Las nuevas tecnologías, una herramienta básica en la traducción de especialidad

Mª TERESA DE CUADRA GARCÍA
Mª JOSÉ FEU GUIJARRO
ASUNCIÓN SÁNCHEZ VILLALÓN
Universidad de Castilla - La Mancha

Las nuevas tecnologías son el elemento que ha marcado cambios estructurales importantes en nuestra forma de vivir y trabajar. La presente comunicación es una reflexión sobre algunos de los aspectos que han repercutido directamente en los planteamientos en nuestra investigación al abordar la tarea de un diccionario bilingüe contrastivo basado en áreas de conocimiento sobre la cata de vinos[1].

En primer lugar hablaremos de las nuevas tecnologías y la sociedad de la información como elementos básicos en los cambios producidos en la comunicación entre los seres humanos, su papel en la divulgación del conocimiento, y su efecto en la tarea de los profesionales que trabajan en traducción y terminología al contemplarse ambas en un contexto comunicativo.

La lingüística de corpus es una ciencia que se ha desarrollado gracias a la innovación tecnológica. SINCLAIR (1991) en su libro *Corpus, Concordance, Collocation* resume la evolución de la lingüística de corpus del siguiente modo:

> *Thirty years ago when this research started it was considered impossible to process texts of several million words in length. Twenty years ago it was considered marginally possible but lunatic. Ten years ago it was considered quite possible but still lunatic. Today is very popular* (1991: 1).

Por tanto, el segundo aspecto que trataremos es la metodología utilizada en la confección de un corpus especializado del léxico de la cata

---

[1] El presente artículo forma parte del proyecto del *Diccionario Bilingüe (Inglés-Español), Contrastivo, Basado en Áreas de Conocimiento* que dirige la Dra. M.ª José FEU GUIJARRO, UCLM.

bilingüe (inglés-español) para abordar los problemas terminológicos y de traducción que se presentaban en el sector vitivinícola.

La competencia cognitiva adquirida mediante la realización de corpus y/o la realización de varios cursos de cata por parte del equipo de investigación nos permiten afirmar que la terminología de la cata es una reformulación discursiva del léxico de varias especialidades científicas (botánica, enología, edafología, ciencias químicas, agronomía, proceso de vinificación...) para describir las sensaciones organolépticas a través de metáforas léxicas e imágenes que reflejan el substrato cultural y la idiosincrasia de un pueblo en ocasiones difícil de trasladar a la lengua meta en las traducciones. En la práctica de la traducción, se presentan problemas de notas de cata como: «huele a jara en una tarde de calor», difíciles de trasladar a un contexto cultural distinto al mediterráneo.

Finalmente expondremos cómo la lingüística de corpus y computacional ligadas al desarrollo de las nuevas tecnologías han permitido crear un contexto comunicativo basado en generación del lenguaje natural e hipermedia adaptativa dinámica, que facilita la interacción comunicativa dependiente del Field (léxico de especialidad), Tenor (experto, aficionado...) y Mode (lenguaje escrito, oral o hipertexto).

1 Nuevas tecnologías y la sociedad de la información

1.1 Nuevos modos de comunicación.

Las nuevas tecnologías están integradas en el modo de vivir de los seres humanos de principios del siglo XXI: la lectura de los códigos de barras en los supermercados, disponer de teléfonos móviles con imagen, los gps de los coches, hacer la compra desde casa por Internet, etc. forman parte de una rutina diaria que no nos cuestionamos.

En el desarrollo de las nuevas tecnologías la informática es esencial, su importancia en la vida diaria o para el ejercicio de cualquier profesión no se pone en duda, ya que nos facilita el trabajo mediante la automatización de procesos, aumenta los recursos y ofrece posibilidades insospechadas en la investigación y las comunicaciones en todos los campos del conocimiento humano.

Entre los aspectos en que las nuevas tecnologías han modificado nuestras vidas hemos de destacar por un lado su influencia en las comunicaciones humanas y por otro su sentido integrador al favorecer la creación de equipos de trabajo multidisciplinares y multilingues.

Al abordar el aspecto de las comunicaciones humanas y las nuevas tecnologías hemos de remitirnos forzosamente a Internet. Si nos referimos

a datos históricos es ya clásico remontarse a los años sesenta como punto de partida de Internet. Como consecuencia de la Guerra Fría, los altos mandos, obsesionados con la seguridad de los sistemas de información y comunicación en caso se guerra total, decidieron encargar el cometido a una empresa (no se les ocurrió nombrar a un comité). La agencia ARPA (*Defense Advanced Research Projects Agency*) puso la primera piedra de Internet.

En 1964, la empresa MND, con las ideas de Paul BARAN, propuso crear una red de comunicaciones descentralizada de modo que siguiera funcionando aunque alguna parte fuese destruida por el enemigo. Cada ordenador estaba conectado con los demás por distintas vías, enviando la información en paquetes independientes que podían tomar distintas rutas, según la integridad y la fluidez de la comunicación, hasta llegar al destino y recomponer allí la información.

En 1984, la NSF (*National Science Foundation*: fundación nacional de la ciencia), creó su propia red, la NSFnet, parar comunicar a los científicos. Cada red se conectaba a la red más próxima y se comunicaban de una red independiente a otra, formándose la «red de redes».

En 1988 esta red tomó conciencia de su existencia y se denominó Internet. En 1989, ARPAnet dejo paso definitivamente a Internet, libre del espíritu militar del que nació. En 1989 en el CERN (Centro Europeo de Investigación Nuclear) Tim BERNERS-LEE desarrolló la WWW (*World Wide Web*), que se basa en un sistema de visualización y acceso de información organizada en documentos (o páginas).

Por su propia naturaleza Internet es una gran red de *nodo*s interconectados por diferentes caminos con un *lenguaje común* especificado en *protocolos*, un acuerdo que permite el intercambio de bits de un usuario a otro, sin entrar a valorar contenidos. Fue creada con esa finalidad: establecer un protocolo (TCP/IP) para que un conjunto de nodos interconectados puedan esquivar los nodos que fallan y encontrar la manera de que los bits enviados desde un extremo de la Red lleguen a su destino por una ruta alternativa.

