

GESTIÓN TERMINOLÓGICA BASADA EN EL CONOCIMIENTO Y GENERACIÓN DE RECURSOS DE INFORMACIÓN SOBRE EL CÁNCER: EL PROYECTO ONCOTERM¹

Clara Inés López Rodríguez²

Maribel Tercedor Sánchez²

Pamela Faber Benítez²

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 todos los demás componentes y constituye el vínculo entre términos en diferentes lenguas (inglés, español y alemán). El sistema está destinado tanto a profesionales de la comunicación y de la salud como a pacientes, utilizando tanto la información extraída de diccionarios y de corpus de textos especializados como la proporcionada por expertos. En la gestión terminológica basada en el conocimiento, es muy importante la identificación de la organización conceptual subyacente al campo de especialidad y de la multidimensional de las representaciones conceptuales.

1. INTRODUCCIÓN

La necesidad de dar a conocer conceptos especializados relativos a una enfermedad a los pacientes y a sus familiares responde al deseo de la población de acceder a la información. En este contexto, Internet constituye el marco de difusión y actualización de

1 Oncoterm: Sistema bilingüe de información y recursos oncológicos (PB98-1342)

2 Departamento de Traducción e Interpretación. Universidad de Granada. España

información idóneo. Knight (2003) describe la tendencia al cripticismo en los artículos científicos que lleva a soluciones como:

- Explicaciones on-line de los artículos (*Science*)
- Revisión de artículos en lenguaje asequible al científico no experto (*Nature, Science, Development, Journal of Cell Science*)
- Hipervínculos desde el término hacia glosarios o sitios web con mayor 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 se llevó a cabo entre 1999 y 2002 por un equipo multidisciplinar formado por investigadores de las Universidades de Granada, Málaga y Valladolid, y del Hospital Universitario Virgen de las Nieves de Granada*. El principal resultado del proyecto ha sido una base de datos con información sobre el cáncer en Internet a varios niveles.

En este artículo describimos los objetivos de Oncoterm y mencionamos brevemente sus fundamentos teóricos para una gestión de la terminología basada en el conocimiento. Asimismo, explicamos la metodología utilizada en la extracción de información sobre el cáncer y en la creación de un marco conceptual de referencia donde encajan todos los términos de nuestra base de datos. Por último, resumimos los resultados del proyecto, disponibles en Internet desde la página <http://www.ugr.es/local/oncoterm>.

* El grupo de investigación que participó en OncoTerm está compuesto por los siguientes miembros: Pamela Faber (investigadora principal), Catalina Jiménez Hurtado, Manuel Jiménez Raya, Miguel Vega Expósito, Carlos Márquez Linares, Clara López Rodríguez, María Isabel Tercedor Sánchez y Claudia Seibel de la Universidad de Granada; Chantal Pérez Hernández y Antonio Moreno Ortiz de la Universidad de Málaga; Mercedes García Quesada de la Universidad de Valladolid; Ricardo Mairal Usón de la Universidad Nacional de Educación a Distancia; y Mercedes Rodríguez del Castillo y Pedro Ballesteros García del Hospital Universitario Virgen de la Nieves. También colaboraron en el proyecto como becarias Arianne Reimerink y M^a Dolores Cerrato Escobar.

1.1. Objetivos del proyecto Oncoterm

Aunque son muchas las revistas, bases de datos y sitios web 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 desde 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 2.4.) y, en particular, del EVENTO MÉDICO ONCOLÓGICO
- Especificar un lenguaje definicional conciso, consistente y controlado que permita la ubicación de un concepto dentro de su categoría y determinar sus relaciones con el resto de 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 la estructuración de la información, su almacenamiento y asegurara una interfaz usuario-sistema de fácil utilización. En este sentido se desarrolló Ontoterm®, una aplicación informática que permite manejar y gestionar grandes cantidades de conceptos y términos
- Especificar esquemas categoriales (*category templates*) que reflejen la organización interna de cada categoría del conocimiento sobre oncología, de modo que sirviera de eje de desarrollo de la base de datos

- Incluir de forma sistemática información pragmática de utilidad para el médico y el paciente dentro de las entradas terminológicas

1.2. 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 Mingorance (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 colocacionales 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. Asimismo, 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.

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, como se aprecia en la Figura 5, donde el concepto COLONOSCOPY evoca otros conceptos, entre otros: FIBEROPTIC ENDOSCOPY, DIAGNOSTIC PROCEDURE, INVASIVE DIAGNOSTIC TEST, COLON, COLORECTAL CANCER O COLONOSCOPE.

