Sobre la creación y el mantenimiento de una base de datos terminológica multilingüe en una organización internacional[1]

NELSON VERÁSTEGUI
Unión Internacional de Telecomunicaciones [2]

Introducción

Fundada en París en 1865 con el nombre de Unión Telegráfica Internacional, la Unión Internacional de Telecomunicaciones (UIT) recibió su nombre actual en 1934 y es desde 1947 un organismo especializado de las Naciones Unidas.

La UIT ofrece a sus 190 Estados Miembros y más de 600 Miembros de los Sectores (empresas científicas e industriales, operadores de los sectores público y privado, empresas de radiodifusión y organizaciones regionales e internacionales) un marco para cooperar en la mejora y la utilización racional de las telecomunicaciones en todo el mundo.

En suma, la misión de la Unión cubre las siguientes áreas:

- Área técnica: promueve el desarrollo, la explotación eficiente, la utilidad y la disponibilidad general de los servicios e instalaciones de telecomunicaciones.
- Área de desarrollo: promueve el desarrollo de las telecomunicaciones en los países en desarrollo y la extensión de los beneficios de las telecomunicaciones a todo el mundo.
- Área de política: promueve la adopción de un enfoque más amplio de los temas de telecomunicación en la economía y la sociedad mundiales de la información.

---

1 Las opiniones aquí expresadas son las del autor y no reflejan necesariamente las opiniones de la Unión Internacional de Telecomunicaciones. Además, el autor es el único responsable de la exactitud de las informaciones contenidas en el presente artículo.

2 Sección de Terminología, Referencias y Ayudas Informáticas para la Traducción (STRAIT)

En la UIT, la Secretaría General gestiona los aspectos administrativos y financieros de las actividades de la Unión y, concretamente, la prestación de servicios de conferencias y de información, la planificación estratégica a largo plazo y otras actividades relacionadas con la Unión (comunicación, asuntos jurídicos, finanzas, personal y servicios comunes).

En cumplimiento de su misión, el Departamento de Conferencias y Publicaciones (C&P) proporciona el apoyo constante y los instrumentos que necesitan los Miembros para reunirse, comunicarse, intercambiar opiniones y difundir información sobre actividades de telecomunicaciones, de conformidad con la Constitución y el Convenio de la UIT. En particular, sus servicios dan apoyo a las conferencias y reuniones dentro y fuera de Ginebra y facilitan servicios lingüísticos y de documentación.

La UIT ha iniciado una transición gradual hacia la plena utilización de todos sus idiomas oficiales y de trabajo. Las secciones de traducción al inglés, árabe, chino, español, francés y ruso produjeron en 2005 del orden de 60 000 páginas (es decir unos 19 millones y medio de palabras o 3 200 documentos originales), casi 20 % de las cuales hacia el español. Aproximadamente 5 % de los documentos originales vienen en este idioma y 77 % en inglés. El personal lingüístico permanente y temporero del Departamento C&P comprende en los seis idiomas, fuera de los servicios de interpretación y dependiendo de la carga de trabajo, unas 78 personas en la traducción, 88 en la dactilografía y 8 en la Sección de Terminología, Referencias y Ayudas Informáticas para la Traducción (STRAIT). Estos últimos datos corresponden a un período de conferencia que normalmente consume muchos recursos.

La traducción en la UIT es esencialmente técnica, con un pequeño porcentaje de documentos administrativos o jurídicos. Puede haber textos sorprendentes, como los que tratan de protección de mediciones radioastronómicas en relación con la vida extraterrestre o el siguiente extracto que concierne la biología.

> *Termitas y hormigas*
> (en *Protección de los cables y las plantas de telecomunicaciones contra los ataques biológicos,* Recomendación UIT-T L.46)
> Unos orificios irregulares que pueden extenderse a lo largo de toda la circunferencia y que dan la sensación de que la cubierta ha sido pelada, son producidos normalmente por termitas y hormigas. Un aspecto mate en una superficie roída es consecuencia típica de un ataque de termitas que producen ese efecto con sus cortas y poderosas mandíbulas. Unas marcas de raspados transversales en el cable son características de las hormigas, que tienen unas mandíbulas largas y afiladas que producen esos cortes. El principio del ataque se caracteriza por la agrupación

local de numerosos orificios de pequeño tamaño. Las hormigas y las termitas pueden secretar un líquido ácido que corroe el plomo.

