

UVIC



ACCESSIBLE
TECHNOLOGIES

In Translation and Interpreting
Vic, 30-31 March, 2006

ISBN-13: 978-84-934995-2-5 ISBN-10: 84-934995-2-8
<http://tinyurl.com/gxqvw>

iDISC
Information
Technologies

 cpsl
TECHNOLOGIES

 thera

 caixaManresa

UVIC

**Translating and Interpreting
Conference:
Accessible Technologies
30 & 31 March 2006**



[Introduction](#)

[Programme](#)

[Abstracts](#)

[Communications](#)

[Plenary Sessions](#)

[Group Session](#)

[Presenting platforms](#)

[Photo Gallery](#)

[Resources](#)

[Video streaming](#)

[Contacts](#)

iDISC
Information
Technologies



Weblog Design: Josep M. Frigola, 2006

ISBN-13: 978-84-934995-2-5 - ISBN-10: 84-934995-2-8

Duoxiu Qian

1. Duoxiu Qian (Beihang University, Beijing, China): [A Computer-aided Approach for Accuracy in Specialized Translation.](#)

Anna Matamala

2. Anna Matamala (UAB): [Les noves tecnologies en l'ensenyament de la traducció audiovisual.](#)

José Ramón Biau Gil

3. José Ramón Biau Gil (URV): [Accessible Technologies vs. Professional Tools.](#)

Presentation: 

Heidi Salaets

4. Heidi Salaets (Lessius Hogeschool, Antwerp, Belgium): [Digital Learning Environments \(DLE\) - Free or Affordable Software in Translation and Interpreting Training.](#)

Short bio: 

Ingrid Cáceres Würsig, Pilar Castaño, Claudia Künschak and Celia Rico

5. Ingrid Cáceres Würsig, Pilar Castaño, Claudia Künschak and Celia Rico (Universidad Europea de Madrid): [Materiales electrónicos para el aprendizaje de la traducción. El proyecto ELECTRA.](#)

Short bio: 

Presentation: 

Francesc Fernández

6. Francesc Fernández (UPF): [Aprendizaje cooperativo y autorregulado de la traducción general \(alemán-español\): ejemplo de un caso práctico de trabajo autónomo por proyectos.](#)

Short bio: 

Presentation: 

Anexo: 

Elena Errico and Mara Morelli

7. Elena Errico (Università di Modena e Reggio Emilia) and Mara Morelli

(Università di Cagliari): [Modalidades no presenciales y destrezas comunicativas para la interpretación.](#)

Olga Torres Hostench, A. Aguilar-Amat, M. Ferrús, B. Mesa, R. Piqué and P. Sánchez-Gijón

8. Olga Torres Hostench, A. Aguilar-Amat, M. Ferrús, B. Mesa, R. Piqué and P. Sánchez-Gijón (UAB): [La funció de les TIC aplicades a la traducció en la transformació del perfil de traductor/a al nou perfil de proveïdor de serveis de traducció.](#)

Short bio: 

María Amparo Jiménez Ivars

9. María Amparo Jiménez Ivars (UJI): [Interconectividad y docencia continua en el marco de la formación virtual del intérprete.](#)

Inge Gorm Hansen

10. Inge Gorm Hansen (Copenhagen Business School): [Video-enhanced interpreter training - Aiming for improved performance and increased motivation in a learner-driven self-study environment.](#)

Short bio: 

Christophe Declercq

11. Christophe Declercq (Imperial College London): [IT needs for a translation student.](#)

Short bio: 

Victoria Soler Puertes, Amparo Alcina Caudet and Anna Estellés Palanca

12. Victoria Soler Puertes, Amparo Alcina Caudet and Anna Estellés Palanca (UJI): [La digitalización de textos para la elaboración de un corpus lingüístico electrónico: una experiencia de trabajo en equipo con estudiantes.](#)

Short bio: 

Presentation: 

