

ALONSO, A.; CABRÉ, T.; DE YZAGUIRRE, L.; TEBÉ, C. (2002) "La utilización de corpus paralelos alineados en la docencia de la traducción y de los lenguajes de especialidad ". En: Iglesias, L.; Doval, S. (ed.) *Studies in Contrastive Linguistics. Proceedings of the Second International Contrastive Linguistics Conference*. Santiago de Compostela: Publicacions de la Universidade de Santiago de Compostela, p. 71-82. ISBN: 84-9750-027-X. (CL).

LA UTILIZACIÓN DE CORPUS PARALELOS ALINEADOS EN LA DOCENCIA DE LA TRADUCCIÓN Y DE LOS LENGUAJES DE ESPECIALIDAD

Araceli Alonso, Mireia Martínez, Noemí Polo, Gal·la Puertas, Lluís de Yzaguirre,
Carles Tebé, Teresa Cabré

Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra

1. Introducción y objetivos

El alineador elaborado en el Institut Universitari de Lingüística Aplicada (IULA) de la Universitat Pompeu Fabra consiste en una herramienta que permite alinear las frases de un documento con su traducción.

Aunque su desarrollo no ha finalizado, el alineador da, en la actualidad, buenos resultados al nivel de frase y unos resultados bastante satisfactorios al nivel de alineación léxica. En estos momentos, se ha desarrollado una técnica de retroalimentación del alineador a partir de diccionarios extraídos automáticamente de los textos alineados. Esta técnica permite obtener un material muy valioso que podría usarse para el análisis contrastivo del texto de partida (lengua A) con el texto de llegada (lengua B), o bien, dos textos de llegada en dos lenguas distintas, facilitando la localización de las principales dificultades del texto, que podrían servir tanto para la formación de traductores como para la enseñanza de los lenguajes de especialidad.

La aportación de los autores de esta comunicación es:

- Presentar el estado actual de desarrollo del alineador del Institut Universitari de Lingüística Aplicada.
- Exponer cómo la técnica de retroalimentación del alineador a partir de diccionarios extraídos automáticamente de los corpus alineados permite obtener un material muy valioso para el análisis contrastivo de textos alineados.

- Mostrar cómo este material puede utilizarse en la docencia de la traducción y de los lenguajes de especialidad.

2. El alineador del IULA

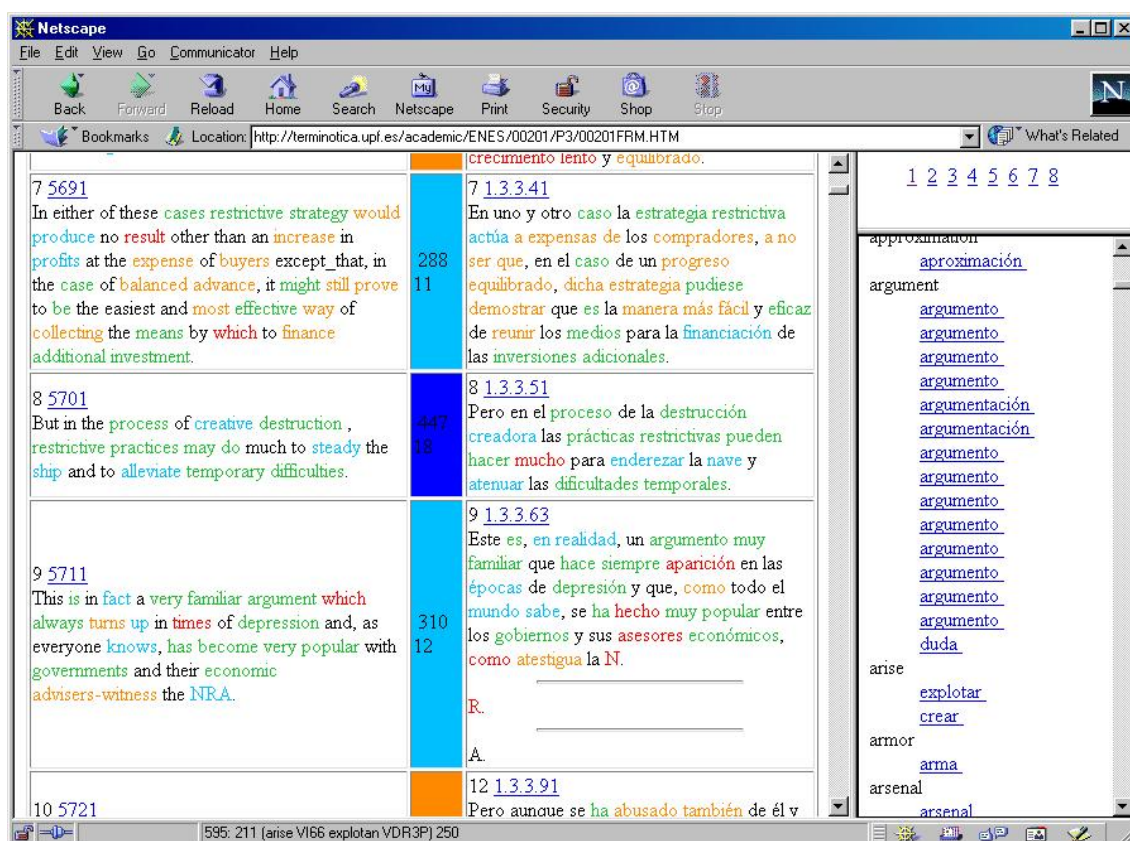
Concebimos la alineación automática de traducciones como la aplicación de un sistema capaz de alinear dos textos, uno de los cuales es el original y el otro, su traducción a otra lengua, con el objetivo de vincular cada frase del texto original con la frase correspondiente de la traducción, basándose en el grado de similitud entre ellas.

