

Utilización de un corpus paralelo trilingüe (inglés-español-italiano) para el análisis de la terminología especializada en productos sanitarios para diagnóstico *in vitro*

Cristina M. CASTILLO RODRÍGUEZ
Universidad de Málaga

Como citar este artículo:

CASTILLO RODRÍGUEZ, Cristina M. (2008) «Utilización de un corpus paralelo trilingüe (inglés-español-italiano) para el análisis de la terminología especializada en productos sanitarios para diagnóstico *in vitro*», en PEGENAUTE, L.; DECESARIS, J.; TRICÁS, M. y BERNAL, E. [eds.] *Actas del III Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación. La traducción del futuro: mediación lingüística y cultural en el siglo XXI. Barcelona 22-24 de marzo de 2007*. Barcelona: PPU. Vol. n.º 1, pp. 419-429. ISBN 978-84-477-1026-3. Versión electrónica disponible en la web de la AIETI:

<http://www.aieti.eu/pubs/actas/III/AIETI_3_CMCR_Utilizacion.pdf>.

Utilización de un corpus paralelo trilingüe (inglés-español-italiano) para el análisis de la terminología especializada en productos sanitarios para diagnóstico *in vitro*

Cristina M. Castillo Rodríguez
Universidad de Málaga

1. Introducción

Actualmente en Internet, podemos encontrarnos con una gran variedad de herramientas útiles para el análisis de determinados aspectos lingüísticos de textos escritos en distintas lenguas. Con la ayuda de los programas de gestión de corpus, podemos analizar de forma rápida la frecuencia de aparición de una determinada palabra en un contexto dado, tanto en el texto original (TO) como en los textos meta (TM). En el presente trabajo, hemos utilizado uno de estos programas de gestión de corpus (WordSmith Tools v. 3.0) con el fin de analizar algunos aspectos terminológicos de un corpus trilingüe (inglés-español-italiano) que hemos extraído de Internet y que está especializado en un campo específico de la farmacología, concretamente textos que describen productos sanitarios para diagnóstico *in vitro*.

2. Descripción del programa de gestión de corpus

Con el fin de realizar el análisis de ciertos aspectos terminológicos así como de las siglas y las unidades de medida, hemos extraído de Internet un corpus de textos especializados en farmacología, en particular un corpus de productos sanitarios para diagnóstico *in vitro*. Este corpus está formado por un subcorpus de diez textos escritos originalmente en lengua inglesa, un subcorpus de diez textos traducidos al español y, por último, un subcorpus de otros diez textos traducidos al italiano.

Para analizar los tres subcorpus, hemos utilizado el programa de gestión de corpus WordSmith Tools v. 3.0 ya que se trata de un conjunto de programas integrados que analiza cómo se comportan las palabras en los textos. Además, contiene una serie de herramientas básicas con las que se puede comprobar el uso de las palabras en textos propios: a) la herramienta WordList, que permite visualizar una lista de todas las palabras o grupo de palabras (*word-clusters*) dentro de un texto, así como ordenarlas por orden alfabético o de frecuencias; b) la herramienta de concordancias, Concord, que permite alinear a izquierda y a derecha, lo que hace posible observar cualquier palabra o unidad fraseológica en su contexto así como la clase de relación que mantiene con el resto de los términos; c) y la herramienta KeyWords¹, con la que pueden encontrarse las palabras claves dentro de un texto determinado.

3. Análisis y explotación del corpus trilingüe especializado

En primer lugar, el formato en .pdf y .doc de los textos del corpus requiere su conversión a formato .txt —texto plano— para el análisis mediante el programa *WordSmith Tools*. Seguidamente, utilizamos la herramienta WordList para extraer listas de palabras por orden alfabético y por orden de frecuencias a partir de diez textos seleccionados en el subcorpus inglés, en el español y en el subcorpus italiano.

¹ La herramienta Keyword no se ha utilizado en esta investigación ya que a los resultados que llegaría esta herramienta los hemos conseguido con WordList restringiendo la inclusión de los términos generales, números y ciertas palabras gramaticales que no son objeto de estudio en este trabajo.