Anna Estellés Palanca, Amparo Alcina Caudet and Neus Andreu

13. Anna Estellés Palanca, Amparo Alcina Caudet and Neus Andreu (UJI): [El uso de moodle en la enseñanza virtual de la terminología.](#)

Short bio: 
Presentation: 

Patricia Rodríguez Inés and Olivia Fox Kennedy

14. Patricia Rodríguez Inés and Olivia Fox Kennedy (UAB): [AVISO: CAMBIO DE AULA... y de metodología de trabajo. Enseñanza de la traducción y digitalización.](#)

Short bio: 

Balma Forés Barberà, Vicent Montalt i Resurrecció and Maite Sánchez Safont

15. Balma Forés Barberà, Vicent Montalt i Resurrecció and Maite Sánchez Safont (UJI): [Explotación de Moodle para la enseñanza de la traducción médica a distancia: el caso de Tradmed de la Universidad Jaume I.](#)

Short bio: 

Paula Ramalho Almeida

16. Paula Ramalho Almeida (Instituto Politécnico do Porto, Portugal): [Subtitle Workshop and Xvid avi format: Freeware at the service of audiovisual translation.](#)

Manuel Mata Pastor

17. Manuel Mata Pastor (UCM): [Las ventajas de la precariedad: una experiencia en la formación de traductores en localización.](#)

Short bio: 
Presentation: 

Juan de Dios Torralba Caballero

18. Juan de Dios Torralba Caballero (UCO): [¿Cómo mejorar la docencia en Filología inglesa?](#)

Annalisa Sandrelli and Jim Hawkins

19. Annalisa Sandrelli (Università di Bologna) and Jim Hawkins (Melissi Multimedia Ltd.): [Computer Assisted Interpreter Training \(CAIT\): what is the way forward?](#)

Presentation: 
Play video clip: 

Marcos Cánovas Méndez

20. Marcos Cánovas Méndez (Universitat de Vic): [Portafolios electrónicos en una asignatura de informática aplicada a la traducción](#)

Presentation: 

Josep M. Frigola Sèculi

21. Josep M. Frigola Sèculi (Universitat de Vic): [Aplicacions didàctiques de la plataforma IRC-Catalunya.TV](#)

Presentation: 

La digitalización de textos para la elaboración de un corpus lingüístico electrónico: una experiencia de trabajo en equipo con estudiantes

Victoria Soler Puertes, Amparo Alcina Caudet y Anna Estellés Palanca
Universitat Jaume I

Resumen: En este trabajo se describe la experiencia de digitalización que se ha llevado a cabo en equipo para la creación de un corpus electrónico especializado. Se analizan los aspectos técnicos de la digitalización y se abordan las dificultades que fueron surgiendo durante el transcurso del proceso de digitalización, que se organizó en torno a un protocolo. Previamente se decidió la organización de los materiales y recursos necesarios, así como el uso del entorno virtual de trabajo colaborativo BSCW.

Terminamos con una valoración final de los beneficios que esta tarea ha supuesto para los estudiantes y el corpus resultante. En ella, analizamos el grado de aceptación y estudiamos de qué manera han progresado en el manejo de las herramientas utilizadas a lo largo de toda la tarea de digitalización.

Palabras clave: digitalización, corpus electrónico, didáctica, informática aplicada a la traducción.

1. *Introducción*

Actualmente gran parte del trabajo del traductor gira en torno al uso de herramientas y recursos electrónicos. Normalmente los clientes suelen trabajar con los textos origen y destino en formato electrónico. Sin embargo, no siempre es así y una de las tareas que deberá dominar el traductor es saber digitalizar los textos que le lleguen en formato impreso para poder sacar el máximo provecho de las herramientas informáticas.

