

## Terminología basada en el conocimiento para la traducción y la divulgación médicas: el caso de ONCOTERM\*

Clara Inés López Rodríguez,\*\* Pamela Faber Benítez\*\*\* y Maribel Tercedor Sánchez\*\*\*\*

**Resumen:** Oncoterm es un proyecto de investigación interdisciplinar sobre terminología médica cuyo objetivo principal es la elaboración de un sistema de información sobre el subdominio biomédico de la oncología en el que los conceptos están vinculados a una ontología. Dicha estructura conceptual subyace a los demás componentes del sistema y constituye el vínculo entre términos en diferentes lenguas (inglés, español y alemán). Asimismo, la identificación de un marco conceptual facilita la comprensión de textos sobre el cáncer en diferentes lenguas, no sólo por parte de los profesionales de la salud, sino también por parte de los traductores, los periodistas y los pacientes.

### Knowledge-based terminology for medical translation and journalism: the interdisciplinary Oncoterm project

**Abstract:** Oncoterm is an interdisciplinary research project on medical terminology whose main objective is the elaboration of an information system on the biomedical subdomain of oncology in which the concepts are linked to an ontology or a hierarchy of concepts. This conceptual structure underlies the other components of the system and is the link between terms in different languages (English, Spanish, and German). Accordingly, the identification of a conceptual framework facilitates the comprehension of oncology texts in different languages not only for health professionals, but also for translators, journalists, and patients.

**Palabras clave:** gestión de terminología basada en el conocimiento; sistema de información sobre oncología, traducción médica y divulgación; representación conceptual; análisis de corpus. **Key words:** Knowledge-based terminology management; information system on oncology, medical translation and science popularisation; knowledge representation; corpus analysis.

**Panace@ 2006; 7 (24): 228-240**

### 1. Introducción

Uno de los objetivos estratégicos de la Fundación Española para la Ciencia y la Tecnología es el de impulsar «la divulgación del conocimiento en materia de ciencia y tecnología, con el fin de crear esa necesaria cultura científica y tecnológica entre los ciudadanos». Lamentablemente, gran parte de la cultura científica que recibe el público general proviene de los medios de comunicación, lo cual trae consigo limitaciones en cuanto a la fiabilidad y la relevancia científica real de la información publicada. En este sentido, como afirma Semir (2003) apoyándose en Balzac, «la relación de un periodista con la verdad es perversa porque, para un periodista, todo aquello que es probable es verdad».

Teniendo en cuenta que *cáncer* es una de las diez palabras relacionadas con la ciencia que más frecuentemente se citan en los medios españoles de comunicación escrita (Núñez, 2003),<sup>a</sup> la cautela al presentar conceptos especializados y avances terapéuticos relativos al cáncer ha de ser máxima para evitar generar falsas expectativas acerca de la curación de la enfermedad. Si a esto se suman el hecho de que la oncología es un campo prolífico en investigaciones publicadas

principalmente en inglés y el deseo de los pacientes y sus familiares de estar informados sobre la enfermedad, es importante generar recursos de información sobre el cáncer, en inglés y español, redactados de forma objetiva, clara y concisa y accesibles para el público general. En este contexto, Internet constituye el marco idóneo de difusión y actualización de información.

En respuesta a esta demanda de la sociedad, se ha desarrollado «Oncoterm: Sistema bilingüe de información y recursos oncológicos», un proyecto que nació con el objetivo de cubrir un vacío de información en español destinada a los pacientes del cáncer. El proyecto lo llevó a cabo entre 1999 y 2002 un equipo multidisciplinar formado por investigadores de las Universidades de Granada, Málaga y Valladolid y el Hospital Universitario Virgen de las Nieves de Granada. Su principal resultado ha sido una base de datos en Internet con información sobre el cáncer, destinada a diferentes usuarios: profesionales de la comunicación y de la salud, pacientes y público general.

En este artículo<sup>b</sup> describimos los objetivos de Oncoterm y sus fundamentos teóricos para la gestión de la termino-

\* Oncoterm: Sistema bilingüe de información y recursos oncológicos (PB98-1342), financiado por el Ministerio de Educación y Ciencia. Su investigadora principal es Pamela Faber Benítez.

\*\* Profesora de Traducción Científica y Técnica, Facultad de Traducción e Interpretación de la Universidad de Granada (España). Dirección para correspondencia: [clarairl@ugr.es](mailto:clarairl@ugr.es).

\*\*\* Catedrática de Terminología, Facultad de Traducción e Interpretación de la Universidad de Granada (España).

\*\*\*\* Profesora de Traducción Científica y Técnica, Facultad de Traducción e Interpretación de la Universidad de Granada (España).

logía basada en el conocimiento. Explicamos asimismo la metodología que seguimos en la elaboración de recursos terminológicos, que se fundamenta en: a) la extracción de conocimiento mediante el análisis de corpus multilingües referentes al campo de especialidad; b) la creación de un marco conceptual de referencia donde encajan todos los términos y recursos de nuestra base de datos, y c) la identificación de la multidimensionalidad de las representaciones conceptuales, es decir, del hecho de que un concepto pueda clasificarse desde más de una perspectiva (Bowker y Meyer 1993).

Por último, resumimos los resultados del proyecto, disponibles en Internet en la página <[www.ugr.es/local/oncoterm](http://www.ugr.es/local/oncoterm)>.

### 1.1. Objetivos del proyecto Oncoterm

Aunque son muchas las revistas, bases de datos y portales sobre enfermedades que ofrecen información específica para el médico y el paciente, la mayoría toma el inglés como lengua base. En este sentido, Oncoterm ha supuesto un enfoque diferente, pues se ha trabajado a partir de textos redactados en español, con el objetivo principal de crear una base de datos terminológica en español, con correspondencias en inglés y alemán, con el fin de proporcionar recursos de información para médicos, pacientes, traductores y redactores de textos especializados. El proyecto se sustentó también sobre los siguientes objetivos específicos:

- Crear un corpus de textos médicos tanto en español como en inglés y alemán y desarrollar un protocolo de extracción de información de dichos textos.
- Elaborar un inventario de relaciones conceptuales específicas del EVENTO MÉDICO (véase apartado 3.4.) y, en particular, del EVENTO MÉDICO ONCOLÓGICO.
- Especificar un lenguaje para las definiciones terminográficas conciso, coherente y controlado que permita la ubicación de un concepto dentro de su categoría y determinar sus relaciones con el resto de las categorías. Este lenguaje es aplicable no sólo al subdominio de la oncología, sino también a otras especialidades médicas y a otras lenguas.
- Configurar una base de datos terminológica articulada en torno a la estructura hallada en la definición de los términos. En la base de datos terminológica, los conceptos están vinculados a una ontología, que constituye el vínculo entre términos en diferentes lenguas.
- Crear una aplicación informática específica que permitiera estructurar y almacenar la información y asegurara una interfaz usuario-sistema de fácil utilización. En este sentido se desarrolló Ontoterm<sup>®</sup>, una aplicación informática<sup>°</sup> que permite manejar y gestionar grandes cantidades de conceptos y términos y exportar dicha información a formato html, con vistas a su publicación en la web de Oncoterm.

- Especificar esquemas por categoría (*category templates*) que reflejasen la organización interna de cada categoría del conocimiento sobre oncología, de modo que sirviese de eje de desarrollo de la base de datos.
- Incluir de forma sistemática información pragmática de utilidad para el médico y el paciente en las entradas terminológicas (Seibel y Jiménez, 2002).

## 2. Fundamentos teóricos

En la codificación del conocimiento especializado, nuestro grupo de investigación sigue el modelo lexemático funcional (MLF) y la semántica de marcos. El MLF de Martín Mingo-rance (1984, 1989, 1995; Faber y Mairal, 1999; Faber, López y Tercedor, 2001) es una teoría de base léxica que facilita la representación de relaciones conceptuales y sintagmáticas en el lenguaje general y especializado. Este modelo ofrece una metodología para la organización de conceptos mediante la información extraída de definiciones lexicográficas/terminográficas estructuradas sistemáticamente.

La organización léxica que el MLF propone para el lexicón se basa parcialmente en la distinción de Saussure entre relaciones paradigmáticas (las que establece una palabra con otras de la misma categoría gramatical capaces de conmutarse por ella en un contexto lingüístico dado) y sintagmáticas (las que establece con otras que aparecen en la misma oración). Esta distinción es pertinente porque se encuentra en nuestra organización conceptual, previa e independientemente del sistema lingüístico (Nelson, 1985: 179). El MLF integra la lexemática de Coseriu para el estudio del eje paradigmático y, para la descripción del eje sintagmático o combinatorio, la gramática funcional de S. C. Dik, que encaja las unidades léxicas en marcos predicativos y estudia su significado mediante la descomposición léxica gradual (cfr. Martín Mingo-rance, 1998, y Faber y Mairal, 1999).