Terminología

Los traductores de la UIT tienen acceso directo a la terminología de telecomunicaciones que se ha acumulado en bases de datos desde hace más de 30 años y que contiene equivalentes a todo tipo de términos, en particular, a aquellos muy recientes y quizá efímeros. De ahí que en ellas figuren tanto entradas terminológicas oficiales como otras en vías de revisión, a veces incompletas, pero también útiles por razones prácticas. Es el resultado principalmente del análisis de documentos de trabajo y de publicaciones de la UIT, de contribuciones de los traductores y de ficheros terminológicos proporcionados por otras oficinas de la UIT.

Desde el exterior, la dirección de la página es <http://www.itu.int/terminology/index.html>. La base cubre las necesidades de los terminólogos, traductores, intérpretes y redactores ya sean de la UIT o de otras organizaciones internacionales.

De acceso libre a través de Internet, TERMITE 6L (68 500 entradas aproximadamente; su nombre es un acrónimo por terminología de telecomunicaciones para los seis idiomas) contiene todos los términos que figuran en los glosarios precedentemente publicados y es objeto de una actualización constante. Aunque se centra esencialmente en la terminología de telecomunicaciones relacionada con las normas técnicas, la base de datos abarca también otras especialidades técnicas y temas administrativos y financieros de interés para la estructura y el funcionamiento de la UIT.

Alrededor de 30 % de las entradas son específicas del Sector de Normalización de las Telecomunicaciones (UIT-T, antiguamente CCITT), 16 %, del Sector de Radiocomunicaciones (UIT-R, antiguamente CCIR); el resto es específico de la UIT o del área de las telecomunicaciones. Los términos están consolidados en conceptos o entradas multilingües de las cuales alrededor del 25 % tienen más de tres idiomas, 9,5 % cubre cinco idiomas, mientras que casi 2 000 entradas tienen los seis idiomas oficiales (2,81 %). Unas pocas llevan también equivalentes en otros idiomas como italiano, alemán y portugués.

El número total de términos, sinónimos y abreviaturas, a mediados de mayo del 2006, era de 262 936, con la siguiente distribución por idioma: inglés 96 751, francés 68 366, español 62 519, árabe 15 027, ruso 10 542 y chino 9 731.

La estructura de la base es de tipo relacional, organizada por conceptos, y sus componentes con relaciones complejas N a N.

La vida útil de cada ficha es muy variable, pues se mantienen mientras los términos sean válidos y corrientes. Cerca de 4 000 conceptos se modifican al año (ya sea por creación, cambio o supresión).

Ejemplo de ficha terminológica

Serial number: MC0274

Spanish: satélite de adquisición de datos – *Reference*: UIT- R, Rec. V.573-4 (2000), H34

English: data collection satellite – *Reference*: ITU- R, Rec. V.573-4 (2000), H34

Arabic: **ساتل لتجميع البيانات** – *Reference*: **الاتحاد الدولي للاتصالات، قطاع الاتصالات الراديوية، التوصية** V.573-4 (2000)

Chinese: 数据收集卫星 – *Reference*: ITU- R, Rec. V.573-4 (2000), H34

French: satellite de collecte de données – *Reference*: UIT- R, Rec. V.573-4 (2000), H34

Russian: спутник сбора *данных – *Reference*: MKKP Pek.573-2

Definitions - Notes

English definition: A satellite whose main purpose is the collection of data from stations on the Earth or in the atmosphere of the Earth, and subsequent forwarding of those data to one or more earth stations. It may also provide for communication in the other direction. – *Reference*: ITU-R, Rec. V.573-4 (2000), H34

French definition: Satellite destiné à la collecte de données provenant de stations situées à la surface de la Terre ou dans son atmosphère, en vue de retransmettre ces données à une ou plusieurs stations terriennes. Un tel satellite peut aussi établir des communications dans l'autre sens, d'une station terrienne vers les stations à terre. – *Reference*: UIT- R, Rec. V.573-4 (2000), H34

Spanish definition: Satélite destinado a la adquisición de datos procedentes de estaciones situadas en la superficie de la Tierra o en su atmósfera, para retransmitir esos datos a una o varias estaciones terrenas. Este satélite puede también establecer comunicaciones en el otro sentido, es decir, de una estación terrena hacia las estaciones en Tierra. – *Reference*: UIT- R, Rec. V.573-4 (2000), H34

Chinese definition:

这种卫星的主要目的是从地球上的或在地球大气中的各站收集数据，而后把这些数据转发给一个或多个地球站。它也可供其他方向上的通信用 – *Reference*: ITU- R, Rec. V.573-4 (2000), H34