La digitalización de textos es un proceso aparentemente fácil, pero que conlleva el manejo de diferentes herramientas informáticas, formatos de texto, distinción de imagen y texto. Durante el proceso pueden tener lugar dificultades relacionadas con el hardware, el software o el idioma de trabajo. Con la intención de formar a los estudiantes de Traducción e Interpretación en las tareas propias de la digitalización de textos, diseñamos

una experiencia que les permitiera, por un lado, aprender poco a poco enfrentándose a diferentes dificultades y por otro, que resultara motivadora y gratificante.

La experiencia que presentamos a continuación forma parte del proyecto de innovación educativa CREC: Creación de Recursos lingüísticos electrónicos, en el marco del cual también se han desarrollado otros aspectos de formación en informática aplicada a la traducción y terminología, como la lematización o la búsqueda y evaluación de referencias bibliográficas (cf. Alcina, Soler y Estellés, 2006). Por otra parte, los textos digitalizados han servido para el desarrollo del corpus TXTCeram en el marco del proyecto de investigación TXTCeram.

En este artículo, en primer lugar, exponemos los aspectos técnicos que implica la digitalización de documentos y los problemas que plantea. En segundo lugar, describimos la experiencia llevada a cabo con los estudiantes. Finalmente, valoramos los resultados de la experiencia, tanto en lo que se refiere a los estudiantes como a los recursos obtenidos.

Con todas estas tareas se pretende motivar a los estudiantes a que utilicen el ordenador y los programas de tecnologías aplicadas a la traducción fuera de las clases y, al mismo tiempo, ayudarles a superar la «barrera psicológica» que para muchos de ellos representa el uso de los ordenadores. Además, estas destrezas le serán exigidas en un futuro profesional donde se requiere un conocimiento amplio de las herramientas informáticas de ayuda al traductor porque, como afirma Arevalillo (2003: 237), el traductor debe “estar al día de los últimos programas, tecnologías y procesos de trabajo, puesto que pueden resultar cruciales al enfrentarse al mercado laboral real”.

2. La digitalización en el mundo de la traducción

En sentido amplio, el proceso de digitalización consiste en reconocer los caracteres de un documento que tenemos en formato papel para transformarlo a formato digital. De este modo, se podrán utilizar con las herramientas informáticas necesarias para permitirnos mejorar los recursos y ahorrar tiempo. Además de estas ventajas, nos encontramos ante el hecho de que vivimos en una era digital en la que, casi de manera imperceptible para muchos, tenemos cualquier tipo de tecnología al alcance de nuestra mano. Por esta razón, en el entorno de trabajo del traductor también se percibe una revolución tecnológica. En 1996, O'Hagan ya preveía este desarrollo tecnológico que, aplicado a la traducción, lo bautizó con el nombre de «teletraducción».

Teletranslation will exploit the global networking capabilities of telecommunications technology to bring language service providers (using both human and computer resources) and their costumers together. In doing so, it will help overcome many otherwise insoluble language problems. (O'Hagan, 1996: 13)

Este tipo de traducción, acuciado por la globalización, se ha convertido en el perfil del nuevo traductor profesional y, por ende, ha fomentado la práctica de la digitalización. Su empleo a lo largo de todo el proceso de traducción es mucho más común de lo que pueda parecer a primera vista.

Uno de los usos que se le puede aplicar y que resulta ser de gran ayuda para la traducción especializada es la creación de un corpus como herramienta de documentación a partir de textos paralelos (cf. Sánchez-Gijón, 2004). Para manejar toda esta documentación sobre una misma temática en formato electrónico, es necesario utilizar programas de gestión de corpus que nos servirán “para obtener datos fiables y reales sobre las frecuencias léxicas tanto de términos como de candidatos a términos, de

fraseología terminológica, de realizaciones alternativas para un mismo movimiento retórico, de preferencias morfosintácticas, etc.” (cf. Corpas, 2002: 157).