El eje paradigmático del lexicón del MLF codifica la configuración de conceptos en el eje de selección, organizándolos onomasiológicamente en una jerarquía de dominios y subdominios. Todos los miembros de un dominio conceptual comparten una serie de propiedades. Es asimismo un factor determinante en el eje sintagmático, en el que se codifica su potencial combinatorio. La convergencia de estos dos ejes es la base de la estructura conceptual propuesta en el MLF.

Por otra parte, en el MLF se concibe nuestro lexicón mental como una compleja red en la que cada nodo es un concepto, y los conceptos están interconectados por diferentes tipos de relaciones (Iris, Litowitz y Evens, 1988: 263). Las redes semánticas se dividen en dominios; los dominios, en marcos, y los marcos, en conceptos interrelacionados.

Por este motivo, en nuestro proyecto se combina el MLF con los principios de la semántica de marcos. La semántica de marcos (Fillmore, 1982; Fillmore, Johnson y Petruck, 2003) define *marco* como 'una representación esquemática' o 'la representación de un sistema de conceptos relacionados de tal manera que la utilización de un único concepto evoca o activa todo el sistema conceptual'. En este sentido, en nuestra base de datos cada concepto se conecta mediante

hipervínculos con otros conceptos superordinados y subordinados.

### 3. Metodología

La metodología que seguimos pretende acercarse y amoldarse a los procesos mentales que tienen lugar en el cerebro del traductor científico-técnico que debe asimilar y gestionar en un tiempo récord su escaso conocimiento sobre un campo de especialidad y su terminología a partir de un encargo de traducción concreto. Asimismo, frente a algunos recursos terminológicos sobre salud que son una mera transferencia de términos en inglés hacia el español, nuestra metodología pretende que el público general y los profesionales de la comunicación aumenten su comprensión sobre el cáncer.

Faber (2004) sintetiza las diferentes fases (no lineales) por las que pasan la adquisición y la representación del conocimiento especializado por parte del traductor: *a)* comprensión de la estructura conceptual general, *b)* especificación de conceptos y elaboración de redes, *c)* análisis textual, *d)* recopilación de textos electrónicos y análisis de corpus y *e)* consulta de diccionarios y de expertos.

Por tanto, en Oncoterm, una de nuestras prioridades metodológicas es la representación del conocimiento y el análisis de textos tanto en inglés como en español con la metodología empírica que ha aportado la lingüística de corpus desde la década de los setenta. En este sentido, nuestro enfoque es el de la gestión terminológica basada en el conocimiento (Faber, 1999), es decir, un enfoque en el que los términos están asociados a un marco conceptual de referencia independiente de la lengua que sirve de cimiento para la base de datos terminológica. El marco conceptual se plasma en una ontología, es decir, un repositorio de conceptos que incluye entidades, sus propiedades, funciones y limitaciones, los procesos a los que afectan estas entidades y las relaciones entre las entidades.

Como se sostiene en Faber, López y Tercedor (2001), el conocimiento especializado es una subdivisión de nuestro sistema de conocimiento general y, por tanto, se compone de esquemas cognitivos flexibles que permiten la manipulación de ideas y la construcción de hipótesis. Cualquier concepto especializado debe poder encajar en un sistema formado por *eventos* (actividades, acciones, sucesos o situaciones), *objetos* (entidades que no son ni acciones ni propiedades) y *relaciones* que los vinculan entre ellos. Por ejemplo, el concepto BRONCHOSCOPY es un evento (procedimiento diagnóstico) y está relacionado no sólo con objetos como BRONCHIUM (mediante la relación AFFECTS-BODY-PART) o ENDOSCOPE (mediante la relación USES), sino también con eventos como BIOPSY (mediante la relación PURPOSE) o BRONCHIOLALVEOLAR-LUNG-CANCER (mediante la relación DIAGNOSTIC-TEST-FOR).

Este sistema conceptual debe quedar reflejado en la definición terminográfica de cada unidad terminológica (García de Quesada, 2001), que se basa en la consulta de diccionarios y expertos, la extracción de información del corpus textual en soporte electrónico y el análisis de imágenes. A conti-

nuación ilustramos con el vocablo *adenocarcinoma* cómo utilizamos las definiciones de diccionarios especializados (tabla 1) para segmentar, extraer, clasificar y relacionar la información que, a menudo, se presenta de forma caótica y poco sistemática en los diferentes diccionarios médicos (Oncoterm, 2002).

**adenocarcinoma** A malignant tumor with cells arranged in a glandlike pattern (*The HarperCollins Illustrated Medical Dictionary*).

**adenocarcinoma** a cancer made up of abnormal gland cells on the lining or inner surface of an organ. It can develop in virtually any part of the body. Adenocarcinomas may develop in the lung, pancreas, breast, prostate, esophagus, stomach, vagina, urethra, and small intestine, among others (*The Cancer Dictionary*).

**adenocarcinoma** A malignant growth of glandular tissue. This tissue is widespread throughout the body's organs and the tumours may occur, for example, in the stomach, ovaries, and uterus. Adenocarcinomas may be subdivided into those that arise from mucous or serous secreting glandular tissue (*Black's Medical Dictionary*).

**Tabla 1.** Definición de adenocarcinoma en tres diccionarios médicos

Al analizar de forma sistemática las definiciones y clasificar la información que aportan, encontramos que, para explicar bien el concepto *adenocarcinoma*, habría que incluir etiquetas que indican:

- la pertenencia a una categoría conceptual: malignant tumor, cancer, malignant growth.
- forma percibida: cell arrangement in glandlike pattern
- lugar:
  - on the lining or inner surface of any organ, e.g. lung, pancreas, breast, prostate, esophagus, stomach, vagina, urethra, and small intestine
  - where there is glandular tissue, e.g. stomach, ovaries, uterus
- subclasificación o tipo:
  - mucous secreting glandular tissue
  - serous secreting glandular tissue

Al hacer este tipo de análisis sistemático, se puede llegar a una estructuración lingüística que refleja la organización conceptual del subdominio especializado donde se ubica el término *adenocarcinoma*, una organización que pone de manifiesto que uno de los parámetros para estructurar el concepto *malignant neoplasm* es el tipo de tejido en el que se asienta, es decir, su derivación histológica, como se percibe en la tabla 2.

<p><b>MALIGNANT NEOPLASMS</b>  <b>DIMENSION: histologic derivation</b></p>
<p><b>Epithelial tissue (tissue lining a surface/cavity)</b> [<i>italics show most common site</i>]</p>
<p><b>carcinoma</b> malignant neoplasm of abnormal epithelial tissue.  <b>squamous cell carcinoma</b> carcinoma of squamous epithelial cells. <i>Bladder, anus and vulva.</i>  <b>oat cell carcinoma</b> very malignant, undifferentiated, small cell carcinoma. <i>Lung, esophagus.</i>  <b>basal cell carcinoma</b> carcinoma arising in the basal layer of the skin / structures derived from basal cells. It invades locally, frequently forming a rodent ulcer. <i>Skin, canthus of eye, tip of nose, chin, lip [...],</i></p>
<p><b>Epithelial tissue (glandular epithelium)</b></p>
<p><b>adenocarcinoma</b> malignant neoplasm made up of abnormal gland cells on the lining or inner surface of an organ. <i>Breast, bronchi, digestive tract, pancreas, endocrine glands, prostate.</i>  <b>cystadenocarcinoma</b> adenocarcinoma, most frequently occurring as a partially solid mass with a cystic pattern. <i>Salivary gland, ovaries, breast, thyroid.</i>  <b>serous cystadenocarcinoma</b> rare cystadenocarcinoma with frequently bilateral loculations and cysts containing transudate. <i>Ovaries.</i>  <b>pseudomucinous cystadenocarcinoma</b> cystadenocarcinoma with cell stratification and cysts containing viscid fluid. <i>Ovaries</i>          [...].</p>

**Tabla 2.** Aplicación de jerarquías terminográficas para ubicar el concepto adenocarcinoma en relación con otros tipos de cáncer

