

*SISTEMAS DE BASES DE DATOS
PARA TRADUCTORES*

Ángel Francisco Zazo Rodríguez
Universidad de Salamanca

FUENTES DE INFORMACIÓN Y TÉCNICAS
DOCUMENTALES APLICADAS A LA
TRADUCCIÓN

Facultad de Traducción e Interpretación
Universidad de Valladolid

Soria, Septiembre de 1997

Introducción

Un sistema de base de datos consiste en un conjunto de datos relacionados entre sí, y un conjunto de programas para tener accesos a esos datos. Esta definición básica ha ido evolucionando a través de los años, y ciertamente, puede encerrar un sin fin de aplicaciones informáticas que no se corresponden con la mejor idea de bases de datos. Una definición más correcta podría ser¹:

"Colección o depósito de datos integrados, con redundancia controlada y con una estructura que refleje fielmente las interrelaciones y restricciones existentes en el mundo real; los datos, que han de ser compartidos por diferentes usuarios y aplicaciones, deben mantenerse independientes de éstas, y su definición y descripción, únicas para cada tipo de datos, han de estar almacenadas junto a los mismos. Los procedimientos de actualización y recuperación, comunes y bien determinados, habrán de ser capaces de conservar la integridad, seguridad y confidencialidad del conjunto de los datos."

Las aplicaciones y procedimientos deben ser capaces de guardar y recuperar información de las bases de datos de forma conveniente y eficiente. En general, los sistemas de bases de datos manejan gran cantidad de información. El sistema de base de datos incluye tanto la definición de la estructura para el almacenamiento de la información, como los mecanismos para su manejo.

En este sentido debemos distinguir lo que se denomina **sistema de gestión de base de datos** (SGBD), de los ficheros donde se almacenan los datos. Un SGBD es un paquete software, generalmente cerrado y comercial, que consiste en un conjunto de programas, procedimientos, lenguajes, etc., que suministran los medios necesarios para describir, recuperar y manipular los datos almacenados en la base, manteniendo su integridad, confidencialidad y seguridad. El SGBD establece la estructura de los ficheros de datos, y se encarga de almacenar la información en ellos.

El SGBD pretende ofrecer a los usuarios una visión abstracta de la información, de modo que se le oculten detalles de cómo son llevadas a cabo ciertas operaciones: de almacenamiento, mantenimiento, estructuras y algoritmos de búsqueda.

Las bases de datos cuentan con un lenguaje de búsqueda o consulta, formado por una serie de instrucciones que el usuario utiliza para extraer de la base de datos la información que necesita. Este lenguaje suele estar formado por comandos y, a menudo, permite la utilización del álgebra booleana.

¹ De Miguel, A. y M. Piattini: *Concepción y diseño de bases de datos*. Madrid, RA-MA, 1993. p. 46

Bases de datos y terminología

La terminología es esencial dentro del trabajo del traductor, sobre todo si éste está dedicado a traducciones técnicas. El lenguaje técnico es un lenguaje en pleno desarrollo, pues los nuevos productos, conceptos y tecnologías exigen ser definidos por algún término. El crecimiento de la terminología no se detiene, entramos en unos índices porcentuales que, para una lengua desarrollada, se convierten en una progresión geométrica.

La terminología es, en cierto modo, el elemento definitorio e individualizador de la traducción técnica. Para la traducción de un texto técnico es impensable no recurrir a diccionarios especializados y bases de datos terminológicas que resuelvan las dudas al traductor, pues aunque éste sea un experto en un tema determinado, es muy difícil dominar completamente un lenguaje que no cesa de renovarse. Las necesidades actuales de la terminología, en creación, mantenimiento, información y traducción, superan en mucho las posibilidades físicas de los terminólogos, y es imprescindible recurrir al ordenador para solucionar un amplio conjunto de tareas.

Las bases de datos terminológicas son aquellas que contienen términos, definiciones, traducciones, y cualquier otra información descriptiva de los términos. La elaboración de glosarios, diccionarios y nomenclaturas en soporte electrónico, y el desarrollo de cada vez más sencillos métodos de consulta han favorecido enormemente el acceso a la información específica por cualquier usuario interesado, por lo que, además, no precisa de conocimientos informáticos avanzados. Los sistemas de gestión de bases de datos terminológicas (SGBDT) actuales favorecen la consulta ágil de la información y la actualización inmediata de los datos.