Las dimensiones de la Red son asombrosas: se habla de más de quinientos mil millones de páginas de información, pero es muy probable que las comunicaciones que las personas intercambian en los foros, chats y «grupos de noticias» sobrepasen con creces esta cifra. Cada minuto se envían en el mundo cinco millones de correos electrónicos (MILLÁN 2004).

Como consecuencia directa de la incorporación de las nuevas tecnologías están surgiendo nuevos modos de comunicación. Podríamos hablar de un nuevo género epistolar con nuevas formas: por un lado tenemos el e-mail recordatorio o conminatorio, sin encabezamiento y de

una sola línea, que convive con el correo electrónico largo tan extenso como las cartas que se escribían el siglo pasado. Por otro, los mensajes con emoticonos, acrónimos, palabras nuevas, el uso de MSM de los teléfonos móviles. Se han creado unas nuevas formas de «oralidad por escrito» que permiten meditar lo que se dice, almacenar los contenidos poder recuperar a posteriori los mensajes propios y ajenos en pocas palabras: nuevas formas de comunicación basadas en el desarrollo de las nuevas tecnologías.

La conclusión inmediata es que las nuevas tecnologías son una nueva herramienta para la comunicación entre los seres humanos, que a su vez han establecido nuevas formas de relación en las que, quizás con más frecuencia de lo deseable, el «contacto digital» está sustituyendo al contacto humano.

Si nos basamos en la reflexión metacomunicativa que considera la comunicación como un proceso social de permanente y mutua interacción entre palabras, gestos, movimientos, miradas, situaciones, y espacios personales (CEREZO ARRIAZA 1997: 46), podríamos decir que hemos aumentado el contexto situacional de la comunicación posibilitando la «comunicación permanente», una comunicación abierta 365 días al año durante 24 horas. El tiempo y la distancia han dejado de ser un obstáculo, los bits viajan ininterrumpidamente por la Red y con los bits los sentimientos, las dudas, los rencores, los amores, las noticias, la información científica...

Las nuevas tecnologías y especialmente Internet han facilitado el acceso a la información de un modo masivo, el derecho a la información no se discute, pero ante la avalancha de datos a la que nos enfrentamos a diario, el problema ahora no es la disponibilidad de la información, sino cómo gestionar el inmenso volumen de contenidos que recibimos constantemente, el modo de asimilar e integrar toda la información que nos llega.

Si lo analizamos despacio nos daremos cuenta de que la mayor parte de las operaciones intelectuales que utilizan la herramienta de la Red, no pretenden sólo «recuperar información»: intentan construir un conocimiento. La distinción entre información y conocimiento radica en que la información es algo externo, rápidamente acumulable, se puede automatizar y es inerte. El conocimiento, por el contrario, es el proceso de interiorizar y estructurar la información, es un proceso humano que conduce a la acción (MILLÁN 2004).

Podríamos resumir diciendo que la era tecnológica conlleva un nuevo modelo de relaciones basado en el intercambio de información que nos permite avanzar en el conocimiento, conocer las cosas por sus principios

y causas, estudiar las nuevas formas de adaptación al medio, nuevas formas de expresión y dotar a los otros hombres de herramientas cada vez más sofisticadas que facilitarán sus condiciones sociales.

Al tratar el aspecto integrador de las nuevas tecnologías nos referimos a que las nuevas tecnologías han servido para reconciliar posturas y afirmar la multidisciplinaridad y el multilinguismo como herramientas básicas para el progreso. La multidisciplinaridad implica un acercamiento entre las distintas ciencias para lograr objetivos comunes y es imprescindible para crear, compartir y explotar la gestión del conocimiento más sencilla y eficiente.

El multilinguismo, entendido como formación lingüística, tanto en lengua materna (óptima redacción), como en una segunda o segundas lenguas se ha convertido en una necesidad en la sociedad de la información y en la gestión del conocimiento. En la sociedad de la información son imprescindibles los profesionales capaces de conectar con éxito al usuario con los contenidos venciendo las barreras del lenguaje.

La Sociedad de la Información centra sus esfuerzos en las tecnologías de futura generación en las que los ordenadores y redes estarán integradas en el entorno cotidiano permitiendo el acceso a multitud de servicios y aplicaciones mediante interfaces amigables. Este enfoque sitúa al individuo, al usuario, como núcleo del desarrollo futuro de una sociedad basada en el conocimiento para todos.

1.2 La divulgación del conocimiento

Internet y las nuevas tecnologías han tenido un papel fundamental en los cambios sufridos por la terminología de especialidad y en los trabajos lexicográficos y de traducción de los últimos años.

Son varios los aspectos que merecen una mención especial, entre ellos y en primer lugar la difusión del conocimiento especializado y en segundo lugar la democratización del conocimiento, consecuencia directa del primer paso, que ha tenido repercusiones directas en la labor de terminólogos, lexicógrafos y traductores. A continuación pasaremos a analizar estos puntos con más detalle.

Una de las consecuencias de las nuevas tecnologías ha sido la difusión al gran público de conocimientos científicos que con anterioridad pertenecían en modo exclusivo a la comunidad científica que los desarrollaba. Hace unos años hubiese resultado extraño que temas como el genoma humano, el ADN, la célula madre, la clonación de seres vivos,

etc. se difundiesen en las noticias de televisión o radio con la naturalidad con que se hace actualmente.

La divulgación masiva de la información de determinadas especialidades científicas a través de los medios de comunicación ha tenido varias consecuencias:

En primer lugar se produce una reformulación discursiva del léxico de especialidad para poder acercar esas disciplinas científicas al gran público.

En segundo lugar, y como consecuencia del primer punto, se produce una banalización del término en el proceso de reformulación que conlleva la pérdida de la precisión propia de la especialidad del saber científico.

Y en tercer lugar la divulgación de la ciencia a niveles diferentes, difusión en los medios de información, tratamiento automático de la información, etc. producen nuevos escenarios comunicativos que han dado lugar a variaciones denominativas que surgen como consecuencia de la adaptación de las formas comunicativas.

En este sentido las nuevas tecnologías han supuesto una democratización del conocimiento y han creado nuevas situaciones comunicacionales en las que el léxico de especialidad, anteriormente relegado a ámbitos muy concretos, ha abandonado su ambiente restringido para formar parte de la competencia de los hablantes en su proceso de adquisición de conocimientos especializados.

