

# ¿Cómo traducen traductores y profesores de idiomas? Estudio de corpus

PATRICIA RODRÍGUEZ-INÉS

*Universitat Autònoma de Barcelona, Bellaterra, España*

[patricia.rodriguez@uab.es](mailto:patricia.rodriguez@uab.es)

## RÉSUMÉ

Le groupe PACTE de l'Université autonome de Barcelone (UAB) s'est consacré pendant plus d'une décennie à la recherche en matière de compétence traductrice (CT). Quant à la démarche de traduction, notre recherche – empirique et expérimentale – a non seulement confirmé le modèle de CT proposé par PACTE, mais elle a également mis de l'avant des données sur la spécificité de cette compétence et sur les différences entre la traduction directe et la traduction inverse. Dans cette étude sur le produit de traduction, les données sont observées sous un autre angle, à savoir celui de la linguistique de corpus. Le présent article analyse le corpus de traductions issu de l'expérience de PACTE en matière de CT, qui se compose de traductions directes et inverses réalisées d'une part par des traducteurs professionnels et, d'autre part, par des professeurs de langues étrangères. Il s'agit en l'occurrence d'une analyse de l'utilisation d'éléments tels que les calques, les emprunts, les notes des traducteurs, les parenthèses, le rapport type/occurrence, les expressions courantes, etc., cette analyse n'ayant pas révélé de grandes variations entre les deux groupes de traducteurs ni en matière de directionnalité. Par contre, le degré de similitude entre les traductions, qui a également été mesuré, a montré de nettes différences.

## ABSTRACT

PACTE, a research group from the Universitat Autònoma de Barcelona, has been carrying out empirical and experimental research on translation competence (TC) for over a decade. In relation to the translation process, our research has both validated our TC model and yielded data on the specificity of the competence and on the differences between direct and inverse translation. We have now studied the translation product from the perspective of corpus linguistics. A corpus comprising direct and inverse translations performed by professional translators and foreign-language teachers as part of a PACTE experiment on TC forms the basis for the analysis presented here, which revolves around the use of certain features. Our findings show that there are no major differences between the two experimental groups and in terms of directionality where the use of many of the features in question (including calques, loanwords, translator's notes, brackets, type/token ratio and sentence length) is concerned. In contrast, however, measuring the degree of similarity among translations has revealed clear differences.

## MOTS-CLÉS/ KEYWORDS

corpus de textes traduits, compétence traductrice, traducteur professionnel, professeur de langue étrangère, directionnalité  
translated text corpus, translation competence, professional translator, foreign language teacher, directionality

## 1. Introducción

Tanto el proceso como el producto de la traducción son objetos de estudio de muchos grupos y proyectos de investigación (Cordiall, CroCo, CTP, TransComp, etc.). El grupo PACTE también se ha interesado por ambos elementos, si bien es cierto que ha dedicado más tiempo y esfuerzo a estudiar el proceso, dada su complejidad (PACTE 2008; 2009; 2011a; 2011b). Una vez procesados los datos de todas las variables del experimento sobre competencia traductora (CT) llevado a cabo por el grupo, se ha podido volver la mirada al producto y, más concretamente, se ha podido comprobar si analizar el producto proporcionaría más información o información diferente de la que habíamos obtenido al analizar el proceso. De esta forma, la pregunta inicial que PACTE se hacía para estudiar el proceso, es aún aplicable para el estudio del producto, es decir, ¿qué hay de especial en la manera en que los traductores traducen? Y si nos centramos en el universo experimental del experimento de PACTE, ¿traducen de manera diferente los traductores profesionales y los profesores de lenguas extranjeras? Y si fuera así, ¿cómo se puede observar esta hipótesis? ¿Qué rastros dejarían esas diferencias entre grupos en la superficie de los textos traducidos? El objetivo de este artículo es estudiar estas diferencias en el corpus de PACTE a través del análisis de elementos como calcos, préstamos, notas del traductor, paréntesis, ratio palabra/forma (o ratio tipo/caso, por type/token ratio) y frecuencias de palabras o expresiones, además de observar la similitud entre traducciones.

Tras una breve introducción sobre la investigación de PACTE en CT y sobre la lingüística de corpus y su presencia en los estudios del proceso traductor, se intentará dar respuesta a todas las preguntas planteadas previamente y a algunas más. Se seguirá cada paso que normalmente se contempla en un estudio de corpus para finalmente ofrecer una posible interpretación de los resultados obtenidos.

## 2. Marco teórico

La competencia traductora (CT) ha sido estudiada por muchos y varias son las propuestas de modelos de CT (Wilss 1976; Bell 1991; Neubert 2000; Kelly 2005; etc.), si bien es cierto que sólo las de Alves y Gonçalves (2007) y de PACTE se han validado empíricamente. A continuación se presenta brevemente la investigación de PACTE con el fin de enmarcar el presente estudio de corpus.

### 2.1. La investigación de PACTE sobre competencia traductora

El grupo PACTE (Proceso de Adquisición de la Competencia Traductora y Evaluación) lleva a cabo una investigación de tipo holístico, empírico y experimental sobre la CT y su adquisición en traducción escrita. Se interesa tanto por el proceso como por el producto en traducción directa e inversa y trabaja con 6 pares de lenguas: inglés, francés y alemán combinados con español y catalán. El grupo ha terminado la investigación sobre la CT y ha validado su modelo componencial de CT (PACTE 2003). Dado que la trayectoria del grupo es larga y la envergadura de su investigación grande, aquí proporcionaremos sólo la información necesaria para comprender el marco en el que se ha llevado a cabo el estudio de corpus que se describe. La página web del grupo PACTE<sup>1</sup> proporciona mucha más información, publicaciones y materiales para aquellos que estén interesados en conocer su investigación en más detalles.

Para investigar la CT, PACTE llevó a cabo pruebas piloto y exploratorias (PACTE 2002; 2005a; 2005b) para, finalmente, realizar un experimento en el que se compararon dos grupos que pertenecen al mismo universo experimental, el de profesionales de las lenguas: 35 traductores profesionales y 24 profesores de lenguas extranjeras sin experiencia en traducción (PACTE 2008; 2009; 2011a; 2011b).

### *2.1.1. Definición de competencia traductora*

En el grupo PACTE definimos la competencia traductora como el sistema subyacente de conocimientos necesarios para traducir. Asimismo, creemos que la CT: (a) es un conocimiento experto; (b) es, sobre todo, un conocimiento procedural, es decir, no declarativo; (c) está compuesto por varias sub-competencias que están interrelacionadas; y (d) incluye un componente estratégico que es de especial importancia. En nuestro modelo (PACTE 2003), la CT se compone de cinco subcompetencias (bilingüe, extralingüística, conocimientos sobre la traducción, instrumental y estratégica), así como de componentes psicofisiológicos.

Dado que todos los bilingües conocen dos lenguas y también pueden tener conocimientos extralingüísticos, consideramos que las subcompetencias propias de la CT son la estratégica, la instrumental y los conocimientos sobre la traducción. Nuestra investigación se centra, por lo tanto, en estas tres subcompetencias.

### *2.1.2. Diseño de la investigación*

Lo que siguen son los elementos principales del diseño experimental de PACTE, que se explican brevemente con el fin de contextualizar el estudio de corpus que aquí se presenta.

#### *2.1.2.1. Hipótesis*

Nuestra hipótesis general es que el grado de experticidad en traducción (es decir, la competencia traductora) se refleja tanto en el proceso como en el producto de la traducción. Dado que es de esperar que traductores con experiencia tengan un alto grado de experticidad en traducción, la definición de experticidad en nuestro estudio se basa en: (a) años de experiencia como traductor; (b) la traducción como principal fuente de ingresos; y (c) experiencia en la traducción de una amplia gama de textos.

#### *2.1.2.2. Sujetos*

El universo del que se tomó la muestra de sujetos es el de profesionales que trabajan con lenguas extranjeras. De este universo, y después de que los sujetos llenaran un cuestionario para confirmar que cumplían con los criterios establecidos, se seleccionaron dos grupos experimentales para el experimento, traductores profesionales (35) y profesores de idiomas (24). Para evitar la presencia de variables extrañas, se seleccionó a traductores que no estuvieran especializados en ningún campo en particular y que hubieran trabajado como traductores profesionales por un periodo de tiempo equivalente. Todos los profesores de lenguas extranjeras tenían un mínimo de 5 años de experiencia docente en la Escuela Oficial de Idiomas, institución pública de enseñanza de idiomas modernos para adultos. Todos los sujetos debían ser hablantes nativos de español o catalán y trabajar como profesionales con el inglés, el francés o el alemán como lenguas extranjeras. Los traductores que participaron en el estudio tenían una media de 7,51 años de experiencia como traductores; el porcentaje medio

de sus ingresos provenientes de la traducción era de un 86,43 %, y habían traducido una gran variedad de textos hacia su lengua materna. A los sujetos se les pagó por sus traducciones a precio de mercado, como si fuera un trabajo de traducción real.

#### *2.1.2.3. Tareas experimentales e instrumentos*

Los sujetos llevaron a cabo las siguientes tareas: (1) traducción directa; (2) cumplimentar un cuestionario sobre los problemas encontrados en la traducción; (3) traducción inversa; (4) cumplimentar un cuestionario sobre los problemas encontrados en la traducción; (5) cumplimentar un cuestionario sobre conocimientos de la traducción; (6) participar en una entrevista retrospectiva.

Cada una de estas tareas aportó datos para el análisis, sobre todo del proceso; las traducciones directa e inversa proporcionaron los textos que se analizan en este artículo. Con el fin de obtener más datos o completar los obtenidos con las tareas anteriores también se grabaron las acciones de los sujetos en pantalla mientras traducían en tiempo real por medio del programa Camtasia, además de hacer una observación directa de los sujetos.

Los instrumentos utilizados en el experimento fueron variados, desde textos originales, que ya se habían probado en ocasiones anteriores para observar qué problemas planteaban, hasta cuestionarios diversos, grabaciones de lo que hace el sujeto en la pantalla del ordenador y plantillas de observación directa. En PACTE (2009) se puede encontrar una definición detallada de todas las variables y sus indicadores, así como de los instrumentos.