Uno de los dolores de cabeza de intérpretes y traductores son las siglas, acrónimos y abreviaturas tan comunes en documentos técnicos y que muchas veces no se traducen. Por este motivo la base contiene casi 12 000 entradas monolingües en inglés. Mientras más corta es su longitud, más posibilidades de interpretación existen. CE, por ejemplo, puede corresponder según el contexto a términos tan variados como: comunicación establecida, elemento de conexión, campo de estado, código de enclavamiento, método de capacitación equivalente, Comisión de Estudios, Consejo Ejecutivo, Comisión Europea, Comunidades Europeas o Consejo de Europa. Otro ejemplo sería CP, en inglés, del que tenemos más de veinte fichas que van desde «equipo de proceso de canal, incluido el equipo de terminación de señalización» hasta «trayecto de los caracteres», por ejemplo.

Hay muchas abreviaturas que no cambian entre el inglés y el español (AARL, ACCH, ACELP, ADPCM, ADSL, ATM, BPKM o UNESCO); casos como estos aparecen en casi 8 000 entradas. Puede ser casual o adrede para facilitar la identificación. A veces por razones históricas o desconocidas un acrónimo no corresponde evidentemente al término entero, como por ejemplo en COMTELCA, que se usa igual en varios idiomas y corresponde a Comisión Técnica Regional de Telecomunicaciones, en español, y a *Regional Technical Committee for Telecommunications*, en inglés, pues el -CA del final del acrónimo tiene reminiscencias de un antiguo título que incluía la palabra Centroamérica.

Los siguientes son ejemplos de cambio de sigla:

- *IPSAS: International Public Sector Accounting Standards*, que corresponde a NICSP: Normas Internacionales de Contabilidad para el Sector Público
- *SIDA: Swedish International Development Cooperation Agency*, que corresponde a ASDI: Agencia Sueca de Cooperación Internacional para el Desarrollo.

La terminología evoluciona paulatinamente y las traducciones y neologismos que se pudieron encontrar en un comienzo son reemplazados por otros más apropiados o más corrientes. Por ejemplo, desde que apareció el concepto informático de *plug and play* a mediados de los noventa, ha pasado mucha agua bajo el puente. Se trata de una tecnología que permite a los usuarios añadir nuevas posibilidades a sus PC insertando tarjetas o dispositivos adicionales, sin preocuparse de complicados problemas de configuración del sistema, pues el proceso es automático y hay menos rechazos que antes, cuando esta operación compleja era solo

manual. En la UIT he encontrado varias traducciones como «enchufe activo», «sin instalación» o «sin preparativos». En Google, encuentro otras como la literal «enchufe y opere», «enchufa y va», la abreviatura PnP, «listo para cable digital» y hasta la más cómica, «enchufa y reza». Probablemente lo que más se diga en nuestro idioma es «plújanplai» o «plagan plei», aunque no se escriba así.

Otro problema frecuente en la traducción del inglés al español es el de la interpretación de apilamientos de adjetivos y sustantivos pues en nuestro idioma tenemos que añadir preposiciones y marcas de género o plural. Por ejemplo, *lower order virtual container access* se ha traducido como 'acceso a contenedor virtual de orden inferior' en español y *accès pour conteneurs virtuels de niveau inférieur* en francés, que no es exactamente la misma idea. Para *open platform card management*, la traducción ha sido 'gestión de tarjetas de plataforma abierta' (¿qué es lo abierto: la plataforma o la gestión?). La traducción de *discrete cosinus transform* es más variable, pues tenemos al menos tres posibilidades, además de la sigla DCT, es decir: 'transformación discreta en coseno', 'transformada discreta bidimensional de coseno' y 'transformada de coseno discreta'. En francés hay una versión más: *transformée cosinusoïdale discrète*. Al poner la información de diferentes idiomas en una sola ficha terminológica, estos problemas de interpretación saltan a la vista y pueden ser corregidos.

En ciertos casos, las posibilidades terminológicas son increíblemente prolíficas, como por ejemplo en 'magnetoscopio', 'VCR', 'magnetoscopio de casete', 'magnetoscopio videocasete', 'VTR', 'grabador de vídeo', 'videograbador'. En inglés también hay varios términos para el mismo aparato: *video tape recorder, VTR, video recorder, video tape machine, video cassette recorder, VCR, video cassette home recorder, VHR, video tape recorder*. La desaparición de esta tecnología acabará con el problema terminológico.