Además, con esta herramienta podemos observar que el análisis estadístico de los textos en inglés muestra un total de 44.636 ocurrencias (*tokens*) a partir de 3.076 palabras-tipo (*types*)²; esto supone una riqueza léxica del 6,89% (*type/token ratio*). Por otro lado, si utilizamos la misma herramienta para el subcorpus español, el análisis estadístico nos da como resultado 52.548 ocurrencias a partir de 3.714 palabras-tipo, con una riqueza léxica del 7,07%. Por último, el subcorpus italiano nos arroja 48.444 ocurrencias, 3.780 palabras tipo y una riqueza léxica del 7,8%.

A continuación se muestran las capturas de pantalla correspondientes a los análisis estadísticos arrojados por el programa en los tres subcorpus (inglés, español e italiano):

N	1	2	3	4	5	6	7
Text File	OVERALL	ANAELI~1.TXT	ANTISS~1.TXT	BORREL~1.TXT	EBNA-1~1.TXT	EBV-EA~1.TXT	EBV-VC~1.TXT
Bytes	274.895	24.914	18.786	29.509	26.929	28.448	30.806
Tokens	44.636	3.999	3.080	4.750	4.383	4.638	5.111
Types	3.076	953	846	1.215	1.071	1.093	1.155
Type/Token Ratio	6,89	23,83	27,47	25,58	24,44	23,57	22,60

Cuadro 1. Análisis estadístico de los textos del subcorpus inglés

N	1	2	3	4	5	6	7
Text File	OVERALL	TPOIGG~1.TXT	MYCOPL~1.TXT	MPOIGG~1.TXT	HPYLOR~1.TXT	EBV-VC~1.TXT	EBV-EA~1.TXT
Bytes	321.394	32.444	33.859	31.284	36.008	36.006	33.579
Tokens	52.548	5.495	5.548	5.194	5.715	5.935	5.472
Types	3.714	1.244	1.455	1.168	1.496	1.350	1.282
Type/Token Ratio	7,07	22,64	26,23	22,49	26,18	22,75	23,43

Cuadro 2. Análisis estadístico de los textos del subcorpus español.

N	1	2	3	4	5	6	7
Text File	OVERALL	ANAELI~1.TXT	ANTISS~1.TXT	BORREL~1.TXT	EBNA-1~1.TXT	EBV-EA~1.TXT	EBV-VC~1.TXT
Bytes	310.665	28.734	21.704	33.907	29.944	33.431	32.870
Tokens	48.444	4.445	3.378	5.258	4.607	5.193	5.026
Types	3.780	1.143	1.007	1.394	1.255	1.272	1.303
Type/Token Ratio	7,80	25,71	29,81	26,51	27,24	24,49	25,93

Cuadro 3. Análisis estadístico de los textos del subcorpus italiano.

Para llevar a cabo un análisis más exhaustivo de las palabras más frecuentes en los tres subcorpus, hemos restringido de la lista de palabras por orden de frecuencias la inclusión de ciertas palabras gramaticales y otras de uso muy general —como las preposiciones, artículos y conjunciones, entre otros—, así como las cifras, puesto que no son objeto de este trabajo de investigación.

Además, la herramienta WordList permite extraer las palabras más frecuentes establecidas en secuencias candidatas a constituir unidades fraseológicas o clusters —denominación utilizada en el programa—. Una secuencia candidata a constituir una unidad fraseológica (*cluster*) puede estar formada por dos o más palabras. En nuestro caso, hemos seleccionado que la herramienta extraiga *clusters* que contengan dos componentes y también hemos restringido la inclusión de palabras que no son susceptibles de ser analizadas con la herramienta Concord. Nos referimos a grupos de palabras como *water and*, *tests are*, *they were*, etc., puesto que el análisis de estos

² *Tokens* es el término que designa el número de palabras que la herramienta WordList encuentra durante el procesamiento del texto, mientras que *type* se refiere al número de lexemas que presenta el texto. Por ejemplo, bajo el lexema *swim* se englobarían todas las derivaciones morfológicas de dicha palabra-tipo: *swim*, *swimming*, *swam*, *swims*.

grupos de palabras no es objeto de estudio de este trabajo.