Siguiendo a Baker (1995), consideramos un corpus *paralelo* como un conjunto de textos “*originally written in a language A alongside their translations into a language B*”. Hablamos de *corpus alineado* cuando éstos disponen de vínculos explícitos entre cada frase del texto origen y cada frase del texto traducción (alineación oracional), o bien entre cada palabra del texto origen y cada palabra del texto traducción (alineación léxica).

El alineador que se ha desarrollado en el Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra permite, tanto alinear frases, como también llevar a cabo una alineación léxica. A diferencia de otras herramientas de alineación, basadas en criterios estadísticos, el alineador del IULA toma las decisiones basándose en la utilización de información lingüística añadida (lemas y etiquetas), es decir, es un programa dependiente de herramientas de marcaje lingüístico para las lenguas que se quieran alinear (De Yzaguirre *et al.* 2000a; Morel, J. *et al.* 1998; Vivaldi *et al.* 1996). En este caso, funciona a partir de textos procesados con la cadena de herramientas del Corpus Técnico del IULA (Bach *et al.* 1997) en inglés, catalán y castellano, ya que son las lenguas de las que se dispone actualmente de marcaje lingüístico. Esto implica que el alineador no compara únicamente la forma de la lengua A con la forma de la lengua B, sino también, sus lemas y etiquetas morfológicas. Así pues, para establecer un par léxico, se da prioridad primero a la coincidencia de lemas, después a la coincidencia de etiquetas y finalmente, a la coincidencia ortográfica.

El programa trabaja en tres niveles: a) nivel de palabra, b) nivel de frase, y c) nivel de documento. En el primer nivel, se mide el grado de similitud de dos palabras; en el segundo nivel, se globalizan los resultados del nivel de palabra para cada pareja de frases; finalmente, en el tercer nivel, se establece una estrategia para decidir qué frase de la lengua A es comparada con qué frase de la lengua B, siguiendo los modelos clásicos de cálculo de longitud de frase, de número de palabras y de posición en el documento.

El resultado es una versión hipertextual de la alineación de dos textos, tal y como se puede observar en la siguiente pantalla¹:



¹ El programa también permite obtener un fichero en formato SGML, que especifica la vinculación entre los párrafos de una lengua a otra mediante los identificadores de los documentos alineados. Para más información sobre el funcionamiento del alineador véase De Yzaguirre (2000b), De Yzaguirre *et al.* (2000c, 2001).