La estructura básica de este tipo de bases de datos contienen registros con los siguientes campos (son los campos, efectivamente, de una ficha terminológica tradicional):

- Término
- Referencia de la fuente
- Área temática
- Marca de ponderación
- Categoría gramatical
- Definición del término
- Referencia de la definición
- Contexto
- Referencia del contexto
- Remisión interlingüística (a varios lenguas)
- Referencia de la remisión
- Tipo de remisión
- Remisión intralingüística
- Fecha
- Autor
- Notas

Existe un gran número de SGBDT. Los hay tanto para PC (la mayoría ya diseñados para el entorno Windows) como para Macintosh o para plataformas Unix. En la revista *Terminómetro*, nº 16, de octubre de 1994 se tiene una lista exhaustiva de programas de gestión terminológica. Aunque ya han pasado tres años, la mayoría de ellos siguen funcionando. No vamos a citar a ninguno por no quitar crédito a otros. No obstante, entre las características fundamentales podemos señalar que casi la mitad de los programas han sido subvencionados por instituciones gubernamentales, universitarias o explícitamente dedicadas a la terminología.

Muchos de los programas de gestión terminológica permiten la creación de glosarios o diccionarios electrónicos a partir de los datos incorporados en el sistema. La mayoría además trabaja con al menos dos lenguas (una de ellas casi siempre el inglés), pero es normal encontrarse con tres o más. Casi todos los programas incluyen los campos antes reseñados, e incluso algunas permiten incorporar imágenes a la ficha terminológica. Algunos también integran relaciones hipertextuales, que posibilitan pasar rápidamente de una ficha a otra seleccionando la palabra que corresponde con un término. Lo normal es encontrarse con sistemas diseñados para funcionamiento monousuario, y sólo unos pocos pueden funcionar en red.

Ejemplos de bases de datos

MULTITERM

MultiTerm realmente es un SGBD textual en texto libre, pero principalmente es utilizado como un sistema de gestión de bases de datos terminológicas. Las entradas en la base de datos se realiza en texto libre. Cada entrada es un concepto, y se pueden definir traducciones para 20 lenguas distintas. Además se pueden realizar búsquedas en cualquier dirección. Dispone de varios campos para información descriptiva de cada término (definición, contexto, ejemplo y uso), y otros para información adicional (fuente o área temática del término, usuarios, etc.)

MultiTerm está configurado para funcionar hipertextualmente, y soporta información gráfica (ficheros de imágenes) en la ficha terminológica. Permite asimismo realizar referencias cruzadas, y acceder a las mismas simplemente pinchando con el ratón.

Por último, una de sus buenas características es que soporta funcionamiento en red, con restricciones a usuarios, y con una buena integridad de la base de datos.

MultiTerm está distribuido por TRADOS GmbH, Stuttgart, Alemania. Se puede obtener una versión demo del programa accediendo a <http://www.trados.com>.

En la figura 1 se presenta un ejemplo de la ficha terminológica del término "icono de aplicación". La figura 2 muestra los criterios de visualización y formateado de los campos de las entradas a la base de datos. La figura 3 muestra la pantalla de edición para modificar la información introducida en la base de datos.