En cada definición la relación TIPO-DE está representada por el *genus* o etiqueta conceptual. Por ejemplo, la tabla indica que *serous cystadenocarcinoma* y *pseudomucinous cystadenocarcinoma* son tipos de *cystadenocarcinoma*. *Cystadenocarcinoma*, a su vez, es un tipo de *adenocarcinoma*, que es una *neoplasia maligna*. Con este tipo de análisis llegamos a una jerarquía en la que cada concepto se define mediante uno semánticamente más simple; éste, a su vez, se reduce a uno más simple y así respectivamente, como se ve a continuación:

```
{serous cystadenocarcinoma}@→
{cystadenocarcinoma}@ → {adenocarcinoma}@→
{malignant neoplasm}@→{neoplasm}@ →
{event}@ → {ALL}
```

Tras consultar los diccionarios y elaborar jerarquías terminográficas, analizamos nuestro corpus<sup>d</sup> textual con la herramienta de análisis léxico *Wordsmith Tools*<sup>e</sup> para generar líneas de concordancia (tabla 3 y apartado 3.2.2.). Y así, por ejemplo, las líneas de concordancia en torno al término *adenocarcinoma* nos muestran otros parámetros de clasificación en los que encaja dicho término: PARTE DEL CUERPO AFECTADA, ESTADIFICACIÓN, TIPO DE CÉLULA QUE LO COMPONE. Estos parámetros de clasificación ponen de manifiesto la multidimensionalidad del concepto *cáncer*, un aspecto que se tratará con más detalle en el apartado 3.4.

De todas formas, el análisis de corpus no es más que una de las etapas no lineales de la metodología empleada en el proyecto Oncoterm:

- recopilación de corpus en soporte electrónico;
- lista de conceptos y términos;
- creación de un gestor de información;
- elaboración de un marco conceptual de referencia mediante la elaboración de jerarquías terminográficas;
- elaboración de entradas terminológicas.

### 3.3.1. Recopilación del corpus en soporte electrónico

Entre los criterios de selección del corpus, Pérez Hernández (2000) destaca los siguientes: cantidad, calidad, simplicidad, documentación, pertenencia al dominio de especialidad, fecha de producción y condición lingüística del texto, actualidad, tipo textual, nivel de tecnicidad y receptores del texto.

En Oncoterm se combinó un corpus en inglés y español sobre oncología de aproximadamente 32 millones de palabras, de las cuales la mayor parte está en inglés (28 771 714). En la compilación del corpus se utilizaron: *a*) textos extraídos de Internet; *b*) enciclopedias, manuales y publicaciones médicas en CD-ROM, y *c*) textos escaneados por su riqueza de vocabulario para las definiciones terminográficas (manuales para médicos y estudiantes de Medicina) y por su relevancia entre especialistas, según MEDLINE. Se descartaron textos de fuentes no fiables y, teniendo en cuenta que los usuarios potenciales de nuestro proyecto son profesionales de la comunicación y de la salud y público general interesado en la oncología, incluimos textos redactados por especialistas para un público con distintos niveles de conocimiento especializado.

Así, se recopilaron páginas web de las principales organizaciones internacionales contra el cáncer,<sup>f</sup> que ofrecen, tanto a los profesionales de la salud como al público general, folletos informativos, resúmenes y publicaciones especializadas. También se seleccionaron artículos experimentales y resúmenes provenientes de publicaciones médicas y oncológicas<sup>g</sup> y fragmentos relevantes de manuales destinados a oncólogos y estudiantes de medicina de nivel avanzado.<sup>h</sup> Por último, para reflejar la comunicación entre el especialista y un público lego, se recopilaron no sólo los monográficos que aparecen en las enciclopedias y las publicaciones de divulgación semiespecializada y general,<sup>i</sup> sino también folletos de salud pública para pacientes y familiares disponibles en los sitios web antes mencionados.

PARTE DEL CUERPO AFECTADA			
1	ent Option Overview	Small Intestine	<b>Adenocarcinoma</b> Small Intestine Lymph
2	ic radiation in stage I	endometrial	<b>adenocarcinoma</b> with high-risk attributes
3	edict prognosis in stage I	endometrioid	<b>adenocarcinoma</b> of the uterine corpus
4	sk factors associated with esophageal		<b>adenocarcinoma</b> are less clear. The prese
5	in situ adenocarcinoma	papillary	<b>adenocarcinoma</b> adenocarcinoma, intest
6	: Prediction of prognosis for prostatic		<b>adenocarcinoma</b> by combined histologi
7	: Conservative management of rectal		<b>adenocarcinoma</b> by radiotherapy. Internat
8	N Renal cell cancer, also called renal		<b>adenocarcinoma</b> , or hypernephroma, can of
9	r curative resection for	colorectal	<b>adenocarcinoma</b> . Lancet 340(8818): 502-50
10	treated with taxol for advanced ovarian		<b>adenocarcinoma</b> . Journal of Clinical
ESTADIFICACIÓN (I)			
1	entially Curatively Resected Stage C1-3		<b>Adenocarcinoma</b> of the Colon
2	Lymphadenectomy in Surgical Stage C		<b>Adenocarcinoma</b> of the Prostate
3	C. Papillary or villoglandular	D.	<b>Adenocarcinoma</b> with squamous
4	chiectomy in Patients with Stage D2		<b>Adenocarcinoma</b> of the Prostate
5	Resection in Patients With Stage II		<b>Adenocarcinoma</b> of the Colon
6	in Following Resection for Stage II/III		<b>Adenocarcinoma</b> of the Colon
7	e for Surgically Resected Stage III		<b>Adenocarcinoma</b> of the Colon
8	uvant therapy of Dukes' A, B, and C		<b>adenocarcinoma</b> of the colon with
9	owing radiation therapy for stage C		<b>adenocarcinoma</b> of the prostate:
10	with surgical treatment of stage D1		<b>adenocarcinoma</b> of prostate:
ESTADIFICACIÓN (II)			
2	chett TR, Lieskovsky G, et al.: Primary		<b>adenocarcinoma</b> of bladder. Urology 3
3	iously untreated patients with advanced		<b>adenocarcinoma</b> of the pancreas: re
4	carcinoma and poorly differentiated		<b>adenocarcinoma</b> of unknown primary site:
5	275 patients with clinically localized		<b>adenocarcinoma</b> of the prostate. Jour
6	treated with taxol for advanced ovarian		<b>adenocarcinoma</b> . Journal of Clinical
7	il in patients with advanced pancreatic		<b>adenocarcinoma</b> . Journal of Clinical
TIPO DE CÉLULA			
1	ometrioid (75%-80%)	A. Ciliated	<b>adenocarcinoma</b> B. Secretory adenoca
2	mucinous adenocarcinoma	clear cell	<b>adenocarcinoma</b> signet-ring cell carci
3	ity of cases) mucinous (colloid)		<b>adenocarcinoma</b> signet ring adenocar
5	papillary adenocarcinoma	mucinous	<b>adenocarcinoma</b> adenosquamous carcinom
6	olloid) adenocarcinoma	signet ring	<b>adenocarcinoma</b> scirrhus tumors
7	carcinoma spindle cell variant		<b>adenocarcinoma</b> acinar papill
8	o DES and may coexist with a clear cell		<b>adenocarcinoma</b> , although it rarely pro

Tabla 3. Líneas de concordancia en torno al término adenocarcinoma, ordenadas según distintos parámetros de clasificación

### 3.2. Lista de conceptos y términos

Para elaborar una lista preliminar de los conceptos más importantes en oncología, estudiamos tanto los recursos terminográficos como algunos de los textos recopilados (véase 3.1.), hicimos consultas a oncólogos y analizamos el corpus de textos para extraer listas de frecuencia lematizadas, líneas de concordancia y palabras clave. También se fue elaborando con las lecturas un borrador del marco conceptual subyacente al dominio.

#### 3.2.1. Listas de frecuencia

Las palabras más frecuentes en un texto o en un corpus constituyen las representaciones lingüísticas de los conceptos más importantes de ese campo de especialidad. En la lista de frecuencia (tabla 4), extraída con *Wordsmith Tools* a partir del subcorpus en lengua inglesa de *Oncoterm*, destacan palabras de la lengua general que denominan principalmente los conceptos PATIENT, CELL, TREATMENT, DISEASE, RESEARCH, CAUSE y EFFECT. De ahí es posible inferir que estas

son algunas de las áreas conceptuales en torno a las cuales se articula el conocimiento especializado del subdominio de la Oncología (López Rodríguez 2001).