3. La técnica de retroalimentación del alineador

En el curso de desarrollo del alineador se ha demostrado que el programa mejoraba notablemente su eficiencia si era capaz de aprender de sus alineaciones, materializándose el aprendizaje en un diccionario de pares léxicos prealineados generado a partir de la alineación léxica.

En principio, el programa al alinear dos textos genera un diccionario automáticamente. Este diccionario se crea a partir de la indexación y acumulación de todas las alineaciones léxicas hechas por el ordenador. Una vez realimentado el alineador con el diccionario y alineados los textos, hemos podido observar que el programa cometía entre el 10 y el 40% de errores de alineación léxica y, por tanto, la calidad del diccionario era bastante insatisfactoria. Si mejorábamos la calidad del diccionario, la alineación, como consecuencia, también mejoraba².

En este momento es cuando se hacía necesaria la intervención de un humano, y tratándose de textos paralelos, la intervención de un traductor era muy beneficiosa para comprobar si los pares léxicos alineados eran realmente equivalentes o, por el contrario, el programa había realizado una alineación léxica errónea. Sin embargo, este procedimiento resultaba muy costoso y largo, por lo que diseñamos una técnica para ordenar los pares léxicos según la probabilidad de que fueran erróneos con el conjunto de informaciones generadas por el alineador. A esta técnica la denominamos Cálculo de la Probabilidad Biléxica Decreciente (PBD)³. De esta forma, podíamos optimizar la intervención del traductor que revisara el diccionario.

Una vez alineados los textos y generado el diccionario, entramos en una base de datos los datos necesarios ordenados por la PBD para que el traductor pudiera proceder al análisis de los pares léxicos y de los textos alineados, y así poder hacer una validación tanto de la alineación léxica como del funcionamiento general del programa.

² Para observar un ejemplo de esta mejora con textos alineados en inglés-castellano, inglés-catalán y catalán-castellano del ámbito de la economía, véase la página <http://terminotica.upf.es/academic/>.

³ El cálculo de la Probabilidad Biléxica Decreciente (PBD) es un mecanismo que sirve para medir la probabilidad de que un par léxico establecido por el alineador sea correcto.

Este material almacenado en la base de datos, junto con la alineación de los textos resultó ser útil no sólo para la validación del funcionamiento del programa sino también como material para usar en la docencia de la traducción y de los lenguajes de especialidad, ya que permitían al alumno detectar problemas y errores de traducción frente a errores propios del programa, observar las diferencias estructurales y fraseológicas entre la lengua de partida y la lengua de llegada, etc. A continuación, haremos una descripción de cómo se haría servir este material para la docencia de la traducción.

4. Aplicación del alineador en la docencia

El uso de corpus alineados y la técnica que anteriormente citamos para mejorar el sistema de alineación desarrollado en el IULA constituye un material muy valioso para la docencia de la traducción especializada. El alumno deberá, a partir de la alineación léxica, hacer un análisis de la traducción, detectar los errores y, por último, hacer una propuesta de mejora de la misma.

Para ilustrar este material y explicar los pasos que el alumno ha de seguir, presentaremos datos y ejemplos extraídos de un texto en inglés de economía del libro *Capitalism, Socialism and Democracy* de J. A. Schumpeter, junto con su texto paralelo en castellano⁴.

En primer lugar, se le pide al alumno que, tras haber aplicado el alineador a dos textos paralelos de un área temática específica, haberse generado el diccionario automáticamente y haber aplicado por segunda vez el alineador, observe la versión hipertextual resultado de la alineación⁵.

A continuación, la tarea del alumno consistirá en revisar los datos obtenidos al generar el diccionario. Estos datos, tal como mencionamos anteriormente, están entrados en una base de datos. Dicha base de datos contiene, por un lado, los pares léxicos alineados

⁴ El mismo corpus alineado se ha analizado en inglés-catalán y en castellano-catalán. Los resultados se pueden ver en la página <http://terminotica.upf.es/academic/>.