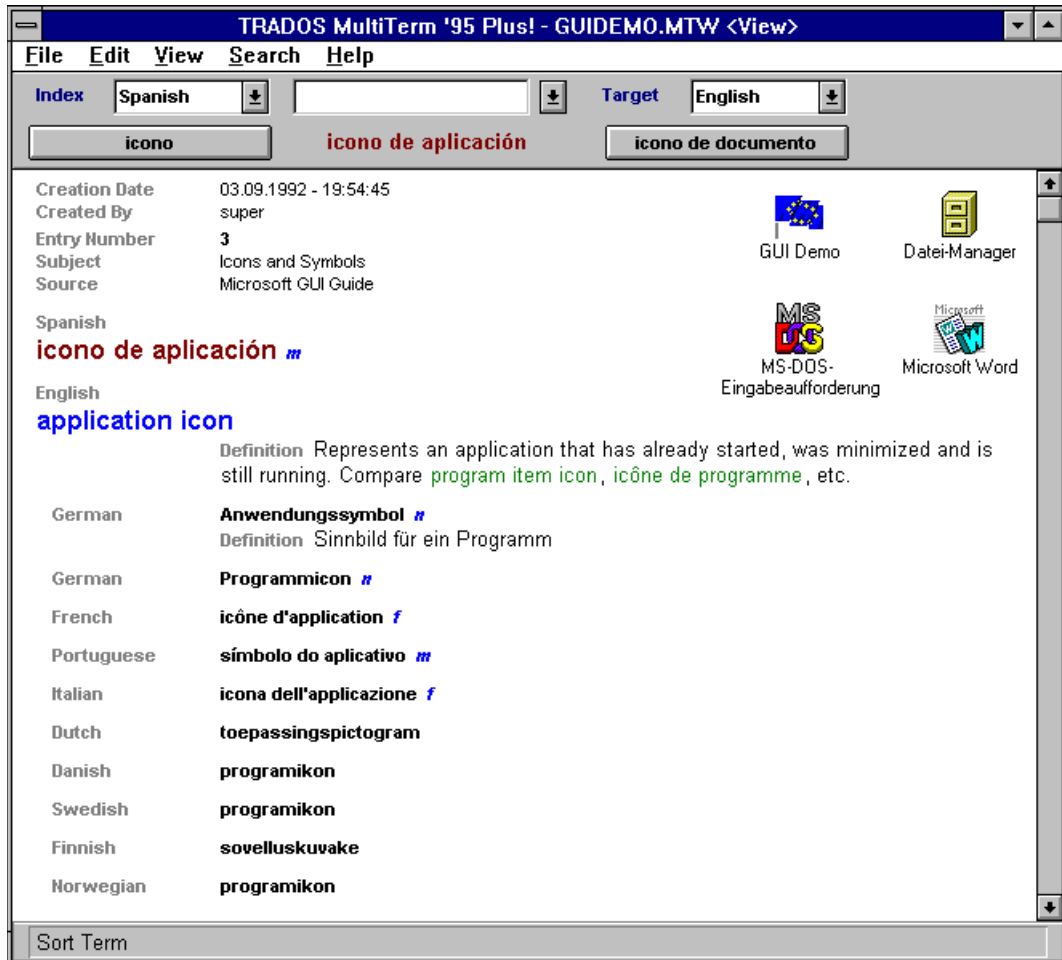


Figura 1. Ficha terminológica de MultiTerm.

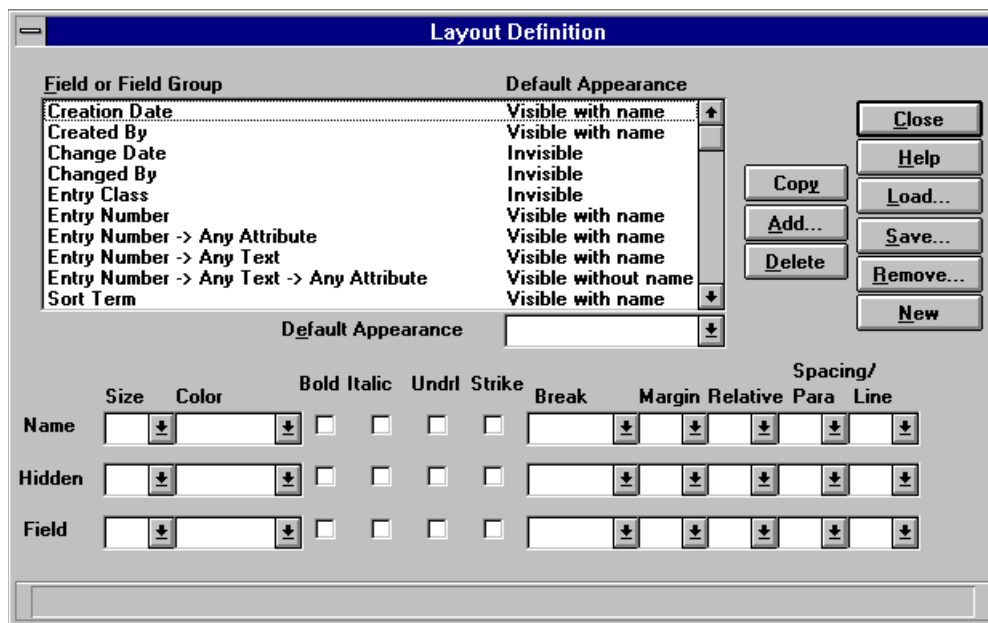


Figura 2. Visualización de campos.