#### *2.1.2.4. Variables*

Para nuestro estudio de la CT se estableció una variable independiente, la experticidad en traducción, y se seleccionaron seis variables dependientes para estudiar, que son: conocimientos sobre la traducción, eficacia del proceso de traducción, toma de decisiones, proyecto traductor, identificación y resolución de problemas de traducción y, finalmente se añadió el uso de recursos instrumentales. Para operacionalizar estas variables se seleccionaron 18 indicadores y, entre éstos, el más importante es el de la aceptabilidad de las traducciones ofrecidas por los sujetos, dado que refleja la calidad de las mismas. La *aceptabilidad* es un indicador transversal que se cruza con indicadores de todas las variables del estudio para determinar la relación entre los resultados obtenidos en dichos indicadores y la calidad de las traducciones de los sujetos. La aceptabilidad, según el estudio de PACTE (2009) hace referencia a la calidad del producto. Se define en términos de si una solución transmite de forma efectiva el sentido del texto original, la función de la traducción y utiliza la lengua de forma apropiada. Cada solución de traducción es clasificada como aceptable, no aceptable o semi-aceptable por dos o tres miembros de PACTE. Se realizan las posibles combinaciones de estas categorías y se asignan valores numéricos para dar como resultado un solo valor numérico indicativo de la calidad de una solución de traducción.

### *2.2. Lingüística de corpus*

Hasta la fecha, muchas áreas relacionadas con el estudio de la lengua se han beneficiado del uso de corpus (McEnery y Wilson 1996), es decir, de un enfoque empírico basado en el análisis de grandes cantidades de textos, lo cual ha dotado a los lingüis-

tas de cimientos más firmes para poder hacer afirmaciones sobre la lengua que no estén basadas en su intuición o en ejemplos prefabricados.

Aunque todas las ramas de los estudios sobre la traducción se han beneficiado de adoptar la metodología de corpus (Olohan 2004), las ramas descriptiva y aplicada son las que más lo han desarrollado hasta ahora. En la rama aplicada, algunas de las áreas que más han aprovechado el uso de corpus son el procesamiento del lenguaje natural (Márquez y Padró 1997; Civit 2003; Koehn 2005; Wilks 2008), la lexicografía (Sinclair 1985; Teubert 1996; Hanks 2008; Royle, Richardson *et al.* 2009) y, hasta cierto punto, la enseñanza de lenguas (Johns 1991; Partington 1998; Aston 2000; Bowker y Pearson 2002) y la enseñanza de la traducción (Rodríguez-Inés 2008), aunque también hay aspectos muy concretos, como el de la detección de plagios, que ha hecho un buen uso de los corpus (Barrón-Cedeño, Potthast *et al.* 2010; Muhr, Kern *et al.* 2010<sup>2</sup>). La última es precisamente una aplicación que se ha utilizado en el estudio que se describe aquí que, si bien no buscaba detectar plagios, sí perseguía medir la similitud entre traducciones.

### 2.2.1. *La investigación en proceso basada en corpus*

Si bien es cierto que cuando se utilizan corpus para estudiar, por ejemplo, la calidad de las traducciones hechas por estudiantes (Granger 1999), los universales de la traducción (Laviosa-Braithwaite 1997; Mauranen y Kujamäki 2004) o el estilo de un traductor (Baker 2000; Winters 2007; Saldanha 2010), de alguna manera se está investigando el producto y el proceso, lo que presentamos aquí es un trabajo sobre el producto llevado a cabo dentro de una investigación centrada específicamente en el proceso traductor y la competencia traductora. Con una filosofía similar destaca el corpus CORPRAT (Corpus on Process for the Analysis of Translations), que forma parte de un corpus mayor conocido como CORDIALL (Corpus of Discourse for the Analysis of Language and Literature), creado por el Núcleo de Estudos da Tradução (NET) de la Faculdade de Letras de la Universidade Federal de Minas Gerais (Brasil):

[CORPRAT is] a subcorpus of novice and expert translators' texts including log files with data on the translation process recorded online (i.e. keystroke logging through TRANSLOG©), audio files gathered by means of concurrent and retrospective verbal protocols as well as their corresponding transcriptions, image files recorded through proxy monitoring systems, and text files with translation product. (Pagano, Magalhães *et al.* 2004: 145)

El grupo brasileño no solo es pionero en la construcción de un corpus como CORDIALL, que comprende otros subcorpus para el estudio de las decisiones y estrategias de los traductores, sino también en el procesamiento y en la triangulación de datos. Su nueva herramienta, Litterae, permite trabajar con datos relacionados con el proceso desde una perspectiva que se asemeja a la metodología de corpus ya que permite alinear protocolos de Translog con algo similar a *unidades de traducción*.

Otro grupo dedicado a la investigación en el proceso de la traducción, TransComp<sup>3</sup>, tiene un repositorio de materiales y datos resultantes del estudio longitudinal sobre adquisición de la competencia traductora que está llevando a cabo. Aunque lo que tiene TransComp no es un corpus en el sentido estricto de la palabra, y no se ha tratado aplicando la metodología de corpus, ya se han sacado datos de su repositorio de materiales, y no sólo sobre el proceso, sino también sobre el producto (Bayer-Hohenwarter 2010).

Por último, la metodología de corpus formaba parte principal del proyecto *CroCo*<sup>4</sup> sobre las propiedades lingüísticas de las traducciones, como por ejemplo la explicitación, y también intentó combinar la investigación sobre el proceso y el producto (Alves, Pagano y Neumann 2010). Del proyecto surgió un corpus anotado bilingüe de originales y traducciones con textos comparables en cuanto a registro en inglés y alemán, y del análisis de este corpus existen numerosas publicaciones (Vela, Neumann *et al.* 2007; Steiner 2008; Hansen-Schirra, Neumann *et al.* 2007; 2012; Čulo, Hansen-Schirra *et al.* en prensa).

El estudio basado en corpus que presentamos aquí es modesto en tanto que no debería ser interpretado como resultados sistemáticos que se pueden generalizar para toda una población (por ejemplo traductores o profesores de lenguas extranjeras) o dirección (traducción directa o inversa). Lo que pretendemos es presentar algunas búsquedas potencialmente interesantes y sus resultados, algunos de los cuales podrían apuntar alguna tendencia. Al mismo tiempo, proporcionamos algunas ideas sobre elementos sobre los que se pueden hacer comparaciones entre dos conjuntos de textos y derivamos algunas implicaciones para la enseñanza de la traducción.

### 3. Metodología

La lingüística de corpus es una metodología de estudio de la lengua con base empírica y cualquier estudio de corpus, independientemente del tamaño de éste, sigue ciertos pasos. Respecto al uso de corpus en investigación, Partington comenta el proceso que normalmente tiene lugar:

A researcher has an intuition about language, checks this against the data the corpus provides, and this checking process frequently suggests other avenues of research to be taken, often entirely unsuspected at the start of the process [...] (Partington 1998: 1).

Más detalladamente, se puede especificar que la aplicación de la metodología de corpus implica las siguientes fases:

1. Establecimiento del objeto de estudio e hipótesis
2. Construcción del corpus
3. Procesamiento del corpus
4. Análisis
5. Interpretación de resultados y reformulación de hipótesis

#### 3.1. Establecimiento del objeto de estudio e hipótesis

El estudio de corpus que se presenta aquí está basado en las traducciones resultantes del experimento de PACTE, cuyo objeto de estudio, como ya se ha indicado, era la competencia traductora. Nuestra principal hipótesis sigue siendo que el grado de experticidad en traducción se refleja (tanto en el proceso como) en el producto de la traducción. Respecto al producto nos preguntamos si se pueden observar diferencias de comportamiento entre los grupos experimentales o la dirección en traducción en el uso de elementos como los calcos, los préstamos, las notas del traductor, etc. Si nos fijamos en la similitud entre textos nos preguntamos si las traducciones hechas por los traductores son más o menos parecidas entre sí que las hechas por los profesores, o si las traducciones directas son más o menos parecidas entre sí que las inversas. Y si nos fijamos en frecuencias nos preguntamos cuáles son las agrupaciones de pal-

abras más frecuentes en cada combinación de lenguas y grupo experimental y si se observa alguna diferencia.

### 3.2. Construcción y descripción del corpus

Nuestro corpus está compuesto por los textos traducidos durante el experimento de PACTE sobre CT. Como ya se ha explicado antes, se pidió a 35 traductores y 24 profesores de lenguas extranjeras que tradujeran un texto del inglés, francés o alemán al español o catalán<sup>5</sup> (traducción directa) y otro texto del español al inglés, francés o alemán (traducción inversa). En el Anexo 1 se pueden encontrar los textos de partida que se utilizaron para el experimento y comprobar, además, que los textos en inglés, francés y alemán son paralelos, es decir, pertenecientes al mismo tema (un virus informático) y género (noticia de prensa) y de extensión similar (150 palabras aproximadamente). El texto original en español también es de una extensión similar, pero con distinto tema (lugar turístico en Cataluña) y género (descripción en página web turística).

El corpus puede definirse, según los criterios propuestos por Corpas Pastor (2001: 157 y ss) como un corpus textual paralelo multilingüe, periódico o cronológico, no anotado ni documentado. Dependiendo de los criterios de división, el mismo corpus puede tratarse como dos subcorpus con dos componentes cada uno:

- Traducciones directas y traducciones inversas, si interesa observar fenómenos relacionados con la direccionalidad.
- Traducciones realizadas por traductores y traducciones realizadas por profesores de idiomas, si interesa observar fenómenos relacionados con los grupos experimentales.

El corpus contiene aproximadamente 23 000 palabras (*tokens*). A pesar de ser un corpus relativamente pequeño, su potencial radica en su homogeneidad, en el conocimiento profundo de los textos que lo componen y las circunstancias en que fueron producidos. Más que del número de palabras, creemos más interesante hablar de que el corpus se compone de 113 textos traducidos, distribuidos de la siguiente forma.

CUADRO 1

Distribución de las traducciones según los grupos experimentales y la direccionalidad

Sujetos	Pares de idiomas	Número de textos	
		Subtotal	Total
Profesores	EN-ES	8	45
	ES-EN	9	
	FR-ES	7	
	ES-FR	8	
	DE-ES	6	
	ES-DE	7	
Traductores	EN-ES	14	113
	ES-EN	15	
	FR-ES	10	
	ES-FR	11	
	DE-ES	9	
	ES-DE	9	

### 3.3. Tratamiento del corpus

Una vez que se establecieron las preguntas clave de este estudio se pudo comprobar que el corpus no necesitaría de un procesamiento complejo para poder proporcionar datos relevantes. Es necesario indicar que las traducciones al catalán fueron eliminadas para mantener la homogeneidad en el componente de traducciones directas y así sólo tener una lengua de llegada, el español<sup>6</sup>.

