



UNIVERSITAT  
JAUME·I

# Qualitat educativa i millora docent a l'Espai Europeu d'Educació Superior

Actes de la VI Jornada de millora educativa  
i V Jornada d'harmonització europea  
de la Universitat Jaume I

Editors:

Leonor Lapeña Barrachina  
Miguel Ángel Fortea Bagán

## Bloc 3.2 PLANIFICACIÓ I DISSENY DE LA DOCÈNCIA

- 3.2.1 [Creació de recursos lingüístics electrònics per a la Informàtica aplicada a la traducció i terminologia. Una experiència de treball en equip amb estudiants.](#) *Alcina Caudet, M<sup>a</sup> Amparo; i Vidal Castellet, Nuria.* Traducció i Comunicació. Universitat Jaume I
- 3.2.2 [Les fitxes de vocabulari del CAL: adaptació al marc europeu comú de referència.](#) *Codonyer Sòria, Pilar; Mulet Vilar, Irma; Ribés Falcó, Carmina; Soldevila Blasco, M. Josep; Keith Cañadas, Miquel Àngel; i Monfort Manero, M<sup>a</sup> Carmen.* Filologia i Cultures europees. Universitat Jaume I
- 3.2.3 [La interdisciplinarietat en la investigació psicopedagògica](#) *Ferrandez Berrueco, Reina; Moliner García, Odet; Altava, Vicenta; Sidro Tirado, José Juan; Aragón Portolés, Fabián; i Sales Cigues, Auxiliadora.* Educació. Universitat Jaume I
- 3.2.4 [Anàlisi de les diferents situacions didàctiques que es produeixen en el desenvolupament del Pràcticum III de la especialitat d'educació física en els centres de primària de la província de Castelló.](#) *Garcia Antolín, Roberto J.; Valls Martínez, Vicente Javier; Oliver Guasp, M<sup>a</sup> Jesús; Hernando Domingo, Carlos; Beilmunt, Hector; Alonso Salt, Manoli; i Torres Navarro, Miguel Ángel.* Educació. Universitat Jaume I
- 3.2.5 [Taller de "xanques": recreant-se a les altures. Disseny i creació de material docent.](#) *Garcia Antolín, Roberto J.; Valls Martínez, Vicente Javier; Torres Navarro, Miguel Ángel; i Viciado Selma, Saul.* Educació. Universitat Jaume I
- 3.2.6 [Anàlisi de la realitat de les festes recreatives de carrer a la província de Castelló: organització d'activitats i dinamització, per a la millora i canvi dels programes de les assignatures O39 i O51.](#) *Garcia Antolín, Roberto J.; Valls Martínez, Vicente Javier; Torres Navarro, Miguel Ángel; i Viciado Selma, Saul.* Educació. Universitat Jaume I
- 3.2.7 [Base de dades per a Orientadors Escolars-BDOE](#) *Gil Beltrán, José Manuel; Marzal Varó, Andrés; Sanz Esbrí, Jesús; Manzana Sorribes, María José; i Peris Beltrán, Rosa María.* Psicologia bàsica, clínica i psicobiologia. Universitat Jaume I
- 3.2.8 [Introducció de la metodologia digital destinada a l'ensenyament-aprenentatge i autoaprenentatge de la interpretació de llengües.](#) *Jimenez Ivars, Amparo; Blasco Mayor, M<sup>a</sup> Jesus; Jennings, Steve; Brehm Cripps, Justine; i González, Víctor.* Traducció i comunicació. Universitat Jaume I

# **CREACIÓN DE RECURSOS LINGÜÍSTICOS ELECTRÓNICOS PARA LA INFORMÁTICA APLICADA A LA TRADUCCIÓN Y TERMINOLOGÍA. UNA EXPERIENCIA DE TRABAJO EN EQUIPO CON ESTUDIANTES**

Amparo Alcina  
Victoria Soler  
Núria Vidal  
Carme Martí  
Dpt. de Traducció i Comunicació  
Universitat Jaume I

En este artículo se describe cómo se ha llevado a cabo el proyecto CREC con los estudiantes de Informática aplicada a la traducción y Terminología de la titulación de Traducción e Interpretación. Este proyecto pretende crear recursos lingüísticos electrónicos útiles para la docencia y la investigación mediante el trabajo en equipo de estudiantes, colaboradoras y profesoras.

Los estudiantes tenían la posibilidad de aumentar su calificación final de ambas asignaturas mediante la participación voluntaria en alguna de las tareas que se propusieron: digitalización de textos, búsqueda y evaluación de referencias bibliográficas, alineación de textos, lematización verbal y creación de ficheros stoplist. Además, la colaboración en este proyecto les ayuda a mejorar en el manejo de los ordenadores, aprender el funcionamiento de programas informáticos y utilizar el ordenador fuera del horario de clase.

## **1. INTRODUCCIÓN**

El proyecto “Creación de recursos lingüísticos electrónicos para la Informática aplicada a la traducción y Terminología. Una experiencia de trabajo en equipo con estudiantes” –en adelante CREC– se diseñó para la creación de diferentes recursos lingüísticos electrónicos. Estos recursos, en concreto, la creación de corpus lingüístico electrónico, listados de lemas y palabras gramaticales, memorias de traducción y listados de términos, se han elaborado mediante el trabajo coordinado de estudiantes, colaboradores y profesores.