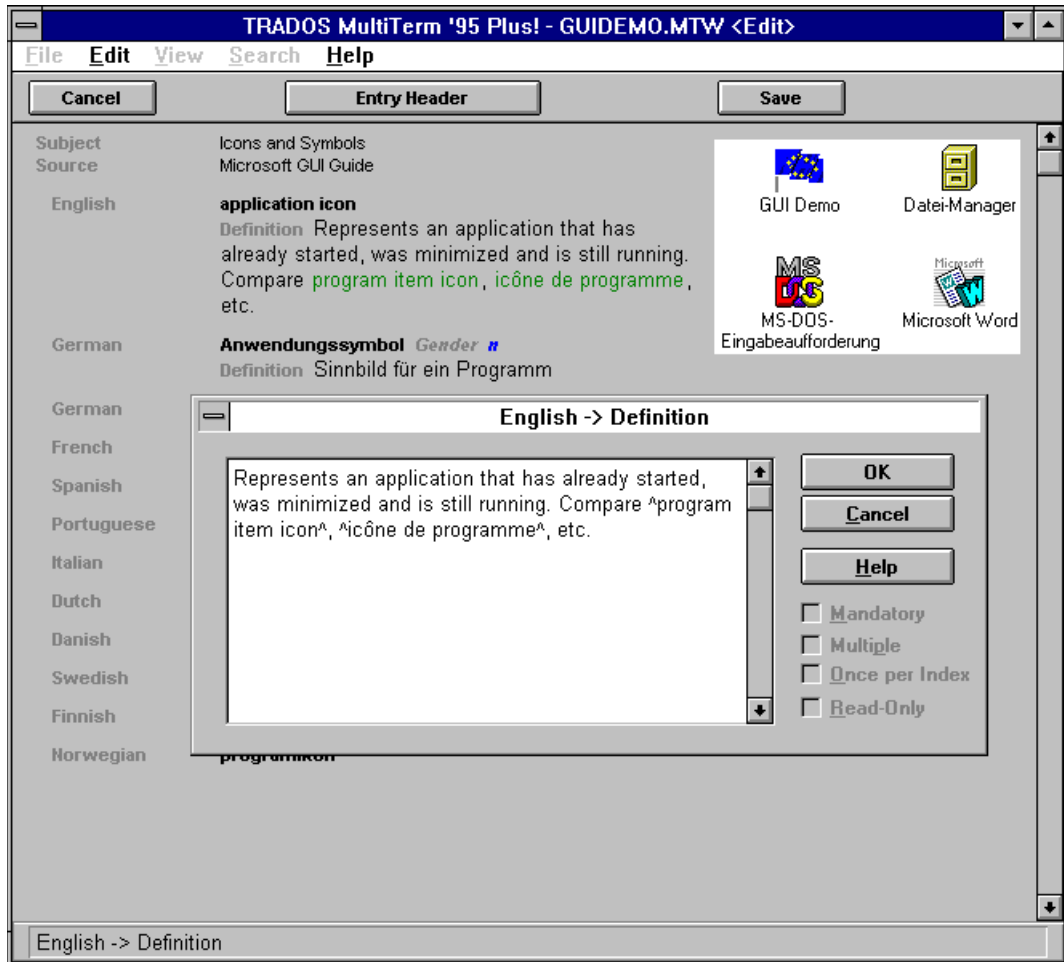


Figura 3. Modificación de la entrada "icono de aplicación".

TERM.MDB

La base de datos TERM.MDB ha sido creada para iniciar a los alumnos de la Licenciatura en Traducción e Interpretación de la Universidad de Salamanca (asignatura Informática Aplicada a la Traducción) en la creación de bases de datos terminológicas. Está diseñada con una estructura plana, utilizando MS-Access, aunque no se aprovechan las características relacionales de este SGBD. Cada registro consta de los campos indicados en la figura siguiente:

Tabla: Términos		
Nombre de campo	Tipo de datos	Descripción
Entrada	Texto	Término
Referencia	Texto	Abreviatura de la fuente
AT	Texto	Abreviatura de Área Temática
MP	Texto	Marca de ponderación
CG	Texto	Categoría Gramatical
Definicion	Memo	Definición del término
RF-Def	Texto	Abr. de la Ref. de Definición
CO1	Memo	Contexto real 1
RF-CO1	Texto	Abr. de la Ref. de Contexto real 1
CO2	Memo	Contexto real 2
RF-CO2	Texto	Abr. de la Ref. de Contexto real 2
CO3	Memo	Contexto real 3
RF-CO3	Texto	Abr. de la Ref. de Contexto real 3
Frances	Texto	Remisión Interlingüística Frances
RF-frances	Texto	Abr. de Ref. Francés
Ingles	Texto	Remisión Interlingüística Inglés
RF-ingles	Texto	Abr. de Ref. Inglés
Aleman	Texto	Remisión Interlingüística Alemán
RF-aleman	Texto	Abr. de Ref. Alemán
Otro	Texto	Remisión Interlingüística en otro idioma: (idioma) remisión
RF-otro	Texto	Abr. de Ref. Otra
IN1	Texto	Tipo de Remisión Intralingüística 1
REM-IN1	Texto	Remisión Intralingüística 1
IN2	Texto	Tipo de Remisión Intralingüística 2
REM-IN2	Texto	Remisión Intralingüística 2
IN3	Texto	Tipo de Remisión Intralingüística 3
REM-IN3	Texto	Remisión Intralingüística 3
IN4	Texto	Tipo de Remisión Intralingüística 4
REM-IN4	Texto	Remisión Intralingüística 4
FECHA	Fecha/Hora	Fecha (dd/mm/aa)
AUTOR	Texto	Autor o Grupo
NOTA	Memo	Notas
NOTATRAB	Memo	Notas de trabajo
ILUSTRACION	Objeto OLE	Ilustración