Para poder realizar los cálculos relacionados con las frecuencias y también comparar documentos del corpus automáticamente fue necesario aplicar el método que se conoce como extracción de raíces (*stemming*)<sup>7</sup>. Por medio de este método se extraen las raíces de los términos que aparecen en los documentos utilizando reglas. La extracción de raíces se hace para aumentar la probabilidad de que palabras con el mismo significado o lexema pero con distinta flexión se agrupen bajo la misma raíz y, como consecuencia, contribuyan a que se observe una mayor similitud. Es necesario apuntar, sin embargo, que no todas las decisiones que toma un extractor de raíces (*stemmer*) basado en reglas tienen una base lingüística firme o incluso correcta, ya que son decisiones automáticas, si bien es cierto que se han observado mejoras significativas en aquellos casos donde se utiliza respecto a casos donde no se utiliza. Aunque la extracción de raíces es muy similar a la lematización<sup>8</sup>, el primero es un proceso impreciso, aunque más rápido que el segundo. Como ejemplo, la extracción de raíces decidió eliminar la última *o* de la palabra española *correo*, con lo que después computó *corre* y *correo* como raíces diferentes y nos obligó a tener este hecho en cuenta a la hora del recuento de casos.

El modelo de recuperación utilizado para comparar documentos es un modelo de espacio vectorial en el que se ha tenido en cuenta el peso de términos y se ha aplicado la extracción de raíces basada en reglas y específico para cada lengua, eliminación de palabras vacías (sólo para 1-gramas), así como conversión de mayúsculas a minúsculas y eliminación de signos diacríticos para normalizar el texto<sup>9</sup>. Este modelo dio como resultado una hoja CSV (*Comma Separated Values*) que mostraba, para cada una de las carpetas del corpus (por ejemplo Profs/DE/inversa), la distribución de las frecuencias de los 1-gramas, 2-gramas, 3-gramas y 4-gramas<sup>10</sup>. Gracias a ello, también fue posible calcular los grados de similitud entre distintas combinaciones de textos.

FIGURA 1

Distribución de frecuencias de 1-gramas, 2-gramas, 3-gramas y 4-gramas en la combinación de lenguas ES>DE

Profs/DE/inverse (7 texts)					Trad/DE/inverse (9 texts)				
1-gram	2-gram	3-gram	4-gram	1-gram	2-gram	3-gram	4-gram	1-gram	
20 novella	15 plana novella	7 domenech i grau	6 Jahre spater von d	27 novella	18 plana novella	9 domenech i grau	9 pere domenech i grau		
15 plana	10 von d	7 la plana novella	6 pere domenech i grau	19 plana	12 novella ist	9 pere domenech i	8 plana novella ist ein		
13 la	7 spater von	6 pere domenech i	5 plana novella ist ein	12 stil	10 des palau	8 novella ist ein	5 ist ein altes landgut		
10 stil	7 novella ist	6 spater von d	4 auf einer kleinen eb	12 palau	9 i grau	8 plana novella ist	5 novella ist ein alt		
7 garraf	7 la plana	6 die plana novella	4 domenech i grau erworb	11 landgut	9 pere domenech	7 des palau novella	5 auf einer kleinen eb		
7 eklekt	7 domenech i	6 Jahre spater von	4 sich auf einer klein	10 olivella	9 domenech i	6 der regierung d	3 der regierung des kon		
7 landgut	7 i grau	5 novella ist ein	3 novella die plana novella	10 garraf	9 palau novella	6 regierung des kon	4 befindet sich auf ein		
7 verschied	6 in d	5 auf einer klein	3 zehn jahre spater von	9 eben	8 ist eklekt	5 ist ein alt	4 eben in der mitt		
7 spat	6 Jahre spat	5 plana novella ist	3 spater von der alfonsin	9 eklekt	8 auf ein	5 sich auf ein	4 eine legende von zugellos		
7 legend	6 auf ein	5 einer kleinen eb	3 plana novella die plana	9 per	8 von d	5 auf einer klein	4 die das vermogen d		
7 domenech	6 pere domenech	4 in der mitt	3 die plana novella di	9 la	8 ist ein	5 garraf und gehort	4 in amerika reich geword		
7 grau	6 einer klein	4 garraf und gehort	3 die plana novella ist	9 domenech	7 regierung d	5 einer kleinen eb	4 legende von zugellosigkeit und		
6 alt	6 die plana	4 sich auf ein	3 ebene in der mitt	9 grau	7 Jahre spat	5 ein altes landgut	4 zehn jahre spater wurd		
6 bezeichnet	5 ist eklekt	4 der architektonische stil	3 wurde 10 jahre spater von	8 vermog	7 die das	4 legende von zugellos	4 ein altes landgut das		
6 palau	5 la trona	4 es befindet sich	3 kleinen ebene in d	8 regier	6 aber ein	4 das vermogen d	4 in der mitt d		
6 per	5 ist ein	4 i grau erworb	3 einer kleinen ebene in	8 spat	6 des kon	4 altes landgut das	3 auf eine kleine eb		
6 gehort	5 befindet sich	3 das vermogen d	3 wurde zehn jahre spat	8 erklart	6 der regier	4 in der mitt	3 garraf und gehort d		
6 olivella	5 kleinen eb	3 von der alfonsin	3 befindet sich auf ein	7 aufall	6 des amerikan	4 la plana novella	3 des palau novella ist		

### 3.4. Análisis

El trabajo con corpus, aunque se tengan hipótesis y objetivos claros, siempre implica cierto grado de ensayo y error, probar intuiciones respecto a lo que puede arrojar el corpus como resultados, observar datos y discernir si son relevantes o no y reformular búsquedas para recoger nuevas intuiciones o hipótesis. Además, como dice Baker (2004: 183), «in corpus work, as in any other type of research, the real challenge lies in two things: one is how a researcher might select features to focus on, and the other is how he or she might interpret what they find in their data». Para responder a la primera de las cuestiones, en nuestro estudio se da el caso de que, al haber participado en el experimento de PACTE como investigadora, los textos originales y los traducidos ya nos eran familiares y, por lo tanto, ya teníamos intuiciones o pistas sobre qué rasgos diferenciales podían encontrarse en la superficie de los textos en el corpus. Además, resultados procesuales del experimento (PACTE, en prensa) nos llevaron a interesarnos por determinados aspectos. Por ejemplo, el hecho de que los profesores habían obtenido valores más bajos en la aceptabilidad de sus traducciones de los *Puntos Ricos*, es decir, problemas de traducción prototípicos preseleccionados, nos llevó a querer analizar más allá de dichos puntos y considerar bien otras expresiones, bien el texto completo. El hecho de que los traductores habían conseguido una mejor aceptabilidad que los profesores en las traducciones inversas a pesar de que ninguno de los dos grupos tenía experiencia a la hora de traducir en esta dirección nos llevó a interesarnos por la comparación de las traducciones entre sí. Finalmente, el interés por observar elementos como los calcos o préstamos, que podrían estar relacionados con las habilidades de documentación, vino, por un lado, del hecho de que los traductores utilizaron más recursos de documentación y realizaron más búsquedas, así como búsquedas más complejas, y por otro lado, del hecho de que los profesores tendieron a copiar las soluciones que aportaba el diccionario bilingüe en mayor medida que los traductores. Hemos dividido los elementos de análisis en dos focos en función de lo que se quiere observar y las herramientas y técnicas utilizadas para tal propósito.

#### 3.4.1. Foco A

El primer foco de análisis lo forman palabras, expresiones, signos de puntuación y algunos cálculos estadísticos en los que se sospechaba que se observarían diferencias entre los dos grupos experimentales. Para la extracción de concordancias y listas de palabras se utilizaron dos programas de análisis de corpus ampliamente conocidos, WordSmith Tools v. 5<sup>11</sup> y AntConc v. 3.1.2w<sup>12</sup>, mientras que para el recuento de palabras y frases y algunos cálculos estadísticos como la *type/token ratio*, la diversidad léxica y la longitud media de frase se utilizó el analizador online ADA (Adelex Analyser<sup>13</sup>).

A continuación se presentan los elementos que fueron analizados en esta primera parte ya que no requerían de ningún tratamiento de texto en particular y se podían hacer pruebas en sólo una parte del corpus para comprobar si había alguna tendencia. Además, para retomar la primera cuestión que planteaba Baker en su cita, se dan algunas razones por las que se ha elegido analizar estos elementos.

### 3.4.1.1 *Palabras susceptibles de ser traducidas como calcos*

Un *calco* es la «adopción del contenido semántico de una palabra o expresión extranjera, traduciendo su significado mediante unidades lingüísticas propias de la lengua de recepción<sup>14</sup>». Aunque algunos calcos terminan por ser totalmente asimilados por la lengua de llegada (por ejemplo súper-hombre, rascacielos, etc.), normalmente pasan un periodo de tiempo en el que su estatus no está claro. Algunos calcos conviven con otras palabras o expresiones que se consideran más correctas, lo cual no garantiza que los primeros no consigan instalarse en la lengua de llegada. A los estudiantes de traducción se les insiste en la importancia de evitar falsos amigos y calcos para que su discurso traducido suene natural y fluido en la lengua de llegada. Hay pruebas de que los traductores, según Bernardini y Ferraresi (2011), tienden a evitar estos elementos en sus traducciones, y otras (Tercedor 2010) que muestran cómo estudiantes de traducción utilizan menos cognados, un concepto que puede compartir características con el calco, que estudiantes de idiomas sin formación en traducción. El calco es, por tanto, un fenómeno del que nuestros sujetos deberían ser muy conscientes, especialmente los traductores.

### 3.4.1.2 *Palabras susceptibles de ser traducidas como préstamos*

Un *préstamo* es un «elemento, generalmente léxico, que una lengua toma de otra<sup>15</sup>». Al igual que ocurre con los calcos, los préstamos pueden entrar en una lengua y llegar a formar parte de ella. Al principio, sin embargo, pueden tener un estatus de corrección incierto e incluso mostrar variantes ortográficas. Se podría pensar que, dado que los traductores podrían estar más acostumbrados a trabajar con distintos campos de especialidad, quizás estén también más acostumbrados a ver préstamos en los textos originales y a utilizarlos en sus traducciones, es decir, a verlos como algo natural.

### 3.4.1.3 *Notas del traductor*

Son notas donde el traductor explicita algo que puede ser de difícil comprensión para el lector. Dado el conocimiento de este recurso, así como la sensibilidad hacia él que los traductores pueden tener frente a los profesores, se pensó que podrían observarse diferencias entre los grupos experimentales. Asimismo, también podría darse un uso distinto según la direccionalidad de la traducción dada la diferencia temática y de género entre el texto para traducir en directa o en inversa.