Los destinatarios directos son los estudiantes de licenciatura de las asignaturas de Informática aplicada a la traducción y Terminología, que dispondrán de estos recursos lingüísticos electrónicos para practicar las destrezas y habilidades que forman parte de los objetivos de las asignaturas, y podrán, mediante su creación, mejorar su propia calificación.

También los estudiantes de doctorado y últimos cursos que participen como coordinadores y que tienen interés en iniciarse en la investigación obtendrán recursos a partir de los cuales será más productiva su investigación.

El proyecto ha contribuido especialmente: 1) a la práctica continuada de destrezas en el uso de las nuevas tecnologías fuera del horario de clase; 2) a introducir al estudiantado en procesos de trabajo en equipo; 3) a aumentar su confianza frente al ordenador y 4) a sensibilizar al estudiantado en la importancia de la calidad en los procesos de trabajo y en sus resultados.

## 2. OBJETIVOS

La finalidad del proyecto es mejorar la habilidad de los estudiantes de Traducción e Interpretación en el uso de las nuevas tecnologías mediante la creación de recursos lingüísticos electrónicos. Estos recursos, además, se utilizarán para la docencia y la investigación.

Los objetivos que se pretende alcanzar con esta acción son:

1. Motivar a los estudiantes de cursos actuales y futuros a mejorar las destrezas y habilidades en el uso del ordenador y concretamente de los programas de tecnologías aplicadas al lenguaje, la terminología y la traducción. Resultará más motivadora la realización de los trabajos por las siguientes acciones: 1) los recursos creados se utilizarán en las clases de las asignaturas; 2) los estudiantes que participen en el proyecto se beneficiarán de una puntuación extra por su colaboración; 3) los recursos creados se pondrán a disposición del público a través de una página Web.
2. Mejorar las expectativas de nota de los estudiantes. Si los estudiantes tienen medios para mejorar su evaluación final, haciendo que ésta no refleje sólo el trabajo final o el examen final, sino que tienen otros medios, se esforzarán más en el conjunto de la asignatura. Se ha propuesto para este curso que los estudiantes puedan obtener hasta tres puntos extra en la calificación de cada una de las asignaturas. Cada una de las tareas propuestas significará 0,5 puntos que se añadirá a su nota final.
3. Aprovechar conjuntamente los recursos electrónicos útiles para las asignaturas de Informática aplicada a la traducción y Terminología. De este modo, se contribuirá a coordinar las necesidades y la docencia de ambas asignaturas, estrechamente relacionadas por sus contenidos y destrezas.
4. Promover el uso de las nuevas tecnologías. La realización de gran parte de los trabajos de las asignaturas consiste en la realización de tareas con el ordenador, y con este proyecto pretendemos reforzar el uso del ordenador fuera del horario de las clases prácticas.
5. Crear recursos electrónicos útiles para traductores y terminólogos. Tanto nuestros estudiantes como los licenciados en traducción podrán consultar estos recursos en sus clases de traducción especializada. La creación de recursos lingüísticos electrónicos y su difusión conllevará un mayor uso de los programas tanto en estas asignaturas como en otras que también pudieran precisar del uso de estos recursos.
6. Favorecer la visibilidad de nuestra titulación y de nuestros licenciados en el ámbito nacional e internacional. La necesidad de recursos lingüísticos electrónicos es cada día más evidente debido a la continua ampliación y profundización de conocimientos en los ámbitos científico y tecnológico, y también por la creciente necesidad de comunicación multilingüe. La creación de recursos puede atraer la atención de traductores, terminólogos y profesores de otras universidades, con lo que podemos conseguir beneficios de carácter profesional/laboral para los alumnos. También las empresas con las que tenemos convenios podrán conocer el tipo de trabajos que pueden desarrollar los alumnos en prácticas o los licenciados, dado que existe una gran demanda de licenciados que

dominen las tecnologías del lenguaje, de la terminología y de la traducción.

7. Mejorar la metodología docente en el proceso de aprendizaje de los distintos programas informáticos pertenecientes al ámbito de las tecnologías del lenguaje, de la terminología y de la traducción. De este modo, se pretende que en próximos cursos las demostraciones de los programas y sus propias prácticas sean más reales, ya que contarán con una gama de recursos para su aplicación y uso.
8. Formar a una estudiante en métodos específicos de organización y desarrollo de recursos lingüísticos electrónicos (corpus electrónicos, memorias de traducción, alineación de textos).

### **3. METODOLOGÍA Y DESARROLLO**

El proyecto se desarrolló en cinco fases:

- a) Planificación y preparación
- b) Desarrollo de tareas
- c) Unificación de materiales
- d) Aplicación
- e) Difusión de la experiencia

En primer lugar, se procedió a la planificación detallada de todo el proceso y establecimiento del calendario y pautas de trabajo entre todas las participantes del proyecto. Se realizaron reuniones y consultas de manera constante. Sin embargo, el grueso del trabajo se ha realizado con los estudiantes de las asignaturas de Informática aplicada a la traducción y Terminología, que han sido los verdaderos protagonistas de este proyecto. El diálogo entre los estudiantes y todas las participantes del proyecto ha sido constante ya que, el propósito del mismo es la creación de recursos lingüísticos electrónicos de manera tal que los participantes únicamente sirvan de meros guías para que los estudiantes den forma a los productos.

A continuación explicamos con más detalle cada una de las fases.