N	Palabra	Frec.	%	LEMAS
1	PATIENTS	175 546	0,61	patient (23 100), patient's (2115)
2	CELLS	144 895	0,50	cell (65 679), cell's (75)
3	JOURNAL	97 597	0,34	
4	ARTICLE	96 714	0,34	
5	STUDY	81 291	0,28	
6	CANCER	81 160	0,28	cancer's (12)
7	TREATMENT	62 967	0,22	treatments (4012)
8	RESULTS	61 455	0,21	result (7130)
9	USING	61 122	0,21	used (28 489)
10	GROUP	54 309	0,19	groups (17 340)
11	EFFECTS	52 790	0,18	effect (25 848)
12	TUMOR	50 779	0,18	tumors (17 569), tumour (4870), tumor's (41), tumour's (4)
13	DISEASE	49 773	0,17	diseases (6231), disease's(10)
14	PROTEIN	49 555	0,17	proteins (13 187)
15	SIGNIFICANTLY	47 642	0,17	significant (23 819)
16	INCREASED	44 978	0,16	increased (17 955), increases (215), increasing (64)
17	CLINICAL	42 253	0,15	
18	CASE	40 768	0,14	cases (26 873)
19	YEARS	40 575	0,14	year (14 623)
20	THERAPY	39 985	0,14	therapies (1528), therapeutic (5451), therapeutical(145), therapeutics (238), therapic (2)

**Tabla 4.** Cabecera de la lista de frecuencia lematizada del corpus en inglés de Oncoterm

La identificación de palabras clave contribuyó al modelado conceptual de la ontología que elaboramos y tiene consecuencias terminográficas. Por ejemplo, *treatment* o *therapy* son la parte nuclear o *definiens* de las definiciones de

*chemotherapy*, *radiation therapy* o *radiotherapy* en distintos diccionarios médicos. En consecuencia, a partir de los lemas más frecuentes de un corpus, es posible identificar las categorías conceptuales sobre las que se fundamenta la definición de los términos del texto.

### 3.2.2. Uso de líneas de concordancia

En nuestro grupo de investigación nos servimos de la líneas de concordancia para extraer conocimiento sobre el campo de especialidad e identificar colocaciones, unidades fraseológicas y de patrones sintácticos (Tercedor, 1999). Las líneas de concordancia presentan las ocurrencias de una palabra en su contexto lingüístico, de forma que la palabra en cuestión aparece en el centro de cada línea, y a ambos lados figuran las palabras con las que suele aparecer en los textos.

Además, las concordancias proporcionan un esquema básico de relaciones conceptuales sobre el cual puede modelarse el resto del subdominio. Las concordancias en la figura 1 ilustran que en la estructuración del subdominio RADIATION THERAPY hay que tener en cuenta la diferencia entre los conceptos INTERNAL RADIATION THERAPY y EXTERNAL RADIATION THERAPY, y que tanto *intracavitary radiation therapy* como *interstitial radiation therapy* son tipos de radioterapia interna. Sin embargo, la radioterapia externa, que es el tipo de radioterapia más común, tiene un abanico de distinciones conceptuales más amplio, como el tipo de radiación utilizada, la trayectoria del haz, la parte del cuerpo enfocada y la extensión del campo enfocado, así como la dosis e intensidad de la radiación. Queda, pues, patente la multidimensionalidad del término *radiation therapy* (véase figura 1, pág. 234).

### 3.3. Creación de un gestor de información

Uno de los aspectos esenciales era contar con un gestor de información para almacenar datos y que proporcionara asimismo la interfaz con el usuario en la web. Ontoterm (<www.ontoterm.com>) es una herramienta desarrollada a tal efecto por Antonio Moreno Ortiz, de la Universidad de Málaga (Moreno Ortiz, 2000a, 2000b; Moreno y Pérez, 2000). La estructuración conceptual utilizada en Ontoterm depende de un recurso de representación de conocimiento ya existente: la ontología Mikrokosmos (Mahesh y Nirenburg, 1995; Viegas et al., 1999).

La arquitectura de Ontoterm tiene dos módulos principales: un editor de ontologías y un editor de la base de datos terminológica. En el *editor de ontologías* se introducen los conceptos y se construye la estructura conceptual. Cada concepto aparece con una descripción (en la que se especifica su concepto superordinado y, por tanto, su lugar dentro de la ontología), propiedades, relaciones con otros conceptos y herencia. Con este programa, los conceptos especializados quedan integrados en los niveles más específicos de la ontología. Por ejemplo, conceptos muy importantes en el dominio de la oncología, como TREATMENT o DIAGNOSTIC-TEST, quedan ubicados como hipónimos de MEDICAL-SERVE (figura 2, pág. 235).

<b>RADIATION_THERAPY</b>
<b>RADIATION_SOURCE_LOCATION</b>
1 516-520, 1981. Bagshaw MA: External <b>radiation therapy</b> of carcinoma of prosta 2 r near the tumor. Also called internal <b>radiation therapy</b> or implant radiation.
<b>9 INTERNAL_RADIATION_THERAPY</b>
3 ] 4. Laser therapy or interstitial <b>radiation therapy</b> for endobronchial 4 ntimeters thick: 1. Intracavitary <b>radiation therapy</b> . In most instances, 6,
<b>9 EXTERNAL_RADIATION_THERAPY</b>
<b>9 BEAM_TRAJECTORY</b>
5 c radiosurgery and stereotactic <b>radiation therapy</b> . stereotaxis (
<b>9 BEAM_TYPE</b>
6 il and intraoperative electron beam <b>radiation therapy</b> on the outcome of pati 7 radiosensitizers, or particle-beam <b>radiation therapy</b> . [14-17] 4. Isotre 8 to chemotherapy.[7-9] Fast neutron beam <b>radiation therapy</b> or accelerated hyper 9 ltileaf collimator.[25] Proton-beam <b>radiation therapy</b> is also under investig 10 al with mixed-beam (neutron/photon) <b>radiation therapy</b> , compared to standard
<b>RADIATION_BEAM_TARGET</b>
<b>9 BODY_PART</b>
11 m the NWT3-3 demonstrate that abdominal <b>radiation therapy</b> does not provide signi 12 metastases are identified, whole brain <b>radiation therapy</b> (30 gray in 2 gray f 13 s incorporating chemotherapy plus chest <b>radiation therapy</b> are listed below. Th 14 pts to lower the dose of craniospinal <b>radiation therapy</b> to 2,340 cGy have resu 15 tive chest wall and regional lymph node <b>radiation therapy</b> are undergoing reassess 16 Jett JR, McGinnis WL, et al.: Thoracic <b>radiation therapy</b> alone compared with
<b>9 FIELD_COVERAGE</b>
17 of whole pelvis versus small-field <b>radiation therapy</b> for carcinoma of prost 18 us 2 months of ABVD plus extended-field <b>radiation therapy</b> is being conducted be 19 arbell NJ, Silver B, et al.: Wide-field <b>radiation therapy</b> with or without chemot
<b>RADIATION_ADMINISTRATION</b>
<b>9 DOSAGE</b>
20 ged remissions. The need for low-dose <b>radiation therapy</b> is under study.[1] The 21 pse after initial wide-field, high-dose <b>radiation therapy</b> have a good prognosis 22 using accelerated superfractionated <b>radiation therapy</b> for advanced squamous 23 e.[6,11-18] 3. Novel fractionation <b>radiation therapy</b> clinical trials are un 24 l evaluation: 1. Hyperfractionated <b>radiation therapy</b> to improve tumor contr
<b>9 INTENSITY</b>
25 ned 5-fluorouracil and supervoltage <b>radiation therapy</b> of locally unresectabl 26 years' experience with megavoltage <b>radiation therapy</b> . Cancer 37(6): 2605-26 27 eutron-beam irradiation or orthovoltage <b>radiation therapy</b> may be used to pal

Figura 1. Parametrización del conocimiento pertinente al concepto RADIATION THERAPY

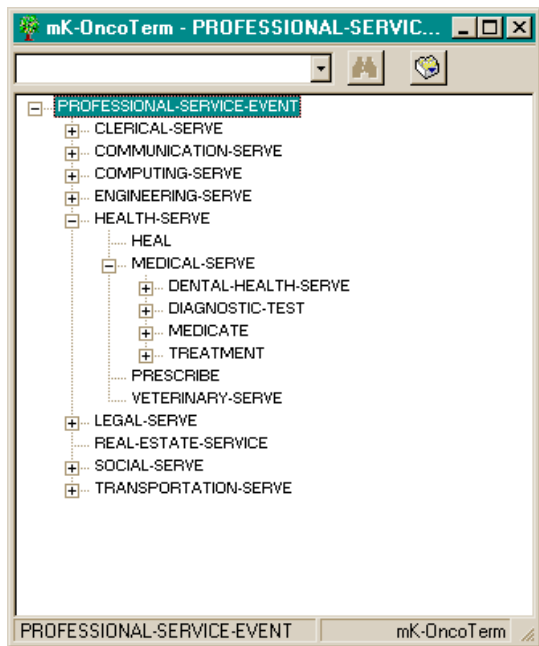


Figura 2. Visualización del concepto HEALTH-SERVE en la ontología

El segundo módulo de Ontoterm es el *editor de la base de datos terminológica*, donde se elaboran las entradas terminográficas vinculadas a los conceptos de la ontología. En este módulo se describe el término o términos que representan cada concepto mediante las categorías de datos de la norma ISO 12620 (figura 3). La información contenida en la base de datos terminológica también puede representarse de forma más gráfica en formato HTML.