1.3 Repercusiones de la divulgación del conocimiento científico en la terminología especializada y la traducción.

Las consecuencias de la innovación tecnológica en el desarrollo de nuestra tarea profesional son innegables y requieren un proceso de adaptación al nuevo medio que, aunque en un principio creó una cierta desconfianza entre los profesionales, se ha impuesto de un modo abrumador, ya que no sólo ofrece muchas más ventajas que inconvenientes, sino que el hecho de no «subirnos al carro» de la innovación nos deja automáticamente fuera del nivel de competencia profesional al que todos aspiramos.

La pregunta que nos planteamos ahora es: ¿cuáles son las repercusiones en la terminología especializada y la traducción de esta divulgación del conocimiento científico?

En primer lugar nos referiremos a la terminología especializada para ver en qué manera las nuevas tecnologías y la sociedad de la información han podido afectar a las dos funciones básicas de la terminología, que son:

1 La función de representación del conocimiento especializado.
2 La función de la transferencia de ese conocimiento.

Haciendo un repaso histórico hemos de recordar que la terminología como disciplina organizada fue establecida por WÜSTER en los años cuarenta siguiendo los postulados del Círculo de Viena (Normas ISO). La idea principal que WÜSTER defendía era la de una terminología prescriptiva, como disciplina independiente, que se fundamentaba en el acuerdo entre los profesionales, cuyo objetivo era evitar la ambigüedad y garantizar la univocidad entre concepto y término en disciplinas científicas como la ingeniería. (CABRÉ 2002: 9-23), pero este sueño de WÜSTER de mantener la terminología prescriptiva como una disciplina al margen de la lexicografía se vio truncado por dos motivos:

1 El estudio de la terminología en su contexto comunicativo.
2 La influencia de las nuevas tecnologías en la divulgación y democratización del conocimiento especializado.

Siguiendo a CABRÉ (2002: 14) diremos que la terminología normalizada que considera las unidades terminológicas como unívocas y monosémicas, de carácter denotativo y estático, que representa un concepto establecido en todas las lenguas y culturas, y que se circunscribe a un solo dominio especializado, ha quedado relegada a la comunicación entre profesionales de alto nivel de especialidad.

Al mismo tiempo que el análisis del conocimiento especializado en el contexto comunicativo ha demostrado, por un lado, que los términos no son el único tipo de unidades que representan el conocimiento especializado (unidades lingüísticas y no lingüísticas: morfológicas, sintácticas, léxicas), y por otro, que los términos en el contexto pierden el carácter monosémico que presentan en los glosarios y adquieren distintos significados.

Existe pues una diferencia muy significativa entre estos dos tipos de terminología que CABRÉ (1999) denomina terminología *in vitro* y terminología *in vivo*:

El estudio de la terminología de especialidad en el contexto comunicativo en que se produce, pone de manifiesto una serie de puntos que se podrían concretar en los siguientes apartados:

1 Reconocer que junto con los términos, hay unidades lingüísticas y no lingüísticas que representan el conocimiento especializado.

2 Admitir que la terminología que recogen los diccionarios presenta en ocasiones diferencias cognitivas importantes respecto a las que aparecen en los textos.

3 El acceso de varios colectivos de otras especialidades que se ocupan de la difusión del conocimiento especializado (documentalistas, redactores, traductores)

4 La aparición de subespecialidades técnicas motivadas por el desarrollo tecnológico de las disciplinas científicas en las que la transdisciplinaridad es esencial.

Como muy bien explica Teresa CABRÉ (2002), las consecuencias del estudio de la terminología de especialidad en su contexto comunicativo nos permiten analizar la influencia que esta ciencia recibe desde otros ámbitos del conocimiento en su contacto con otras especialidades científicas y como resultado de su utilización por otros profesionales, que pueden concretarse en tres aspectos.

1 Aspectos Cognitivos
Porque se ha cuestionado la uniformidad e independencia de la terminología especializada, para defender la integración cognitiva diversificada en función de los hablantes.

2 Aspectos Lingüísticos
Porque se rechaza el concepto del término especializado como elemento aislado del concepto general y se acepta que forme parte de la competencia de los hablantes en la medida en que adquieren conocimientos especializados.

3 Aspectos Sociales
Porque el estudio de los términos en su contexto comunicativo cuestiona el planteamiento de la terminología prescriptiva de la univocidad entre concepto y término y defiende que la variación discursiva que surge del contexto comunicativo en que se produce presenta, en ocasiones, variaciones cognitivas significativas.

En el apartado de la influencia de las nuevas tecnologías hemos de valorar dos aspectos que, aunque están relacionados, ya que ambos se vinculan a la innovación tecnológica, han repercutido en la terminología especializada de dos modos diferentes: los avances en Lingüística Computacional (LC) como parte de la Inteligencia Artificial (IA) y Lingüística Informática, que han facilitado nuevas herramientas en la

representación del léxico especializado y los aspectos que han contribuido a la transferencia de ese conocimiento y a aumentar la divulgación mediática del léxico de especialidad.

Estos dos aspectos han repercutido directamente en nuestra tarea profesional como lexicógrafos y/o traductores

Desde el punto de vista de la traducción, la tarea del traductor se ha visto muy favorecida por la innovación de las nuevas tecnologías, que se materializa en un aumento de recursos y herramientas informáticas que han roto con el aislacionismo y la proverbial soledad que hasta hace pocos años dificultaba tremendamente el trabajo diario.

No es el objeto de esta ponencia pero a nadie se le escapa que la existencia de foros profesionales, comunidades virtuales (TransList, TermXchange), herramientas para traducción automática (MT) como Systran, Logos Translation System, Transtype, traducción asistida (CAT), que proporcionan a su vez memorias de traducción como Trados o herramientas de localización de software, bases de datos terminológicas, herramientas terminológicas como Multiterm de Trados, Termstar de Star, TermWatch de Atril asociadas a CAT y otras no asociadas como Eurodicautom, Avalon Glossary Assistant, etc. han cambiado el panorama laboral, aunque las competencias del traductor en algunos aspectos básicos siguen siendo las mismas.