En el caso del subcorpus español y del italiano, hemos tenido que crear dos listas de candidatos a unidades fraseológicas (UF) o *clusters*, una formada por secuencias constituidas por dos palabras, y la otra lista formada por secuencias de tres. Esto se debe a que en inglés la mayoría de las unidades fraseológicas de dos componentes, analizados con la herramienta de Concord, son traducidos al español y al italiano de dos formas: algunas unidades fraseológicas inglesas serán traducidos al español y al italiano por unidades fraseológicas integradas por dos componentes, mientras que otras UF serán traducidas por sintagmas formados por tres componentes. Un ejemplo ilustrativo de traducción de ambos tipos de unidades fraseológicas son: *Mean Absorbance* por *Absorbancia Media* y *Assorbanza Media* (UF de dos palabras en subcorpus inglés = UF de dos palabras en subcorpus español e italiano); *Sample Diluent* por *Diluyente de Muestra* y *Diluyente per Campione* (UF de dos palabras en subcorpus inglés = UF de tres palabras en subcorpus español e italiano).

Con respecto a la herramienta Concord del programa WordSmith Tools, hay que destacar que se trata de una herramienta muy útil, puesto que permite al usuario tener una idea en cuanto al uso de una palabra en el contexto de varios ejemplos, de forma que se pueda comprender mejor el significado incluso en el caso de los neologismos. Además, un diccionario puede mostrar los significados de una palabra, sin embargo, no muestra cómo se utiliza dicha palabra en profundidad.

El primer paso consiste en seleccionar los textos del corpus; en nuestro caso, hemos seleccionado todos los textos originales escritos en inglés, es decir, diez textos planos del subcorpus inglés. El segundo paso consiste en especificar una palabra de búsqueda en los textos que queremos analizar. En este caso queremos analizar las concordancias de algunas de las palabras más frecuentes extraídas de las listas de palabras que hemos guardado previamente con la herramienta WordList, tanto del subcorpus inglés como del español y del italiano. Analizaremos y observaremos cómo se comportan dichas palabras en su contexto original, cómo han sido traducidas y cómo se comportan en los TM, así como la coincidencia o no en el número de palabras en los tres subcorpus.

En primer lugar, seleccionamos la lista de palabras correspondiente al subcorpus inglés; la primera palabra que vamos a analizar es el término *well*. Para este término encontramos concordancias para 138 entradas —resultado de búsqueda de los diez textos seleccionados—. Estas 138 entradas pueden visualizarse en una ventana que muestra todos los posibles usos de esta palabra en el contexto de diez textos científicos escritos originalmente en inglés. Entre estos posibles usos se incluyen tanto las palabras, como los adverbios, locuciones, etc. No obstante, hay que añadir que con el mismo término en su forma plural, esto es *wells*, se evita esta ambigüedad, puesto que no existe en el subcorpus inglés otra palabra gramatical igual.

Si hacemos lo propio con *pocillo* —traducción de *well* al español como sustantivo— en el subcorpus de textos en español, observamos que el número de entradas no coincide con el resultado obtenido anteriormente. Esta vez la herramienta Concord nos da concordancias para 102 entradas de la palabra *pocillo*. Esto se debe a lo mencionado en el párrafo anterior: la traducción de *well* como sustantivo corresponde a *pocillo* en español, por lo que se excluyen del cómputo de concordancias en español las traducciones de *well* como adverbio y locución. Lo mismo ocurre con el subcorpus italiano, donde podemos encontrar 102 entradas para la palabra *pozzetto* (traducción de *well* como sustantivo).