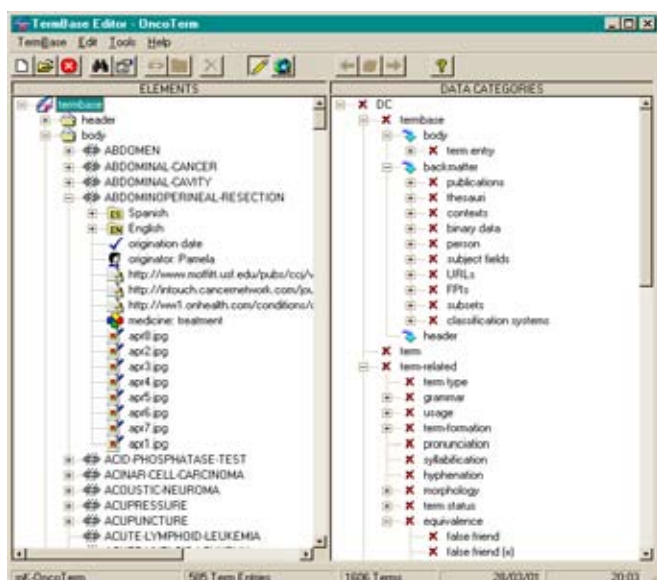


Figura 3. Editor de la base de datos terminológica

### 3.4. Elaboración de un marco conceptual de referencia

Para elaborar un marco conceptual de referencia es preciso establecer jerarquías de términos mediante el análisis

de las definiciones lexicográficas y del corpus (apartado 3), dado que partimos de la premisa de que la información contenida en los diccionarios y los textos especializados constituye una red léxico-conceptual que necesariamente tiene correspondencia con el conocimiento expresado. Por ejemplo, si quisiéramos definir el término *intrapertitoneal radiation therapy*, tendríamos que consultar las definiciones de diferentes diccionarios especializados y elaborar líneas de concordancia conceptuales que mostraran los parámetros en torno a los cuales se organiza el concepto *RADIOTHERAPY* (véase la figura 1). A partir de ahí se elaboraría una jerarquía terminográfica como la que elaboramos para *adenocarcinoma*.

Con dicho método, pudimos identificar estas grandes categorías conceptuales: FACTOR DE RIESGO, PARTE \_ DEL \_ CUERPO, TUMOR, SÍNTOMA, PROCEDIMIENTO DIAGNÓSTICO, INSTRUMENTO, TRATAMIENTO, MEDICAMENTO, EFECTO SECUNDARIO, ESPECIALISTA Y HOSPITAL. La mayoría de estas categorías son extrapolables a otros dominios médicos (Faber, 1999: 99; Faber y Mairal, 1999) y se integran en un diagrama que Faber (2002) denomina *EVENTO MÉDICO ONCOLÓGICO* (figura 4). Los términos superordinados de cada categoría pertenecen a la lengua general, aunque sus niveles más específicos son propios de la terminología de la medicina y la oncología.

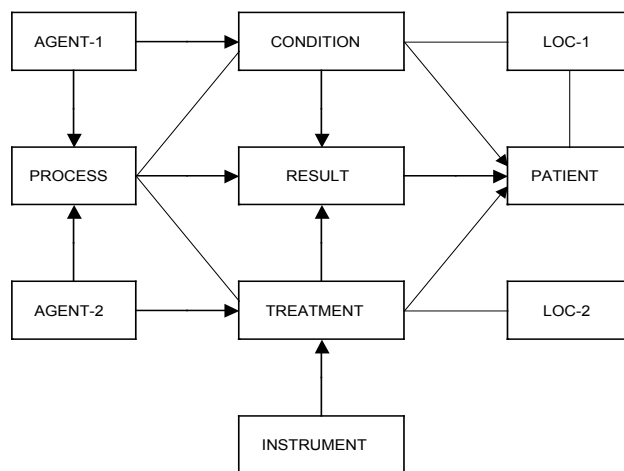


Figura 4. Evento médico oncológico

La estructura interna de cada categoría conceptual quedó representada por un conjunto de tipos de información sobre las características de un concepto (su intensidad) y sobre las entidades en el mundo real que pueden pertenecer a esta categoría (su extensión). A partir de estas categorías conceptuales y de sus respectivos esquemas por categoría, se fue modelando la ontología, insertando cada concepto en el marco conceptual *EVENTO MÉDICO* y estableciendo relaciones entre conceptos; la relación principal fue la de TIPO \_ DE (en inglés, IS \_ A). Estas relaciones son tanto jerárquicas (TIPO \_ DE, PARTE \_ DE...) como no jerárquicas (TIENE \_ FUNCIÓN, AFECTA \_ A, etc.), como se percibe en la tabla 5, en relación al concepto *COLONOSCOPY*:



Conceptual Structures	
IS A	FIBEROPTIC-ENDOSCOPY
ANCESTORS	ALL
	DIAGNOSTIC-PROCEDURE
	ENDOSCOPY
	EVENT
	FIBEROPTIC-ENDOSCOPY
	HEALTH-SERVE
	INVASIVE-DIAGNOSTIC-TEST
	MEDICAL-SERVE
	PROFESSIONAL-SERVICE-EVENT
	SERVICE-EVENT
SOCIAL-EVENT	
WORK-ACTIVITY	
RELATIONS	AFFECTS-BODY-PART : COLON
	DIAGNOSTIC-TEST-FOR : COLO-RECTAL-CANCER
	INSTRUMENT : COLONOSCOPE
	PURPOSE : BIOPSY

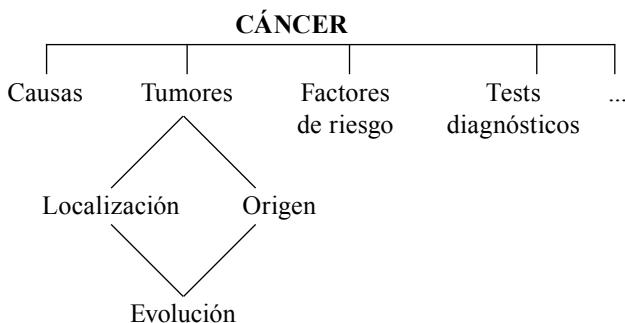
**Tabla 5.** Relaciones en torno al concepto COLONOSCOPY

Por último, en la elaboración del marco conceptual y en cualquier labor terminológica aplicada a la traducción es esencial profundizar en la multidimensionalidad a través de la presencia de unidades léxicas que resaltan distintas características de un concepto para representarlo en el texto.

**3.4.1. La multidimensionalidad en el campo de la oncología**

La naturaleza de los campos científicos y técnicos es multidimensional, puesto que existen distintas formas o dimensiones de concebir la realidad. Serán factores como la función, el lector al que va destinado el texto o el grado de especialización de un texto los que brinden mayor relevancia a unas características que a otras. A través del análisis de las concordancias podemos llegar a reconocer las dimensiones más relevantes en el texto original (TO) y así llegar al texto traducido (TT) adecuado para un encargo determinado.

Y así, si en oncología analizamos con líneas de concordancia las características de los tumores, se mostrará la multidimensionalidad conceptual del área (figura 5) y se facilitará la comprensión y la traducción del TO.



**Figura 5.** Esquema de la activación de la dimensión TUMORES en relación con las otras dimensiones de la oncología (Tercedor Sánchez, 1999)

Según este esquema, los tumores se clasifican atendiendo a criterios de localización, origen y evolución. De todas formas, desde el punto de vista pragmático, la clasificación en función del origen está menos presente en los textos de divulgación general.