2. METODOLOGÍA

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 profesionales de la comunicación aumenten su comprensión sobre el cáncer. 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 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 relaciones entre 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 tanto con objetos como BRONCHIUM (mediante la relación AFFECTS-BODY-PART) o ENDOSCOPE (mediante la relación USES) como con eventos tal y 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 extracción de información del corpus textual en soporte electrónico, de la consulta a diccionarios y a expertos, y del análisis de imágenes. El corpus se analiza con la herramienta de análisis léxico *Wordsmith Tools*³, aunque esta 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 Programa creado por Mike Scott [<http://www.lexically.net/wordsmith/index.html>] de la University of Liverpool y distribuido por Oxford University Press.

2.1. Recopilación del corpus en soporte electrónico

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 definicional (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, de la salud y público general interesado en la oncología, incluimos textos redactados por especialistas para un público con distintos niveles de conocimiento especializado.

Y así, se recopilaron páginas web de las principales organizaciones internacionales contra el cáncer⁴, que ofrecen, tanto a 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 oncológicas⁵, y fragmentos relevantes de manuales destinados a oncólogos y estudiantes de medicina de nivel avanzado⁶. Por último, para reflejar la comunicación entre el especialista y un público lego, se recopilaron no sólo monográficos que aparecen en

4 Entre estos centros de información oncológica destacamos CancerNet, CancerBacup, Medscape, MedicineNet, Oncoweb, Virtual Hospital, Alcasa, Atheneum y Diario Médico.

5 En inglés, se incluyen textos publicados en *British Medical Journal*, *Lancet*, *New England Journal of Medicine*, *Cancer*, *CANCERLIT*, *C-A. A Cancer Journal for Clinicians*. Los textos en español provienen de *Medicina Clínica*, *Revista Clínica Española*, *Neoplasia*, *Revisiones 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*.

6 *Harrison's Principles of Internal Medicine*, *Cancer: Principles and Practice of Oncology*, *Medicina Interna de Farreras-Rozmán*, *Cancer. Principios y Práctica de Oncología y Oncología Médica-Guía de Oncología Médica*.

enciclopedias y publicaciones de divulgación semiespecializada y general⁷, sino también folletos de salud pública para pacientes y familiares disponibles en los sitios web antes mencionados.

2.2. Lista de conceptos y términos

Para elaborar una lista preliminar de los conceptos más importantes en oncología, estudiamos tanto recursos terminográficos como algunos de los textos recopilados (véase 2.1.), hicimos consultas a oncólogos y analizamos el corpus de textos para extraer listas de frecuencia lematizadas, líneas de concordancia y palabras clave. Con las lecturas también se fue elaborando un borrador del marco conceptual subyacente al dominio.

2.2.1. 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 (Figura 1), 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. 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.

⁷ 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 también 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, tales como *Oncolink*, *Reuters Health*, la revista *TIME*, *QUO* y suplementos de periódicos (*Blanco y negro*, *El Mundo*, *El Semanal* o *Ideal*).

N	PALABRA	FREC.	%	LEMAS
1	PATIENTS	175.546	0,61	patient(23100),patient's(2115)
2	CELLS	144.895	0,50	cell(65679),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(28489)
10	GROUP	54.309	0,19	groups (17340)
11	EFFECTS	52.790	0,18	effect(25848)
12	TUMOR	50.779	0,18	tumors(17569),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(13187)
15	SIGNIFICANTLY	47.642	0,17	significant(23819)
16	INCREASED	44.978	0,16	increased (17955) , increases (215), increasing (64)
17	CLINICAL	42.253	0,15	
18	CASE	40.768	0,14	cases(26873)
19	YEARS	40.575	0,14	year(14623)
20	THERAPY	39.985	0,14	therapies(1528),therapeutic(5451),therapeutic al(145), therapeutics(238), therapic(2)

Figura 1. Cabecera de la lista de frecuencia lematizada del corpus en inglés de Oncoterm

2.2.2. 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 aparecen las palabras con las que suele aparecer en los textos.

Asimismo, las concordancias proporcionan un esquema básico de relaciones conceptuales sobre el cual puede modelarse el resto del subdominio. Las concordancias en la Figura 2 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, la extensión del campo enfocado, así como la dosis e intensidad de la radiación. Queda pues patente la multidimensionalidad del término *radiation therapy* en el campo de la oncología, dado que se puede clasificar atendiendo a las diferentes perspectivas mencionadas arriba.