También hay que tener en cuenta que, mediante el empleo del signo *, este programa permite la introducción de ecuaciones de búsqueda que obvian las desinencias de género

y número, así como de tiempos verbales. De esta forma, si empleamos la ecuación de búsqueda «positiv*» en la lista de palabras del subcorpus inglés, sólo dará resultados para la palabra *positive*, con concordancias para un total de 220 entradas. De forma inconsciente, el traductor buscaría *positivo* en el listado de palabras del subcorpus español. Sin embargo, las concordancias para este adjetivo dan como resultado un total de 92 entradas, lo cual no coincide con el número de entradas anteriormente obtenido. En cambio, al introducir la misma ecuación de búsqueda en la lista de palabras del subcorpus español, es decir, «positiv*», nos da como resultado el número de concordancias obtenido anteriormente para *positive*, puesto que se incluyen los adjetivos *positiva* —20 entradas—, *positivos* —29 entradas—, *positivas* —78 entradas—, así como *positive*, en el subcorpus español —1 entrada—, este último extraído de las referencias bibliográficas. Si hacemos lo propio en el subcorpus italiano, observamos los siguientes registros: 98 entradas para *positivo*, 4 entradas para *positiva*, 111 entradas para *positivi* y 7 entradas para *positive* (en los que va incluido *positive* en inglés, extraído igualmente de las referencias bibliográficas del subcorpus). En total los tres subcorpus dan como resultado 220 entradas para *positiv**.

Por otro lado, en nuestro corpus hemos encontrado que en inglés puede haber dos términos para designar uno sólo en español. Es el caso de *suero*, cuyas entradas ascienden a un total de 352, mientras que en inglés se utilizan dos términos *serum* y *sera*, forma plural del anterior. Esto se debe a que en nuestro subcorpus español no se utiliza la forma plural de *suero*, sino que se tiende a parafrasear como, por ejemplo: *clinically characterized sera*, ha sido traducido por *muestras de suero clínicamente caracterizadas*. Esta es la razón por la que el número de entradas correspondientes a *suero* es mayor al de *sera* y *serum*. En el caso del subcorpus en italiano no se produce este problema ya que sí se utiliza la forma plural, es decir, encontramos entradas tanto para *siero* como para *sieri*. Sin embargo, hay que destacar que en el subcorpus italiano también se tiende a parafrasear. Utilizando el ejemplo anterior, *clinically characterized sera*, ha sido traducido por *sieri campioni clinicamente noti*.

Respecto al ejemplo que acabamos de analizar, viene a colación especificar que debido a las distintas paráfrasis que se producen en el subcorpus español, el número total de concordancias para *muestras* o *muestra* supera con creces a sus equivalentes en inglés, es decir, *samples* y *sample* —382 entradas para ambos términos en inglés frente a las 625 entradas en español y a las 520 entradas en italiano—.

Si seleccionamos, además, la opción *cluster*³ de la barra de herramientas, Concord nos muestra también los patrones fraseológicos que se repiten en las concordancias realizadas. De esta forma, podemos observar el número de frecuencias de cada patrón fraseológico; en nuestro caso, queremos ver cuántas veces aparece *muestras de suero* en el subcorpus español. Por otro lado, hay que tener en cuenta que las unidades fraseológicas *blood specimen* y *blood specimens* dan lugar a las traducciones de las unidades fraseológicas en español *muestra de sangre* y *muestras de sangre* y en italiano *campione di sangue* y *campioni di sangue*. Por eso, el número de entradas para *muestras* y *muestra* así como para *campioni* y *campione* es mucho mayor que el número de entradas recogidas para los términos *sample* y *samples*. Por último, hemos encontrado en nuestro corpus un uso más de la palabra *muestra* que, además de un sustantivo, también es la tercera persona del singular en el presente del indicativo.

En el listado de palabras del subcorpus inglés, también nos encontramos con el caso de *off*. Por sí solo *off* carece de sentido, por lo que tendríamos que analizarlo en su

³ Véase más adelante el análisis realizado a los *clusters*, obtenidos de las listas de palabras mediante la herramienta WordList.

contexto. Concord nos muestra que *off* siempre va acompañado de *cut*. Se trata, por tanto, de una palabra compuesta, *cut-off*, cuya traducción en los textos españoles aparece como discriminatorio, mientras que en italiano optan por mantenerlo en inglés. Por otro lado, si buscamos con la opción cluster de la barra de herramientas, podemos encontrar el contexto exacto en que se utiliza este adjetivo en el subcorpus inglés. Esto es, el adjetivo *cut-off* acompaña la mayoría de las veces a palabras como *value*, *Standard*, *Calibrator* y *Control*.