En caso de que no tengamos ambos textos en formato electrónico, la digitalización también permite, con ayuda de un programa de alineación, la posibilidad de alimentar memorias de traducción a partir de traducciones ya existentes. De este modo, no será necesario crear una memoria de traducción desde cero sino que es posible reutilizar trabajo anterior (cf. Austerlühl, 2001: 135).

Otra aplicación muy interesante para un traductor es la extracción de estadísticas del proyecto mediante una herramienta de traducción asistida. Con ella es posible analizar el número exacto de palabras, el porcentaje de repeticiones, el texto reutilizable, el número de palabras nuevas, etc. (cf. Arevalillo, 2003: 255) para la elaboración de un presupuesto.

La formación del traductor en todos estos aspectos tecnológicos está en manos de los centros de enseñanza de Traducción. Son ellos los encargados de diseñar una propuesta curricular que se adecúe a la realidad laboral y que ayude a los recién licenciados en su inserción laboral (cf. Corpas y Varela, 2003). Sin embargo, muchas veces, las universidades van a remolque de la continua evolución que sufren las tecnologías de la traducción, y es por ello que se deben intentar cubrir estas lagunas con actividades y cursos fuera del horario académico, como la presente experiencia de digitalización.

3. Aspectos técnicos del reconocimiento óptico de caracteres (OCR)

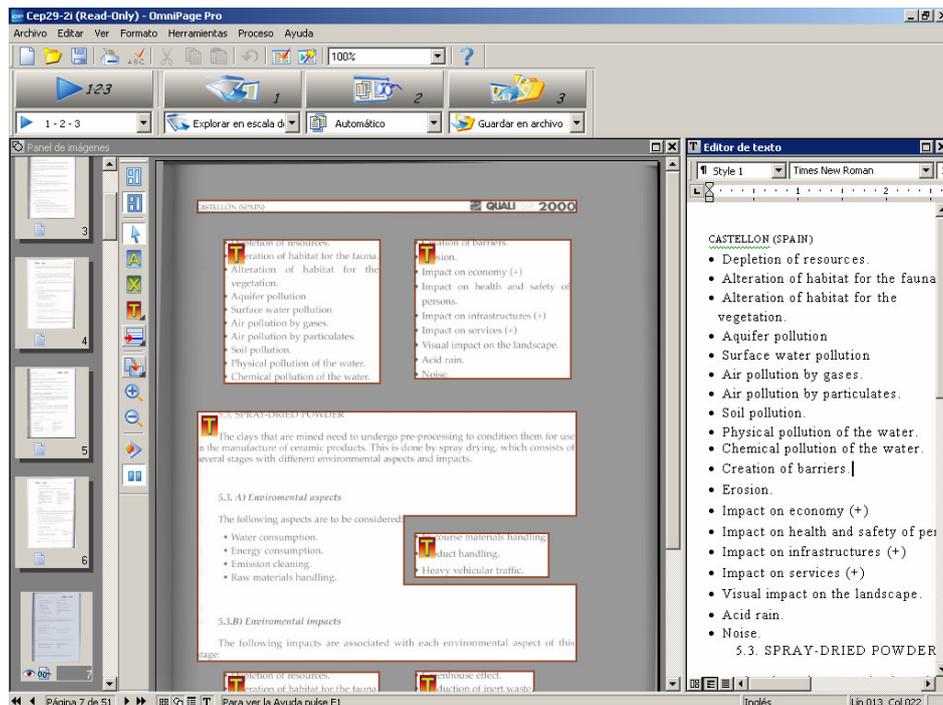
Para que podamos convertir los textos de formato papel a formato electrónico de forma que sean comprensibles para el ordenador son necesarios un escáner y un programa de reconocimiento óptico de caracteres (OCR, del inglés *Optical Character Recognition*).