Propiedades del campo	
Tamaño del campo	75
Formato	
Máscara de entrada	
Título	
Valor predeterminado	
Regla de validación	
Texto de validación	
Requerido	No
Permitir longitud cero	No
Indexado	Sí (Sin duplicados)

La descripción del campo es opcional. Le ayuda a describir el campo y también se presenta en la barra de estado cuando selecciona este campo en un formulario. Presione F1 para buscar ayuda sobre descripciones.

Figura 4. Base de datos TERM.MDB

La estructura de la base de datos no está optimizada, pues para la mayoría de los registros, muchos de los campos están vacíos. No obstante es una buena forma de iniciar a los estudiantes en el mundo del diseño de bases de datos terminológicas.

La entrada de datos se realiza con el formulario que se presenta a continuación (fig. 5). Los alumnos introducen el término y la información pertinente.

Entrada de términos	
Entrada: <input type="text" value="canon"/>	AT: <input type="text" value="griego"/> MP: <input type="text" value="LEA"/>
Referenc.: <input type="text" value="HAA"/>	CG: <input type="text" value="m"/>
Definición: Norma que establece las relaciones armónicas entre las diversas partes de una obra, tomando como referencia una determinada unidad de medida o módulo.	RF-Def: <input type="text" value="MBB"/>
CO1: "El tamaño de las columnas responde a una proporción o canon, tomando como unidad de medida o módulo, el diámetro del fuste en su base, de igual manera que en la	RF-CO1: <input type="text" value="HAA p. 67"/>
CO2: "Se otorga en la escultura griega la absoluta primacía a la representación del cuerpo humano [...]. Existe, en todo caso, un canon cuyo módulo suele ser la cabeza, y un ritmo	RF-CO2: <input type="text" value="HAA p. 70"/>
CO3: " Con Lisipo se alcanza la estilización, al alargar el canon de las esculturas, según vemos en el Apoxiomenos..."	RF-CO3: <input type="text" value="HAA p.73"/>
Frances: <input type="text" value="canon (m)"/>	RF-frances: <input type="text" value="RB"/>
Ingles: <input type="text" value="canon"/>	RF-ingles: <input type="text" value="OX"/>
Aleman: <input type="text" value="das Kanon"/>	RF-aleman: <input type="text" value="BH"/>
Otro: <input type="text"/>	RF-otro: <input type="text"/>
IN1: <input type="text" value="sin."/>	REM-IN1: <input type="text" value="proporción"/>
IN2: <input type="text"/>	REM-IN2: <input type="text"/>
IN3: <input type="text"/>	REM-IN3: <input type="text"/>
IN4: <input type="text"/>	REM-IN4: <input type="text"/>
FECHA: <input type="text" value="1/05/97"/>	AUTOR: <input type="text" value="MBB"/>
NOTA: <input type="text" value="Etimología: gr. Kanôn"/>	
NOTATRAB: <input type="text"/>	
ILUSTRAC.: <input type="text"/>	
<input type="button" value="Inicio"/> <input type="button" value="Anterior"/> Registro: 11 de 89 <input type="button" value="Siguiente"/> <input type="button" value="Fin"/>	

Figura 5. Formulario de entrada de datos.

Por último, se presenta también la salida de un informe que produce un glosario con los datos de la base.

Entrada	Definición	Frases	Inglés	Alemán	Otro
canon	Norma que establece las relaciones armónicas entre las diversas partes de una obra, tomando como referencia una determinada unidad de medida o módulo.	canon (m)	canon	das Kanon	
capitel	Parte superior de la columna sobre la que descansan el arquiteabo, decorada según el orden arquitectónico al que pertenece.	chapiteau (m)	capital	das Kapitäl	
capitel compuesto	Capitel con ábaco curvo y fino, volutas y hojas de acanto.	chapiteau composite (m)	composite capital		
capitel corintio	Capitel formado por dos cueros con hojas de acanto superpuestas y volutas en los ángulos.	chapiteau corinthien (m)	Corinthian capital		
capitel dórico	Capitel formado por ábaco liso de forma cuadrada y equino circular.	chapiteau dorique (m)	Doric capital		
capitel jónico	Capitel rectangular con dos volutas, que se une al fuste por medio de molduras circulares.	chapiteau ionique (m)	Ionian capital		

Figura 6. Glosario de términos.

