación "local").
- S.- DR= (and (|email I 1| HYPERONYME |substance I 1|)
(|email I 1| CARACTERISTIQUE |liquide I 2|)
(|email I 1| CARACTERISTIQUE |vitri fiable I 1|)
(|email I 1| POSSESSEUR |faïence I 1|))
Esmalte es una **sustancia líquida vitri fiable** de la **procelana**

c) PETICION DE PROPIEDADES DE UN CONCEPTO

Especificación:

Datos: $C \in \text{concepto}$; $A \in \text{atributo}$; $T \in \text{diccionario}$;

Resultados: $V \in \text{lista-valores}$

Sintaxis:

petición-propiedades= "DPRO" "(" concepto-thesaurus "," atributo
"," diccionario "," lista-valores ")".

Ejemplos:

1. T.- DPRO (|ambre I 1|, |couleur I 1|, LPPL, ?LV)
¿De qué **color** es el **ámbar**?
- S.- LV= ((CARACTERISTIQUE (|jaune I 1|)))
Jaune.
2. T.- DPRO (|ambre I 1|, ?P, LPPL, ?LV)
Propiedades del **ámbar**.
- S.- LV= ((HYPERONYME(|résine I 1|))
(CARACTERISTIQUE(|fossile I 1|
|jaune I 1||translucide I 1|)))
Se trata de una **resina**, **fósil**, **amarilla** y **translúcida**.

Producción del texto objetivo

d) BUSQUEDA TESAURICA DE CONCEPTOS

El objetivo de esta función es obtener la lista de conceptos que satisfacen unas determinadas restricciones. Especificación:

Datos: $R \in \text{expresión-restricciones}$; $T \in \text{diccionario}$; $L \in \text{lengua}$

Resultados: $LC = \langle C_1, C_2, \dots, C_n \rangle$, lista de conceptos.

Sintaxis:

búsqueda-tesaúrica=
"RTHS" "(" expresión-restricciones "," diccionario ","
lengua-resultado "," variable-concepto-resultado ","
lista-conceptos ")".

Ejemplos:

1. T.- **RTHS** ((and (?X HYPERONYME |instrument I 1|)
(?X OBJECTIF |mesurer I 1|)),
LPPL, francés, ?X, ?LC)
Nombres en francés de **instrumentos** utilizados para **medir**.
S.- LC= (|baromètre I 1| |dynamomètre I 1| |télémètre I 1|...)
Barómetro, dinamómetro, telémetro, ...
2. T.- **RTHS** ((and (?X HYPERONYME |consumer I 1|)
(?X AGENT |feu I 1|)),
LPPL, euskara, ?X, ?LC)
S.- LC= (|izeki I 1|, |kiskali I 1|)
quemar, calcinar.

e) BUSQUEDA DE RELACION ENTRE CONCEPTOS

Especificación:

Datos: $C_1, C_2 \in \text{concepto}$; $T \in \text{diccionario}$

Resultados: $R \in \text{expresión-relación}$, R denota un camino entre C_1 y C_2 .

Sintaxis:

búsqueda-relación=

"DRAP" "(" concepto-thesaurus "," concepto-thesaurus "," diccionario
"," expresión-relación)".

Ejemplos:

1. T.- **DRAP** (|guêpe I 1|, |miel I 1|, LPPL, ?R)
¿Cuál es la relación entre **guêpe** y **miel**?
S.- R= ((and (and (|guêpe I 1| HYPERONYME |insecte I 1|)
(|insecte I 1| HYPONYME |abeille I 1|)
(|abeille I 1| PREDICATION |produire I 1#n|)
(|produire I 1#n| THEME |miel I 1|))
(and (|abeille I 1| CARACTERISTIQUE |hyménoptère I 1|)
(|guêpe I 1| CARACTERISTIQUE|hyménoptère I 1|))))
*La miel es **producida** por la **abeja**, que es un **insecto**
himenóptero como la **avispa***

f) PETICION DE DIFERENCIAS ENTRE CONCEPTOS

Especificación:

Datos: $C_1, C_2 \in \text{concepto}$; $T \in \text{diccionario}$

Resultados: $D \in \text{expresión-diferencia}$

Precondición: $C_1 D C_2$ (C_1 y C_2 pertenecen a la misma taxonomía)

Sintaxis:

petición-diferencia=

"DDIF" "(" concepto-thesaurus "," concepto-thesaurus "," diccionario
"," expresión-diferencia)".

Ejemplos:

1. T.- **DDIF** (|laver I 1|, |balayer I 1|, LPPL, ?DIF)
Diferencia entre **lavar** y **barrer**.
S.- DIF= ((and (and (|laver I 1| INSTRUMENTAL |liquide I 1|)
(|balayer I 1| INSTRUMENTAL |balai I 1|))))
Se **lava** con un **líquido**, pero se **barre** con un **escoba**

g) VERIFICACION DE DEFINICION DE UN CONCEPTO

Datos: $C \in \text{concepto}$; $D \in \text{expresión-definición}$; $T \in \text{diccionario}$; $L \in \text{lengua}$

Resultados: $true$ sii **def**(C) @ D (además devuelve **def**(C))

Precondición: C y D en la misma lengua

Sintaxis:

```
verificación-definición-relacional=
"VDEF" "(" concepto-thesaurus "," expresión-definición ","
diccionario "," lengua-resultado ","
resultado-verificación "," definición ")".
```

Ejemplos:

```
1. T.- VDEF ( |cantonner I 1|,
              (and (|cantonner I 1| HYPERONYME |établir I 1|)
                    (|cantonner I 1| THEME |soldat I 1|)
                    (|cantonner I 1| LOCATIF |ville I 1|)),
              LPPL, francés)
      ¿acantonar significa establecer soldados en una
      población?
S.- True,def(|cantonner I 1|) =
      (and (|cantonner I 1| HYPERONYME|installer I 1|)
            (|cantonner I 1| HYPERONYME |placer I 1|)
            (|cantonner I 1| HYPERONYME |établir I 1|)
            (|cantonner I 1| THEME |troupe I 3|)
            (|cantonner I 1| LOCATIF |localité I 1|))
      Sí, acantonar es instalar (situar, establecer) tropas
      en una localidad.
```

h) VERIFICACION DE PROPIEDADES DE UN CONCEPTO

Datos: $C \in \text{concepto}$; $T \in \text{diccionario}$; $L \in \text{lista-pares-atributo-valor}$

Resultados: *true* sii todos los pares atributo-valor de la lista L se dan en el concepto C.

Sintaxis:

```
verificación-propiedades=
"VPRO" "(" concepto-thesaurus "," diccionario ","
lista-pares-atributo-valor "," resultado-verificación ")".
```

Ejemplos:

```
1. T.- VPRO (|or I 1|, LPPL, ((|couleur I 1| |rouge I 2|)))
      ¿Es rojo el color del oro ?
S.- False
2. T.- VPRO (|souris I 1|, LPPL, ((|classe I 1| |rongeur I 2|)))
      ¿Pertenece ratón a la clase de los roedores?
S.- True
```