#### **3.1. Fase de planificación y preparación**

Antes de que comenzara el curso académico 2004-2005, se empezó a organizar por una parte, cuáles eran los recursos que se necesitaba elaborar, y por otra parte el modo en el que se iban a llevar a término todas las tareas que se propondrían a los alumnos a lo largo del curso.

Los aspectos que nos propusimos dejar claros desde un principio eran los siguientes:

- Se establecieron unos horarios de atención dentro de los cuales los estudiantes han podido acudir al laboratorio de traducción donde las colaboradoras en el proyecto les han guiado y les han servido de apoyo para la realización de las tareas.
- En un primer momento se estableció que los estudiantes debían entregar los trabajos ya acabados en un plazo de diez días después de haberse comprometido a hacerlos. De este modo, nos asegurábamos de que los estudiantes no postergaran todos los trabajos para el final de cada semestre. Sin embargo, una vez comenzamos el curso nos dimos cuenta de que la mayoría de ellos, aunque no lo entregasen exactamente diez días

después, sí que solían terminar los trabajos en un plazo relativamente corto. Por eso, al poco tiempo ya no le dimos tanta importancia a las fechas de entrega sino que, únicamente establecimos una fecha tope de entrega al final de cada semestre, que coincidía con las fechas de los exámenes de Informática aplicada a la traducción y Terminología.

- Para continuar con el propósito de familiarizar a los estudiantes con el uso de las herramientas informáticas, se utilizó el entorno virtual BSCW (<http://bscw.uji.es>) como escenario de trabajo colaborativo para el intercambio de trabajos. Para ello se creó una carpeta con el nombre «CREC CREación de REcursos lingüísticos electrónicos» en el entorno virtual y se invitaba a los estudiantes a participar como miembros. Dentro de esta carpeta se crearon otras subcarpetas para cada una de las tareas en las que los estudiantes han añadido todos los trabajos que han ido realizando. En la captura de imagen siguiente podemos observar la división en subcarpetas.

The screenshot shows the BSCW web interface. At the top, there is a navigation bar with 'Archivo', 'Edición', 'Ver', 'Opciones', 'Ir a', and 'Ayuda'. Below this is a toolbar with icons for home, public, portap, papira, dirs, and agenda. The main content area shows the current location: 'Su ubicación: :al084341 / CREC Creación de REcursos lingüísticos electrónicos'. Below this is a toolbar with actions: 'visto', 'enviar', 'copiar', 'link', 'recortar', 'borrar', and 'archivar'. The main content area displays a folder structure with 8 entries. The entries are listed in a table with columns: Nombre, Size, Compartido, Nota, Calific., Propietario, Fecha, Eventos, and Acciones.

Nombre	Size	Compartido	Nota	Calific.	Propietario	Fecha	Eventos	Acciones
CREC Creación de REcursos lingüísticos electrónicos							8 entries	
Búsqueda y evaluación de textos Cerámica	5				alcina	2005-01-15 16:54		
Búsqueda y evaluación de textos Copa América	7				alcina	2005-01-03 18:43		
Corpus electrónico Cerámica	129				alcina	2005-07-12 20:01		
Corpus electrónico Copa América	10				alcina	2005-06-02 16:52		
Lematización catalán	50				alcina	2005-06-23 13:16		
Lematización español	50				alcina	2005-06-20 12:47		
Memoria de Traducción	1				alcina	2005-02-01 11:49		
Stolist catalán y español	2				al084341	2005-04-27 17:08		

- Aunque no se elaboran los protocolos de cada tarea en este momento, sí que se decide que el formato de la documentación que se les entregará a los alumnos sobre los pasos a seguir en todos el proceso será mediante protocolos. En la fase de elaboración se especificará cada uno de ellos.
- En todas las tareas se ha intentado que la carga de trabajo requerida para la obtención de 0,5 puntos fuera equivalente a 10 de horas de trabajo.
- Como uno de los temas que se va a tratar a lo largo de todo el proyecto es el de la Cerámica, se entra en contacto con el Instituto de Tecnología Cerámica (ITC) de la Universitat Jaume I con el fin de proveernos de materiales que puedan ser utilizados para la creación de recursos. Gracias a este contacto hemos podido alimentar el corpus electrónico de la Cerámica con referencias especializadas en tema y que no hemos podido encontrar en la biblioteca de la universidad.

- Todas las tareas que se han llevado a cabo finalmente a lo largo del curso han necesitado de una fase de preparación previa a su elaboración. El proceso que ha tenido lugar en todas ellas se explica detalladamente más abajo dentro de los apartados destinados a cada una de las tareas.
- Las profesoras explican en clase a los alumnos las tareas previstas en las que podían participar a lo largo de todo el curso.
- Se solicitó al Servei de Informàtica el alias [crec@uji.es](mailto:crec@uji.es) para tener informadas a las colaboradoras del proyecto. Este alias se utilizó también para las consultas de alumnos o entrega de trabajos.

### **3.2. Fase de desarrollo de tareas**

En esta fase detallaremos los pasos seguidos en el desarrollo de las tareas que se han llevado a cabo durante el proyecto. También analizaremos los problemas surgidos en algunas de ellas y la aceptación de los trabajos por parte de los estudiantes.

La metodología de trabajo que se ha seguido con todos los estudiantes durante el curso ha sido siempre la misma. Primero acudían al laboratorio de traducción para que una de las colaboradoras les explicara el trabajo en grupos reducidos de estudiantes y les entregara un paquete con la tarea. Luego, los estudiantes tenían la posibilidad de hacerlo en el laboratorio de Traducción o en sus casas. En caso de que les surgieran dudas, siempre podían acudir al laboratorio en las horas de atención a los estudiantes de las colaboradoras o realizar las consultas por correo electrónico.