FORMANTE.MDB

También es una base de datos creada con MS-Access. En este caso, sí aprovecha las características relacionales de este SGBD. La base de datos se enmarca en un trabajo terminológico desarrollado por los Becarios de Colaboración del Departamento de Traducción de la Universidad de Salamanca, a los que hemos dado nuestro apoyo.

El sistema consta de 6 tablas, y se han diseñado las diferentes relaciones para crear una base de datos de formantes.

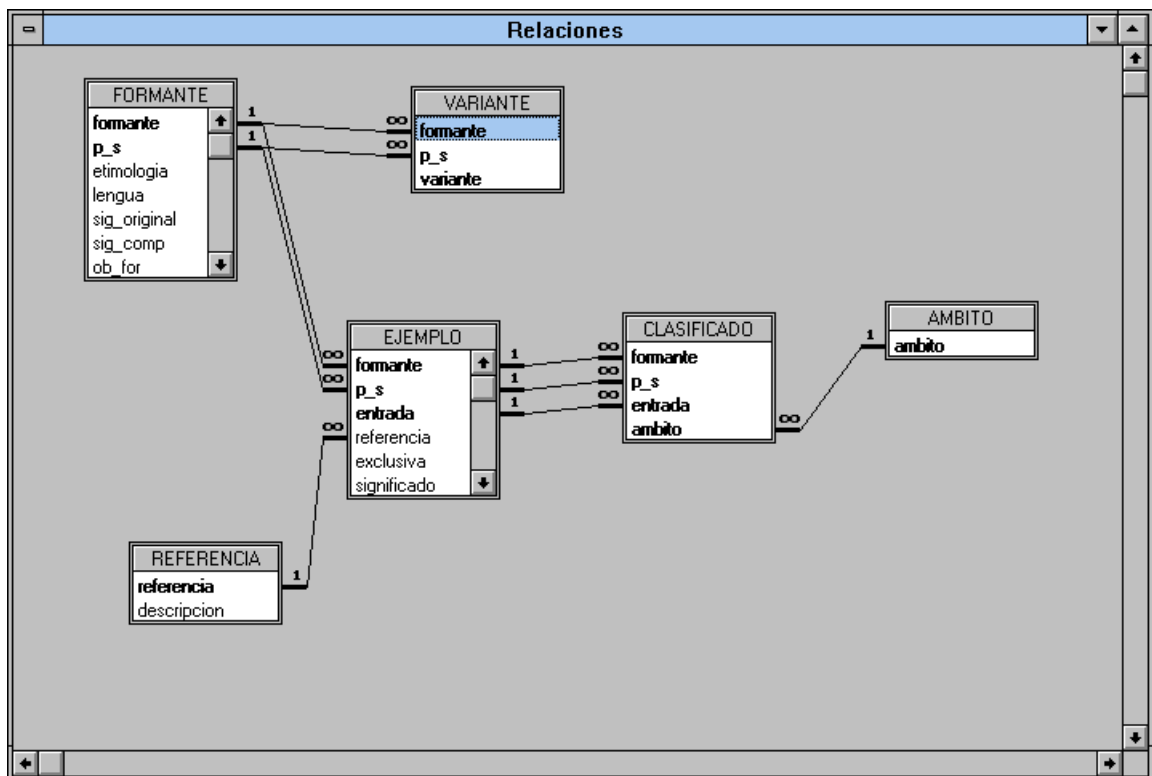


Figura 7. Relaciones en la base de datos FORMANTE.MDB

Esta base de datos se ha diseñado con la capacidad de disponer de una entrada de formante, que puede disponer de varias variantes, y cada uno de ellos con varios ejemplos. Se controla el ámbito (área temática) y las referencias como ejercicio de control de autoridades, para no tener dos ámbitos que designen el mismo área temática.

En la figura 8 se muestra el formulario de entrada de datos. Éste se ha dividido en dos partes, uno para la información de los datos relativos al formante (con sus variantes), y la otra para la entrada de ejemplos. Se dispone de dos botones para introducir nuevas referencias a la base de datos, y nuevos ámbitos.

INTRODUCCION DE FORMANTES

FORMANTE: P/S: variante:

Etimologia: Lengua:

Sig. original:

Sig. en compuestos:

OBSERVACIONES GENERALES

Formales:

De signif.:

Otras:

AUTOR/A: FECHA:

Ejemplos para:

Entrada: Ref.

▶ **Ámbito** ¿Exclusiva de un ámbito?:

Registro: 1 de 1

Significado:

Observac.:

Autor/a: **Fecha:**

▶ **Ámbito** **Referencia**

Registro: 1 de 1

Figura 8. Formulario de entrada.