La traducción de *smooth earth* tiene unas ocho posibilidades combinando 'tierra' o 'suelo' con el adjetivo 'liso', 'regular', 'plano' o 'uniforme'. El término *boresight* tiene muchas traducciones: 'punto de intersección del eje del haz con la Tierra', 'dirección de visado', 'referencia de puntería', 'dispositivo de puntería', 'referencia de alineación óptica', 'alineación', 'eje de puntería', 'línea de mira', 'visual', 'línea de visual', 'eje de visado'. Hasta el tristemente famoso *mobbing* o *harassment in the workplace* tiene varias traducciones: 'acoso moral en el trabajo', 'acoso moral', 'hostigamiento profesiona'l, 'maltrato psicológico', 'acoso laboral', 'acoso psicológico'.

Por eso es importante indicar todos los términos correctos posibles, pero dando como principal en cada idioma el que se considera más correcto o el que se ha decidido usar en los textos de cada idioma. La presencia de los sinónimos españoles es más útil para un traductor inglés que puede encontrase con ese término y menos para el traductor español que se contenta con el término principal. Eso no impide que en ciertos casos el traductor decida usar un sinónimo si es más apropiado según el contexto. Para facilitar esta decisión se incluye una referencia al documento de donde se extrajo el término.

Con el paso del tiempo las fichas de este tipo se van limpiando o completando con indicaciones de uso. Por ejemplo se desaconseja explícitamente los términos 'abonado UPT', 'agrupación funcional', 'semafón' o 'dígito de relleno', pues han sido reemplazados por 'cliente del servicio UPT', 'grupo funcional', 'buscapersonas' y 'dígito de justificación', respectivamente.

El nombre de organismos también cambia y es importante que esta información se indique explícitamente para facilitar las referencias. Tenemos por ejemplo la Agencia Telegráfica de Información de Rusia (ITAR-TASS) que antiguamente era Agencia TASS, o más cerca de nosotros la Oficina de Normalización de las Telecomunicaciones (TSB) que era la Secretaría del CCITT (Comité Consultivo Internacional Telegráfico y Telefónico) y que forma parte de la UIT.

Las marcas regionales también son un elemento importante de información. El *bug* informático se traduce normalmente por 'error de programación', pero es bueno saber que en Argentina se le llama también 'polilla' y en México 'basura'. Se habla de 'ordenador' en España, de 'computador' en Chile y Colombia, pero la palabra más usada es 'computadora'. El adjetivo *backup*, que se aplica a procedimientos y materiales destinados a ser usados en ciertos casos de anomalía de funcionamiento, se suele traducir por 'de seguridad' en España y *de* 'respaldo' en Hispanoamérica.

Refundición de las bases de datos lingüísticos existentes

Se está procediendo a unificar las diferentes bases terminológicas en los seis idiomas de la UIT para dar aplicación a la Resolución 115 (Conferencia de Plenipotenciarios, Marrakech, 2002) y ya se han realizado considerables progresos en este sentido. Siguiendo el asesoramiento proporcionado por los Grupos Asesores Sectoriales, dicha refundición se está efectuando paulatinamente, con el fin de integrar los seis idiomas y armonizar conceptos y plataformas técnicas, dentro de los recursos

disponibles y sin entorpecer por ello el funcionamiento y la utilización diarios de las bases de datos existentes.

Desde el 2006 hay solo una base en el Departamento C&P, pues en realidad el contenido actual de TERMITE 6L, que se beneficia de una operación de limpieza en curso para suprimir redundancias y duplicaciones, es una refundición de la vieja base TERMITE y de una más nueva ACRoTermite. El sistema actual de captura de datos detecta automáticamente los términos duplicados reduciendo así casi completamente el riesgo de ese tipo de error. Esta fusión ha tomado varios años y ha sido progresiva. En los años ochenta había una base por cada glosario que se publicaba (más de diez); a mediados de los 90, había prácticamente cuatro bases, una para inglés, francés y español, una para el árabe, inglés y francés, una para chino e inglés y otra para inglés y ruso; de 1999 al 2004, hubo dos bases, TERMITE, para inglés, francés y español, y ACRoTermite para inglés, árabe, chino y ruso.