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

Figura 2. Parametrización del conocimiento pertinente al concepto radiation therapy

2.3. Creación de un gestor de información

Uno de los aspectos esenciales era contar con un gestor de información para almacenar datos, que proporcionara asimismo la interfaz con el usuario en la web. Ontoterm [<http://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.

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.

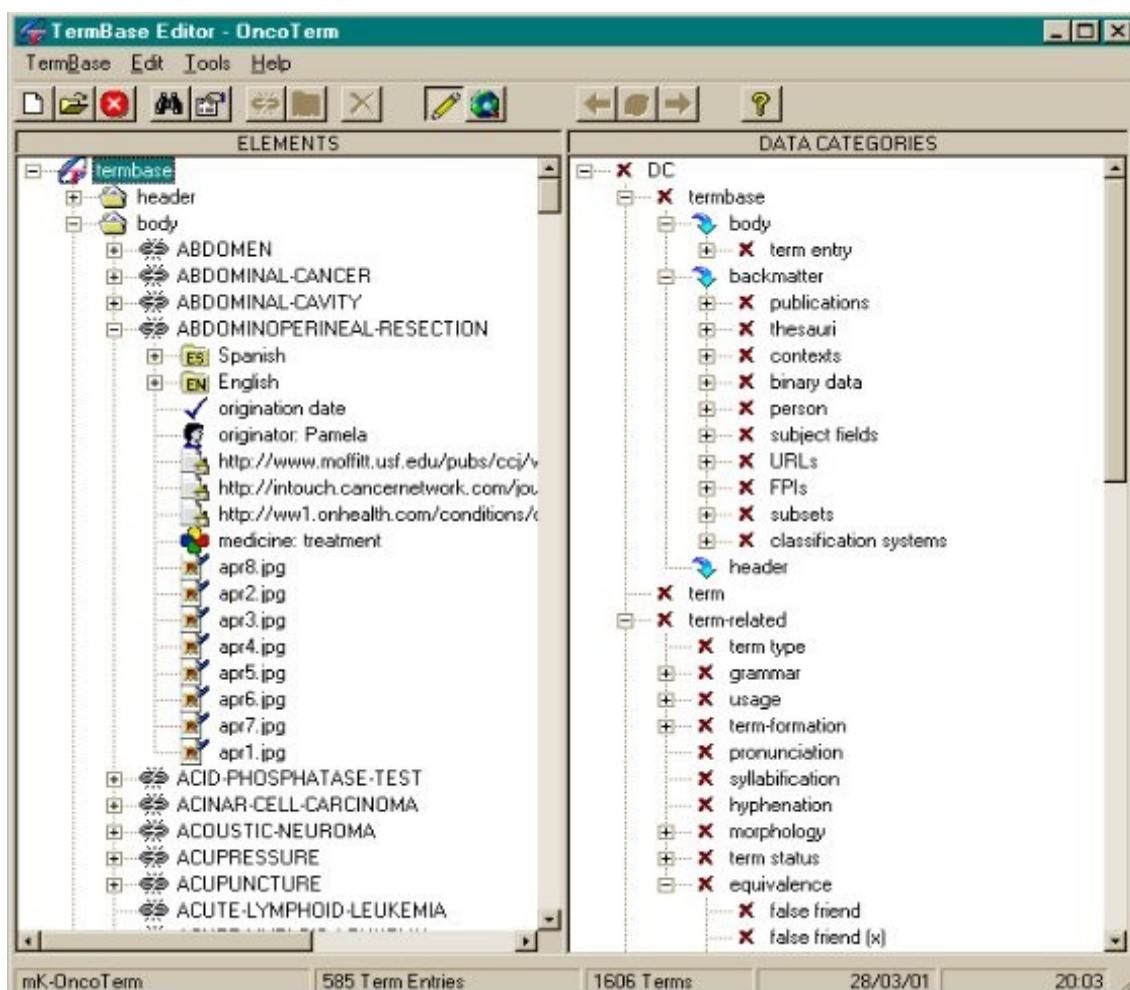


Figura 3. Editor de la base de datos terminológica

2.4. Elaboración de un marco conceptual de referencia

Para elaborar un marco conceptual de referencia es preciso establecer jerarquías de términos mediante el análisis de definiciones lexicográficas y del corpus, dado que partimos de la premisa de que la información contenida en diccionarios y textos especializados constituye una red léxico-conceptual que necesariamente tiene correspondencia con el conocimiento expresado. Por ejemplo, si quisiéramos definir el término *intraoperative 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

Figura 2). A partir de ahí se elaboraría una jerarquía en la que cada concepto se define mediante uno semánticamente más simple; este a su vez, se reduce a uno más simple y así respectivamente, como se ve a continuación:

{intraoperative radiation therapy}@ → {intracavitary radiation therapy}@ →
 {internal radiation therapy}@ → {radiation therapy}@ → {treatment}@ →
 {event}@ → {ALL}

Por tanto, términos como *intraoperative radiation therapy* e *internal radiation therapy* están en una relación de subordinación con *treatment* y *event*. Esto facilita la ubicación del campo especializado dentro de un contexto más amplio.