### 3.4.1.4 *Paréntesis*

Los paréntesis pueden cumplir funciones variadas como proporcionar información adicional, sinónimos o equivalentes en otra lengua, encerrar números o letras, etc. Se podría sospechar que quizás habría diferencias en la manera en que traductores y profesores utilizarían este signo, como por ejemplo en la frecuencia de uso o la función según la direccionalidad y el género textual.

### 3.4.1.5 *Ratio palabra/forma y densidad léxica*

En la lingüística de corpus, «[...] a ‘type’ [palabra o tipo] is the base form of a word, as you might find it listed in a dictionary, whereas a ‘token’ [forma o caso] is the actual word, each occurrence of which can be counted» (Hughes 1996: 163). El *ratio palabra/*

*forma* (type/token ratio) es «the ratio between the actual number of different words in a text and the total number of words (...) and is an indicator of lexical density or richness» (Corpas Pastor y Seghiri Domínguez 2008). Un concepto similar es la *diversidad léxica*, que mide la variedad de vocabulario independientemente de la extensión del texto. En nuestro estudio, y dado que el ratio palabra/forma depende mucho de la longitud del texto, también ofrecemos el índice de diversidad léxica (Yule 1944; citado en Oakes 1998: 204), tal como lo calcula el analizador online ADA. Ya que se sospechaba que no habría grandes variaciones entre los textos del corpus dada su limitada extensión y poco margen para la creatividad o la ampliación, sólo se hicieron estos cálculos con las traducciones inversas al inglés a modo de prueba.

#### 3.4.1.6 *Longitud de frase*

Es de todos conocido que las lenguas tienen cierta preferencia por las frases largas o cortas dependiendo de los recursos con los que cuenten (género, número, desinencias verbales, etc.). Aparte de este hecho, el número de palabras que se encajan en una frase puede ser indicativo de la dificultad textual, pero quizás también, cuando se traduce, de la adhesión a las estructuras o vocabulario del texto o frase original. Sin embargo, como en el caso anterior, en esta ocasión se sospechaba que las diferencias no serían relevantes debido a la limitada extensión y poco margen de maniobra de los textos, y por ello sólo se hizo este cálculo con las traducciones inversas al inglés a modo de prueba.

#### 3.4.2. *Foco B*

Gracias al proceso de extracción de raíces descrito antes se pudo medir la similitud entre los textos traducidos por traductores en comparación con los traducidos por profesores, así como la similitud entre las traducciones directas en comparación con las inversas, y la frecuencia de las agrupaciones de 1, 2, 3 o 4 palabras en cada combinación de lenguas. Así, esta parte del análisis combina el uso de técnicas de corpus para la detección de plagios y la identificación humana de fenómenos destacables, esto último sobre todo en las listas de n-gramas que fueron generadas.

### 3.5. *Interpretación de resultados y reformulación de hipótesis*

#### 3.5.1. *Palabras susceptibles de ser traducidas como calcos*

En nuestro estudio se identificaron varios calcos potenciales (EN: firm, Trojan horse, launch...a connection; gain access; FR: **soumissions**; ES: americano, geografía comarcal de Cataluña). Aquí mostramos el ejemplo de Trojan horse, que podría traducirse al español como *troyano*, *caballo de Troya* o *caballo troyano*, si bien las dos últimas expresiones son calcos. La primera opción, *troyano*, es la traducción más apropiada y utilizada en el campo de los virus informáticos en español, lo cual se puede observar perfectamente en la siguiente búsqueda a través de Google dentro de sitios web de conocidas empresas antivirus, así como en la base de datos de prensa Nexis® UK.

## CUADRO 2

## Frecuencia de equivalentes en español de Trojan horse en fuentes en Internet

Sitios Web	Equivalentes de Trojan horse		
	<i>troyano</i>	<i>caballo de Troya</i>	<i>caballo troyano</i>
<i>Google España</i> <sup>16</sup>	1 190	68	2
<i>Panda Security España</i> <sup>17</sup>	29 700	40	0
<i>AVG España</i> <sup>18</sup>	42	1	0
<i>McAfee España</i> <sup>19</sup>	13	0	0
<i>Nexis® UK: Spanish Language News</i> (from 1 January 2000 to 15 August 2011) <sup>20</sup>	( <i>troyano + virus</i> ) 1 179	( <i>caballo de troya + virus</i> ) 224	( <i>caballo troyano + virus</i> ) 9

Respecto a cómo manejaron la traducción de este término los dos grupos experimentales, 5 de 8 (62,5 %) profesores de inglés escogieron el calco *caballo de Troya* como equivalente de *Trojan horse*. Un profesor eligió *caballo troyano* y otro *programa tapadera*. Solo un profesor (12,5 %), PI9, que es el que tiene mayor puntuación en aceptabilidad de resultados entre los profesores, eligió el equivalente más apropiado, *troyano*. Mientras tanto, 8 de 14 (57,1 %) traductores escogieron *troyano*, 4 eligieron *caballo de Troya*, uno *caballo troyano* y uno más eliminó la referencia por completo.

Si ponemos esto en una tabla, las diferencias se vuelven aún más evidentes:

## CUADRO 3

## Porcentaje de sujetos que optaron por un calco o no en su traducción EN&gt;ES

Sujetos	Traducción	
	<i>troyano</i>	<i>caballo de Troya + caballo troyano</i>
Traductores	57,1 %	35,7 %
Profesores de idiomas	12,5 %	75 %

Si, además de mirar las cifras, nos fijamos en las concordancias de *troya\**, se puede observar perfectamente que sólo el profesor PI9 evitó el calco y utilizó *troyano* (aunque hay que decir que, un poco más adelante en la frase, este mismo sujeto cae en un calco en una expresión diferente ya que traduce *gain access* como *ganar acceso*).

FIGURA 2

Concordancias de *troya\**

KWC	File
un programa tipo caballo de Troya que podría permitir al pirata hacerse con el control remoto	TR_PI1_1.txt
instala un programa Caballo de Troya el cual podría permitir a un pirata informático	TR_PI3_1.txt
la una especie de caballo de Troya informático que podría hacer que un pirata tomar	TR_PI4_1.txt
virus instala un caballo de Troya que permite al hacker manipular el ordenador	TR_PI6_1.txt
instala un programa Caballo de Troya con el que el pirata puede controlar a distancia	TR_PI7_1.txt
un programa llamado caballo troyano que permitiría al pirata informático tomar control	TR_PI8_1.txt
. El virus también instala un troyano que permitiría a los piratas <i>ganar acceso</i> remoto	TR_PI9_1.txt
imágenes variables, usando el troyano ya instalado. Según las informaciones de los fabricantes. También instala un troyano que permite al pirata informático tomar el control	TR_TRI3_1.txt
de las. También instala un virus troyano que permitiría a un pirata hacerse con el control	TR_TRI10_1.txt
un poco, también instala un troyano, gracias al cual el pirata puede manejar con toda	TR_TRI12_1.txt
instala un programa "Caballo Troyano" que permite al hacker controlar el ordenador informáticas. También introduce un troyano que permite el control a distancia de los ordenadores. También un programa "Caballo de Troya" que permite al hacker tomar el control remoto de	TR_TRI13_1.txt
También instala un programa troyano que posibilitaría a un pirata hacerse con el control	TR_TRI15_1.txt
instala un programa caballo de Troya, que puede permitir que un pirata informático se	TR_TRI1_1.txt
programa llamado caballo de Troya, que permite que el pirata informático controle	TR_TRI2_1.txt
contraseñas. También instala un troyano que permite a los piratas informáticos controlar	TR_TRI3_1.txt
. También instala un programa troyano que permite a un hacker el acceso y control remoto	TR_TRI4_1.txt
contraseñas. También instala un virus troyano que permite al hacker controlar remotamente las t	TR_TRI5_1.txt
	TR_TRI6_1.txt
	TR_TRI8_1.txt
	TR_TRI9_1.txt

### 3.5.2. *Palabras susceptibles de ser traducidas como préstamos*

En los textos originales de nuestro estudio se identificaron varios préstamos potenciales, que son keylogger, hacker, backdoor, e-mail e indiano.

En este caso no se trata tanto de si el equivalente es correcto o no, sino de observar cuál ha sido la elección de los sujetos en cuanto a utilizar un préstamo o un equivalente en la lengua de llegada. Para más información, el Libro de estilo de *El País*, diario que supuestamente encargaba las traducciones directas, indica que «las palabras no castellanas se escriben en cursiva, salvo las excepciones recogidas en el Libro de estilo, y desde luego, con la acentuación, el género o los plurales que les corresponden en su idioma original<sup>21</sup>».

A continuación se muestran los porcentajes de sujetos que eligieron utilizar el préstamo en sus traducciones.

CUADRO 4

Porcentaje de sujetos que optaron por los préstamos keylogger o hacker en su traducción EN>ES

Sujetos	Préstamos	
	<u>keylogger</u>	<u>hacker</u>
Traductores	35,7 %	42,8 %
Profesores de idiomas	25 %	25 %

Dos (25 %) son los profesores que eligieron usar el préstamo keylogger. Uno de ellos es PI9, el mismo que obtuvo el mayor índice de aceptabilidad entre los profesores y cuyo comportamiento, como podemos comprobar, se asemeja en ocasiones al del grupo de traductores. Aparte de este hecho, los datos parecen apuntar hacia una ligera tendencia de los traductores a utilizar más el préstamo, como habíamos pensado inicialmente. Sin embargo, en las traducciones directas del alemán se da una situación diferente y particular ya que en el texto original en alemán aparecía una palabra en inglés y todos los profesores optaron por mantenerla en su traducción al español.

CUADRO 5

Porcentaje de sujetos que optaron por el préstamo backdoor en su traducción DE>ES

Sujetos	Préstamo
	<u>backdoor</u>
Traductores	55,5 %
Profesores de idiomas	100 %

También se comprobaron otros casos de préstamos para verificar si existía alguna tendencia. Si bien es cierto que pueden tratarse como dos casos diferentes, en el siguiente caso agrupamos los resultados de la misma palabra, e-mail, que aparecía tanto en el texto inglés como en el alemán.

CUADRO 6

Porcentaje de sujetos que optaron por el préstamo email/e-mail en su traducción EN/DE&gt;ES

Sujetos	Préstamo
	email / e-mail
Traductores	17,4 %
Profesores de idiomas	21,4 %

De nuevo, los datos no permiten extraer conclusiones respecto al uso de este préstamo ni hacer generalizaciones sobre los grupos.

Incluso se comprobó un préstamo que se daba en las traducciones inversas, *indiano*, que por su carácter cultural podría haber pasado sin cambios a la traducción.