El objetivo es que a fines de 2006 se cuente únicamente con dos bases de datos (TERMITE 6L y términos y definiciones del UIT-R y del UIT-T), ambas en los seis idiomas. Esto permitirá proporcionar un apoyo terminológico sólido y plenamente operacional a los usuarios de la UIT (delegados, traductores, editores, ingenieros, etc.). Ulteriormente, habrá que definir y ejecutar un proyecto detallado para unificar todos los elementos, definiendo las plataformas, la estructura y los conceptos, así como la gobernanza y la implementación en el marco de la estrategia global de la Unión y, particularmente, en el contexto de su estrategia IT.

La fusión de las bases de datos del UIT-R y el UIT-T responde a necesidades técnicas, mientras que TERMITE 6L tiene por objeto facilitar el trabajo de traducción. Las bases de datos son distintas, por lo que será difícil armonizar sus objetivos. En efecto, los términos y definiciones del UIT-R y del UIT-T, que aparecen en secciones específicas de las recomendaciones, son recopilados de manera automática y sistemática y representan unos 23 000 términos, casi todos en inglés, francés y español solamente, mientras que TERMITE 6L se enriquece por medio de un análisis y selección terminológica de todo tipo de textos de la UIT con miras a completar la base actual.

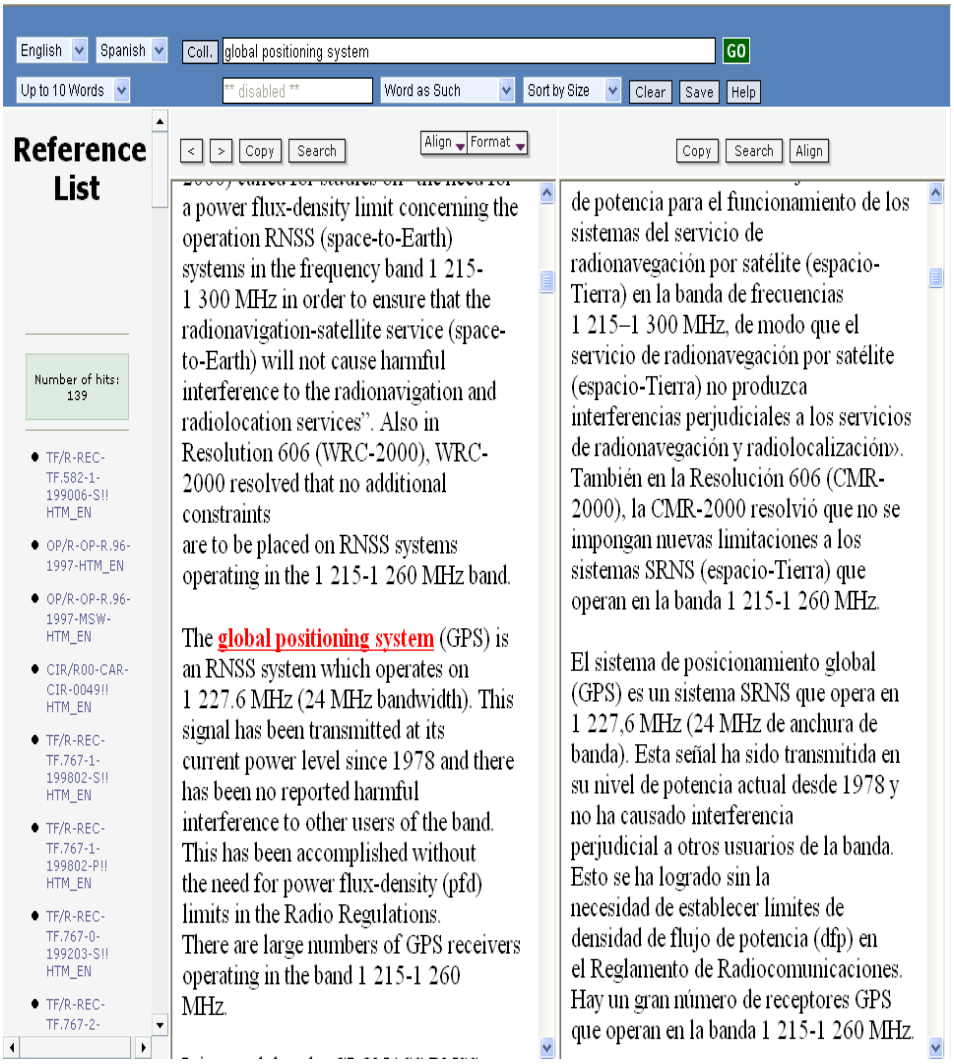
Este trabajo parece muy sencillo si se compara al realizado en los últimos años en las instituciones europeas que refundieron tres grandes bases de datos terminológicas (Eurodicautom, TIS y Euterpe) para crear una sola llamada IATE, que contiene cerca de 1,4 millones de conceptos, 7 millones de términos de los cuales alrededor de 1,5 millones en inglés en todos los ámbitos de trabajo y estudio de las instituciones de la Unión Europea. El trabajo de limpieza y consolidación está también en curso.