Algunos autores (Alonso Ramos et al., 1995) han mostrado que la co-ocurrencia léxica de los lexemas está influida por la semántica de los colocadores. Tal y como se muestra en el caso de la clasificación de los tumores, algunos colocadores *exigen* la combinación con una determinada base: «tumores **precancerosos**» y «tumores **sólidos**». Aportamos ejemplos que muestran las formas fraseológicas relevantes en la descripción de los tumores en las distintas dimensiones.↓

- localización:  
[...] casos de **cáncer de colon y rectal** que sus colegas, en tanto que sus índices de **cáncer de pulmón** no
- origen:  
Los **cánceres epiteliales**, las especies más comunes, se denominan carcinomas.
- evolución:  
in situ and 17 **invasive carcinomas** among 40 rats).
- ORIGEN + LOCALIZACIÓN (muy frecuente en textos especializados):  
[...] de los tumores humanos, incluidos los **carcinomas de colon**, páncreas y pulmón se encuentran [...]
- origen + evolución:  
Lymphomas of the gastrointestinal tract are the most common type of **primary extranodal lymphomas**, accounting for 5 to 10% of all non-Hodgkin's lymphomas.
- localización + evolución:  
**Lobular invasive carcinoma** and **ductal invasive carcinoma** were given a histologic grade 1. **Primary intestinal lymphomas** represent about 15-20% of gastrointestinal lymphomas.
- evolución + origen + localización:  
A case of **chronic myelomonocytic leukemia** with a reciprocal translocation and other numerical and structural abnormalities is described.

**3.5. Elaboración de las entradas terminológicas**

Las entradas terminológicas se rellenaron en Ontoterm a partir de los conceptos de la ontología, de forma que cada concepto quedó vinculado a uno o más términos por lengua. En la elaboración de las entradas se siguieron los siguientes pasos:

- elaboración de definiciones coherentes e informativas a partir de la aplicación de esquemas por categoría;
- información gramatical básica;

- inserción de líneas de concordancia representativas;
- banco de recursos (imágenes, URL, etc.).

En este momento, se elaboran definiciones consensuadas de cada lexema mediante el análisis del corpus y de las entradas de varios diccionarios. La definición terminográfica tiene que presentar de forma explícita el lugar que ocupa el término en la jerarquía. De esta forma, *intraperitoneal radiation therapy* se define en función de su hiperónimo inmediatamente superior (*intracavitary radiation therapy*), y así sucesivamente, hasta llegar a RADIATION THERAPY, que es un subdominio de TREATMENT (figura 6).

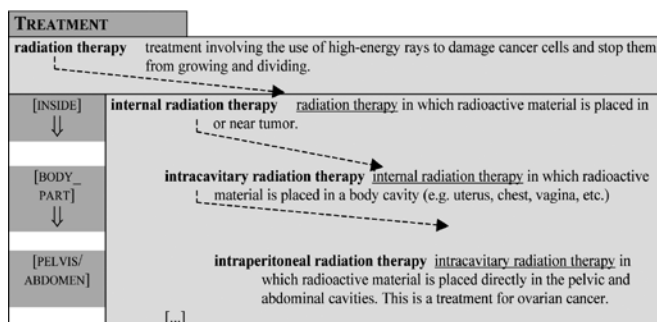


Figura 6. Jerarquía conceptual de RADIATION THERAPY reflejada en las definiciones de los términos (adaptado de Faber, López y Tercedor, 2001: 184)

En las definiciones se percibe que cada concepto posee un esquema por categoría que sirve de modelo de la categoría conceptual en cuestión y que servirá para dar coherencia a las definiciones de términos semánticamente relacionados. El esquema básico para la categoría conceptual TREATMENT aparece reflejado en la tabla 6.

CONCEPTUAL CATEGORY	CONCEPTUAL RELATION
TREATMENT	IS-A
	USES-INSTRUMENT
	HAS-FUNCTION
	HAS-LOCATION

Tabla 6. Esquema categorial de TREATMENT (Faber, López y Tercedor, 2001: 186)

Si se aplica el esquema básico de TREATMENT a los términos que pertenecen al subdominio RADIATION THERAPY como *intraperitoneal radiation therapy*, se percibe que hay una herencia de todos los valores generados a partir de las relaciones que configuran el esquema. Más concretamente, se puede observar que *intraperitoneal radiation therapy* IS-A *internal radiation therapy* IS-A *radiation therapy* IS-A *treatment*. Igualmente, *intraperitoneal radiation therapy* AFFECTS *abdomen/pelvis*, con lo que se hereda el esquema *radiation therapy* AFFECTS *body part* (figura 7).

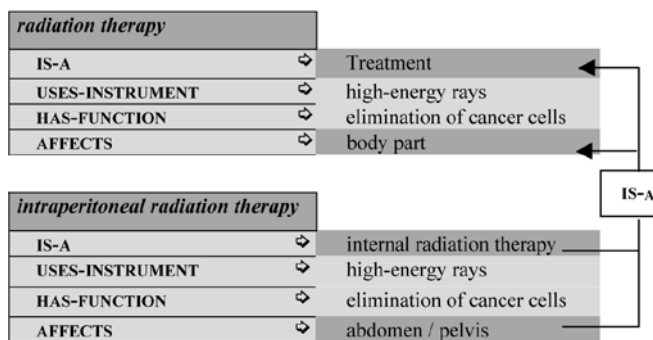


Figura 7. Aplicación del esquema categorial TREATMENT a radiation therapy y intraperitoneal radiation therapy (ib.)

Después de elaborar las definiciones, se introduce la información gramatical básica y se insertan las líneas de concordancia representativas y los recursos multimedia que enriquecen las representaciones lingüísticas de los conceptos.

#### 4. Resultados del proyecto

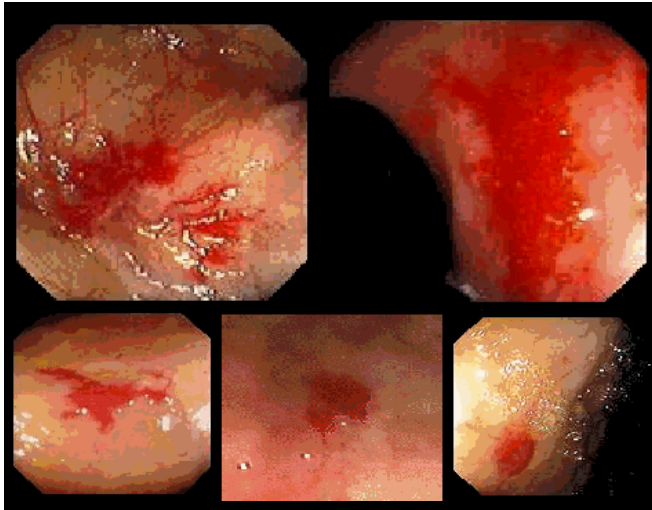
En Oncoterm (2002) y Faber, López y Tercedor (2001) se resumen los principales resultados del proyecto; en esencia, la representación de la estructura conceptual del dominio de la ONCOLOGÍA MÉDICA, sus conceptos pertinentes y sus características e interrelaciones. Esta estructura conceptual constituye el vínculo entre términos en diferentes lenguas, ya que se trata de información no específica de ninguna lengua. Asimismo, con la gestión terminológica basada en el conocimiento se identificaron esquemas de la estructura interna de cada categoría, que se aplicaron a conceptos más específicos, y se confeccionaron definiciones terminográficas concisas y sistemáticas. Estas definiciones reflejan la estructura conceptual del dominio de la oncología.

Los resultados del proyecto están ubicados en Internet (<www.ugr.es/~oncoterm>) y son de utilidad para los profesionales de la salud, investigadores, pacientes y familiares, así como para los traductores y redactores de textos especializados, además de ser una herramienta que contribuye a hacer más eficaz el trabajo, dado que facilita el acceso en línea, en muy poco tiempo, a la información solicitada y asegura su calidad, gracias al sistema de búsqueda y organización conceptual, que permite recuperar no sólo el término objeto de la búsqueda, sino también aquellos relacionados conceptualmente con este.

La web de Oncoterm cuenta con una base de datos jerarquizados con 1896 conceptos relacionados con el cáncer y 4033 términos en español, inglés y alemán. Aunque existen servidores con una función parecida, estos se centran principalmente en la lengua inglesa; Oncoterm desarrolla la información sobre la enfermedad tomando el español como lengua base, partiendo de textos y recursos originales en esta lengua y ofreciendo correspondencias en inglés y alemán. De este modo se facilita la comprensión de la información especializada en lengua española y el acceso a la información de servidores en otras lenguas.