Como hemos anticipado previamente, la traducción de *cut-off* como adjetivo en el subcorpus español es discriminatorio. Al analizar este adjetivo con la opción *cluster*, observamos que también va acompañado de palabras tales como *valor*, *estándar*, *calibrador* o *control*. Si analizamos *cut-off* en el contexto del subcorpus en italiano nos damos cuenta de que también acompaña a palabras como *standard*, *valore*, *controllo* y *calibratore*.

Otros ejemplos de unidades fraseológicas podemos encontrarlas en los siguientes términos de ambos subcorpus, como *human sera* – *suero humano*, *siero umano*; *comparative method* – *método comparativo*, *metodo comparativo*; *positive control* – *control positivo*, *controllo positivo*; *negative control* – *control negativo*, *controllo negativo*; *room temperature* – *temperatura ambiente*, *temperatura ambiente*; *sodium azide* – *ázida sódica*, *sodio azide*; *equivocal results* – *resultados dudosos*, *risultati dubbi*, entre otros.

En lo que concierne a las unidades fraseológicas formadas por dos componentes y traducidas por UF de tres componentes tanto en subcorpus español como italiano, resaltamos como referencia palabras como *sample diluent* – *diluyente de muestra*, *diluyente per campione*; *wash solution* – *solución de lavado*, *soluzione di lavaggio*; *performance characteristics* – *características de rendimiento*, *caratteristiche di rendimento*; *correlation coefficient* – *coeficiente de correlación*, *coefficiente di correlazione*; *bubble formation* – *formación de burbujas*, *formazione di bolle*; *absorbance values* – *valores de absorbancia* – *valori di assorbanza*, entre otros.⁴

Queremos aclarar que ciertas unidades fraseológicas formadas por dos componentes en el subcorpus inglés a veces se traducen por unidades fraseológicas de cuatro componentes en el subcorpus español y en el subcorpus italiano. Un ejemplo de este tipo de unidades fraseológicas es el caso de *horseradish peroxidase*, cuya traducción al español en nuestro subcorpus hemos encontrado que equivale a *peroxidasa de rábano picante*, así como *wash concentrate* traducido en el subcorpus italiano por *soluzione di lavaggio concentrata*.

Mediante la herramienta Concord, se pueden identificar claramente las convenciones textuales tanto para el texto original como para los traducidos. Con esto, nos referimos a los títulos de los apartados ya establecidos en los documentos técnicos de productos sanitarios para diagnóstico *in vitro*, tales como *Summary of procedure* – *Resumen del procedimiento* – *Schema della procedura*, *Intended use* – *Objetivo de su utilización* – *Scopo del test*, *Summary and explanation* – *Resumen y explicación* – *Sommario e descrizione*, *Principle of the procedure* – *Principio del procedimiento* – *Principio del test*, *Reagents* – *Reactivos* – *Reagenti*, *Precautions* – *Precauciones* – *Avvertenze e precauzioni*, *Specimen collection* – *Recogida de muestras* – *Raccolta dei campioni*, *Procedure* – *Procedimientos* – *Procedimenti*, *Quality control* – *Control de calidad* – *Controllo di qualità*, *Results* – *Resultados* – *Risultati*, *Limitations* – *Limitaciones* – *Limitazioni*, *Expected values* – *Valores predictivos* – *Valori attesi*, *Performance*

⁴ Por cuestiones obvias de espacio no hemos podido incluir todas las unidades fraseológicas analizadas en ambos subcorpus. Por ello, hemos ofrecido una breve muestra de traducciones de tales términos.

characteristics – Características de rendimiento – Caratteristiche di rendimento, References – Bibliografía – Bibliografia.

4. Análisis y traducción de siglas y unidades de medida en los subcorpus

En el corpus de productos sanitarios para diagnóstico *in vitro*, además de los términos especializados de un determinado dominio conceptual, abundan también las unidades de medida diversas (*ml*, *mg*, *nm*, etc.), así como las siglas referidas a enfermedades, síndromes, antígenos, anticuerpos, entre otros.