El trabajo se estructuró en forma de paquetes de trabajos voluntarios. Cada trabajo voluntario del estudiante exigía el dominio de diferentes contenidos y habilidades relacionados con las asignaturas. Por cada trabajo, que se estimó en 10 horas de trabajo del estudiante medio, el estudiante obtenía 0,5 puntos. De esta manera, se conseguía que los estudiantes practicasen de manera extensiva e intensiva el uso de ciertos programas informáticos que les resultarán imprescindibles en el ejercicio de su profesión. Para cada tipo de trabajo voluntario, los estudiantes disponían de un protocolo y de los materiales necesarios para su realización. Todos los trabajos eran revisados individualmente, ya que un error en un fichero podía hacer fallar el procesamiento de la información en una fase posterior.

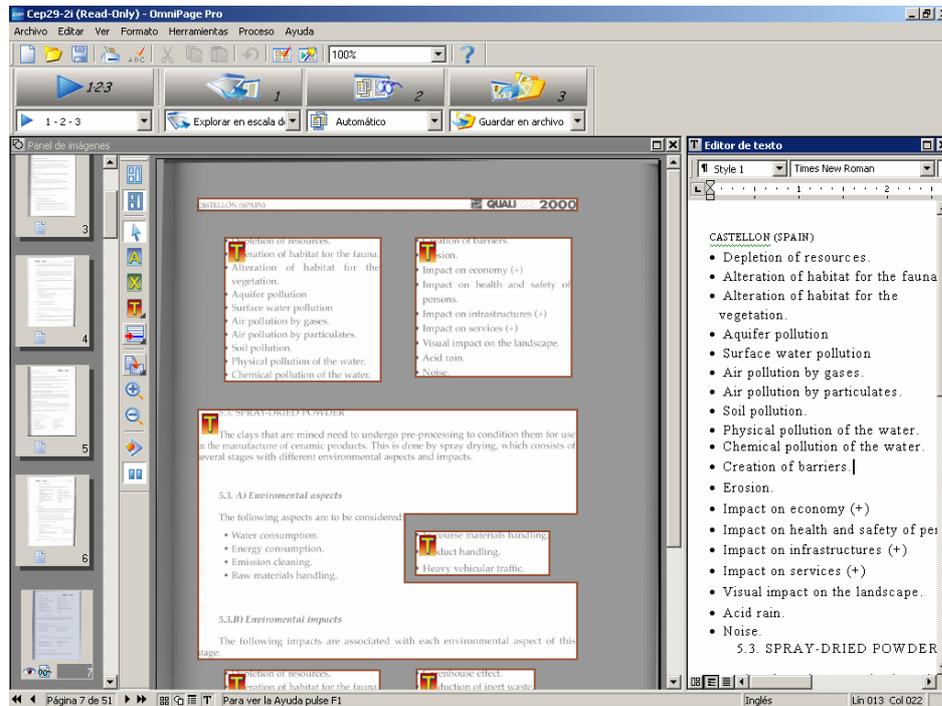
Una vez terminados los trabajos, los subían a la carpeta del BSCW preparada para cada tarea y enviaban un correo electrónico a [crec@uji.es](mailto:crec@uji.es) para avisar de la entrega de cada trabajo. Entonces, las colaboradoras procedían a revisar que el trabajo era correcto ya que, de no ser así, se les enviaba a los estudiantes un correo con las indicaciones pertinentes para su corrección.

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.

A continuación explicamos algunos aspectos individuales de cada tarea.

## A) Digitalización de textos para la creación de corpus electrónicos

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. Para ello, es necesario el uso de un escáner y de un programa de reconocimiento óptico de caracteres (OCR, del inglés Optical Character Recognition) como el Omnipage, que hemos utilizado en este proyecto y me muestra a continuación.



Se determinó que a cada estudiante se le entregaría 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 identifica cada paquete. Por ejemplo, CE001-1e corresponde al ámbito de la Cerámica (AM para las referencias de la Copa América), escrito en español (e = español, c = catalán, i = inglés) cuya referencia concuerda con el número 1. Los textos pertenecían al ámbito de la cerámica y las referencias se habían obtenido a través del Instituto de Tecnología Cerámica.

Una vez escogido el sistema con el que se van a identificar las obras, procedimos a elaborar el protocolo de digitalización que servirá de guía para los alumnos. 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
- 9) Revisión de los trabajos por parte de las colaboradoras

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, con los que no estaban acostumbrados a lidiar.

Sin duda, la digitalización es una de las tareas que más éxito ha tenido entre los estudiantes a pesar de las dificultades mencionadas. Sin embargo, una vez aprendieron como funcionaban estas opciones de Word, han seguido eligiendo esta tarea para conseguir puntuación extra.

### **B) Búsqueda y evaluación de referencias bibliográficas**

Para poder alimentar el corpus con más documentos, era necesaria la búsqueda de más referencias con las que alimentar el corpus. Por eso, aunque no estaba prevista dentro de nuestros objetivos en un primer momento, pensamos que ésta sería una buena manera de involucrar a los estudiantes también en el proceso de búsqueda de referencias para los corpus. Esta tarea, al mismo tiempo, les ayudaría a aprender a valorar la fiabilidad de los recursos que se pueden encontrar en Internet. Para la búsqueda y evaluación de referencias se les entregó a los estudiantes un protocolo y las fichas que tenían que rellenar con las que era posible evaluar la fiabilidad de las obras seleccionadas.