CUADRO 7

Porcentaje de sujetos que optaron por el préstamo *indiano* en su traducciones inversas (ES>EN y FR/DE)

Sujetos	Préstamo
	<i>indiano</i>
Traductores	25,7 %
Profesores de idiomas	20,8 %

Sin embargo, no se pudo observar ni un uso abundante ni unas diferencias relevantes entre grupos.

### 3.5.3. Notas del traductor

Dado el género periodístico del texto para la traducción directa, en este caso las notas del traductor hubieran sido totalmente excepcionales. Sin embargo, el texto para la traducción inversa, dado su género y carga cultural, sí se prestaba a este recurso. Así, se pudo comprobar que dos profesores y un traductor introdujeron una nota para aclarar el término *indiano*. A continuación ofrecemos dichas notas tal y como las escribieron los sujetos. No se observa ninguna diferencia de uso entre los mismos, aparte de su mayor o menor corrección gramatical y estilística.

- PI4: This term refers to a Spanish migrant that made a fortune somewhere in Latin America in that period.
- PI6: A Spaniard who has made good in America
- TRI13: Indiano (from “las Indias”, South America): A wealthy man in 19th century Catalonia who became rich in America, mainly Cuba.

### 3.5.4. Paréntesis

El siguiente cuadro muestra los datos obtenidos de todo el corpus respecto al uso de los paréntesis según el grupo y la direccionalidad.

CUADRO 8

Porcentaje de sujetos que utilizan paréntesis en sus traducciones

Sujetos	Traducción directa	Traducción inversa
Traductores	18,2 %	45,7 %
Profesores de idiomas	4,8 %	29,2 %

Si detallamos estas cifras, mientras que 6 traductores de 33 (18,2 %) utilizaron paréntesis en la traducción directa, sólo 1 profesor de 21 (4,8 %) lo hizo. En cuanto a las traducciones inversas, 16 traductores de 35 (45,7 %) y 7 profesores de 24 (29,2 %) utilizaron este signo.

Como se desprende de los datos, los traductores tienen una cierta tendencia a utilizar este signo de puntuación más que los profesores, y en general, se observa una mayor presencia en inversa que en directa. Una explicación plausible para el segundo hecho es, como ya hemos apuntado, que el texto para traducir a la lengua extranjera tenía muchos elementos de la cultura de origen que exigirían dar más información. Éste es precisamente una de las funciones con las que se han utilizado los paréntesis en las traducciones inversas, además de proporcionar equivalentes de términos en la lengua de partida o de llegada. Los paréntesis en las traducciones directas también sirvieron este último propósito, como en el ejemplo backdoor (*puerta trasera*).

### 3.5.5. *Type/token ratio y densidad léxica*

La type/token ratio sólo se calculó con las traducciones inversas al inglés a modo de prueba para comprobar si se podía observar alguna tendencia. Como se sospechaba, no se encontró ninguna diferencia entre los grupos experimentales, lo cual puede deberse al hecho de que los textos originales eran muy cortos y no permitían mucho juego en cuanto a creatividad, extensión, etc.

En los Anexos 2 y 3 se muestran dos tablas con todos los sujetos o archivos ordenados primero según la type/token ratio y después según la diversidad léxica en orden decreciente. Se puede comprobar que no aparece ningún patrón.

### 3.5.6. *Longitud media de frase*

Además de observar el número de frases en cada texto y el número de palabras en la frase más larga y más corta, se calculó la longitud media de frase. No se puede decir que emergiera un patrón claro de comportamiento, lo cual se puede comprobar en el Anexo 4.

### 3.5.7. *Similitud*

En el cuadro 9 los números 1, 2, 3 y 4 representan agrupaciones de 1, 2, 3 o 4 palabras (o gramas) y los decimales indican el grado de similitud entre los textos que se comparan. Si tomamos la primera fila de datos de la primera columna como ejemplo, 0,66 es el grado de similitud entre los 6 textos traducidos del alemán al español tomando palabras individuales como base de la comparación. Si los mismos textos se comparan en base a agrupaciones de 2 palabras, el grado de similitud es de 0,35, si son agrupaciones de 3 palabras es de 0,15, y si se toman en cuenta agrupaciones de 4 palabras el grado de similitud es de 0,09. Obviamente, cuanto más larga es la agrupación, más bajo será el grado de similitud ya que es más difícil que las mismas agrupaciones estén presentes en varios textos.

## CUADRO 9

Grados de similitud entre textos basada en n-gramas (de 1 hasta 4) y separada según par de lenguas, direccionalidad y sujetos

Direccionalidad	Profesores					Traductores				
	Entre N archivos (N)	n-gramas (n)				Entre N archivos (N)	n-gramas (n)			
		1	2	3	4		1	2	3	4
DE-ES directa	6	0,66	0,35	0,15	0,09	9	0,72	0,42	0,20	0,12
ES-DE inversa	7	0,54	0,23	0,10	0,05	9	0,52	0,24	0,11	0,05
EN-ES directa	8	0,73	0,44	0,22	0,14	14	0,74	0,48	0,25	0,16
ES-EN inversa	9	0,53	0,36	0,17	0,1	15	0,57	0,36	0,15	0,08
FR-ES directa	7	0,74	0,57	0,37	0,28	10	0,70	0,48	0,26	0,18
ES-FR inversa	8	0,63	0,41	0,27	0,17	11	0,66	0,40	0,25	0,15

Si nos centramos primero en la columna de los profesores, se puede observar que los números más altos son siempre los de las traducciones tanto desde el francés (de 1-grama hasta 4-gramas: 0,74; 0,57; 0,37; 0,28) como al francés (0,63; 0,41; 0,27; 0,17). Lo mismo ocurre en la columna de los traductores, especialmente si nos fijamos en las agrupaciones de 3 y 4 palabras. En otras palabras, las traducciones hechas por traductores y profesores del francés al español y viceversa son las más parecidas entre sí en comparación con las traducciones ES<>EN y ES<>DE.

Sin embargo, si miramos los datos con más atención, éstos revelan que el nivel de similitud más alto entre las traducciones se da en aquellas realizadas por profesores y del francés al español. Así, se repiten grandes segmentos y el grado de similitud es notablemente más alto (4: 0,28) que en ninguna otra combinación de lenguas o grupo de sujetos.

A continuación se muestran unas concordancias que ilustran cuán similares son los textos traducidos por los profesores del francés al español.

FIGURA 3

Concordancias de *hombro* en el subcorpus de traducciones directas FR>ES hechas por profesores

KWIC	File
R.B, EL VIRUS INFORMÁTICO QUE MIRA POR ENCIMA DEL HOMBRE A SUS VÍCTIMAS. El gusano arrebata información c	TR_PP1_1.txt
ar.B, el virus informático que lee por encima del hombro de sus víctimas El gusano sustrae información co	TR_PP2_1.txt
AR.B, EL VIRUS INFORMÁTICO QUE LEE POR ENCIMA DEL HOMBRE DE SUS VÍCTIMAS. El gusano roba informaciones co	TR_PP3_1.txt
AR.B, EL VIRUS INFORMÁTICO QUE LEE POR ENCIMA DEL HOMBRE DE SUS VÍCTIMAS El gusano roba informaciones con	TR_PP4_1.txt
AR.B, el virus informático que lee por encima del hombro de sus víctimas El gusano se apodera de las info	TR_PP6_1.txt
AR.B, EL VIRUS INFORMÁTICO QUE LEE POR ENCIMA DEL HOMBRE DE SUS VÍCTIMAS. El gusano roba informaciones c	TR_PP7_1.txt
PAR.B EL VIRUS INFORMÁTICO QUE LEE POR ENCIMA DEL HOMBRE DE SUS VÍCTIMAS. El gusano roba informaciones co	TR_PP8_1.txt

Una posible explicación para el hecho de que las traducciones del francés al español y viceversa siempre sean las más parecidas entre sí es la proximidad que existe entre las dos lenguas. Se puede esbozar la hipótesis de que los sujetos, y en particular los profesores, se adhieren más al texto original y no exploran por completo las posibilidades de la lengua de llegada, con lo que se limitan ellos mismos el número de traducciones posibles<sup>22</sup>.

Continuando con el análisis de similitudes entre textos, otra posible comparación es entre traducciones directas e inversas, sin distinguir entre los grupos experimentales. Lo que se compara aquí es, por ejemplo, todas las traducciones hechas del

alemán al español y ver lo parecidas que son entre sí. El cuadro 10 muestra los datos resultantes de esta comparación.

CUADRO 10

Grados de similitud entre textos basada en n-gramas (de l'hasla 4) y separada según par de lenguas y direccionalidad

Direccionalidad	Profesores y traductores			
	n-gramas (n)			
	1	2	3	4
DE-ES directa	0,69	0,38	0,18	0,10
ES-DE inversa	0,52	0,23	0,10	0,05
EN-ES directa	0,73	0,46	0,23	0,15
ES-EN inversa	0,57	0,36	0,15	0,09
FR-ES directa	0,69	0,50	0,29	0,20
ES-FR inversa	0,64	0,40	0,25	0,16

Como se puede observar, los números de las filas que corresponden a las traducciones inversas son siempre más bajos que los de las traducciones directas. En otras palabras, las traducciones inversas son siempre menos parecidas entre sí, independientemente del par de lenguas o grupo de sujetos implicados.

Este hecho, el que las traducciones inversas sean siempre más variadas entre sí que las directas, puede deberse a que cada uno escribe en la lengua extranjera dentro de los límites de sus posibilidades. Según la competencia lingüística del sujeto, éste expresará la misma idea de una forma u otra. Pero también puede haber otra posible explicación. La aceptabilidad de las soluciones a los puntos ricos en las traducciones inversas fue mucho más bajo que en las traducciones directas, lo cual se traduce en que a la gama de posibles traducciones *correctas* habría que añadir una amplia gama de traducciones *incorrectas*.

### 3.5.8. Frecuencias destacables

Las listas de n-gramas con las agrupaciones de palabras más frecuentes en cada combinación de lenguas y grupo experimental facilitaron la observación de patrones o irregularidades en el uso de dichas agrupaciones.

A modo de ejemplo, en la lista de agrupaciones (de 1 palabra) en español proveniente de las traducciones directas del inglés, se observó que *máquina/s* aparecía 13 veces en 8 textos de profesores y 12 veces en 14 textos de traductores. En el texto original machine en el sentido de ordenadores aparece 4 veces. Con estos datos se puede afirmar que, aunque no es una opción del todo incorrecta, los profesores han optado por la traducción más pegada al original, sin tener tanto en cuenta el contexto. Los traductores, que probablemente ya se habían encontrado en su profesión con textos donde aparecía la palabra machine en diferentes contextos, quizás tenían ya almacenados en su memoria diferentes equivalentes y sabían que la opción más parecida al original, *máquina*, no era la mejor en este caso.