**COLONOSCOPY**

**check date:** 24/08/2001  
**checker:** Pamela  
**subject field: medicine:** diagnostic procedure  
**origination date:** 07/12/2000  
**originator:** CATI  
**URL:** [http://www.sages.org/sg\\_asgepub1030.html](http://www.sages.org/sg_asgepub1030.html)  
**originator:** Arianne  
**origination date:** 06/02/2001  
**figure:**



**URL:** <http://www.murrasaca.com/nt4.htm>  
**originator:** Arianne  
**origination date:** 08/02/2001

Spanish	
<b>colonos- copia</b>	<p><b>term type:</b> main entry term  <b>number:</b> singular  <b>gender:</b> feminine  <b>reliability code:</b> 10  <b>part of speech:</b> noun  <b>definition:</b> endoscopia de partes del colon (colonoscopia parcial) o del colon entero (colonoscopia total) con un endoscopio flexible, el colonoscopio. El instrumento mide 1m (para colonoscopia parcial) o 2m (colonoscopia total). En muchos casos, la flexibilidad del instrumento permite inspeccionar el colon entero, fotografíar partes sospechosas y tomar muestras de tejido para una biopsia. Este método sirve para la detección del carcinoma de colon. Dieta durante las 24 h previas; se realiza previa administración de un purgante; eventualmente con sedación. (es)  <b>origination date:</b> 28/01/2002  <b>originator:</b> Arianne</p>

context (x): E\_colSCO.txt

N	Concordance
1	Derivarse de un cribado masivo con colonoscopia, como tampoco la compleji
2	en adultos de más de 50 arios con colonoscopia comple- ta en los positivos,
3	Hepáticas múltiples), se solicitó colonoscopia con biop- sia de la muestra,
4	Paget. El TEGD e intestinal y una colonoscopia fueron normales. En la eco
5	ones. Con posterioridad se realizó colonoscopia hasta ángulo esplénico don
6	lon requiere de la realización de colonoscopia o de una simple rectosco
7	colon dilatado. Por su parte, la colonoscopia permite la visualización
8	sobre la conveniencia de efectuar colonoscopia total a personas asintomát

**German**

<b>Kolos- kopie</b>	<p><b>term type:</b> main entry term  <b>reliability code:</b> 10  <b>part of speech:</b> noun  <b>number:</b> singular  <b>gender:</b> feminine  <b>definition:</b> Endoskopie von Teilen des Dickdarms (partielle Koloskopie) oder des gesamten Dickdarms (totale Koloskopie) mittels eines flexiblen Faserglas- Endoskops, dem Koloskop. Länge des Instruments: 1m bei partieller u. 2m bei totaler Koloskopie. Durch die Flexibilität des Geräts gelingt es in vielen Fällen, den gesamten Dickdarm zu inspizieren, verdächtige Stellen zu fotografieren und durch Biopsie Materi- al für histologische Untersuchungen zu gewinnen. Dieses Verfahren dient zur Diagnose des Kolonkar- zinoms. Vorherige 24-st. Diät, vorherige Gabe eines Abführmittels, eventuell unter medikamentöser Sedie- rung. (de)</p>
<b>Dick- darms- piege- lung</b>	<p><b>definition:</b> Spiegelung des gesamten oder teil- weisen Dickdarms mit einem Spezial-Instrument, dem Koloskop. Mit dieser Untersuchung wird festgestellt, ob ein Dickdarmkrebs vorliegt. Das Instrument wird durch den After bis zum Grimm- darm vorgeschoben. Mit Möglichkeit zur Entnah- me von Gewebeproben, die unter einem Mikros- kop auf Krebszellen hin untersucht werden. Diät ab 24 St. vor der Untersuchung, vorherige Einnah- me eines Abführmittels, meist ohne Betäubung, Verabreichung eines beruhigenden und schläfrig machenden Medikaments. (de)  <b>gender:</b> feminine  <b>number:</b> singular  <b>part of speech:</b> noun  <b>reliability code:</b> 10  <b>term type:</b> synonym</p>

Figura 8. Fragmento de la entrada en HTML en torno al concepto COLONOSCOPY (<[www.ugr.es/~oncoter/codata/COLONOSCOPY.html](http://www.ugr.es/~oncoter/codata/COLONOSCOPY.html)>)

Al acceder al servidor de Oncoterm, encontraremos una lista de conceptos ordenada alfabéticamente. Si entramos en cualquiera de estos conceptos, por ejemplo, COLONOSCOPY, vamos a encontrar:

- una referencia a la categoría conceptual general a la que pertenece. En este caso, *Subject field: medicine: diagnostic procedure*;
- datos administrativos (p. ej., fecha de creación y revisión de la entrada o terminólogo);
- URL o imágenes que ayudan a que el usuario comprenda mejor ese concepto;
- un cuadro denominado «Conceptual Structures», que refleja el esquema por categoría del subdominio especializado donde se ubica el término *colonoscopy* (FIBEROPTIC-ENDOSCOPY) y los conceptos con los que está más relacionado, entre otros, FIBEROPTIC ENDOSCOPY, DIAGNOSTIC PROCEDURE, INVASIVE DIAGNOSTIC TEST, COLON, COLORECTAL CANCER O COLONOSCOPE;
- definiciones de ese concepto en inglés, español y alemán;
- información terminológica sobre cada una de las denominaciones del concepto: *colonoscopy* (inglés), *colonoscopia* (español), y *Koloskopie* y *Dickdarmspiegelung* (alemán). Esta información comprende: tipo de término, categoría gramatical, número, (género), grado de fiabilidad y líneas de concordancia (*context*).

Ilustramos un fragmento de esta entrada en la figura 8, donde se destacan los términos en español y alemán, y las líneas de concordancia en torno al término *colonoscopia* (véase figura 8 en pág. 237).

## 5. Conclusiones

Oncoterm es un proyecto innovador en el campo de la traducción y la terminología médica por combinar un marco teórico sólido con técnicas de análisis textual, recuperación y representación de información. En este artículo ha quedado patente la utilidad que tiene para el traductor extraer y representar conocimiento a partir de técnicas de corpus, dado que le permite una mayor comprensión del campo de especialidad y sus principales conceptos, así como un mayor conocimiento de su fraseología.

En cualquier caso, es importante que el análisis de corpus esté integrado con otros elementos: una ontología para el modelado y la representación conceptual, una base de datos terminológica vinculada a la ontología, recursos terminográficos, consulta a expertos y análisis de imágenes.

Este proyecto ha sido, asimismo, un punto de encuentro de investigadores en el campo de la traducción, la terminología, la documentación y la medicina, del cual se han derivado investigaciones sobre traducción y terminología médica o sobre la redacción de artículos experimentales en inglés y español.<sup>k</sup> La metodología desarrollada en Oncoterm está siendo aplicada con éxito en el proyecto de I+D «Ingeniería de puertos y cos-

tas: estructuración del conocimiento y generación de recursos terminológicos» (BFF2003-04720). Ha quedado de manifiesto, pues, la utilidad de este marco metodológico en otros dominios del saber.

## Notas

<sup>a</sup> Este trabajo versa sobre la educación científica que ha de tener la ciudadanía y recoge una investigación en la que se analizaron las diez palabras relacionadas con la ciencia que aparecían con mayor frecuencia en los medios de comunicación escrita, en concreto, en *El País*, *El Mundo*, el *ABC* y *La Voz de Galicia*. Pues bien, la tercera palabra más frecuente fue «cáncer».

<sup>b</sup> Una versión simplificada de este artículo ha aparecido en el volumen 2, número 7 (2006) de la publicación electrónica *Revistaesalud.com*.

<sup>c</sup> Ontoterm (<[www.ontoterm.com](http://www.ontoterm.com)>) es una herramienta desarrollada por Antonio Moreno Ortiz, de la Universidad de Málaga.

<sup>d</sup> McEnery y Wilson (1996: 24) definen corpus así: «a finite-sized body of machine-readable text, sampled in order to be maximally representative of the language variety under consideration».

<sup>e</sup> Programa creado por Mike Scott (<[www.lexically.net/wordsmith/index.html](http://www.lexically.net/wordsmith/index.html)>) de la Universidad de Liverpool y distribuido por la Oxford University Press.