Bases de datos terminológicas en línea

Hoy en día casi todos los profesionales de la traducción hacen uso de diferentes herramientas informáticas que le ayudan en su trabajo, entre las que podemos citar procesadores de texto, diccionarios electrónicos mono o multilingües, correctores ortográficos y gramaticales, bases de datos terminológicas, programas de reconocimiento de caracteres o de voz, programas de traducción automática o asistida, etc. A todos estos elementos hay que incluir una de las herramientas informáticas que más está cambiando la forma de trabajo en los últimos años: el acceso a diferentes recursos mediante redes teleinformáticas.

En muchos casos, la rapidez informativa que demanda la sociedad actual sólo puede solventarse con el acceso a recursos por medio de las telecomunicaciones. De hecho, en un mundo de vertiginosos avances tecnológicos, el trabajo de muchos profesionales de la traducción depende directamente de la celeridad con la que pueden obtener información terminológica actualizada.

La velocidad que proporcionan las telecomunicaciones, unido a las herramientas computacionales antes indicadas, y conociendo que la terminología es esencial para el trabajo de muchas traducciones científico-técnicas, todo ello provoca que los bancos de datos terminológicos en línea sean una eficaz fuente de información actualizada, en una escala que no es alcanzable con sistemas locales. Pero además, se brinda la oportunidad de disponer de la fuente misma del término, en el contexto en el que ha sido establecido. El profesional puede acceder a bases de datos bibliográficas donde consultar referencias de artículos de revistas, ponencias a congresos, patentes, monografías, y literatura gris de difícil acceso por métodos tradicionales. El empleo de bases de datos bibliográficas es particularmente útil para resolver problemas terminológicos.

Podemos destacar una serie de actividades que los traductores y profesionales de la lengua pueden realizar utilizando lo que se denomina recuperación de información en bases de datos remotas:

- Búsqueda de documentación técnica muy actualizada, tanto en la lengua origen como destino. Esta documentación puede servir al traductor para familiarizarse con un campo del conocimiento antes del trabajo de traducción, o para extraer nueva terminología. El punto de vista tradicional no permite acceder con la velocidad suficiente a estas fuentes.
- Búsqueda de terminología específica en la lengua destino. Se corresponde con una búsqueda en diccionarios, en la que la traducción correcta depende del contexto. Los diccionarios en línea contienen información muy actualizada, de la que en general carecen los diccionarios impresos en papel.
- Verificación de traducciones, para obtener mayor fiabilidad.
- Compilación de terminología para preparar glosarios, tesauros, etc.

En esta exposición nos vamos a centrar en el acceso a diccionarios y bases de datos terminológicas, cuyo exponente principal es EURODICAUTOM.

EURODICAUTOM

Es una de las primeras bases de datos distribuidas por ECHO y una de las más conocidas. EURODICAUTOM es una base de datos multilingüe que recoge toda la terminología y abreviaturas utilizadas en las diversas áreas de actividad de la Comisión Europea. La base de datos está producida por el Servicio de Traducción de la Comisión y se actualiza constantemente, tanto de los traductores, como de glosarios externos. EURODICAUTOM contiene términos, frases contextuales y abreviaturas en todas las lenguas oficiales de la Unión Europea (Danés, Holandés, Inglés, Francés, Alemán, Griego, Italiano, Portugués, Español, Finés y Sueco). Contiene definiciones de más de 4.500.000 términos y unas 180.000 abreviaturas (marzo de 1997), que se distribuyen entre todas las lenguas. Los términos y abreviaturas se agrupan en dos grandes secciones, que facilita al usuario un trabajo fácil y eficaz.

Terminología

Un ejemplo típico de ficha terminológica consiste en términos, definiciones, código de materias y otras informaciones. Un registro típico aparece como sigue:

```

BE= BTB TY= MAA92 NI= 0000001 DATE = 941103 CF= 4
  CM CE JU
ES VE Unizn Europea
  RF Tratado de la Unizn Europea,Maastricht 1992
PT VE Uniao Europea
  RF Tratado sobre a Uniao Europeia,Maastricht 1992
FR VE Union europzenne
  RF Traitz sur l'Union europzenne,Maastricht 1992;titre
DE VE Europzische Union
  RF Vertrag zber die Europ,ische Union,Maastricht 1992
IT VE Unione europea
  RF Trattato sull'Unione europea,Maastricht 1992
EN VE European Union
  RF Treaty on European Union,Maastricht 1992
NL VE Europese Unie
  RF Verdrag betreffende de Europese Unie,Maastricht 1992
DA VE Den Europziske Union
  RF Traktaten om Den Europziske Union,Maastricht 1992.1