Con este método, pudimos identificar estas grandes categorías conceptuales: FACTOR DE RIESGO, PARTE_DEL_CUERPO, TUMOR, SÍNTOMA, PROCEDIMIENTO DIAGNÓSTICO, INSTRUMENTO, TRATAMIENTO, MEDICAMENTO, EFECTO SECUNDARIO, ESPECIALISTA y HOSPITALRADIOTHERAPY. 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: 8) denomina EVENTO MÉDICO ONCOLÓGICO:

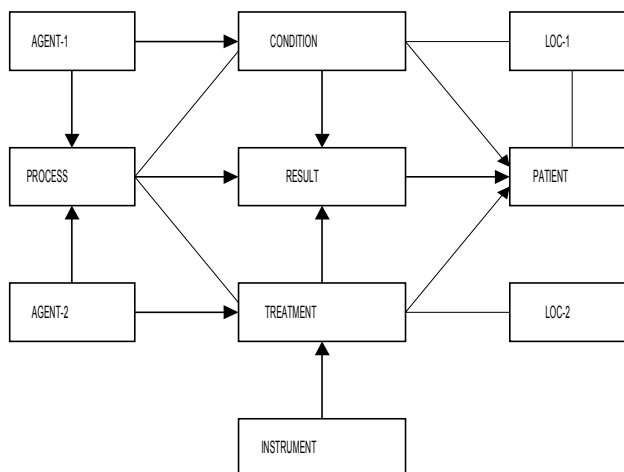


Figura 4. Evento médico oncológico

Como puede constatarse, los términos superordinados de cada categoría pertenecen a la lengua general, aunque sus niveles más específicos son propios de la terminología.

La estructura interna de cada categoría conceptual quedó representada por un conjunto de tipos de información sobre las características de un concepto (su intensión) 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 categoriales, se fue modelando la ontología, insertando cada concepto en el marco conceptual **EVENTO MÉDICO** y estableciendo relaciones entre conceptos, siendo la relación principal 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 Figura 5, en relación al concepto **COLONOSCOPY**:

Conceptual Structures	
ISA	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 : COLORECTAL-CANCER INSTRUMENT : COLONOSCOPE PURPOSE : BIOPSY

Figura 5. Relaciones en torno al concepto colonoscopy

2.5. Elaboración de 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 categoriales
- 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 de 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. Y así, *intraperitoneal radiation therapy* se define en función de su hiperónimo inmediatamente superior (*intracavitary radiation therapy*), y así respectivamente hasta llegar a *RADIATION THERAPY*, que es un subdominio de *TREATMENT*.