Siguiendo a DÍAZ DE LIAÑO (2002: 87) las competencias del traductor son de tres tipos:

En primer lugar podríamos hablar de las *competencias lingüísticas*, es decir un dominio suficiente de la lengua origen y de la lengua meta de la traducción a distintos niveles.

En segundo lugar *competencias traslativas*, es decir, las competencias que permiten llevar a cabo la traslación de una lengua a otra aplicando las ideas desarrolladas en el marco de la teoría de la traducción.

Y en tercer lugar *competencias de gestión de la documentación*, es decir, análisis de la documentación, asociación de ideas que nos permita depurar y filtrar la información.

Por lo tanto, para realizar una traducción adecuada de un texto perteneciente a un dominio especializado, es necesario que el traductor posea competencia lingüística en ambas lenguas, dominio de las convenciones textuales, el lenguaje propio de la especialidad, o dicho en otras palabras, competencia cognitiva sobre el dominio. Esta realidad es aún más compleja en el caso de la terminología de la cata del vino, ya que la competencia cognitiva del traductor debe extenderse a las distintas áreas de especialidad que componen la cata y al substrato cultural e histórico que rodea esta área de conocimiento.

Los textos están sujetos a variables situacionales como el tiempo y el espacio, variables socioculturales, nivel de especialización y grado de formalidad. De la destreza al valorar estas variaciones verticales y horizontales de las unidades léxicas dependerá en gran manera la elección de las unidades léxicas que determinarán el éxito de la producción del texto meta. Es imprescindible una adecuación sintáctica en la composición sintáctica y una adecuación pragmática marcada por el contexto situacional y comunicativo.

2 Confección de un corpus especializado del léxico de la cata bilingüe (inglés-español)

SINCLAIR define el corpus como: «A collection of naturally occurring language text, chosen to characterize a state or variety of a language» (SINCLAIR 1991: 171). Un corpus normalmente contiene millones de palabras, ya que es un hecho aceptado que, dada la creatividad del lenguaje natural, hay una variedad tan enorme de expresiones que es difícil aislar aquellos *recurrent patterns* (patrones recurrentes) que son las claves de la estructura léxica del lenguaje.

La lingüística de corpus es la rama de la lingüística que utiliza corpus para analizar la lengua y determinar sus características, verificar teorías lingüísticas o ensayar aplicaciones de la ingeniería lingüística

Si hacemos una breve revisión de la tipología de los corpora podemos distinguir los *sample corpus*[2], que son los corpora formados por una colección finita de textos —es decir, un corpus cerrado— elegidos meticulosamente y estudiados en detalle. El corpus, una vez elegido, no puede ser modificado.

El *monitor corpus*, que puede definirse como el corpus que reutiliza lenguajes de texto que han sido preparados en formato impreso para otros propósitos de periódicos, revistas, libros y procesadores de textos; también pueden recopilar lenguaje hablado, en principio cuanto más extenso sea, mejor.

Podemos hablar también de «corpus de referencia», que son aquellos corpora que utilizamos para contrastar los resultados de nuestros estudios. Están los «subcorpus», que es cuando trabajamos con parte de un corpus; está el «corpus oral», que es aquel que recoge grabaciones de textos orales..., etc.

---

[2] En el texto original de SINCLAIR (1991: 171) aparece también entre comillas.

2.1 Planteamientos metodológicos básicos

El primer paso es definir cuáles van a ser los criterios de configuración del corpus. Debe incluirse vocabulario propio de la especialidad, pero también léxico relativamente culto, que refleje aspectos socioculturales específicos que son de gran trascendencia en el mundo de la cata en el que la descripción organoléptica va muy ligada al contexto, ya que se establece mediante comparaciones o referencias. Ej.: «rojo picota».

Asimismo es conveniente incluir aquellas voces que perteneciendo al léxico de especialidad han sufrido un proceso de vulgarización y se han incorporado al léxico común. La inclusión de una voz como término de una especialidad científica viene condicionada «por la especialización del concepto al que hace referencia y por el grado de especialización de los textos en que aparece» (GARCÍA PALACIOS *et al.* 2002: 126).

Es importante destacar que sin una pauta en la realización del corpus, en el criterio de selección de los textos, las técnicas de volcado de textos y la indexación, el corpus pierde la representatividad. No se trata de ir eligiendo textos al azar: los textos elegidos deben ser representativos del lenguaje de la especialidad y del uso que de él hacen los especialistas, en nuestro caso los catadores; pero hay datos contextuales como tipo de documento, lugar en el que aparece publicado, público al que se dirige, etc. que ayudan al análisis de contextualidad y dan coherencia al corpus.

Por eso todos los documentos que se incorporen al corpus deben ir debidamente etiquetados e indexados. Establecer las pautas de actuación del equipo de investigación, el modo de anotar incidencias y solventar problemas es esencial para el buen funcionamiento del equipo y evita males mayores, que suelen redundar en pérdida de tiempo y energías.

Como datos fundamentales deberían aparecer por un lado datos sobre el texto: si es un libro, una revista divulgativa o técnica, si pertenece a un programa de televisión; y por otro datos sobre la formación del autor, su trayectoria profesional, su intencionalidad al escribir el documento, etc.

En la lingüística de corpus se tratarán las unidades lingüísticas, palabras o combinaciones de palabras que componen el corpus, como vehículos de conocimiento, ya sea este general o especializado. Se parte de la idea de que el conocimiento es un continuo en el que se integra tanto el área de especialidad como el de otras áreas afines, junto con el conocimiento general expresado por el léxico primario.

Mediante el estudio de los corpora podemos identificar los términos de una o varias palabras, ya que muchos conceptos especializados se representan lingüísticamente por unidades léxicas formadas por más de una palabra, y las variaciones terminológicas. TERCEDOR (2000: 82) afirma

que la terminología tradicionalmente considera las unidades poliléxicas como secuencias fijas de caracteres, pero que en la actualidad se acepta la variación en función de las variaciones terminológicas del contexto en todos los campos especializados y en la lengua general, como una muestra de los distintos puntos de vista y un medio de hacer accesible el conocimiento a distintos tipos de audiencia.