En primer lugar, previo al análisis de tales unidades de medida y siglas, debemos cribar de nuevo las listas de palabras de los tres subcorpus con la herramienta WordList. Esta vez debemos eliminar todas aquellas palabras de uso general y gramaticales, así como todos los términos especializados. Lo único que debemos mantener son las siglas y las unidades de medida.

En lo que concierne a la traducción de las unidades de medida, hay que destacar que existen ciertas unidades que no cambian de un idioma a otro. Estas unidades son las siguientes: *mm* (milímetro), *mg* (miligramo), *ml* (mililitro), *dl* (decilitro), *nm* (nanómetro), *KD* (unidad de medida de proteínas) y *ppm* (unidad de medida de principio activo)⁵. No obstante, el corpus arroja ciertas unidades de medida que varían en la traducción al español. Las unidades a las que nos referimos son: *IU* (inglés, *International Units*) – *UI* (español, *Unidades Internacionales*); *EU* (inglés, *Elisa Units*) – *UE* (español, *Unidades Elisa*); y *EV* (inglés, *Elisa Values*) – (español, *Valores Elisa*). En el subcorpus italiano hemos observado que estas unidades de medida se quedan en su forma original inglesa.

Por otro lado, en el corpus encontramos otras unidades de medida referidas al tiempo, como, por ejemplo, la abreviatura para segundos y minutos tanto en inglés como en español y en italiano: en inglés, estas abreviaturas se representan mediante *sec* y *min*; mientras que en español *min* queda igual, los segundos se representan con *seg* y en italiano mediante *min* y *sec*.

En cuanto a las siglas y abreviaturas contenidas en el corpus, nos encontramos, en primer lugar, el caso de las inmunoglobulinas, que se presentan tanto en el subcorpus original como en los dos subcorpus traducidos de la siguiente manera: *Immunoglobulin A, M, G – IgA, IgM, IgG*; *Inmunoglobulina A, M, G – IgA, IgM, IgG*.

Debido a la diversidad de siglas y abreviaturas las hemos tenido que agrupar en diferentes grupos: a) antígenos, virus, anticuerpos; b) enfermedades y síndromes; c) tests y pruebas.

Entre las siglas que representan los antígenos, virus y anticuerpos en los tres subcorpus, encontramos las siguientes: *ANA* y *ANAs*, en inglés corresponden a *antinuclear antibody* y *antinuclear antibodies*, mientras que en el subcorpus español y el italiano sólo aparece la sigla *ANA* para *anticuerpo antinuclear* y *anticorpo antinucleare* así como para su forma plural *anticuerpos antinucleares* y *anticorpi antinucleari*. En varios de los textos del corpus aparecen los anticuerpos *ANCA* con sus dos variantes, a saber, *cANCA* y *pANCA*. *ANCA* son las siglas de *anti-neutrophil cytoplasmic antibodies*, cuya traducción al español y al italiano, según los datos que arroja el corpus, es *anticuerpos anticitoplasma de neutrófilo* y *anticorpi anticitoplasma dei neutrofili*, manteniéndose así la sigla original en inglés. Las variantes de *ANCA* que hemos mencionado son *cANCA*, *cytoplasmic ANCA*, y *pANCA*, *perinuclear ANCA*.

⁵ En este trabajo de investigación representaremos aquellas unidades que se encuentren en el corpus, aunque existen más unidades de medida que no varían de una lengua a otra.