Otro ejemplo similar se encuentra en la lista de agrupaciones (de 1 palabra) en español proveniente de las traducciones directas del alemán. Aquí se observó que *productor* aparecía 6 veces en 6 traducciones de profesores, mientras que sólo aparecía 3 veces en 9 traducciones de traductores. El término original en alemán es

*Hersteller* y, como es posible que ocurriera en el caso anterior, el primer equivalente que ofrece el diccionario bilingüe es *productor*, si bien en este contexto el término apropiado sería *fabricante*.

Sin embargo, las listas de n-gramas no sólo sirvieron para detectar decisiones erróneas por parte de los profesores. En la lista de agrupaciones (de 2 palabras) en inglés proveniente de las traducciones inversas del español se observó una frecuencia demasiado alta para el término high chair en ambos grupos de sujetos. Aunque high chair puede ser una traducción de *trona*, y así lo muestra cualquier diccionario bilingüe, el equivalente correcto en este caso era throne ya que mantiene el mismo sentido de váter que el original. Al investigar un poco más, se observó que el 46,7 % de los traductores y el 55,5 % de los profesores optaron por el equivalente incorrecto high chair, lo cual no dice mucho en favor de las habilidades de documentación de ambos grupos de sujetos en este caso en particular.

#### 4. Conclusiones

El corpus de traducciones resultante del experimento del Grupo PACTE es un ejemplo del esfuerzo puesto en el diseño de una investigación. El haber cuidado la selección de sujetos, con dos grupos internamente homogéneos en cuanto a criterios profesionales, la validez ecológica, con un entorno experimental tan natural como era posible, la selección de textos, con tres textos paralelos muy similares en distintas lenguas, entre otros elementos, nos han permitido contar con un corpus tremadamente homogéneo y controlado en el que se han podido observar fenómenos que apuntan diferencias o similitudes, bien entre los grupos experimentales, bien en cuanto a direccionalidad.

Así, aunque limitados por el número de casos, se ha podido observar una diferencia clara en el uso de un calco entre los grupos, mientras que el análisis de algunos préstamos en las traducciones no ofreció ninguna tendencia, al igual que en el caso de las notas del traductor. Si bien la frecuencia en el uso de los paréntesis parecía apuntar a una distinción entre grupos, luego se vio que su uso era similar. En la misma línea, los resultados para la type/token ratio y la densidad léxica, así como la longitud de frase, no aportaron ningún dato revelador o distintivo.

En cuanto al análisis para el que se requería un procesamiento especial del corpus, aquí sí se pudo comprobar numéricamente la existencia de diferencias entre grupos y dirección de la traducción. En general, las traducciones del y al francés resultaron ser las más parecidas entre sí en ambos grupos, si bien, en concreto, las traducciones del francés al español hechas por profesores mostraron el índice de similitud más alto, revelando que este grupo en esta dirección se aleja poco del texto original. Por otro lado, las traducciones inversas resultaron ser más diferentes entre sí que las traducciones directas, apuntando quizá a diferentes niveles de competencia lingüística en la lengua extranjera. Por último, gracias al proceso de stemming y la generación de listas de n-gramas se pudieron observar diferencias destacables en las frecuencias de términos o agrupaciones de palabras, lo que podría apuntar a cuestiones relacionadas con la documentación o la capacidad de los sujetos de contextualizar.

Los resultados encontrados, especialmente aquellos en los que se han observado diferencias entre los grupos o la direccionalidad, pueden tener implicaciones para el estudio de la competencia traductora y su adquisición. El uso del calco, la proximidad

o lejanía respecto al texto original, la importancia del contexto o el uso crítico de los diccionarios u otras fuentes de documentación, pueden revelarse como aspectos diferenciales muy concretos. Por último, en cuanto a cuestiones metodológicas, se ha podido comprobar cómo un análisis más sencillo de un corpus sin tratar puede ofrecer algunos resultados (foco A), si bien es cierto que, en este caso, sin la aplicación de ciertas procedimientos, se hubieran perdido datos aún más relevantes y reveladores (foco B).

## NOTAS

1. *Grup PACTE* (Actualizado el 25 de julio de 2013): Consultado el 8 de septiembre de 2013, <<http://grupsderecerca.uab.cat/pacte/es>>.
2. Markus MUHR, Roman KERN, Mario ZECHNER y Michael GRANITZER (2010): External and Intrinsic Plagiarism Detection using a Cross-Lingual Retrieval and Segmentation System, no publicada. *Conference on Multilingual and Multimodal Information Access Evaluation*. Padua, 20-23 de septiembre de 2010 (Actualizado el 22 de septiembre de 2010): Consultado el 8 de septiembre de 2013, <<http://www.uni-weimar.de/medien/webis/research/events/pan-10/pan10-talks/pan10-plagiarism-detection/muhr10-talk-external-and-intrinsic-plagiarism-detection-using-a-cross-lingual-retrieval-and-segmentation-system.pdf>>.
3. *TransComp – The Development of Translation Competence* (Actualizado el 8 de agosto de 2011): Consultado el 8 de septiembre de 2013, <<http://gams.uni-graz.at/fedora/get/container:tc/bdef:Container/get>>.
4. *CroCo – Linguistic Properties of Translations – A Corpus-based Investigation for the Language Pair English-German* (Actualizado el 26 de enero de 2010): Consultado el 8 de septiembre de 2013, <[http://fr46.uni-saarland.de/croco/index\\_en.html](http://fr46.uni-saarland.de/croco/index_en.html)>.
5. Para reducir el ruido en el análisis se eliminaron del corpus los textos traducidos al catalán, que eran sólo cinco.
6. Además, cada texto, que estaba originalmente en formato MSWord, se convirtió a texto plano, y a los archivos .txt resultantes se le dio el mismo nombre que a sus correspondientes .doc. Los nombres de todos los archivos mantienen una lógica. Así, por ejemplo, el nombre de archivo TR\_PI6\_2.txt, permite identificar inmediatamente el tipo de fuente o material (TR = traducción), el tipo de sujeto (P = profesor), la lengua extranjera del sujeto (I|=inglés), el número identificativo del sujeto (6) y si la traducción era directa o inversa (2 = inversa).
7. La extracción de raíces (*stemming*) es una técnica utilizada para aumentar el rendimiento de los sistemas de IR [information retrieval] [...]. El objetivo de los algoritmos de extracción de raíces, o de eliminación de sufijos, es obtener un único término a partir de diferentes palabras con un mismo significado que difieren esencialmente en su morfología. (Ureña López 2002: 24)
8. La lematización se utiliza en RI [recuperación de información] para combinar variantes morfológicas y obtener la raíz morfológica de cada palabra. (Baeza-Yates 2009: 18)
9. La aplicación de la extracción de raíces, la generación de listas de n-gramas y los cálculos comparativos entre textos se encargaron al profesor alemán Martin Potthast, de la Bauhaus-Universität Weimar.
10. Los n-gramas son series de *n* palabras.
11. *WordSmith Tools v. 5*: Consultado el 8 de septiembre de 2013, <<http://www.lexically.net/wordsmith/index.html>>.
12. *AntConc v. 3.1.2w*: Consultado el 8 de septiembre de 2013, <[http://www.antlab.sci.waseda.ac.jp/antconc\\_index.html](http://www.antlab.sci.waseda.ac.jp/antconc_index.html)>.
13. *ADA Adelex Analyser v. XX*: Consultado el 8 de septiembre de 2013, <<http://www.ugr.es/~inped/ada>>.
14. REAL ACADEMIA ESPAÑOLA (2001): Entrada *calco*. *Diccionario de la lengua española*, 22<sup>a</sup> ed. Consultado el 8 de septiembre de 2013, <<http://lema.rae.es/drae/?val=calco>>.
15. REAL ACADEMIA ESPAÑOLA (2001): Entrada *préstamo*. *Diccionario de la lengua española*, 22<sup>a</sup> ed. Consultado el 8 de septiembre de 2013, <<http://lema.rae.es/drae/?val=pr%C3%A9stamo>>.
16. *Google España*. Consultado el 8 de septiembre de 2013, <<https://www.google.es/>>.
17. *Panda Security España*. Consultado el 8 de septiembre de 2013, <<http://www.pandasecurity.com/spain>>.

18. AVG *España*. Consultado el 8 de septiembre de 2013, <<http://www.avg.com/es>>.
19. McAfee *España*. Consultado el 8 de septiembre de 2013, <<http://www.mcafee.com/es/>>.
20. *Nexis® UK: Spanish Language News*. Consultado el 8 de septiembre de 2013, <[http://www.lexisnexis.com/uk/nexis/auth/signonform.do;jsessionid=BFA0CC598F3F0A76C147967CF1F38B73.FPc2sZS5bB5L6pr4uoFkQ?localeParams=es\\_ES\\_ukBusiness](http://www.lexisnexis.com/uk/nexis/auth/signonform.do;jsessionid=BFA0CC598F3F0A76C147967CF1F38B73.FPc2sZS5bB5L6pr4uoFkQ?localeParams=es_ES_ukBusiness)>.
21. *Manual de estilo del diario "El País" de España* (Sin fecha de actualización): Consultado el 8 de septiembre de 2013, <<http://blogs.elpais.com/files/manual-de-estilo-de-el-pa%C3%ADs.pdf>>.
22. La tesis doctoral de Massana (en curso), que trabaja con el portugués y el español, también observa el hecho de que la proximidad entre dos lenguas sea un factor que frena a los sujetos a la hora de explorar todas las posibilidades de la lengua de llegada.

## REFERENCIAS BIBLIOGRÁFICAS

ALVES, Fábio y GONÇALVES, Jose Luiz (2007): Modelling translator's competence: relevance and expertise under scrutiny. In: Yves GAMBIER, Miriam SHLESINGER y Radegundis STOLZE, eds. *Translation Studies: doubts and directions. Selected papers from the IV Congress of the European Society for Translation Studies*. Ámsterdam: John Benjamins, 41-55.

ALVES, Fábio, PAGANO, Adriana, NEUMANN, Stella, et al. (2010): Translation units and grammatical shifts: towards an integration of product- and process-based translation research. In: Gregory SHREVE y Erik ANGELONE, eds. *Translation and Cognition*. Ámsterdam: John Benjamins, 109-142.

ASTON, Guy (2000): Corpora and language teaching. In: Lou BURNARD y Tony MCENERY, eds. *Rethinking language pedagogy from a corpus perspective*. Bern: Peter Lang, 7-17.