Las fichas de evaluación recogían la siguiente información:

- a) Datos de la referencia bibliográfica: datos necesarios para la identificación de la fuente ya sea impresa o electrónica.
- b) Datos generales: hacen referencia a la lengua en que está escrito el texto, si existen textos paralelos y si está disponible en formato electrónico.
- c) Fiabilidad y calidad de la referencia: se indagan cuestiones relacionadas con la fiabilidad del autor y del sitio.
- d) Aspectos pragmáticos y grado de especialización del texto: cuestiones relacionadas con el nivel de especialización del texto, el emisor y el receptor.
- e) Aspectos lingüísticos: se analizan las características sintácticas, morfológicas y estilísticas del texto para asegurarnos de su pertenencia a un lenguaje de especialidad y de un grado de especialización.

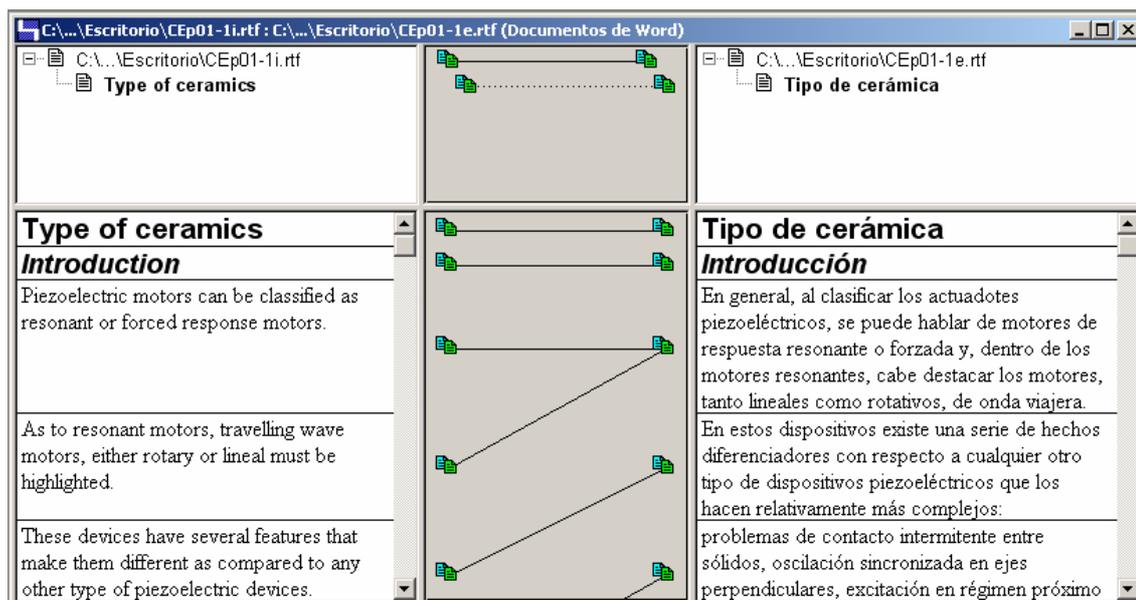
La finalidad de esta es acostumbrar al estudiante de traducción a analizar los distintos aspectos de las fuentes documentales en relación con la información que interesa desde el punto de vista de la terminología y la traducción.

No obstante, la aceptación de esta tarea por parte de los alumnos no ha sido todo lo buena que cabía esperar.

### **C) Alineación de textos**

Con la alineación de textos bilingües pretendemos alimentar una memoria de traducción incorporando textos originales y sus traducciones, de modo que se puedan utilizar en la traducción asistida interactiva con la aplicación Workbench de Trados y en la investigación sobre textos y sus traducciones.

Para realizar la tarea de alineación es necesario disponer de un texto original y su traducción, ambos en formato electrónico (por ejemplo, ficheros del tipo DOC o RTF). En el mercado existen varias herramientas de alineación de textos paralelos pero en este caso utilizaremos la aplicación WinAlign de Trados.



Con esta herramienta, conseguimos segmentar fácilmente las frases del texto origen y del texto meta, y unir cada frase del texto origen con su frase traducida del texto meta. A esas frases, el programa las denomina segmentos. La unión de un segmento origen y un segmento meta se denomina unidad de traducción. Una vez alineados los ficheros, los trabajos de cada alumno pasarán a formar parte de una memoria de traducción conjunta.

En esta tarea, se le entrega a cada estudiante un paquete con los ficheros en ambos idiomas cuya alineación final se identificará con un código del tipo MT-001, donde MT significa «memoria de traducción» y 001 el número de proyecto. Al tratarse de una actividad exclusiva para la asignatura de Informática aplicada a la traducción y al ser ésta del primer semestre, por causas ajenas al proyecto, la alineación de textos no se ha podido llevar a cabo este curso académico 2004/2005. Aún así, se comenzó a dar forma a un protocolo dedicado a esta tarea de alineación de textos que se probó con una única alumna.