⁵ En caso de que los textos sean muy extensos, se podrá dividir el material entre varios alumnos para que hagan el análisis, bien individualmente o por grupos.

ordenados por la PBD⁶, junto con los lemas y etiquetas gramaticales correspondientes, y por otro, una serie de valores estadísticos relacionados con la alineación de cada uno de los pares. Además, la base está formada por otros cuatro campos en blanco que son los que tendrá que completar el alumno. Estos campos son los siguientes: el campo OK, que servirá para especificar si un par léxico es aceptable o no; el campo ID, que servirá para indicar si se incluye la equivalencia de la palabra en un diccionario o no; el campo AC, para indicar el tipo de palabra (si la pareja es aceptable) o el tipo de error (si no lo es); y un campo de notas donde el alumno especificará las observaciones que crea pertinentes.

OK	AC	ID	DICC	LEMA1	ETIQ1	LEMA2	ETIQ2	X1	X2	X3	X4	X5	X6	X7	X8	X9	NOTES
			00199	not	D6	no	D4	06493	00570	00085	00008	00606	00732	00104	00924	00627	
			00199	capitalist	NN6S	capitalista	N56P	01823	00152	00095	00008	00160	00158	00108	00909	00638	
			00199	socialist	NN6S	socialista	JQ6S	08986	00100	00093	00008	00106	00107	00136	00959	00638	
			00199	economic	JA	económico	JQFS	03213	00100	00087	00008	00109	00120	00125	00944	00648	
			00199	social	JA	social	JQ6S	08975	00063	00089	00008	00071	00069	00132	00943	00654	
			00199	case	NN6P	caso	N5MP	01860	00117	00085	00008	00139	00134	00106	00942	00655	
			00199	problem	NN6S	problema	N5MS	07523	00064	00080	00008	00071	00090	00129	00938	00655	
			00199	product	NN6S	producto	N5MS	07582	00040	00089	00008	00042	00048	00129	00949	00655	
			00199	rational	JA	racional	JQ6P	07941	00052	00088	00007	00060	00057	00144	00980	00658	
			00199	important	JA	importante	JQ6S	04802	00044	00086	00008	00051	00051	00127	00932	00660	
			00199	political	JA	político	JQFS	07243	00061	00078	00008	00073	00083	00139	00984	00660	
			00199	capitalism	NN6S	capitalismo	N5MS	01817	00036	00090	00008	00040	00040	00120	00892	00661	
			00199	condition	NN6P	condición	N5FP	02275	00058	00091	00008	00063	00064	00108	00918	00661	
			00199	modern	JA	moderno	JQMP	06178	00033	00091	00007	00038	00034	00136	00931	00662	
			00199	of_course	D6	por supuesto	D4	06622	00057	00086	00008	00089	00063	00123	00913	00663	

Para proceder a la revisión de los datos, el alumno deberá, en primer lugar, comprobar si el par léxico es aceptable. Si lo es, lo indicará en el campo OK de la base de datos mediante una “S” y, si no lo es, mediante una “N”. Por ejemplo:

OK	AC	ID	LEMA1	ETIQ1	LEMA2	ETIQ2	X9
N	O		have	V6R6	partir	V8R6S	00941
N	O		be	H6	burgués	JQMP	00941
N	O		show	V166	haber	VDR3S	00940
S	C		save	V166	ahorrar	V1	00940
N	O		profit	NN6S	precio	N5MP	00940
N	O		have	V6R3S	ver	V1	00940
S	C		effective	JA	efectivo	JQFP	00940
N	O		admit	V6R6	haber	V8R6S	00940
N	O		labor	NN6S	teoría	N5FS	00939
N	O		country	NG6S	política	N5FS	00939
N	O		be	V6R3S	comprender	VDR3P	00939
N	O		be	V166	demostrar	V1	00939
N	O		short-run	JA	malo	JQMP	00939
N	O		will	NN6S	hecho	N5MS	00939
S	C		thing	NN6S	cosa	N5FP	00939
S	C		understand	V6R3S	comprender	V8R6S	00939
N	O		carry	V6A6	ser	VDP3S	00939
N	O		factor	NN6S	renta	N5FS	00938

⁶ El coeficiente de PBD indicará al profesor cuáles son los pares léxicos que deberá revisar el alumno o

En segundo lugar, tomando como corpus de referencia un diccionario⁷, el alumno habrá de constatar si el par de palabras alineadas aparecen como equivalentes en el diccionario. Esta información la incluirá en el campo ID mediante “S” o “N” dependiendo de si se incluye el par léxico en el diccionario o no.