BAEZA-YATES, Ricardo (2009): Retos en la interacción de la recuperación de información y el procesamiento del lenguaje natural. In: Amparo ALCINA, Esperanza VALERO y Elena RAMBLA, eds. *Terminología y Sociedad del Conocimiento*. Bern: Peter Lang, 15-32.

BAKER, Mona (2000): Towards a Methodology for Investigating the Style of a Literary Translator. *Target*. 12(2):241-266.

BAKER, Mona (2004): A corpus-based view of similarity and difference in translation. *International Journal of Corpus Linguistics*. 9(2):167-193.

BARRÓN-CEDEÑO, Alberto, POTTNST, Martin, Rosso, Paolo, et al. (2010): Corpus and Evaluation Measures for Automatic Plagiarism Detection. In: *Proceedings of the Seventh conference on International Language Resources and Evaluation*. (LREC'10, Valletta, 17-23 de mayo de 2010). 4 p. Consultado el 8 de septiembre de 2013, <<http://www.lsi.upc.edu/~albarron/publications/2010/BarronPlagCompetitionLREC.pdf>>.

BAYER-HOHENWARTER, Gerrit (2010): Comparing translational creativity scores of students and professionals: flexible problem-solving and/or fluent routine behaviour?. In: Susanne GÖPFERICH, Fábio ALVES y Inger M. MEES, eds. *New Approaches in Translation Process Research*. Copenhagen Studies in Language Series. Vol. 39. Copenhagen: Samfundslitteratur, 83-111.

BELL, Roger T. (1991): *Translation and Translating*. Londres: Longman.

BERNARDINI, Silvia y FERRARESI, Adriano (2011): Practice, Description and Theory Come Together: Normalization or Interference in Italian Technical Translation? *Meta*. 56(2):226-246.

BOWKER, Lynne y PEARSON, Jennifer (2002): *Working with specialized language: A practical guide to using corpora*. Londres; Nueva York: Routledge.

CIVIT, Montserrat (2003): *Criterios de etiquetación y desambiguación morfosintáctica de corpus en español*. Colección de Monografías de la Sociedad Española para el Procesamiento del Lenguaje Natural -SEPLN-; 3. Alicante: Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). Consultado el 25 de agosto de 2011, <<http://ararat.ujaen.es/sepln/web/wp-content/uploads/2011/02/monografiaCivit.pdf>>.

CORPAS PASTOR, Gloria (2001): Compilación de un corpus ad hoc para la enseñanza de la traducción inversa especializada. *Trans*. 5: 155-184. Consultado el 15 mayo 2011 <[http://www.trans.uma.es/Trans\\_5/t5\\_155-184\\_GCorpas.pdf](http://www.trans.uma.es/Trans_5/t5_155-184_GCorpas.pdf)>.

CORPAS PASTOR, Gloria y SEGHIRI DOMÍNGUEZ, Miriam (2008): Virtual corpora as documentation resources: translating travel insurance documents (English-Spanish). *In: Allison BEEBY, Patricia RODRÍGUEZ-ÍNÉS y Pilar SÁNCHEZ-GIJÓN, eds. Corpus use and translating: corpus use for learning to translate and learning corpus use to translate.* Ámsterdam: John Benjamins.

ČULO, Oliver, HANSEN-SCHIRRA, Silvia, NEUMANN, Stella, *et al.* (en prensa): Querying the CroCo corpus for translation shifts. Beyond corpus construction: exploitation and maintenance of parallel corpora. *In: Silvia HANSEN-SCHIRRA, Stella NEUMANN y Oliver ČULO, eds. Beyond corpus construction: exploitation and maintenance of parallel corpora. International Journal of Corpus Linguistics* (Número especial).

GRANGER, Sylviane (1999): Use of tenses by advanced EFL learners: evidence from an error-tagged computer corpus. *In: Hilde HASSELGÅRD y Signe OKSEFJELL, eds. Out of corpora: Studies in honour of Stig Johansson.* Ámsterdam: Rodopi, 191-202.

HANKS, Patrick (2008): The lexicographical legacy of John Sinclair. *International Journal of Lexicography.* 21(3):219-229.

HANSEN-SCHIRRA, Silvia, NEUMANN, Stella y STEINER, Erich (2007): Cohesive explicitness and explicitation in an English-German translation corpus. *Languages in Contrast.* 7(2):241-265.

HANSEN-SCHIRRA, Silvia, NEUMANN, Stella y STEINER, Erich (2012): *Cross-linguistic Corpora for the Study of Translations. Insights from the Language Pair English-German.* Berlín; Nueva York: Mouton de Gruyter.

HUGHES, Rebecca (1996): *English in Speech and Writing: Investigating Language and Literature.* Londres; Nueva York: Routledge.

JOHNS, Tim (1991): From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *ELR Journal: classroom concordancing.* 4:27-45.

KELLY, Dorothy (2005): *A Handbook for Translator Trainers.* Manchester: St. Jerome.

KOEHN, Philipp (2005): Europarl: A parallel corpus for statistical machine translation. *MT summit.* 5:12-16.

LAIVIOSA-BRAITHWAITE, Sara (1997): Investigating Simplification in an English Comparable Corpus of Newspaper Articles. *In: Kinga KLAUDY y János KOHN, eds. Transferre Necessum Est, Proceedings of the 2<sup>nd</sup> International Conference on Current Trends in Studies of Translation and Interpreting 5-7 September, 1996.* Budapest: Scholastica, 531-540.

MÁRQUEZ, Lluís y PADRÓ, Lluís (1997): A flexible POS tagger using an automatically acquired language model. *In: Proceedings of the 35th annual meeting of the Association for Computational Linguistics: Joint EACL/ACL.* Madrid: EACL/ACL, 238-245.

MASSANA-ROSELLÓ, Gisela (en curso): *La adquisición de la competencia traductora en segundas lenguas extranjeras. Un estudio entre lenguas cercanas: La traducción de portugués a castellano.* Tesis de doctorado. Barcelona: Universitat Autònoma de Barcelona.

MAURANEN, Anna y KUJAMÄKI, Pekka (2004): *Translation universals: Do they exist?* Ámsterdam/Nueva York: John Benjamins.

MCENERY, Tony y WILSON, Andrew (1996): *Corpus linguistics.* Edimburgo: Edinburgh University Press.

NEUBERT, Albrecht (2000): Competence in language, in languages, and in translation. *In: Christina SCHÄFFNER y Beverly ADAB, eds. Developing translation competence.* Ámsterdam: John Benjamins, 3-18.

OAKES, Michael P. (1998): *Statistics for Corpus Linguistics.* Edimburgo: Edinburgh University Press.

OLOHAN, Maeve (2004): *Introducing corpora in Translation Studies.* Londres/Nueva York: Routledge.

PACTE (2002): Exploratory tests in a study of translation competence. *Conference Interpretation and Translation.* 4(2):41-69.

PACTE (2003): Building a Translation Competence Model. *In: Fábio ALVES, ed. Triangulating Translation: Perspectives in process oriented research.* Ámsterdam: John Benjamins, 43-66.

PACTE (2005a): Investigating Translation Competence: Conceptual and Methodological Issues. *Meta*. 50(2):609-619.

PACTE (2005b): Primeros resultados de un experimento sobre la Competencia Traductora. In: *Actas del II Congreso Internacional de la AIETI (Asociación Ibérica de Estudios de Traducción e Interpretación) 'Información y documentación'*. Madrid: Publicaciones de la Universidad Pontificia Comillas, 573-587. CD-ROM

PACTE (2008): First results of a Translation Competence Experiment: 'Knowledge of Translation' and 'Efficacy of the Translation Process'. In: John KEARNS, ed. *Translator and Interpreter Training. Issues, Methods and Debates*. Londres: Continuum, 104-126.

PACTE (2009): Results of the Validation of the PACTE Translation Competence Model: Acceptability and Decision Making. *Across Languages and Cultures*. 10(2):207-230.

PACTE (2011a): Results of the Validation of the PACTE Translation Competence Model: Translation Project and Dynamic Translation Index. In: Sharon O'BRIEN, ed. *Cognitive Explorations of Translation*. Londres/Nueva York: Continuum Studies in Translation, 30-53.

PACTE (2011b): Results of the validation of the Pacte Translation Competence model: Translation problems and Translation Competence. In: Cecilia ALVSTGAD, Adelina HILD y Elisabet TISELIUS, eds. *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies* Benjamins Translation Library. Vol. 94. Ámsterdam/Nueva York: John Benjamins, 317-343.

PACTE (en prensa): *Researching Translation Competence*. Ámsterdam/Nueva York: John Benjamins.

PAGANO, Adriana, MAGALHÃES, Célia y ALVES, Fábio (2004): Towards the construction of a multilingual, multifunctional corpus: Factors in the design and application of CORDIALL. *TradTerm*. 10: 143-161. Consultado el 8 de septiembre de 2013, <<http://myrtus.uspnet.usp.br/tradterm/site/images/revistas/v10n1/v10n1a07.pdf>>.

PARTINGTON, Alan (1998): *Patterns and meanings: using corpora for English language research and teaching*. Ámsterdam/Filadelfia: John Benjamins.

RODRÍGUEZ-INÉS, Patricia (2008): *Uso de corpus electrónicos en la formación de traductores (inglés-español-inglés)*. Tesis de doctorado. Barcelona: Universitat Autònoma de Barcelona.

ROYLE, Phaedra, RICHARDSON, Isabelle, BOISVERT, Sophie, et al. (2009): Is Bigger Better? Corpus and Dictionary Use in the Search for Compounds, Collocations, Derived Forms and Fixed Expressions. *Meta*. 54(3):520-532.

SALDANHA, Gabriela (2010): Translator style: methodological considerations. *The Translator*. 17(1):25-50.

SINCLAIR, John (1985): Lexicographic Evidence. In: Robert ILSON, ed. *Dictionaries, Lexicography, and Language Learning*. Oxford: Pergamon, 81-94.

STEINER, Erich (2008): Explication - towards an empirical and corpus-based methodology. In: Jonathan WEBSTER, ed. *Meaning in Context: Implementing Intelligent Applications of Language Studies*. Londres: Continuum, 235-278.

TERCEDOR, Maribel (2010): Cognates as lexical choices in translation. *Target*. 22(2):177-193.

TEUBERT, Wolfgang (1996): Comparable or parallel corpora? *International Journal of Lexicography*. 9(3):238-264.

UREÑA LÓPEZ, L. Alfonso (2002): *Resolución en la ambigüedad léxica en tareas de clasificación automática de documentos*. Alicante: Club Universitario.