En el proceso de traducción es esencial tener acceso a todas las formas relevantes de una variación terminológica en función de la audiencia y la situación comunicativa que se quiera traducir a una lengua meta.

En el corpus de la cata de vinos este concepto es especialmente importante, porque las unidades lingüísticas proceden de varios campos de especialidad (enología, botánica, ciencias químicas, agronomía) pero a su vez hay otras palabras que, perteneciendo al léxico primario, adquieren una nueva dimensión en el léxico de cata.

En un corpus es importante reflejar la *cobertura lingüística*, es decir, la restricción de la información a los términos que son importantes en el ámbito de la especialidad, aunque en este caso no siempre es fácil establecer los límites entre lenguaje de especialidad y léxico primario.

La aplicación de todos estos criterios tiene como finalidad la creación de un Repositorio Terminológico que se define como una base de datos en la que sea posible almacenar el conocimiento específico de un subdominio de conocimiento (en este caso la cata).

Estamos hablando de la *representatividad del corpus*, que depende claramente de dos aspectos: el modo de recopilación y el diseño, que van íntimamente unidos. Es el eterno dilema: cantidad o calidad. El diseño del corpus depende del objetivo a conseguir y la tipología del corpus y su propósito. Se puede confeccionar un corpus especializado (para un ámbito de especialidad científica) o un corpus de referencia como el *British National Corpus* (BNC) de 100 millones de palabras. PÉREZ HERNÁNDEZ (2000: 12) afirma:

> Se deben recopilar diferentes tipos de información: La orientación onomasilógica de la terminografía exige que la construcción de un repositorio terminológico parta de la organización y sistematización de la estructura conceptual (los conceptos y sus relaciones) del dominio de especialidad, para después determinar cuáles son las unidades terminológicas que los dominan.

Los corpora deben ser herramientas útiles para que los terminógrafos puedan estudiar las unidades de significación especializada en su dimensión conceptual, lingüística y comunicativa.

2.2 Sobre la creación del corpus

Para llevar a cabo nuestro proyecto lexicográfico sobre la cata de vino, se ha elaborado un corpus comparable de textos en español y en inglés y está formado por documentos impresos y corpus oral. De este modo nuestro corpus cumple con los requisitos de *adecuación terminológica*, ya que los textos representan el uso real de la lengua por parte de los especialistas dentro del campo del saber y de *adecuación para la traducción*, ya que el corpus contiene los tipos textuales y géneros representativos del campo de especialidad y está clasificado en función del grado de especialización y contexto comunicativo.

Uno de los puntos más cruciales era establecer los criterios de elección de textos, y para ello solicitamos la opinión de expertos enólogos, catadores, sumilleres, bodegueros, traductores y periodistas —españoles y de otras nacionalidades (ingleses, americanos, sudafricanos)[3]—, con los que hemos contactado a lo largo de estos dos años y que, aparte de ayudarnos con su experiencia y animarnos en nuestra labor, nos han orientado en la incorporación de nuevo material al corpus. Otro criterio fundamental en la confección del corpus fue el grado de especialización de los textos.

El corpus se confeccionó con términos procedentes de tres fuentes fundamentales:

- Texto impreso: libros, revistas, periódicos.
- Textos procedentes de Internet.
- Grabaciones de sesiones de cata, que una vez transcritas se incorporan al corpus en el corpus oral.

Los criterios que establecimos para la elección:

A *Textos en formato impreso y/o electrónico:*

- Libros y/o revistas referentes a la cata. Es decir, manuales especializados de cata y textos divulgativos de revistas de especialidad.

[3] La lista de los profesionales que han apoyado nuestra labor se encuentra en un anexo al final del trabajo y a todos ellos manifestamos nuestro agradecimiento.

- Diccionarios de cata y de vino.

B *Textos procedentes de Internet:*

El requisito indispensable para incluir textos de Internet en el corpus fue que hubiesen sido editados originalmente en inglés por profesionales de origen anglófono. Es cierto que hay notables diferencias entre el inglés australiano, el americano o el inglés británico, pero nos pareció que estas variantes de uso, que por otro lado podían tipificarse al introducir un comentario sobre la procedencia del autor, eran siempre menos peligrosas que introducir «cualquier texto escrito en inglés» de un modo indiscriminado.

El objetivo final de esta decisión no era otro que establecer algún filtro de calidad para evitar incluir involuntariamente algunas de las «traducciones», si pueden llamarse así, que a veces se encuentran en Internet.

C *El corpus oral:*

La gran aportación del corpus oral en nuestro proyecto es que nos ha acercado a la reformulación de los contenidos científico-técnicos que se producen en la pragmática comunicativa de las sesiones de cata, que se establece a partir de la evocación de lugares comunes, de la creación de imágenes que actúan como referentes del marco cognitivo y conceptual que comparten los hablantes (WIERZBICKA 1996: 339).

Las aportaciones de una sesión de cata resultan enormemente enriquecedoras para nuestro corpus por los siguientes motivos:

1 Las sesiones de cata son un *input* perfecto para entender cómo se acerca el usuario a la descripción de los conceptos del vino.
2 Las sesiones de cata permiten la recogida de un corpus oral que puede ser determinante para entender la definición de los conceptos de valoración de los vinos y traducirlos a otras lenguas
3 El material pragmático de la sesiones de cata es un espejo o un reflejo de la mente del usuario y de los caminos que ésta sigue a la hora de buscar y determinar las calidades de los vinos.

El corpus oral está formado íntegramente por sesiones de cata que se graban y luego se transcriben a Microsoft Word (formato .txt). Las grabaciones que constituyen la base de nuestro corpus oral empezaron a recogerse en enero del año 2001 con permiso y con consentimiento

informado de los distintos catadores que llevaban a cabo las sesiones de cata. Las grabaciones se realizaron con grabadoras analógicas por motivos exclusivamente presupuestarios en distintos lugares de España (Castilla-La Mancha, Madrid, Barcelona, Córdoba) a catadores nacionales y extranjeros en español y en inglés respectivamente.