El escáner es un periférico del ordenador que digitaliza de una manera rápida y sencilla imágenes y texto. Se sirve de un foco de luz y una lente que registran los puntos de información de la imagen capturada. Las características principales que los escáneres deben poseer para conseguir resultados satisfactorios son la alimentación automática de papel y una calidad suficiente para escanear imágenes. La primera permite al usuario la posibilidad de automatizar el escaneado de hasta 50 hojas sueltas con el consiguiente ahorro de tiempo. La segunda precisa, como mínimo, de una resolución de 300 ppp y un color de 8 bits en escala de grises.

El programa de OCR analiza los conjuntos de puntos o formas obtenidos como imágenes a través del escáner y los reconoce como caracteres. Finalmente convierte el texto a formato electrónico intentando mantener el diseño original de las páginas.

En el mercado existen varios programas que se pueden utilizar para este fin. Para desarrollar esta experiencia se ha hecho uso de la versión española del programa Omnipage Pro 14.0ⁱⁱ. La empresa Scansoft es la encargada de distribuir el programa y es posible adquirirlo por Internet desde su sitio web <<http://spain.scansoft.com/omnipage>>.

En la siguiente imagen se muestra su interfaz de usuario.



3.1 *Funcionamiento y configuración del programa*

Este programa de reconocimiento de caracteres, Omnipage, ofrece un manejo sencillo e intuitivo que facilita su utilización. De manera muy visual, en la interfaz podemos distinguir los tres pasos que realizará el programa en el proceso de reconocimiento.



Con el primer paso se obtienen los documentos escaneados como imágenes. Después, con el segundo se realiza el reconocimiento de caracteres de las imágenes previamente escaneadas en el paso anterior. Por último, en el tercero se exportan los resultados a un archivo de texto para poder trabajar electrónicamente con el documento.

Es de gran importancia configurar el programa según el idioma en el que se va a realizar la digitalización. Por eso, Omnipage posee una herramienta con la que el usuario

determina la lengua del texto, que ayudará al programa durante el reconocimiento de caracteres y ofrecerá sugerencias durante la corrección. Además de tener incorporados estos diccionarios por idiomas, el programa también incluye la opción de creación de un diccionario del usuario que él mismo elabora y modifica según sus necesidades. Gracias a esta característica, se podrá mejorar el reconocimiento en futuras digitalizaciones.

Por otro lado, durante el reconocimiento de caracteres, el programa detecta automáticamente los lugares en los que existe cualquier elemento e ignora aquellos que no estime pertinentes. Entonces, define el tipo de información que contienen las distintas zonas seleccionadas (texto, tabla y gráfico). No obstante, el usuario también puede modificar o dibujar de nuevo las zonas según crea conveniente.

3.2 *Dificultades en el reconocimiento óptico*

Existen algunas variables que no dependen del usuario y que determinan la obtención de resultados más satisfactorios en el reconocimiento de caracteres. A lo largo de la experiencia comprobamos que, sobre todo en los textos en español, el OCR confundía algunos caracteres y ofrecía reconocimientos como por ejemplo «cle» por «de», la letra «l» en vez de «í» o «rn» cuando debería ser una «m». De estas apreciaciones se desprende que el programa estaba más desarrollado para trabajar con el idioma inglés que con el español, incluso con textos con exactamente la misma calidad de impresión.

El tipo de impresión también condicionaba la calidad de los resultados del OCR. Es decir, utilizar una tipografía poco recargada como las sin serifas en el texto original o usar las impresiones originales en vez de recurrir a fotocopias puede facilitar el reconocimiento de los caracteres de un texto.

4. *Aplicación con los estudiantes*

La experiencia se ha realizado con los estudiantes de las asignaturas de Terminología e Informática aplicada a la Traducción de la licenciatura de Traducción e Interpretación en la Universitat Jaume I de Castellón.

En la primera etapa se reunieron la profesora y las colaboradoras para analizar todo el proceso y establecer las guías de trabajo de la tarea. Nos pusimos de acuerdo para decidir la división del material y el sistema de codificación de la documentación, organizar la entrega del material a los estudiantes, elaborar el protocolo, decidir el modo de entrega de los trabajos o establecer el procedimiento de corrección.