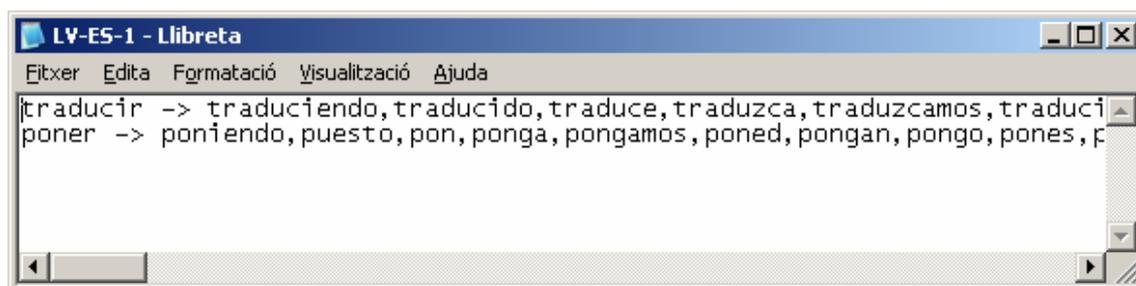
#### D) Lematización de verbos

La lematización verbal consiste en la agrupación de todas las formas conjugadas con su infinitivo. El fichero que se crea con el listado de todos los verbos en español o catalán se introduce en el programa de análisis textual WordSmith para obtener mejores resultados en el estudio de los corpus electrónicos.

Primero nos proveímos de una conjugación verbal española de la que extrajimos el listado de verbos que, mediante la digitalización, dispusimos en un documento de Word. Cada paquete de lematización consistía en la conjugación de 80 verbos irregulares o 120 verbos regulares que se disponían tal y como se muestra en la siguiente imagen.

<i>Proyecto CREC – Lematización verbos</i>				LV-ES-1	
haber	1	abreviar	3	acaudalar	3
ser	2	abreviar	3	acaudillar	3
abalar	3	abribonarse	3	accidental	3
abalear	3	abrillantar	3	accionar	3
aballestar	3	abrochar	3	acecinar	3
abanderar	3	abromar	3	acechar	3
abandonar	3	abroquelar	3	acedar	3
abañar	3	abrumar	3	aceitar	3
abaratar	3	abuchar	3	acelerar	3

Como se puede observar, cada hoja tiene un código en el lado superior derecho, por ejemplo LV-ES-1, que corresponde al trabajo de lematización verbal (LV) del español (ES) con el número uno. Además, cada verbo tiene un número a su lado que corresponde al modelo de conjugación verbal en el que se tienen que basar. A cada estudiante se le provee de una fotocopia de los modelos que necesite en su paquete. Una vez decidido el modo en el que se organizará la tarea, preparamos el protocolo para los verbos españoles. En él se explica el modo de presentación de la tarea como se expone en el ejemplo:



Por otra parte, también se les pide a los estudiantes que comprueben si todos los verbos del paquete aparecen en el diccionario de Real Academia. Esto se debe a que durante el proceso de digitalización del listado de verbos se detectaron diversos errores como “agerrir” en vez de “aguerrir”.

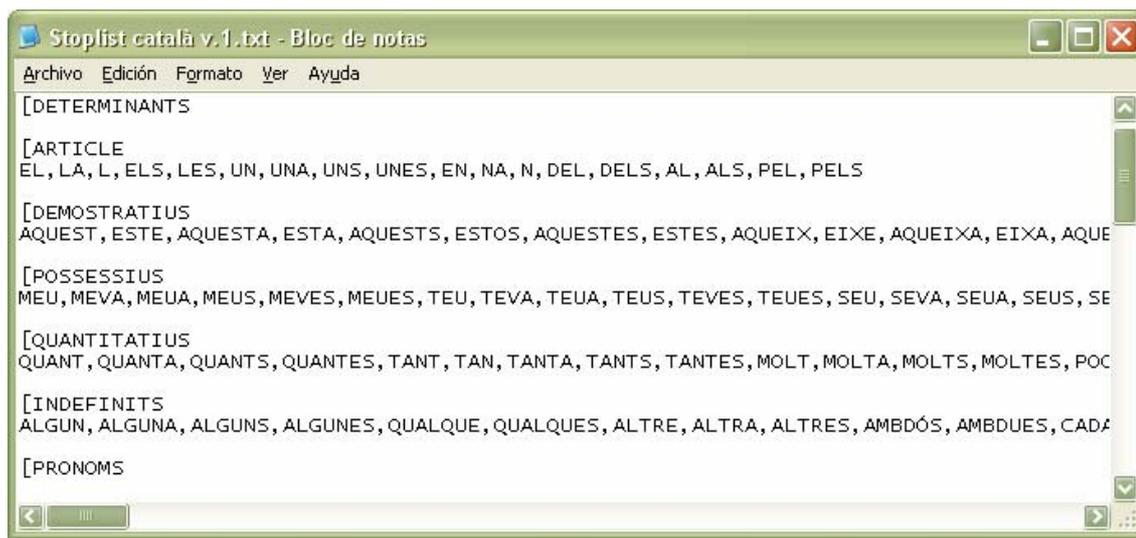
Una vez comenzamos a repartir los paquetes de lematización entre los estudiantes, comenzamos a preparar la lematización de verbos en catalán. Para ello, seguimos el mismo proceso que con los verbos españoles: digitalización y creación del listado para repartir a los estudiantes. Para la elaboración del protocolo de los verbos catalanes, nos basamos en el protocolo para el español, pero con las adaptaciones propias de los verbos en catalán. Precisamente en este momento fue cuando nos dimos cuenta de algunas deficiencias que existían en el protocolo para los verbos españoles y que tendríamos que añadir en todos los ficheros que ya habían entregado los alumnos. Eran las siguientes: 1) en los verbos reflexivos en español, el pronombre se une también al infinitivo formando una única palabra y 2) en todos los verbos, el participio también tendría que haberse conjugado en género y número y no sólo en masculino plural.