Hace un par de años quisimos comparar el contenido de TERMIUM, la base de datos terminológica de los servicios gubernamentales del Canadá, que contiene más de 3 millones y medio de términos en los ámbitos científicos y técnicos, con TERMITE, pero limitándonos al tema de las telecomunicaciones. Tomando al azar una recomendación que nos había permitido introducir o completar un centenar de fichas en TERMITE, nos dimos cuenta buscándolas en TERMIUM que solamente una entrada tenía equivalente español, 6 % estaba también en esta base y el 94 % restante solo estaba en TERMITE. También buscamos todas las entradas de TERMITE con el término inglés *antenna* con uno o más adjetivos, es decir unas 145 fichas y llegamos a la conclusión de que solamente 15 % aparecían en inglés, francés y español en TERMIUM y 27 % en inglés y francés. Por otro lado, las búsquedas en Eurodicautom suelen dar como resultado, en el ámbito de las telecomunicaciones, muchas fichas que la UIT intercambió en los años 80 con esta base y que sin embargo han seguido su vida autónoma en TERMITE.

Por lo tanto, en el campo de las telecomunicaciones los traductores de la UIT tienen interés en consultar primero que todo nuestras bases de datos terminológicas y documentales y luego, solamente si no encuentran una respuesta, recurrir a bases de datos externas.

Conclusión

Hoy los usuarios de TERMITE 6L tienen acceso a otras fuentes de información por la Internet o la Intranet lo mismo que a otro tipo de herramientas informáticas como los «alineadores» de texto o «concordanciadores» que son un tipo de memoria de traducción para buscar una expresión en una base documental y ver el resultado en paralelo con el idioma de salida al lado del idioma de llegada, como por ejemplo, en la siguiente búsqueda de la expresión *global positioning system* en más de 150 000 documentos en seis idiomas:



La limpieza en curso de TERMITE 6L, como consecuencia de la refundición, ha revelado la existencia de muchos duplicados que se explican por varias razones:

- La base es originalmente el resultado de la consolidación casi automática del conjunto de términos que figuraban en los diversos glosarios impresos de la UIT (desde 1980);

- Las contribuciones «masivas» sin filtrado real que se aceptaron en varias épocas;
- Ciertas entradas datan evidentemente de antes de la generalización de la informática individual y del acceso a la Internet y a la Intranet (en esa época se buscaba que los utilizadores encontraran en la base un máximo de términos que podrían aparecer en los textos independientemente del ámbito tratado);
- Una buena parte de los duplicados se puede imputar a la voluntad de las diferentes personas que han intervenido en el enriquecimiento de la base sin haber realmente tomado o tenido el tiempo de proceder a las verificaciones previas exigidas (por ejemplo, si el término o su equivalente existía ya en otro idioma, quizás bajo una forma ligeramente declinada);
- Los sistemas informáticos no efectuaban suficientes controles de coherencia y de detección de duplicados como ahora sucede;
- Ha sucedido que dos fichas diferentes sin término común en un mismo idioma (por ejemplo, una ficha en inglés-ruso y otra en francés-español) pueden de repente aparecer como duplicados cuando una de ellas se actualiza con un término en el idioma de la otra ficha.

Si los documentos de base son de muy buena calidad y la terminología es coherente, se podrían usar las memorias de traducción en vez de compilar la información en una base de datos terminológica. Como la realidad no es así de perfecta (la calidad de los textos es variable, las traducciones no aparecen al mismo tiempo que los originales, el volumen de información es inmenso), el hecho de centralizar la terminología en una base de datos permite ganar tiempo y calidad, pues se evita repetir el análisis y selección de términos que un terminólogo puede hacer para varios idiomas a la vez. Lo que realmente se necesita, por un lado, son sistemas de gestión que faciliten el trabajo del terminólogo ayudándole a evitar el máximo de errores e incoherencias desde el momento mismo de la creación o modificación de una ficha y, por otro lado, interfaces de consulta rápidas y ergonómicas con posibilidades de filtrar los datos según diversos criterios. Ese es en todo caso nuestro objetivo.