VELA, Mihaela, NEUMANN, Stella y HANSEN-SCHIRRA, Silvia (2007): Querying multi-layer Annotation and Alignment in Translation Corpora. In: Matthew DAVIES, Paul RAYSON, Susan HUNSTON y Pernilla DANIELSSON, eds. *Proceedings of the Corpus Linguistics Conference* (CL2007, Birmingham, 27-30 de julio de 2007, artículo nº 97). 18 p. Consultado el 8 de septiembre de 2013, <<http://www.birmingham.ac.uk/Documents/college-artslaw/corpora/conference-archives/2007/97Paper.pdf>>.

WILKS, Yorik (2008): *Machine Translation: Its Scope and Limits. Studies in Natural Language Processing*. Nueva York: Springer.

WILSS, Wolfram (1976): Perspectives and limitations of a didactic framework for the teaching of translation. In: Richard W. BRISLIN, ed. *Translation applications and research*. Nueva York: Gardner, 117-137.

WINTERS, Marion (2007): F. Scott Fitzgerald's *Die Schönen und Verdammten*: A Corpus-based Study of Speech-act Report Verbs as a Feature of Translators' Style. *Meta*. 52(3):412-425.

YULE, George Udny (1944): *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.

## ANEXOS

### Anexo 1: Textos originales

#### 1. Texto original en inglés

##### Email virus strikes in new form

Computer users were warned last night to be on the lookout for an email virus that can steal confidential information and allow hackers to take control of infected machines. The virus, a new variant of the BugBear email worm that infected tens of thousands of computers around the world last October, began to spread rapidly from Australia to Europe and the USA at around 8am yesterday. According to MessageLabs, a Cheltenham-based virus filtering firm which reported about 30,000 infected messages in 115 countries, the propagation rate of BugBear.B almost doubled every hour throughout the morning. There was also a huge surge as US users came online. Like its predecessor, the variant spreads by sending itself as an attachment to every address in an infected machine's email address book. To disguise where it came from, it uses different subject headings. As well as searching for anti-virus software and disabling it, BugBear.B installs a keylogger to record what the user types, which may allow hackers to record confidential information such as credit card details and passwords. It also installs a "Trojan horse" program which could allow a hacker to take remote control of infected machines. Experts analysing the virus code have found that it contains the domain name of about 20 banks and financial institutions. One of the possibilities being investigated is that BugBear.B is able to check whether an infected machine is inside one of these domains and, if it is, launch a dial-up connection that could allow a hacker into a bank's computer system.

*The Guardian* – 6 June 2003

#### 2. Texto original en francés

##### Bugbear.b, le virus informatique qui lit par-dessus l'épaule de ses victimes

##### Le ver dérobe des informations confidentielles

ALERTE rouge sur le front des virus informatiques. Le ver Bugbear.B, une «résurgence» de Bugbear qui avait frappé en septembre 2002, se répand comme une traînée de poudre depuis le 4 juin. Le surlendemain, l'un des leaders de l'édition de logiciels antivirus, Symantec, annonçait la réévaluation du risque de 3 à 4 sur une échelle qui compte 5 niveaux. En moins de quarante-huit heures, l'entreprise a reçu plus de 1 000 soumissions, des communications du virus à ses experts par ses clients. Le 6 juin, l'entreprise MessageLabs, spécialiste du filtrage des messages électroniques, révélait avoir bloqué 150 000 copies du ver.

La principale originalité qui distingue Bugbear.B de ses congénères réside dans son aptitude à enregistrer les caractères tapés sur le clavier de l'ordinateur infecté, de stocker ce butin dans un fichier et d'ouvrir une «porte» dans la machine afin de permettre aux pirates de venir récupérer le fruit de leur forfait. De quoi dérober les mots de passe et les numéros de carte bancaire utilisés par la victime. Le voleur de données peut ensuite exploiter ces informations pour accéder gratuitement à Internet ou à des services en ligne et faire des achats sur la Toile, sans bourse délier.

*Le Monde*, 13 June 2003

### 3. Texto original en alemán

#### **Wurm in der Leitung**

#### **Computervirus mit neuen Tricks**

Ein Computervirus mit dem Namen Mydoom.B verbreitet sich zurzeit mit hoher Geschwindigkeit im Internet. Der digitale Parasit reist mit den E-Mails durch das Datennetz. Aktiviert wird er, wenn Internetnutzer versuchen, den als harmlose Datei getarnten Schädling zu öffnen. Anders als viele bekannte E-Mail-Würmer, die sich als Pornobildchen tarnen, versteckt sich Mydoom hinter technischen Begriffen wie „Mail Delivery System“ oder „Error“. Er versucht Rechner im lokalen Netzwerk, die schon von seiner Vorgängervariante infiziert wurden, zu infizieren, indem er das schon installierte Backdoor benutzt. Nach Angaben führender Hersteller von Antiviren-Software ist er darauf programmiert, vom 1. Februar an die Webseite des Softwareherstellers SCO zu attackieren.

Aktiviert wird der Schädling beim Start des Betriebssystems Windows. Er kopiert sich auch unter verschiedenen Dateinamen in das Download-Verzeichnis der Dateien-Tauschbörse Kazaa. Weiteren Angaben zufolge soll er auch ein Programmteil enthalten, das Tastatureingaben von PC-Nutzern nach Kreditkartennummern und Ähnlichem überwacht. Das Bundesamt für Sicherheit in der Informationstechnik hat eine detaillierte Beschreibung im Internet veröffentlicht: [www.bsi.de](http://www.bsi.de).

### 4. Texto original en español

#### **La Plana Novella**

La Plana Novella es una antigua heredad adquirida por el Indiano Pere Domenech i Grau en 1885 que se encuentra en una pequeña planicie en el centro del Parc Natural del Garraf y pertenece al municipio de Olivella. La Finca fue declarada colonia agrícola 10 años más tarde por el gobierno alfonertino, pero de aquella época perdura una leyenda de desenfreno y dilapidación que hizo desaparecer la fortuna del americano. El estilo arquitectónico del Palacete es ecléctico, es decir que mezcla diferentes estilos. La geografía comarcal de Cataluña lo califica de «Castillo de Bambalinas» como si fuese un decorado de teatro. Sin ningún tipo de duda la construcción estilísticamente más original de Palau Novella es el lavadero gaudiniano, pero una de las piezas más características y llamativas del Palau es el comunitario conocido como “la trona».

#### **Anexo 2: Traducciones ordenadas según el ratio palabra/forma (de más a menos)**

Archivo	Tokens	Types	Type/token ratio %
TR_PI9_2.txt	140	98	70,00
TR_TRI5_2.txt	141	97	68,79
TR_TRI10_2.txt	143	98	68,53
TR_TRI3_2.txt	136	92	67,65
TR_PI5_2.txt	139	94	67,63
TR_TRI8_2.txt	154	104	67,53
TR_PI8_2.txt	144	96	66,67
TR_TRI2_2.txt	142	92	64,79
TR_TRI14_2.txt	164	106	64,63
TR_PI6_2.txt	174	112	64,37
TR_TRI12_2.txt	160	102	63,75
TR_TRI4_2.txt	145	92	63,45
TR_TRI7_2.txt	153	97	63,40
TR_TRI15_2.txt	153	97	63,40
TR_TRI9_2.txt	142	90	63,38

TR_TRI1_2.txt	147	93	63,27
TR_PI2_2.txt	156	98	62,82
TR_PI7_2.txt	142	88	61,97
TR_PI3_2.txt	161	99	61,49
TR_TRI11_2.txt	158	96	60,76
TR_PI4_2.txt	162	98	60,49
TR_TRI6_2.txt	150	90	60,00
TR_PI1_2.txt	155	92	59,35
TR_TRI13_2.txt	194	114	58,76

Anexo 3: Traducciones ordenadas según la diversidad léxica (de más a menos)

Archivo	Tokens	Types	Diversidad léxica
TR_TRI7_2.txt	153	97	250,331
TR_PI4_2.txt	162	98	237,769
TR_PI7_2.txt	142	88	209,284
TR_TRI15_2.txt	153	97	201,632
TR_PI2_2.txt	156	98	198,882
TR_TRI13_2.txt	194	114	192,369
TR_TRI12_2.txt	160	102	191,406
TR_TRI1_2.txt	147	93	188,810
TR_TRI2_2.txt	142	92	186,471
TR_TRI11_2.txt	158	96	185,868
TR_PI1_2.txt	155	92	183,143
TR_TRI9_2.txt	142	90	181,512
TR_TRI3_2.txt	136	92	178,417
TR_TRI6_2.txt	150	90	176,889
TR_TRI14_2.txt	164	106	173,260
TR_PI3_2.txt	161	99	172,061
TR_PI6_2.txt	174	112	170,432
TR_PI8_2.txt	144	96	169,753
TR_TRI5_2.txt	141	97	166,994
TR_TRI8_2.txt	154	104	157,699
TR_PI5_2.txt	139	94	152,166
TR_TRI4_2.txt	145	92	142,687
TR_TRI10_2.txt	143	98	129,102
TR_PI9_2.txt	140	98	110,204

## Anexo 4: Traducciones ordenadas según la longitud media de frase (de menos a más)

Archivo	Nº de frases	Longitud máxima	Longitud mínima	Longitud media
TR_TRI2_2.txt	11	37	1	13,27
TR_TRI4_2.txt	10	21	2	14,7
TR_PI2_2.txt	10	24	3	15,9
TR_PI8_2.txt	9	22	2	16,33
TR_PI4_2.txt	10	44	3	16,5
TR_TRI13_2.txt	12	28	3	16,58
TR_TRI7_2.txt	9	45	1	17,44
TR_PI1_2.txt	9	31	3	17,66
TR_TRI12_2.txt	9	28	3	18,11
TR_PI3_2.txt	9	25	3	18,11
TR_TRI10_2.txt	8	37	3	18,5
TR_TRI15_2.txt	8	38	3	19,5
TR_PI9_2.txt	7	35	3	20,42
TR_TRI14_2.txt	8	38	3	20,87
TR_TRI1_2.txt	7	36	3	21,14
TR_TRI8_2.txt	7	42	3	22,85
TR_TRI11_2.txt	7	44	3	22,85
TR_TRI3_2.txt	6	38	2	23,16
TR_TRI5_2.txt	6	41	3	23,83
TR_PI5_2.txt	6	41	2	23,83
TR_TRI9_2.txt	6	43	2	24,16
TR_PI7_2.txt	6	41	3	24,16
TR_PI6_2.txt	7	48	3	25,14
TR_TRI6_2.txt	6	41	3	25,33