El hecho de haber llevado a cabo desde el inicio del proyecto la validación jurídica de nuestro corpus oral, es decir, los registros de voz con autorización expresa de los actantes nos permitió establecer un convenio de colaboración en febrero del 2003 con el Laboratorio de Lingüística Informática de la Universidad Autónoma de Madrid (LLI-UAM).

El Laboratorio de Lingüística Informática que dirige el Dr. Antonio MORENO SANDOVAL, tiene entre sus objetivos principales la creación de recursos lingüísticos en formato electrónico: gramáticas y lexicones computacionales, bases de datos terminológicas, catálogos digitalizados y todo tipo de corpus, orales y escritos, diacrónicos y sincrónicos, monolingües y multilingües.

Nuestro proyecto ha desarrollado su propio esquema de codificación a partir de los formatos originales de cada equipo, aunque nosotros hemos adoptado el modelo de transcripción prosódica y de validación que el equipo de la Universidad Autónoma de Madrid ha estado desarrollando en el proyecto C-ORAL-ROM[4].

Los puntos esenciales de nuestro trabajo común son:

- La *legalidad de los textos*: contamos con los permisos firmados por todos los participantes. Este requisito es obligatorio pues ha cambiado la legislación y afecta a los derechos de autor (en el caso de conferencias y grabaciones de los medios de comunicación) como al derecho de intimidad.
- La *validación*: Hay distintos niveles de validación. La más importante es la validación interna. Cada texto pasa por cinco etapas: transcripción, revisión, anotación prosódica, revisión de la anotación y alineamiento texto-sonido. Cada etapa la realiza un lingüista diferente a la precedente y varios investigadores intervienen en cada texto.
- La utilización de *XML*: explotamos la universalidad de este lenguaje de marcación, que garantiza la reusabilidad del corpus. XML y sus

[4] C-ORAL-ROM es un corpus oral multilingüe de las lenguas romances oficiales (italiano, francés, portugués y castellano) desarrollado por un consorcio de nueve socios y financiado por la Unión Europea dentro de su V Programa Marco, área Tecnologías para la Sociedad de la Información.

extensiones se han convertido ya en el formato para las tecnologías lingüísticas, ya que a partir de él los textos se pueden convertir en cualquier otro formato textual.

- La *calidad digital en las grabaciones*: se ha reutilizado parte de las grabaciones analógicas anteriores. Las nuevas grabaciones se han realizado con una grabadora digital, que proporciona una calidad excelente que nos permitirá utilizar en un futuro inmediato programas como Winpitch, que permiten la alineación de grabaciones de sonido digital y el texto de la transcripción realizada.

3 El contexto situacional

Otro de los aspectos innovadores de nuestro corpus ha sido la estructuración del contexto situacional *field, tenor* y *mode* (HALLIDAY 1984). Estos mismos criterios son los que se han seguido aplicando desde entonces en la elección de textos tanto en castellano como en inglés. Este tercer componente del corpus ha adquirido un protagonismo especial en nuestro trabajo casi desde el principio y ha marcado de un modo significativo el desarrollo del mismo. Por este motivo vamos a dedicarle especial atención.

Siguiendo la tradición semántico-funcional establecida por HALLIDAY (1984: 64), que define el concepto de registro como « a useful abstraction linking variation of language to variation of social context» y considerando que este contexto social está condicionado por tres variables conocidas como *field, tenor* and *mode,* hemos tenido en cuenta esta clasificación para aplicarla al realizar el etiquetado en la elaboración del corpus, que será la base de nuestro proyecto lexicográfico.

HALLIDAY (1984) al definir los principios funcionales de la pragmática textual afirma: «a text is a product and a process» y añade «a text is an output, something that can be recorded and studied but it is also a process in the sense of a continuous process of semantic choice». Nosotros consideramos que este proceso continuo de elección semántica en función del contexto situacional es esencial a la hora de establecer la adecuación del trabajo lexicográfico.

Sobre este tema MAIRAL USÓN y LÓPEZ CHOLLET (2001: 20) afirman que los diccionarios, en la actualidad, se evalúan o deberían evaluarse pensando en la mente de los usuarios. El hablante cuando ejecuta una búsqueda lexicográfica debe contrastar el contexto que él tiene en la cabeza sobre el concepto con los contextos que ofrece el diccionario (si los ofrece) para determinar cuál es el más adecuado. Y establecer si la elección es correcta es especialmente complicado para el hablante

cuando se trata de otro idioma, a menos que se aporten datos de contextualidad.

BAJO (2000) comenta que el criterio a seguir para valorar una obra lexicográfica es responderse a la pregunta de si se corresponde con la finalidad para la que fue creada, si satisface las necesidades que pretendía cubrir. En la medida que seamos capaces de responder a esta necesidad del usuario habremos acertado en nuestros planteamientos.

A continuación y siguiendo las directrices de HALLIDAY, en cuanto a las tres variables que configuran el contexto situacional que caracteriza a un tipo de registro en particular, vamos a definir o aclarar cuáles son estos tres rasgos fundamentales que configuran un término en su contexto apropiado

Field es definido por HALLIDAY y HASSAN (1985: 31) como «those meanings that express our experience of the world around us» es decir, representa el *experiential component.*

Tenor implica la relación entre los participantes, «the personal relationship involved» (ibid.), o lo que es igual, representa el *interpersonal component,* la relación entre el experto / diseñador y el usuario / cliente.

Mode estudia «what part the language is playing» (ibid.) y no solo analiza la función, sino que informa el tipo de canal utilizado (oral o escrito) y además nos clasifica el tipo de información, llamado *rhetorical mode,* que puede ser persuasivo, expositivo, didáctico o similar.

MARCO (2001), que ha contribuido de un modo significativo al *register analysis* en la evaluación de la calidad de la traducción literaria mediante la justificación expresa del análisis del *register analysis*, señala que esta herramienta «provides the necessary link between a communicative act and the context of situation in which it occurs» (p. 1). En su opinion el *register analysis* es «the most comprehensive framework proposed for the characterisation of context» y una de sus principales ventajas es que aporta un número limitado de variables que puede definir cualquier contexto.