Otros casos en los que se mantiene la sigla de los antígenos, virus y anticuerpos en inglés, aunque varía su significado en español y en italiano son: *EA* (*early antigen*, en inglés, *antígeno temprano*, en español y *l'antigene early*, en italiano), *EBNA-1* (*Epstein-Barr Nuclear antigen-1*, en inglés, *antígenos-1 nuclear del virus del Epstein-Barr*, en español y *antigeni-1 nucleari del virus del Epstein-Barr*), *EBV* (*Epstein-Barr virus*, en inglés, y *virus del Epstein-Barr*, en español y en italiano), *HM-CAP antigen* (*high molecular cell-associated proteins*, en inglés, *proteínas de elevado peso molecular asociadas a la célula*, en español y *proteine di elevato peso molecolare associate alla cellula*), *TPO* (*thyroid peroxidase*, en inglés, *peroxidasa tiroidea*, en español y *perossidasi tiroidea*, en italiano), *VCA* (*viral capsid antigen*, en inglés, *antígeno de la cápside viral*, en español y *antigene capsidico virale*, en italiano). No obstante, hemos encontrado un solo caso de virus en el que la sigla en español cambia con respecto a la sigla inglesa; se trata del virus del sida *HIV* (*human immunodeficiency virus*), que en español es *VIH* (*virus de la inmunodeficiencia humana*), aunque en el subcorpus italiano no varía.

En el bloque que hemos denominado «enfermedades y síndromes», obtenemos del corpus resultados muy diversos: siglas que coinciden con su forma original inglesa, siglas que se han traducido al español y al italiano, siglas que se han traducido al español pero no al italiano y siglas en inglés sin equivalentes en el subcorpus español y en italiano. En el primer caso se encuentra *EM*, que en inglés es *erythema migrans* y en español e italiano es *eritema migrans*. En cuanto al segundo caso, hemos comentado que en el corpus existen siglas que se han traducido al español y al italiano, a saber, *SLE* (*systemic lupus erythematosus*) – *LES* (*lupus eritematoso sistémico* y *lupus eritematoso sistémico*). Por otro lado, hay siglas que sólo se han traducido en español, mientras que en italiano se ha optado por dejar su forma inglesa original aunque con su definición entre paréntesis y traducida al italiano como es el caso de *MP* (*microscopic polyangiitis*) – *PM* (*poliangeitis microscópica*) – *MP* (*poliangoite microscópica*). En el cuarto caso, encontramos dos siglas en los que sólo se ha traducido el nombre de la enfermedad sin la sigla: 1) la sigla *MCTD*, que en inglés es *mixed connective tissue disease*, no se ha traducido en el subcorpus español y en el italiano, puesto que únicamente se ha dejado el nombre de la enfermedad, es decir, *enfermedad mixta del tejido conjuntivo* y *malattia mista del tessuto connettivo*; 2) y la sigla *PSS* (*progressive systemic sclerosis*), sólo se ha traducido por *esclerosis sistémica progresiva* y *sclerosi sistémica progressiva*.

En cuanto a las siglas de tests y pruebas del corpus destacamos los siguientes ejemplos: *EIA* para las tres lenguas (*enzyme immunoassay*, en inglés, *enzimoinmunoensayo*, en español y *test immunoenzimatico*, en italiano); el test *ELISA* también quedaría igual en los subcorpus traducidos (*enzyme-linked immunosorbent assay*, en inglés, *ensayo inmunoabsorbente ligado a enzimas*, en español y *enzyme-linked immunosorbent assay*, en italiano, en el que opta, además, por poner incluso la definición en inglés); *ID* (*immunodiffusion*, en inglés, *inmunodifusión*, en español y *immunodiffusione*, en italiano); y la sigla del test *IF* (*immunofluorescent*, en inglés, *inmunofluorescencia*, en español y *immunofluorescenza*, en italiano). En cambio, para *IFA*, en tanto test denominado *indirect immunofluorescence* se traduce por la sigla española *IFI*, *inmunofluorescencia indirecta*, mientras que en italiano se opta por dejarla en inglés, aunque se traduce por *immunofluorescenza indiretta*. Sin embargo, *IFA* en tanto anticuerpo llamado *indirect fluorescent antibody* quedaría igual en español y en italiano, es decir, *IFA* para los *anticuerpos fluorescentes indirectos* y para *anticorpi fluorescenti indiretti*.