OK	AC	ID	LEMA1	ETIQ1	LEMA2	ETIQ2	X9	NOTES
S	T	N	means	NN66	medio	N5MP	00896	Componente del SN "medios de producción".
S	T	N	industrial	JA	industria	N5FS	00909	Componente del SN "gerencias de industria".
S	T	N	rigidity	NN6S	rigidez	N5FS	00913	Componente del SN "rigidez de los precios".
S	T	N	producer	NG6P	producción	N5FS	00925	Componente de "bienes de producción".
S	T	S	capitalist	NN6S	capitalista	N56P	00638	
S	T	S	socialist	NN6S	socialista	JQ6S	00638	
S	T	S	capitalism	NN6S	capitalismo	N5MS	00661	
S	T	S	socialism	NN6S	socialismo	N5MS	00683	
S	T	S	democracy	NN6S	democracia	N5FS	00698	

El diccionario también lo hará servir para cubrir el campo AC (cuando el par es aceptable). Cuando el par es aceptable, el alumno deberá decidir qué pares léxicos son comunes (AC="C"), cuáles son técnicos (AC="T") y cuáles enciclopédicos (AC="E"). Se considerarán palabras técnicas aquellas que tengan un valor especializado dentro de la disciplina sobre la que tratan los textos y, enciclopédicas, los nombres de persona, de lugares geográficos, obras bibliográficas, etc. El resto de palabras se clasificarán como comunes.

Esta operación tiene un valor didáctico por sí mismo. Uno de los aspectos que creemos que creará problemas al alumno es la distinción entre unidades léxicas (palabras comunes) o unidades léxicas especializadas (palabras técnicas o términos), debido en parte a la frontera difusa existente entre lengua general y lengua especializada.

Tal como cita Cabré (1999:90) *“intuimos que las unidades del lenguaje provistas de referencia e incluidas en la gramática del hablante no son, de entrada y en abstracto, ni palabras ni términos, sino unidades del léxico de la gramática que, en virtud de las características de la situación comunicativa dejan que se seleccione solamente un determinado haz de rasgos del conjunto que las describe. Siguiendo este supuesto, toda unidad dotada de referencia puede ser candidata a término y a palabra. Parece cierto que algunas unidades muy especializadas sólo se realizan como términos, dado que*

alumnos para poder validar la traducción.

⁷ El corpus utilizado como corpus de referencia dependerá del tipo de textos alineados, su temática y su grado de especialización.

siempre se utilizan para denominar el conocimiento especializado, pero ello no impide suponer que algún día puedan banalizarse a través de la difusión del conocimiento.

Esta frontera no es fácil de establecer, pero el futuro traductor debe conocerla por varias razones: en primer lugar, porque como traductor puede permitirse una cierta libertad cuando tratan con palabras de la lengua general, pero no así con unidades léxicas especializadas; en segundo lugar y desde un punto de vista práctico, el alumno ganará tiempo y esfuerzo en la medida en que intuya acertadamente a donde puede recurrir para encontrar un equivalente de una palabra de la lengua de partida en la lengua de llegada; en tercer lugar, cuando el alumno no encuentra un equivalente de una unidad léxica especializada en una fuente convencional, tendrá que averiguar qué recursos existen para poder solucionar el problema terminológico .