HATIM y MASON (1990, 1997) emplean también el *register analysis* como parte de su visión del contexto en la traducción, pese a su afirmación de que hay otro factores contextuales, como los pragmáticos y los semióticos, que trascienden el marco del *register*. Los autores asumen que identificar la pertenencia de un *register* a un texto es una parte esencial del procesamiento del discurso, ya que implica al lector en una reconstrucción del contexto mediante un análisis de «lo que ha ocurrido» (*field*), «quién ha participado» (*tenor*) y «qué medio se ha empleado para transmitir el mensaje» (*mode*). Estas tres variables establecen una transacción comunicativa en el sentido de que proporcionan las

condiciones básicas para que la comunicación se produzca. (HATIM / MASON 1990: 55)

En la práctica de la traducción el contexto situacional es esencial ya que determina el proceso de elección de términos y por lo tanto la adecuación. No se habla igual de vinos en una cata profesional, una cata didáctica, una cata promocional o cuando estamos tomando el aperitivo con los amigos. Pero no es este el único problema, al establecerse la descripción organoléptica del vino en función de referentes, el contexto sociocultural es en ocasiones fundamental para trasladar el contexto a otras culturas. Como ejemplo ilustrativo aportaremos un nota de cata de nuestro corpus oral:

1 Se aprecian aromas complejos de especias, de tabaco, los tostados de la madera, el regaliz, los cueros..
2 Yo aprecio algo de balsámico, pero no eucaliptus... más bien es como monte bajo...no sé ...jara, tal vez... y algo de cuero, como a caza.
1 Sí, hay jara, me recuerda el olor de la jara en una tarde de calor...
3 Eso es la Syrah[5]... a eso huele la Syrah.. es un aroma primario...

Notas de cata como: «huele a jara en una tarde de calor», difíciles de trasladar a un contexto cultural distinto al mediterráneo y la alusión al Syrah requieren competencia cognitiva en variedades de uvas, clasificación de los aromas que se perciben en la cata, etc.

4 Aportación de la lingüística de corpus y la lingüística computacional

La innovación que proporcionan las nuevas tecnologías desarrolladas por la Ingeniería Informática, la Lingüística Computacional y el Procesamiento del Lenguaje Natural nos ofrece recursos antes insospechados a la hora de establecer los contextos situacionales y por tanto la adecuación terminológica de la lexicografía y la traducción.

En los objetivos del VI Programa Marco para las «Tecnologías para la Sociedad de la Información» (IST–Information Society Technology) se promueve:

> el desarrollo de interfaces amigables intuitivas, que puedan interpretar todos nuestros sentidos como el habla, la visión y el tacto y que puedan comprender nuestros gestos en varias lenguas. Deben desarrollarse

[5] Una variedad de uva foránea que se ha introducido en España hace unos años y que está dando unos resultados bastante satisfactorios. Esa variedad es muy apreciada en los vinos australianos.

> conjuntamente tecnologías basadas en el conocimiento más potentes y flexibles que sean *semantic-based and context-aware*. Estas serán las bases de la próxima generación Web y darán acceso a la creación de un contenido digital más creativo y efectivo. (www.cordis.lu).

Siguiendo las recomendaciones de la normativa europea del IST proponemos crear un contexto comunicativo basado en generación del lenguaje natural e hipermedia adaptativa dinámica, en los que la *usabilidad*[6], y las *interfaces amigables*[7] faciliten la interacción comunicativa dependiente del *Field* (léxico de especialidad), *Tenor* (experto, aficionado...) y *Mode* (lenguaje escrito, oral o hipertexto), que hemos desarrollado en la estructura del corpus.

5 Conclusiones

Pensamos que la adaptación en nuestro trabajo lexicográfico al contexto situacional es básica para el desarrollo de los nuevos proyectos lexicográficos basados en las nuevas tecnologías que permiten cambios esenciales en la macro y microestructura y que son sólo posibles gracias al desarrollo de ciencias como la Lingüística Informática y la Ingeniería Lingüística.

La Lingüística Computacional y el Procesamiento del Lenguaje Natural aportan la posibilidad de construir sistemas informáticos que procesen la estructura lingüística y cuyo objetivo sea la simulación parcial de la capacidad lingüística humana.

La Lingüística Computacional es una disciplina aplicada que se ocupa de: «el estudio científico del lenguaje desde una perspectiva computacional»[8]. Su objetivo es proporcionar modelos computacionales para varios tipos de fenómenos lingüísticos. Estos modelos pueden ser «basados en el conocimiento» (*hand-crafted*) o «basado en datos» (*statistical* o *empirical*). Por tanto podemos concluir que «el reto de la

---

6 De nuevo aclarar que no es una traducción personal o un barbarismo. Usabilidad: En este caso seguiremos la definición ISO/IEC 9126: «La usabilidad se refiere a la capacidad de un software de ser comprendido, aprendido, usado y ser atractivo para el usuario, en condiciones específicas de uso».

7 De nuevo insistir, como ya se ha indicado, en que no es traducción personal. Otra posible traducción es interfaz ergonómica, pero está mucho menos difundida que esta otra definición: *Interfaz Amigable:* la interfaz del sistema es totalmente amigable para el usuario cuando los software son implementados sobre sistemas de iconos o botones que facilitan el manejo del sistema. E.I.D Empresa de Identificación y Desarrollo.http://www.eidchile.cl/sistemas.htm.

8 Definición proporcionada por The Association for Computational Linguistics en su página web: http://www.aclweb.org/.

Lingüística Computacional es conseguir que los ordenadores nos entiendan hablando nuestra propia lengua» (MORENO SANDOVAL 1998: 11)

El procesamiento del Lenguaje Natural se basa en la comprensión semántica del lenguaje que permita generar nuevos mensajes. Teóricamente es posible partir de la representación semántica de un texto como base para la generación del lenguaje. En la práctica se ha conseguido la comprensión parcial de los mensajes, y este es un primer paso que permite una primera clasificación fiable de textos, que puede usarse luego para centrarse en los aspectos semánticos más importantes del texto.

También puede usarse, junto con conocimientos estadísticos y lingüísticos, para identificar automáticamente las características lingüísticas de palabras desconocidas, que luego se incorporan al sistema. El reconocimiento de un modelo semántico básico por parte del ordenador sirve, por ejemplo, para relacionar una solicitud de información con su significado básico y es la combinación del análisis y la generación de un modelo semántico básico lo que permite la traducción automática de textos.