5. Conclusiones

La investigación basada en corpus ha supuesto el nacimiento de nuevos métodos de análisis para el traductor especializado, ya que constituye una importante fuente de recursos a la hora de solventar problemas particulares en el proceso traductológico. Como hemos podido observar, el uso de un programa de gestión de corpus se trata de una herramienta muy útil porque nos muestra cómo todas las palabras gramaticales así como los términos especializados se comportan en su contexto. Por tanto, la compilación de este corpus paralelo trilingüe (inglés-español-italiano) especializado en un campo específico de la farmacología permite al traductor:

1. La extracción de datos léxicos y terminológicos para crear una base de datos especializada trilingüe.
2. La elaboración de patrones de traducción de aspectos sintácticos, fraseológicos y discursivos para este tipo de textos.
3. La compilación de un glosario trilingüe de abreviaciones.
4. El uso de este corpus trilingüe tanto en la traducción profesional como en la investigación en la traducción especializada.

Referencias bibliográficas

- Aijmer, K.; Stenström, A. B. (eds.) (2004). *Discourse Patterns in Spoken and Written Corpora*. Ámsterdam: John Benjamins.
- Barnbrook, G., P. Danielsson y M. Mahlberg (eds.) (2005). *Meaningful Texts. The Extraction of Semantic Information from Monolingual and Multilingual Corpora*. Londres: Continuum.
- Bowker, L. (2002). *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.
- Bowker, L. y J. Pearson (2002). *Working with Specialized Language: A practical guide to using corpora*. Londres: Routledge.
- Corpas Pastor, G. (2004a). «La traducción de textos médicos especializados a través de recursos electrónicos y corpus virtuales». En L. González y P. Hernández (eds.). *Las palabras del traductor. Actas del II Congreso Internacional El español, lengua de traducción. 20 y 21 de mayo, Toledo 2004*. Bruselas: Comisión Europea/ESLETRA. 137-164. También disponible en <http://www.toledo.net/index.php>.
- Corpas Pastor, G. (2004b). «Localización de recursos y compilación de corpus vía Internet: Aplicaciones para la didáctica de la traducción médica especializada». En C. Gonzalo García y V. García Yebra (eds.). *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/Libros. 223-257.
- Corpas Pastor, G. (2002). «Traducir con corpus: de la teoría a la práctica». En J. García Palacios y M. T. Fuentes Morán (eds.). *Texto, terminología y traducción*. Salamanca: Almar. 189-226.
- Corpas Pastor, G. (2001). «Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada». *TRANS. Revista de traductología* 5. 155-184.
- Gonzalo García, G. y V. García Yebra (eds.) (2004). *Manual de documentación y terminología para la traducción especializada*. Madrid: Arco/Libros.
- Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. Londres: Addison Wesley Longman.
- Kenny, D. (2005). «Parallel corpora and translation studies: old questions, new perspectives? Reporting that in Gepcolt: a case study». En G. Barnbrook, P. Danielsson y M. Mahlberg (eds.). *Meaningful Texts. The Extraction of Semantic Information from Monolingual and Multilingual Corpora*. Londres: Continuum. 154-165.

- Olohan, M. (2004). *Introducing Corpora in Translation Studies*. Londres: Routledge.
- Pérez Hernández, C. (2002). «Terminografía basada en corpus: principios teóricos y metodológicos». En P. Faber y C. Jiménez (eds.). *Investigar en terminología*. Granada: Comares. 127-166.
- Sinclair, J. (2004). *Trust the Text. Language, Corpus and Discourse*. Londres: Routledge.
- Véronis, J. (2000). *Parallel Text Processing. Alignment and Use of Translation Corpora*. Dordrecht: Kluwer Academic Publishers.
- Zanettin, F. (2002). «Corpora in Translation Practice». Ponencia presentada en el *First International Workshop on Language Resources (LR) for Translation Work and Research*. Las Palmas de Gran Canaria (28 de mayo de 2002). Disponible en <http://www.ifi.unizh.ch/cl/yuste/postworkshop/download.html>.
- Zanettin, F. (2000). «Parallel Corpora in Translation Studies. Issues in Corpus Design and Analysis». En M. Olohan (ed.). *Intercultural Faultlines. Research Models in Translations Studies I. Textual and Cognitive Aspects*. Manchester: St. Jerome. 105-118.