Determinamos que el material que se entregaría a cada estudiante consistiría en un paquete de unas 80 páginas acompañado de una ficha en la que aparecía la referencia bibliográfica, el nombre del estudiante, la fecha de entrega y el código con el que se reconoce cada paquete. Por ejemplo, el código CE001-1e corresponde al ámbito de la Cerámica y a la referencia número 1 cuyo idioma es el español.

El protocolo de digitalización de textos es una guía del proceso que debe seguir cada estudiante para poder conseguir transformar un documento de formato papel a formato digital paso a paso y de manera independiente. Este protocolo se les entregaba, junto con el material preparado en paquetes, tras las explicaciones que se realizaban con grupos reducidos de estudiantes.

Después, cada estudiante individualmente comenzaba a digitalizar su paquete en un ordenador con escáner y con el programa Omnipage configurado correctamente. Las tareas de digitalización se realizan en el laboratorio de Traducción y fuera del horario lectivo. Los pasos que se indicaban en el protocolo y que debían seguir para finalizar con

éxito el trabajo son los siguientes: 1) escaneado del paquete; 2) selección del idioma de trabajo; 3) reconocimiento de caracteres; 4) edición del tipo de zona; 5) exportación del reconocimiento a formato RTF; 6) revisión en Word del reconocimiento; 7) creación de una copia del documento en formato TXT para alimentar el corpus lingüístico; 8) entrega de los trabajos; y 9) revisión de los trabajos por parte de las colaboradoras.

Si en cualquier momento del proceso necesitaban guardar el trabajo realizado para continuar más tarde, en el servidor del laboratorio de traducción se creó una unidad de disco especialmente para el proyecto. Esta unidad era accesible desde cualquiera de los ordenadores conectados entre sí por red local.

Para continuar con el propósito de familiarizar a los estudiantes con el uso de las herramientas informáticas, se utilizó el entorno virtual BSCW como escenario de trabajo colaborativo (cf. Alcina, 2002) para la entrega de trabajos. Para ello se creó una carpeta en el entorno virtual y se invitaba a los estudiantes a participar como miembros. Dentro de esta carpeta cada uno de ellos añadía los archivos digitalizados para ser corregidos por las colaboradoras. Una vez éstas los habían revisado y se habían cerciorado de que el estudiante había corregido los fallos, les comunicaban mediante correo electrónico la obtención de su calificación extra.

En general, el manejo del programa OmniPage no resultaba problemático. En cambio, resultaba de mayor dificultad el uso de Word ya que tenían que reproducir diferentes aspectos del formato (estilos, columnas, bordes y sombreado, espaciado y sangría), tratamiento de tablas, imágenes, notas al pie y cuadros de texto.

Además, también se apreciaba un mal hábito bastante extendido entre los estudiantes. No suelen estar acostumbrados a seguir un protocolo escrito donde se detallan todos los pasos que deben seguir para llevar a cabo una tarea. A veces les resultaba más sencillo

preguntar las dudas a otras personas antes que leer con detenimiento la información que tienen a su alcance. Muchas veces, esta actitud está unida a la barrera psicológica que les produce el ordenador.

5. *Valoración*

Alrededor de 60 estudiantes han participado durante el curso 2004/2005 en esta experiencia de digitalización, cosa que demuestra que la aceptación de este tipo de actividades es elevada. Pero el mayor estímulo para continuar esta tarea los próximos años ha sido la aplicación que los estudiantes han encontrado del OCR para otras situaciones de trabajo. Muchos de ellos han utilizado el Omnipage para convertir en formato electrónico los documentos que los profesores de otras asignaturas entregaban en papel.