De la unión entre los planteamientos lingüísticos del contexto situacional y las posibilidades que ofrece la generación del lenguaje natural al interactuar con el usuario han surgido nuevas posibilidades que con trabajo e imaginación serán las que nos permitan cumplir los objetivos de desarrollar tecnologías del conocimiento «semantic-based and context aware». Esperamos que con estos planteamientos nuestro proyecto reúna estos requisitos y sea capaz de satisfacer las necesidades de adecuación en la tarea terminológica, lexicográfica y de traducción que demanda el usuario del siglo XXI.

Bibliografía

BAJO PÉREZ, E. (2000) *Los diccionarios. Introducción a la lexicografía del español*, Ediciones Trea, Gijón

CABRÉ, M.T. (1999) *La Terminología: representación y comunicación*, Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra, Barcelona

CABRÉ, M.T. (2002) «Entre el lenguaje y el metalenguaje: importancia de la terminología para la comunicación en las lenguas europeas», 9-22 en *Panorama actual de la lexicografía*, Editorial Comares, Granada

CEREZO ARRIAZA, M. (1994, 1997) *Texto, contexto y situación: guía para el desarrollo de las competencias textuales y discursivass,* Ediciones Octaedro, Barcelona

DÍAZ DE LIAÑO ARGÜELLES, J. L. (2002) «Herramientas informáticas de traducción en una empresa privada de servicios lingüísticos», 82-92 en *Actas de* «*El español, lengua de traducción. I Congreso Internacional*», Almagro

GARCÍA PALACIOS, J. / M. T. FUENTES MORÁN (2002) «Los diccionarios de especialidad y el traductor», 117-136 en *Panorama actual de la lexicografía,* Editorial Comares, Granada

HALLIDAY, M. A. K. (1994[2]) *Functional Grammar,* Edward Arnold, London

HASAN, Ruqaiya / M. A. K. HALLIDAY (1985) *Language, Context and Text*: *Aspects of Language in a Social-semiotic Perspective*, Deakin University Press, Geelong [1989[2], OUP, Oxford]

HATIM / MASON (1990, 1997) *Discourse and the Translator*, Longman, NewYork

JUNCAL, J. A. (2000) «Un útil de traducción: las comunidades virtuales», 117-136 en *Terminologie et Traduction* 2*,* Ed. Communautés européennes

MAIRAL USÓN, R. / M. J. LÓPEZ CHOLLET (2001) *Avances tecnológicos en la gestión del léxico y la terminología,* UNED

MARCO, J. (2001) «Register Analysis in Literary Translation. A function Approach», 1-19 en *Babel* 46.1

MILLÁN, J. A. (2004) *La lectura y la sociedad del conocimiento,* http://www.lafactoriaweb.com/articulos/millan19.htm [18.2.2004]

MORENO ORTIZ / C. PÉREZ HERNÁNDEZ (2000) «Reusing the Mikrokosmos Ontology for Concept-Based Multilingual Terminology Databases», 1061-1067 en *Proceedings of the 2nd International Conference on Language Resources and Evaluation,* LREC, 31 May -2 June, Athens

MORENO, A. (2002) «La evolución de los corpus de habla espontánea: la experiencia del LLI-UAM» en *Actas de las II Jornadas de Tecnologías del Habla*, Granada

MORENO SANDOVAL, A. (1998) *Lingüística computacional*, Síntesis, Madrid

PÉREZ HERNÁNDEZ, Ch. (2000) *Explotación de los corpora textuales informatizados para la creación de bases de datos terminológicas* [tesis doctoral], Universidad de Málaga [http://elies.rediris.es/elies18.2002]

SEARLS Doc / David WEINBERGER (2004) *World of Ends. What the Internet is and How to Stop Mistaking it for Something Else* [Last update: 28.4.2003], http://navegante2.elmundo.es/navegante/2004/02/11/weblog/1076521790.html

SINCLAIR, J. (1991) *Corpus, Concordance, Collocation,* OUP, Oxford

TERCEDOR SÁNCHEZ, Maribel / Beatriz MÉNDEZ CEDÓN (2000) «Fraseología y variación terminológica: estudio descriptivo en corpora biomédicos», 82-99 en *Terminologie et Traducción* 2, Communautés européennes

WIERZBICKA, A. (1996) *Semantics. Primes and Universals*, OUP, Oxford / New York

ANEXO

Profesionales del mundo del vino consultados

AGRADECIMIENTOS:

Al Consejo Regulador D.O. La Mancha, en especial a sus presidentes D. Nicolás López Treviño, D. José Fenollar Ávila, D. Gregorio Martín-Zarco López-Villanueva y al secretario del Consejo D. Alfonso Alonso Ramos.

A la Unión Española de Catadores (UEC), destacando a D. Fernando Gurrucharri, (Presidente) y a D. Jordi Vidal (Gerente).

A los enólogos, catadores y bodegueros que han autorizado las grabaciones de la sesiones de cata: D. José Luis Barco, Dña. Pilar Vela, D. Francisco del Castillo García, D. Óscar Pasanau, D. Vicente Dalmau Cebrián, Dña. María Vargas Montoya, D. Ignacio de Miguel Poch, D. José Luis Fuentes Mateos, Mr. John Salvi, Mr. Steven Spurrier, D. Miguel Ángel Pérez González, D. Juan Muñoz Ramos, D. Juan Carlos Muñoz Pinilla, D. Javier Ausas, D. Javier Nieto Sepúlveda, Mr. Samuel Caldwell, Mr. Jeremy Watson, D. Manuel María López Alejandre. D. Ángel González Pinto, D. Custodio López Zamarra, D. José Luis Casado, D. Andrés Proensa, D. Joaquín Vila, D. José Luis González Cledera, D. Luis Miguel Martín, D. Luis Vida, D. Jerónimo Pereda Contreras, Dña. Dolores Núñez Pinto, Dña. Milagros Romero Torres.

A los periodistas: D. Carlos de la Morena, D. Alonso Castro, D. Joaquín Coronel, D. Marcial Pita.

Y a todas aquellas personas del mundo del vino que han facilitado nuestra labor con su amabilidad y buen hacer.