```

Los campos más interesantes son:

```

BE      Productor
DATE   Inserción de la fecha en la base de datos
CM      El código de materia de el término
CF      El código de fiabilidad de la traducción. (Este código clasifica el
        término en la base de datos en función de su fiabilidad. Los
        términos y las abreviaciones se clasifican de 0 a 5. Si la ficha
        tiene un código 5 este es la traducción oficial de la CE.)
VE      Términos
RF      Referencia de la fuente.
DF      Definición del término
NT      Nota explicativa

```

El código de materias "CM" indica bajo que tema el término ha sido clasificado. La lista de las diferentes materias es la siguiente:

AD GESTIÓN EN SECTORES PÚBLICOS Y PRIVADOS	AG AGRICULTURA, PESCA, SILVICULTURA - PROCESOS INDUSTRIALES DE ALIMENTACIÓN
AR ARTE	AS SEGUROS
AT INDUSTRIA NUCLEAR (FÍSICA APLICADA ATÓMICA Y NUCLEAR)	AU AUTOMATIZACIÓN (INCLUYE TELECOMUNICACIONES E INFORMÁTICA)
BA INDUSTRIA DE LA CONSTRUCCIÓN	BZ BOTÁNICA Y ZOOLOGÍA
CE UNIÓN EUROPEA	CH QUÍMICA
CO COMERCIO - CIRCULACIÓN DE BIENES	DE DEFENSA
DI DOCUMENTACIÓN E INFORMACIÓN	DO ECONOMÍA DOMESTICA
EC ECONOMÍA	ED EDUCACIÓN
EL INGENIERÍA ELECTRÓNICA Y DE LA ENERGÍA	EN MEDIO AMBIENTE
FI TEMAS FINANCIEROS - IMPUESTOS - ADUANAS	GE GENERAL
GO EL COSMOS	HI HISTORIA, ETNOLOGÍA, COSTUMBRES Y TRADICIONES
IC INDUSTRIA QUÍMICA	IN INDUSTRIAS VARIAS Y ARTESANOS
JU JURISPRUDENCIA	LA LENGUA Y LITERATURA
MA MATEMÁTICAS	ME MEDICINA
MG INGENIERÍA MECÁNICA	MI MINERÍA
NO NORMALIZACIÓN, MEDIDAS Y PRUEBAS	OO MEDIOS DE COMUNICACIÓN
OR ORGANIZACIONES INTERNACIONALES	PG PUBLICACIONES E IMPRENTA
PH FÍSICA	PO POLÍTICA
RP RELIGIÓN Y FILOSOFÍA	SC COOPERATIVAS
SI INDUSTRIAS SIDEROMETALÚRGICAS	SO HOMBRE Y SOCIEDAD
SP DEPORTES, OCIO Y TIEMPO LIBRE	ST ESTADÍSTICA
TE TÉCNICA E INDUSTRIA EN GENERAL	TR TRANSPORTES
TS TERRENO Y PROPIEDAD	TV RELACIONES LABORALES

Abreviaturas

Un ejemplo típico de abreviaturas es el siguiente:

```
FR AB GSM
  SI systSme paneurop,en GSM mobile cellulaire numzrique
DE SI europaweites zellulares digitales GSM-Mobilfunksystem
IT AB GSM
  SI sistema digitale cellulare paneuropeo;servizio radiomobile
numérico
  NT Servizio pubblico numerico cellulare paneuropeo di radiotelefonía
mobile terrestre.
EN SI pan-European digital mobile cellular GSM system
NL SI paneuropees digitaal cellulair mobiel GSM-systeem
DA SI det digitale celleopbyggede faelleseuropaeiske
GSM-mobilkommunikationssystem
ES SI sistema paneuropeo GSM movil digital celular
PT SI sistema mzvel digital celular pan-europeu
```

Los campos más importante en la lista de las abreviaciones son dos:

SI Significado
NT Nota explicativa

El lenguaje de consulta de EURODICAUTOM no es el CCL como en otras bases de datos de ECHO. Si se accede directamente al distribuidor (por ejemplo, *telnet echo.lu*) se debe pasar al lenguaje CCL y desde allí seleccionar la base de datos (BASE EU92), y para realizar consultas teclear CALL EDIC.

También se ha diseñado un interface de consulta a través de WWW, el URL es el siguiente: **<http://www2.echo.lu/edic/>**.

DICCIONARIOS EN LÍNEA

El desarrollo de Internet en los últimos años ha potenciado la creación de multitud de servicios para traductores. Un caso particular muy extendido son los diccionarios en línea. Existen multitud de ejemplos, sólo hay que utilizar un rastreador de Internet (tipo AltaVista o similar) e interrogar por la palabra *dictionary*. Saldrán decenas miles de documentos con esa palabra.