Sin lugar a dudas, ésta es la tarea que más éxito ha tenido entre los estudiantes de ambas asignaturas. Esto se demuestra con el gran número de trabajos (100 en total) realizados en tan sólo el segundo semestre.

## E) Creación de ficheros stoplist para el español y el catalán

El fichero *stoplist* para el WordSmith nos ayuda en el estudio de los corpus lingüísticos porque permite ignorar todas aquellas palabras gramaticales vacías de significado léxico en el análisis de frecuencias de uso y en la extracción de palabras clave.

Para esta actividad solamente era necesaria la ayuda de una alumna que realizara ambos ficheros *stoplist*. Para eso, se le proporcionaron dos gramáticas, una española y otra catalana, de donde extrajo la información necesaria para la creación de los ficheros. A partir del protocolo que se le entregó, el fichero ha quedado como sigue:



## F) Extracción de terminología con Extraterm

La finalidad de esta tarea es la de trabajar con la aplicación Extraterm del paquete Trados que, a partir de un corpus electrónico dado, es capaz de extraer automáticamente los términos que él mismo cree como tales dentro del área temática sobre la que versa el corpus. Sin embargo, es en la segunda parte del proceso en la que nosotros entramos en juego. Nuestra tarea será la de validar los términos que el programa nos ofrece, es decir, estudiar los resultados y decidir uno por uno si se trata de un término del campo con el que se trabaja o no.

Precisamente aquí es donde empezaron a surgir los problemas de esta tarea. Extraterm no permite que dividir el archivo con la extracción terminológica en varias partes más pequeñas para luego volverlas a unir por lo que, sería necesario trabajar todos con el mismo archivo. No obstante, ésta tampoco resulta ser una solución factible porque, las modificaciones que se le hagan a un mismo archivo desde dos ordenadores a la vez no se quedan guardadas al cerrarse, únicamente se guardarían los cambios de la última persona que cierre el archivo. Otra posibilidad podría ser exportar los resultados a Multiterm para que los estudiantes trabajasen desde esta aplicación pero el inconveniente sería que luego no se podría volver a importar a Extraterm.

A causa de todos estos obstáculos, no nos ha sido posible sacar adelante esta actividad. De todos modos, no se descarta estudiar otras posibles soluciones el curso que viene.

### 3.3. Fase de unificación de materiales

Esta fase sólo es necesaria para dos de las tareas expuestas: la lematización y la digitalización. Para la primera, no se han podido unir todos los pequeños archivos en uno para el español hasta que se terminaron de lematizar todos los paquetes. En la siguiente imagen se puede observar la apariencia de un fragmento del fichero de lemas en el español para el programa WordSmith:



Por otra parte, se ha ido almacenando en un CD, a lo largo del curso, todos los archivos digitalizados para la creación de los corpus de Cerámica y de Copa América. Por tanto, la fase de unificación del corpus se ha ido realizando progresivamente.

### 3.4. Fase de aplicación

Los resultados positivos del proyecto ya son factibles. Algunas de las colaboradoras del proyecto ya han utilizado el corpus de la Cerámica y los ficheros *stoplist* para sus investigaciones en el doctorado. Por ello, creemos que las aplicaciones de los recursos obtenidos se multiplicarán exponencialmente una vez superada la fase de difusión.

### 3.5. Fase de difusión de materiales

Aún no se han publicado en Internet los recursos elaborados pero está previsto que se publiquen próximamente en la siguiente página Web: <http://www3.uji.es/~alcina/crec.htm>

Por otra parte, se ha difundido la experiencia del presente proyecto en dos publicaciones. La primera, que versa sobre la tarea de búsqueda y evaluación de referencias bibliográficas, aparece en un libro dedicado a la documentación en traducción: ALCINA CAUDET, M. A.; V. SOLER PUERTES Y A. ESTELLÉS PALANCA (2005): «Internet como instrumento y recurso para la documentación en terminología y traducción. Hacia las plataformas de recursos electrónicos para el traductor» en DORA SALES (ed.): *La Biblioteca de Babel. Documentarse para traducir*. Granada: Comares.

La segunda, sobre nuestra experiencia en la tarea de digitalización, se ha difundido a través de una comunicación en las X Jornades de Traducció i Interpretació a Vic: *Tecnologies a l'abast*. La comunicación tiene por título «La digitalización de textos para la elaboración de un corpus lingüístico electrónico: una experiencia de trabajo en equipo con estudiantes» y sus autoras son Victoria Soler Puertes, Amparo Alcina Caudet y Anna Estellés Palanca.

## 4. VALORACIÓN DE LOS RESULTADOS

### 4.1. Recursos lingüísticos electrónicos creados

Se ha conseguido un corpus bastante amplio del ámbito de la Cerámica en español, inglés y catalán, otro más pequeño de la Copa América en español, los dos ficheros stoplist para el español y el catalán, el fichero de lematización verbal en español y la mitad del fichero de lematización verbal en catalán.