<sup>f</sup> Entre estos centros de información oncológica destacamos *CancerNet*, *CancerBacup*, *Medscape*, *MedicineNet*, *Oncoweb*, *Virtual Hospital*, *Alcase*, *Atheneum* y *Diario Médico*.

<sup>g</sup> En inglés, se incluyen textos publicados en *British Medical Journal*, *Lancet*, *New England Journal of Medicine*, *Cancer*, *CANCER-LIT*, *C-A. A Cancer Journal for Clinicians*. Los textos en español provienen de *Medicina Clínica*, *Revista Clínica Española*, *Neoplasia*, *Revisión en Cáncer*, *Revista Española de Anestesiología y Reanimación*, *Archivos Bronconeumológicos*, *Revista Española de Enfermedades Digestivas*, *Anales Otorrinolaringológicos Ibero-Americanos*, *Anales Españoles de Pediatría* y *Actas Urológicas Españolas*.

<sup>h</sup> *Harrison's Principles of Internal Medicine*, *Cancer: Principles and Practice of Oncology*, *Medicina Interna de Farreras-Rozmán*, *Cáncer. Principios y Práctica de Oncología* y *Oncología Médica-Guía de Oncología Médica*.

<sup>i</sup> Entre las enciclopedias médicas destinadas a un público no versado en medicina podemos citar *The Merck Manual of Diagnosis and Therapy / Manual Merck en español* y *Mosby's Medical Encyclopedia for Health Consumers*. Hemos sacado asimismo textos de la *Enciclopedia Microsoft Encarta 97*. Los monográficos provienen también de publicaciones de divulgación semiespecializada, como *Scientific American / Investigación y Ciencia*, y de divulgación general, como *Oncolink*, *Reuters Health*, la revista *TIME*, *QUO* y suplementos de periódicos (*Blanco y negro*, *El Mundo*, *El Semanal* o *Ideal*).

<sup>j</sup> Tercedor (1999, apdo. 5.2.2.) profundiza en este tema. Aquí presentamos una selección de los ejemplos citados por esta autora.

<sup>k</sup> En este sentido, se pueden consultar los trabajos de López Rodríguez (2000 y 2002) y Reimerink (2006).

## Bibliografía

Alonso Ramos, M., A. Tutin y G. Lapalme. (1995): «Lexical functions in the Explanatory Combinatorial Dictionary for lexicalization in

- text generation». En P. Saint-Dizier y E. Viegas (eds.): *Computational Lexical Semantics*. Cambridge: Cambridge University, pp. 351-366.
- Faber, P. (1999): «Conceptual analysis and knowledge acquisition in scientific translation», *Terminologie et traduction*, 1999; (2): 97-123.
- (2002): «Investigar en terminología». En P. Faber y C. Jiménez (eds.): *Investigar en terminología*. Granada: Comares, pp. 3-23.
- (2004): «How to Become an Expert in Record Time». En F. Díaz Pérez y A. M. Ortega Cebreros (eds.): *A world of English, a world of Translation. Estudios Interdisciplinarios sobre Traducción y Lengua Inglesa*. Jaén: Servicio de publicaciones de la Universidad de Jaén.
- , C. I. López Rodríguez y M. I. Tercedor Sánchez (2001): «La utilización de técnicas de corpus en la representación del conocimiento médico», *Terminology*, 7 (2): 167-197.
- y R. Mairal Usón (1999): *Constructing a lexicon of English verbs*. Berlín: Mouton de Gruyter.
- Fillmore, C. J. (1982): «Frame semantics». En *Linguistics in the Morning Calm*. Seúl: Hanshin, pp. 111-137.
- , C. R. Johnson y Petruck, M. (2003): «Background to FrameNet», *International Journal of Lexicography*, 16 (3).
- García de Quesada, M. (2001): *Estructura definicional terminográfica en el subdominio de la oncología clínica*. Madrid: CSIC, Elies [<http://elies.rediris.es/elies14>].
- López Rodríguez, C. I. (2000): «Tipologías textuales y géneros en la normalización terminológica y ortotipográfica de la traducción médica», *Terminologie et Traduction*, 3: 78-97.
- (2001): *Tipologías textuales y cohesión en la traducción biomédica inglés-español: un estudio de corpus*. Granada: Universidad de Granada [[www.ugr.es/~dpto\\_ti/profesores/cilr-tesis.html](http://www.ugr.es/~dpto_ti/profesores/cilr-tesis.html)].
- (2002): «Extracción de información conceptual, textual y retórica en terminología: la distribución de verbos en los resúmenes de artículos experimentales». En P. Faber y C. Jiménez (eds.): *Investigar en terminología*. Granada: Comares, pp. 167-195.
- Mahesh, K., y S. Nirenburg (1995): «A situated ontology for practical NLP». En *Proceedings on basic ontological issues in knowledge sharing*. International Joint Conference on Artificial Intelligence (UCAI-1995), agosto de 1995, Montreal (Canada).
- Martín Mingorance, L. (1984): «Lexical fields and stepwise lexical decomposition in a contrastive English-Spanish verb valency dictionary». En Hartmann (ed.): *LEXeter 83: Proceedings of the International Conference on Lexicography*. Tübinga: Niemeyer, pp. 226-236.
- (1989): «Functional Grammar and Lexematics». En J. Tomaszczuk y B. Lewandowska (eds.): *Meaning and Lexicography*. Amsterdam, Filadelfia: John Benjamins, pp. 227-253.
- (1995): «Lexical logic and structural semantics: methodological underpinnings in the structuring of a lexical database for natural language processing». En Hoinkes (eds.), *Panorama der Lexikalischen Semantik*. Tübinga: Gunter Narr, 461-474.
- Moreno Ortiz, A. (2000a): «Managing conceptual and terminological information in a user friendly environment». En *Proceedings of OntoLex 2000*. Workshop on Ontologies and Lexical Knowledge Bases. Septiembre del 2000, Sofía (Bulgaria).
- (2000b): «OntoTerm: un sistema abierto de representación conceptual». *Actas del XVI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)*. Octubre del 2000, Vigo (España).
- y C. Pérez Hernández (2000): «Reusing the Mikrokosmos ontology for concept-based multilingual terminology databases». *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC 2000)*. Junio del 2000, Atenas (Grecia), pp. 1061-1067.
- Nelson, K. (1985): *Making Sense: the acquisition of shared meaning*. Orlando: Academic.
- Nirenburg, S. (ed.) (1987): *Machine Translation: Theoretical and Methodological Issues*. Cambridge: Cambridge University.
- Núñez, Ramón (2003): «Palabras científicas en los titulares de prensa». En Reyes Requera (ed.): *Ciencia, tecnología y lengua española: la terminología científica en español*. Madrid: FECYT, pp. 172-176. [[www.fecyt.es/doc/Ponencias%20CTL.pdf](http://www.fecyt.es/doc/Ponencias%20CTL.pdf)].
- Oncoterm (2002): «Sistema Bilingüe de Información y Recursos Oncológicos». [[www.ugr.es/~oncoterm/oncodesc.htm](http://www.ugr.es/~oncoterm/oncodesc.htm)].
- Pérez Hernández, M. C. (2000): *Explotación de los corpora textuales informatizados para la creación de bases terminológicas basadas en el conocimiento*. Madrid: CSIC, Elies, vol. 18. [<http://elies.rediris.es/elies18>].
- Reimerink, A. (en prensa): *Redactar y traducir artículos de investigación: un programa de software*. Tesis doctoral de la Universidad de Granada.
- Seibel, C., y C. Jiménez Hurtado (2002): «La Pragmática de la terminología: en busca del perfil del usuario». En P. Faber y C. Jiménez (eds.): *Investigar en terminología*. Granada: Comares.
- Semir, Vladimir de (2003): «La comunicación de la ciencia en la sociedad del conocimiento». En Reyes Requera (ed.): *Ciencia, tecnología y lengua española: la terminología científica en español*. Madrid: FECYT, pp. 189-199. [[www.fecyt.es/doc/Ponencias%20CTL.pdf](http://www.fecyt.es/doc/Ponencias%20CTL.pdf)].
- Tercedor Sánchez, M. I. (1999): *La fraseología en el lenguaje biomédico: análisis desde las necesidades del traductor*. Madrid: CSIC, Elies, vol 6. [<http://elies.rediris.es/elies6/>].
- Viegas, E.; K. Mahesh; S. Nirenburg y S. Beale (1999): «Semantics in Action». En Saint-Dizier (ed.): *Predicative Forms in Natural Language and Lexical Knowledge Bases*. Dordrecht: Kluwer, pp. 171-204.