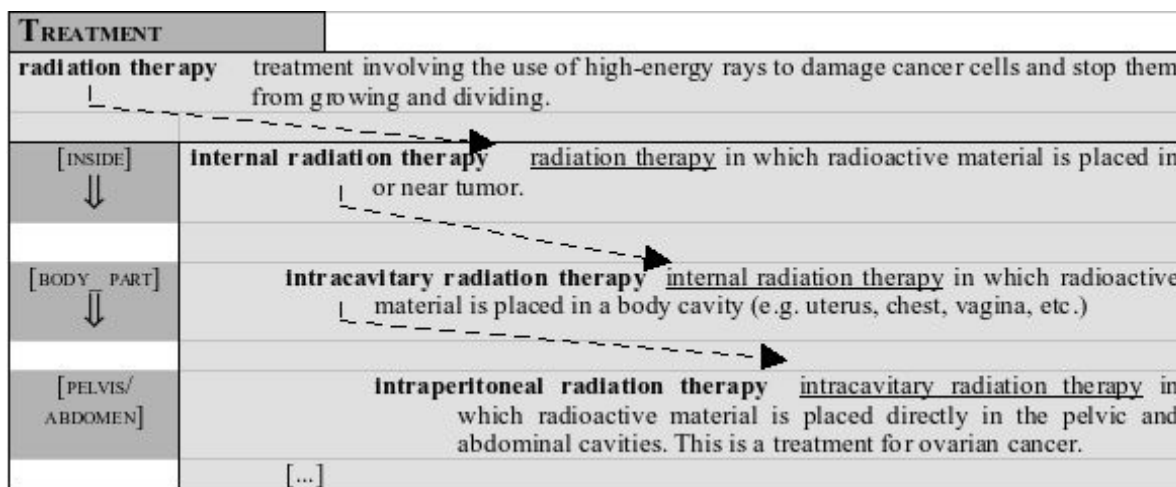


Figura 6. Jerarquía conceptual de radiation therapy reflejada en las definiciones de los términos (adaptado de Faber, López y Tercedor 2001: 184)

En las definiciones se percibe que cada concepto posee un esquema categorial 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 Figura 7. 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 *intraoperative radiation therapy* IS-A *internal radiation therapy* IS-A *radiation therapy* IS-A *treatment*. Igualmente, *intraoperative radiation therapy* AFFECTS *abdomen/pelvis*, con lo que se hereda el esquema *radiation therapy* AFFECTS *body part* (Figura 8).

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

Figura 7. Esquema categorial de treatment (Faber, López y Tercedor 2001: 186)

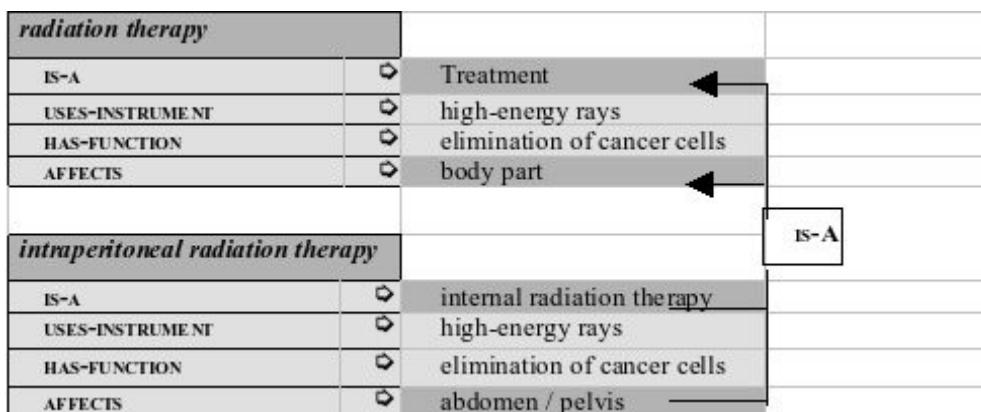


Figura 8. Aplicación del esquema categorial treatment a radiation therapy y intraoperative radiation therapy (ib.)

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

3. RESULTADOS DEL PROYECTO

En oncoterm (2002)⁸ 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, características e

8 Disponible desde: <http://www.ugr.es/~oncoterm/oncodesc.htm>.

interrelaciones. Esta estructura conceptual constituye el vínculo entre términos en diferentes lenguas, ya que se trata de información no específica a ninguna lengua. Asimismo, con la gestión terminológica basada en el conocimiento se identificaron esquemas de la estructura interna de cada categoría y su aplicación a conceptos más específicos, así como a la confección de definiciones terminográficas concisas y sistemáticas que reflejen la estructura conceptual del dominio de la oncología.

Los resultados del proyecto están ubicados en Internet [<http://www.ugr.es/-oncoterm>] y son de utilidad para profesionales de la salud, investigadores, pacientes y familiares, así como para traductores y redactores de textos especializados, además de ser una herramienta que contribuye a hacer más eficiente el trabajo, dado que facilita el acceso on-line a la información solicitada en muy poco tiempo y asegura su calidad, gracias al sistema de búsqueda y organización conceptual que permite recuperar no sólo el término objeto de la búsqueda, sino aquellos relacionados conceptualmente con este. La web cuenta con una base de datos jerarquizados con 1.896 conceptos relacionados con el cáncer y 4.033 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.