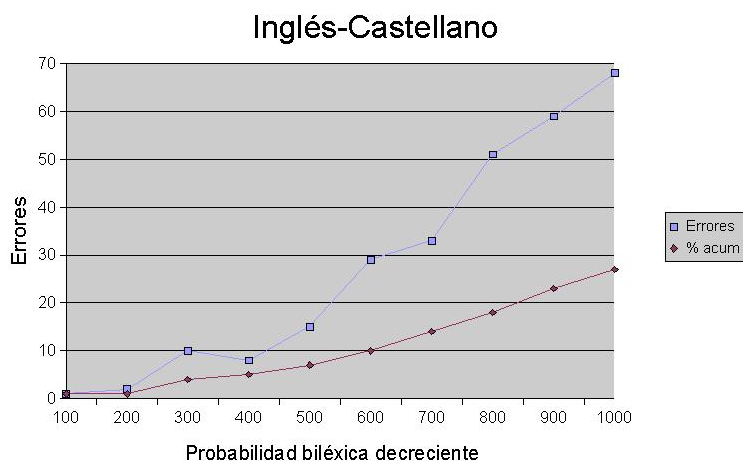
OK	AC	ID	LEMA1	ETIQ1	LEMA2	ETIQ2	X9	NOTES
S	C		interpretation	NN6S	interpretación	N5FS	00798	
S	C		dilemma	NN6S	dilema	N5FS	00802	
S	C		conversion	NN6S	conversión	N5FS	00802	
S	C		association	NN6P	asociación	N5FP	00802	
S	E	N	Sismondi	X	sismondi	?	00802	
S	E	N	von	NN6S	von	?	00801	Componente de "Von Schmolter".
S	C		preference	NN6S	preferencia	N5FP	00801	
S	C		aversion	NN6S	aversión	N5FS	00801	
S	C		institution	NN6P	institución	N5FP	00800	
S	C		constitution	NN6S	constitución	?	00800	
S	E	N	barone	NN6S	barone	?	00800	
S	E	N	paribu	X	paribus	?	00798	
S	T	N	volume	NN6S	volumen	N5MS	00792	Componente del SN "volumen de producción".
S	C		irrational	JA	irrational	JQ6S	00798	
S	C		argument	NN6P	argumento	N5MP	00785	

Seguidamente, se le pedirá al alumno que cubra el campo AC cuando los pares léxicos no son aceptables. Es decir, si el par léxico no es correcto, el alumno está frente a un tipo de error. El alumno habrá de indicar si se trata de un error debido a un defecto del programa (AC="O"), de un error del traductor (AC="t") o un error polilexema⁸ (AC="M").

⁸ Nos referimos a un error polilexema cuando un concepto que se representa en la lengua de partida con un término monolexemático, en la lengua de llegada se representa con un término polilexemático o viceversa. Esto sucede debido a la distancia existente entre las lenguas de trabajo. El alineador no detecta que se trata de dos equivalentes y sólo alinea un componente del término.

OK	AC	ID	LEMA1	ETIQ1	LEMA2	ETIQ2	X9
N	O		may	V666	muy	D4	00828
N	M		apriori	X	priori	?	00824
N	O		be	V6R6	tener	VDR1P	00821
N	t		fatal	JA	fatal	JQ6S	00819
N	O		much	DC	más	D4	00818
N	O		practical	D6	prácticament	D6	00863
N	O		be	V6R6	poder	VDR3P	00817
N	O		precise	D6	precisamente	D6	00839
N	O		see	V166	ser	VJA6S	00813
N	t		compromise	NN6S	compromiso	N5MS	00803
N	O		indirect	D6	indirectament	D6	00800
N	O		concern	NN6P	concern	N5MP	00797
N	t		realization	NN6S	realización	N5FS	00779
N	M		namely	D6	saber	V1	00772
N	t		cost	NN6S	costo	N5MS	00666
N	O		be	V166	haber	VDR3S	00818
N	O		be	V6R3S	hacer	VDR3P	00855

Esta información permite calcular el porcentaje de errores cometidos por el programa, independientemente del tipo de error, según el número de pares léxicos alineados y ordenados por el coeficiente de PBD, tal y como se indica en la siguiente gráfica:



La gráfica muestra, en este caso, que el alineador ha cometido un 30% de errores en las primeras 1.000 palabras. En las 600 primeras palabras hay un 10% de errores, mientras que en las primeras 300 palabras no llega a un 5%.

Por último, se le pedirá al alumno que analice todos los datos obtenidos para poder hacer una validación final de la traducción y una propuesta de mejora de la misma, justificando todas sus decisiones, de manera que se pueda entrar en una discusión sobre el tema con el resto de la clase.