A pesar de que nos habíamos propuesto alcanzar una mayor variedad de recursos lingüísticos, no nos ha sido posible conseguirlos todos debido a la cantidad de horas que se necesitaban para coordinar a todos los estudiantes de 3r. curso que se han prestado voluntarios para colaborar en el proyecto. Esta demanda de trabajos era tan grande que el tiempo de dedicación de los colaboradores en el proyecto se destinaba únicamente a la explicación de las tareas que se pudieron preparar a principio de curso, la entrega de las mismas a los estudiantes y su posterior corrección.

Como la mayoría de los protocolos ya se han generado durante este curso, durante el curso que viene ya no sería necesaria su elaboración sino que, simplemente haría falta una persona que guiase a los estudiantes, que se hiciese cargo de las correcciones de los trabajos.

Los productos que hemos obtenido gracias a este proyecto y los estudiantes que en él han participado son los siguientes:

- Corpus lingüístico electrónico sobre Cerámica en inglés (340.667 palabras), español (2.368.618 palabras) y catalán (12.426 palabras) a partir de 35 referencias distintas.
- Corpus lingüístico electrónico sobre Copa América en español con 275.276 palabras a partir de 3 referencias.
- Fichero stoplist para el español (306 palabras) y fichero stoplist para el catalán (406 palabras).
- Fichero de lematización verbal para el español (5434 verbos, completo)
- Fichero de lematización verbal para el catalán (5500 verbos, en proceso).
- Once proyectos de alineación para la alimentación de una memoria de traducción del ámbito de la cerámica.

Por otra parte, también se han elaborado los siguientes protocolos:

- Protocolo de búsqueda y evaluación de textos (Alcina Caudet y otros, 2005)
- Protocolo de digitalización de textos en español e inglés (Soler Puertes y otros, 2005)
- Protocolo de lematización de verbos del español y catalán
- Protocolo de alineación de textos en dos idiomas

## **4.2. Experiencia con los estudiantes**

El desarrollo de habilidades y destrezas en el manejo del ordenador está directamente relacionado con la cantidad de tiempo que el estudiante se dedica a usarlo, y también con la variedad de actividades para las que lo utiliza. En este sentido, entendemos que el proyecto ha animado a los estudiantes a pasar más horas expuestos al ordenador, al uso de funciones de tipo variado, a enfrentarse a problemas que debe resolver, etc. Especialmente en el caso de la asignatura Informática aplicada a la traducción, la evaluación realizada hasta el presente curso, consistente en la realización de trabajos, se había revelado insuficiente ya que en ocasiones un mismo trabajo podría ser presentado por distintos alumnos. A través de este proyecto, nos aseguramos de que el estudiante ha dedicado un cierto número de horas y, por tanto, se habrá enfrentado inevitablemente a una serie de cuestiones que habrá tenido que resolver.

Por último, cabe señalar la falta de costumbre por parte de los estudiantes de seguir un protocolo escrito donde se detallen todos los pasos que deben seguir para llevar a cabo una tarea. A veces les resultaba más sencillo preguntar las dudas a las colaboradoras antes que leer con detenimiento la información que tenían a su alcance. Muchas veces, esta actitud está unida a la falta de confianza frente al ordenador.

La cooperación de los estudiantes en el proyecto ha sido mayor de la esperada. La idea de poder mejorar su calificación en ambas asignaturas a través de trabajos voluntarios les ha parecido interesante. Prueba de ello son los 60 estudiantes (con uno o varios trabajos voluntarios) que han participado en el proyecto y la cantidad de resultados obtenidos gracias a ellos. Se ha trabajado en un total de 141 paquetes de digitalización, 100 paquetes de lematización, 1 paquete de búsqueda y evaluación, 1 paquete de stoplist y 1 paquete de alineación.

Los resultados en la nota de los estudiantes son también evidentes, ya que la relación entre trabajo realizado y aumento de nota es directo. Hay que observar, no obstante, que algunos trabajos no eran correctos y tuvieron que ser devueltos al estudiante para que fueran subsanados, pero todos (salvo en un caso) lo hicieron.

## **5. PERSPECTIVAS DE FUTURO**

En estos momentos, ya se ha podido comprobar la utilidad del corpus de la Cerámica, del fichero de lematización verbal y del fichero stoplist, de los que se ha hecho uso de ellos con fines investigadores y también en las clases de las asignaturas implicadas.

El proyecto debería continuarse para seguir alimentando los corpus de la Cerámica y Copa América y continuar creando ficheros de lematización de, por ejemplo, sustantivos y adjetivos. Pero, sobre todo, para crear algunos de los productos que nos proponíamos en un principio y que, por razones técnicas, no se han podido elaborar. De ahí que se haya propuesto un segundo proyecto, CREC II.

En general, de cara al estudiantado, haría falta estudiar la diferente distribución de preferencias de los estudiantes en los distintos tipos de tareas y fomentar un mayor equilibrio, ya que los distintos paquetes de tareas enfatizan diferentes tipos de habilidades y destrezas.

## **Bibliografía**

- Alcina Caudet, M.A.; Soler Puertes, V. and Estellés Palanca, A. (2005) Internet como instrumento para la documentación en terminología y traducción . Hacia las plataformas de recursos electrónicos para el traductor especializado. In D. Sales (ed.) *La Biblioteca de Babel: Documentarse para traducir*. Granada: Comares.
- Soler Puertes, V.; Alcina Caudet, M.A. and Estellés Palanca, A. (2005) El trabajo en equipo con estudiantes para la elaboración de un corpus lingüístico electrónico. In *X Jornades de Traducció i Interpretació a Vic: Tecnologies a l'abast*, Vic.