Además de conseguir que vean la utilidad de estas herramientas, también se ha logrado que, al final de curso, todos los estudiantes realizaran los trabajos de manera autónoma. Al principio del curso, la mayor parte requería la presencia de las colaboradoras para que les solucionara todos los problemas que se les planteaban. Sin embargo, conforme iba avanzando el curso, cada vez era menor la necesidad de ayuda hasta incluso llegar a una completa independencia.

Todo esto significa que los objetivos principales de este proyecto docente se han cumplido: adquisición de estrategias de digitalización, aumento de la confianza frente al ordenador y uso del mismo fuera del horario de clase. Por esta razón, durante este curso académico 2005/2006 también se ha continuado desarrollando este proyecto porque estamos convencidos de los beneficios que pueden aportar a los estudiantes como futuros traductores profesionales.

Por último, con un total de 35 referencias utilizadas, se ha conseguido crear un corpus en español (2.368.618 palabras), inglés (340.667 palabras) y catalán (12.426 palabras) sobre el ámbito de la Cerámica.

6. Conclusión

El reconocimiento de caracteres se encuentra estrechamente unido a la traducción en el sentido en que ambos tienen como herramienta principal de trabajo los textos y las palabras. Por este motivo, y por la creciente demanda de traductores con amplios conocimientos informáticos, es imprescindible su inclusión en algún momento de la formación de los traductores. El diseño de nuestro proceso de digitalización con estudiantes ha resultado provechoso y los protocolos como guías de trabajo se han mostrado como herramientas eficaces de aprendizaje.

7. Bibliografía

Alcina Caudet, M. Amparo, 2002. "Tutorizar trabajos terminológicos en un entorno virtual de trabajo colaborativo". *Sendeban* 13: 169-181.

Alcina Caudet, M. Amparo; Victoria Soler Puertes y Anna Estellés Palanca, 2005. "Internet como instrumento y recurso para la documentación en terminología y traducción. Hacia las plataformas de recursos electrónicos para el traductor". *La Biblioteca de Babel. Documentarse para traducir*. Ed. Dora Sales. Granada: Comares. 221-241.

Arevalillo Doval, Juan José, 2003. "Las tecnologías de la traducción: aspectos profesionales". *Entornos informáticos de la traducción profesional: las memorias de traducción*. Eds. Gloria Corpas y M. José Varela. Granada: Atrio. 237-263.

Austermühl, Frank, 2001. *Electronic Tools for Translators*. Manchester: St. Jerome.

Corpas Pastor, Gloria, 2002. "Utilización de corpus multilingües en traducción: introducción al tipologizador textual automático para textos jurídicos". *La Traducción*

científico-técnica y la terminología en la sociedad de la información. Eds. Amparo Alcina y Silvia Gomero. Castellón: Servicio de Publicaciones de la Universitat Jaime I. 155-162.

Corpas Pastor, Gloria y Varela Salinas, M. José, 2003. "Sobre la inclusión de elementos curriculares relativos a entornos informáticos en las licenciaturas de Traducción e Interpretación". *Entornos informáticos de la traducción profesional: las memorias de traducción*. Eds. Gloria Corpas y M. José Varela. Granada: Atrio. 301-328.

O'Hagan, Minako, 1996. *The coming Industry of Teletranslation*. Clevedon, Philadelphia, Adelaida: Multilingual Matters.

Sánchez-Gijón, Pilar, 2004. *L'ús de corpus en la traducció especialitzada*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Scansoft, 2003. *Omnipage Pro 14. User's guide*. Massachussets: Scansoft.

ⁱ Este trabajo ha sido realizado gracias a la financiación del proyecto TXTCeram por la Generalitat Valenciana (GV05/260) y del proyecto CREC por la Unidad de Apoyo Educativo de la Universitat Jaume I.

ⁱⁱ En un primer momento comenzamos a utilizar la versión 12.0, pero al comprobar que provocaba que los ordenadores se bloquearan y ralentizaba el trabajo de los estudiantes, decidimos actualizar la versión del programa.