5. Conclusiones y perspectivas

Después de presentar el estado de desarrollo del alineador, mostrar cómo la técnica de retroalimentación del alineador permite generar unos materiales útiles para el análisis contrastivo de corpus alineados y exponer un procedimiento para la docencia de la traducción especializada a partir de materiales extraídos del alineador y de la retroalimentación del alineador, podemos llegar a las siguientes conclusiones:

Por un lado, el uso de corpus alineados y el estudio de alineaciones léxicas constituye un material muy valioso para la docencia de la traducción, especialmente:

Desde la perspectiva del profesor de traducción, para aproximarse a las especificidades propias del sublenguaje o lenguaje especializado que se esté estudiando, localizar dificultades de traducción entre dos lenguas y poder preparar, de esta manera, un material de base real para sus clases.

Desde la perspectiva del alumno, para sensibilizarse de manera más profunda sobre diferencias léxicas y estructurales entre lenguas distintas en un mismo ámbito de especialidad. Además, un análisis de alineaciones léxicas permite al alumno detectar los problemas de traducción, los errores de traducción frente a los errores del propio programa, así como observar las diferencias estructurales y fraseológicas entre dos lenguas, reflexionar sobre la frontera entre lengua general-lengua especializada, unidad léxica-unidad léxica especializada y ver el grado de especialización de un texto. Asimismo, le permite constatar la información de un diccionario con la información obtenida a partir de textos reales y poder observar, por tanto, las deficiencias de los mismos. Esto le permite además familiarizarse con los recursos usados como fuentes de referencia. Por otro lado, la detección de los diferentes tipos de errores le aporta información al alumno útil para la mejora general del programa.

Por otro lado, el tipo de análisis realizado por el alumno constituye un material provechoso para la validación de traducciones y del propio funcionamiento del programa. Esta validación del programa podría contribuir a la mejora del sistema de

alineación, pues, de hecho, el desarrollo del programa no es definitivo, ya que muchas posibilidades que nos aporta el sistema están todavía por explorar.

Asimismo, el análisis de los datos extraídos de la alineación léxica podría contribuir en la generación automática de diccionarios superespecializados. Actualmente, existen diccionarios especializados de economía, derecho, medicina, física, etc., pero hay pocos que se centren en un aspecto concreto de cualquiera de estas disciplinas y, mucho menos, de las interdisciplinas. Ya sea por motivos económicos, falta de interés o dificultad de la tarea, lo cierto es que aún hay un largo camino por recorrer para mejorar la calidad de las obras lexicográficas. Sin embargo, la alineación de corpus bilingües puede contribuir a la generación de diccionarios bilingües realmente focalizados en un ámbito de especialidad.

Como perspectivas de futuro podemos señalar las siguientes:

Comprobar el funcionamiento del alineador y de la técnica de retroalimentación del mismo para la generación de materiales con textos de un volumen mayor. Utilizaremos para ello un corpus inglés-castellano de medio millón de palabras del campo de la sanidad de la PAHO (Pan-American Health Organization).

Experimentar la propuesta de docencia presentada en una clase.

Profundizar en los datos extraídos de la alineación léxica para producir más materiales útiles para la docencia de la traducción, como por ejemplo, el análisis de la polisemia léxica en los textos especializados.

Hacer un estudio de la posibilidad de generar diccionarios focalizados en un tema automáticamente a partir de corpus alineados y del análisis de la alineación léxica del corpus.

6. Otras informaciones

Para obtener más información sobre el funcionamiento del alineador, así como ejemplos de textos alineados consúltense las páginas siguientes:

<http://www.iula.upf.es/faq/cache/149.html>

<http://terminotica.upf.es/CREL/alinea.HTM>

http://traductica.upf.es/alinea/v_alcala/FRM1.HTM

<http://terminotica.upf.es/academic/>

BIBLIOGRAFÍA

Bach, C.; Saurí, R.; Vivaldi, J.; Cabré, M. T (1997) *El Corpus de l'IULA: descripció*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, (Papers de l'IULA. Sèrie Informes, 17).

Baker, M. (1995) "Corpora in Translation Studies: An Overview and Some Suggestions for the Future Research". *Target* 7(2), pp. 223-43.