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 categorial del subdominio especializado donde se ubica el término *colonoscopy* (FIBEROPTIC-ENDOSCOPY) y los términos con los que está más relacionado
- Definiciones de ese concepto en inglés, español y alemán
- Información terminológica sobre cada una de las denominaciones de ese concepto: *colonoscopy* (inglés), *colonoscopia* (en español), y *Koloskopie* y *Dickdarmspiegelung* (en alemán). Esta información comprende: tipo de término, categoría gramatical, número, (género), grado de fiabilidad y líneas de concordancia (*context*)

4. CONCLUSIONES

Oncoterm es un proyecto innovador en el campo de la terminología médica por las técnicas de análisis, recuperación y representación de información utilizadas. Emplea un gestor de información terminológica que almacena información en una auténtica base de conocimiento, relacionando conceptos y estableciendo relaciones entre estos. Son muchas las asociaciones que han facilitado enlaces a Oncoterm por parte de sociedades médicas del territorio español e internacional como la Sociedad Asturiana de Patología Respiratoria, la Asociación Gallega contra enfermedades neuromusculares.

Desde que se finalizó el proyecto en 2002 se han detectado algunas incorrecciones e hipervínculos externos que han dejado de funcionar. Con esto se pone de manifiesto la situación de muchos recursos de información sobre salud que, por falta de financiación y medios, no se actualizan. Esperamos contar con apoyo humano y técnico para implementar un sistema tipo Wikipedia para que especialistas y pacientes puedan hacer sugerencias de mejora.

Por último, este proyecto ha sido 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. Esta última faceta ha culminado con una aplicación

de software para redactar artículos conforme a las normas de estilo de las principales revistas médicas (Reimerink 2006).

5. REFERENCIAS BIBLIOGRÁFICAS

1. 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. 3-23.
—, 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*. Berlin: Mouton de Gruyter.
2. Fillmore, C. J. 1982. "Frame semantics". En *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co, 111-137.
— C. R., Johnson y Petruck, M. 2003. "Background to FrameNet". En *International Journal of Lexicography*. vol. 16 (3). Oxford Journals.
3. García de Quesada, M. 2001. *Estructura definicional terminográfica en el subdominio de la oncología clínica*. Madrid: CSIC/Elies. [<http://elies.rediris.es/elies14>].
4. Knight, Jonathan. 2003. "Scientific literacy: clear as mud". *Nature* 423, 376-378 (22 May 2003) | doi: 10.1038/423376a.
5. López Rodríguez, C. I. 2001. *Tipologías textuales y cohesión en la traducción biomédica inglés-español: un estudio de corpus*. Granada: Editorial Universidad de Granada. [http://www.ugr.es/~dpto_ti/profesores/cilr-tesis.html].
6. 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), August 1995, Montreal, Canada.

7. Martín Mingorance, L. 1984. "Lexical fields and stepwise lexical decomposition in a contrastive English-Spanish verb valency dictionary". En Hartmamann (ed.), LEXeter 83: Proceedings of the International Conference on Lexicography. Tübingen: Niemeyer, 226-236.
- 1989. "Functional Grammar and Lexematics". En Tomaszczyk, J. y B. Lewandowska (eds.), Meaning and Lexicography. Amsterdam / Philadelphia: John Benjamins, 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. Tubinga: Gunter Narr, 461-474.
8. Moreno Ortiz, A. 2000a. "Managing conceptual and terminological information in a user friendly environment". Proceedings of OntoLex 2000. Workshop on Ontologies and Lexical Knowledge Bases. Septiembre 2000, Sofía, Bulgaria.
9. Moreno Ortiz, A. 2000b. "OntoTerm: un sistema abierto de representación conceptual". Actas del XVI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). Octubre 2000, Vigo, España.
10. Moreno Ortiz, A. 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 2000, Atenas, Grecia, 1061-1067.
11. Nirenburg, S. (ed.) 1987. Machine Translation: Theoretical and Methodological Issues. Cambridge: Cambridge University Press.
12. Pérez Hernández, M.C. 2000. Explotación de los corpora textuales informatizados para la creación de bases terminológicas basadas en el conocimiento. Tesis doctoral. Universidad de Málaga. [<http://elies.rediris.es/elies18>].
13. Reimerink, A. (en prensa). Redactar y traducir artículos de investigación: un programa de software. Tesis doctoral de la Universidad de Granada.
14. Seibel, C. y C. Jiménez Hurtado. 2002. "La Pragmática de la terminología: en busca del perfil del usuario". En En P. Faber y C. Jiménez (eds.) Investigar en terminología. Granada: Comares.

15. 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/>].
16. Viegas, E., Mahesh, K., Nirenburg, S. & Beale, S. 1999. "Semantics in Action". En Saint-Dizier (ed.), *Predicative Forms in Natural Language and Lexical Knowledge Bases*. Dordrecht: Kluwer, 171-204.