Brown, P.F.; Lai, J.C.; Mercer, R.L. (1991) "Aligning Sentences in Parallel Corpora". *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*. Morriston, NJ.: University of California.

Cabré, M. T. (1999) "Elementos para una teoría de la terminología: hacia un paradigma alternativo". En: *La Terminología: representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra. pp. 69-92.

Chang, J.S.; Chen, M.H. (1997) "An Alignment Method for Noisy Parallel Corpora based on Image Processing Techniques". En: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference- of the European Chapter of the Association for Computational Linguistics*. San Francisco, Madrid: UNED.

De Yzaguirre, Ll. (2000a) "[L'etiquetador PALIC i el desambiguador AMBILIC](#)", comunicació presentada en las Jornades del Centre de Referència en Enginyeria Lingüística (CREL). IEC. Barcelona, 4-5 de abril del 2000.

De Yzaguirre, Ll. (2000b) "[Validador de traduccions, una eina de paral·lelització](#)", comunicació presentada en las Jornades del Centre de Referència en Enginyeria Lingüística (CREL). IEC. Barcelona, 4-5 de abril del 2000.

De Yzaguirre, Ll.; Ribas, M.; Vivaldi, J.; M.T. Cabré (2000c) "[Some Technical Aspects About Aligning Near Languages](#)", comunicació presentada en la LREC-2000, Atenas, 31 mayo-2 junio del 2000. En: Gabrilidou, M. *et al.* (Ed.) *Second International Conference on Language Resources and Evaluation. Proceedings*, vol I. Atenas: National Technical University of Athens Press. pp. 545-548.

De Yzaguirre, Ll.; Ribas, M.; Vivaldi, J.; Cabré, M. T. (2001) "[Alineación automática de traducciones: descripción y usos en los ámbitos de la profesión, de la docencia y de la investigación traductológica](#)", comunicació presentada en los IV Encuentros Alcalaínos de Traducción, Alcalá de Henares, 17-18 de febrero del 2000. En: Valero, C. y De la Cruz, C. (Ed.) *Traducción y nuevas tecnologías. Herramientas auxiliares del traductor. Encuentro en torno a la traducción 4*. Alcalá de Henares: Universidad de Alcalá. pp. 391-398.

De Yzaguirre, Ll.; Matamala, A.; Cabré, M.T. (en prensa) "[El lematizador PALIC del IULA \(UPF\)](#)", comunicació presentada en el XVIII Congreso de AESLA. Barcelona, 4-6 de mayo del 2000.

De Yzaguirre, Ll.; Matamala, A.; Bach, C.; Castillo, N.; Ustrell, E. (en prensa) "[AMBILIC, el desambiguador lingüístico del Corpus del IULA](#)", comunicació presentada en el XVIII Congreso de AESLA. Barcelona, 4-6 de mayo del 2000.

Gale, W.A.; Church, K.W. (1991) "A Program for Aligning Sentences in Bilingual Corpora". *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*. Morriston, NJ.: University of California.

Johansson, S.; Oksefjell, S. (eds.) (1998) *Corpora and Cros-Linguistic Research. Theory, Method and Case Studies*. Amsterdam y Atlanta (GA): Rodopi.

Kenny, D. (1998) "Corpora in translation studies". En: Baker, M. (Ed.) et al. *Routledge Encyclopedia of Translation Studies*. London: Routledge.

Melamed, D. (1997) "A Portable Algorithm for Mapping Bitext Correspondence". En: *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*. San Francisco, Madrid: UNED.

Morel, J.; Torner, S.; Vivaldi, J.; De Yzaguirre, Ll.; Cabré, M.T. (1998) *El corpus de l'IULA: Etiquetaris*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. (Papers de l'IULA. Sèrie Informes, 18). [Segunda edició, revisada y corregida].

Vivaldi, J.; De Yzaguirre, Ll.; Solé, X.; Cabré, M.T. (1996). *Marcatge estructural i morfosintàctic del corpus tècnic amb l'estàndard SGML*. Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra. (Papers de l'IULA. Sèrie Informes, 1).