



Comisión de Terminología

# Cómo traducir con corpus y no morir en el intento

Una nutrida concurrencia tuvo el privilegio de contar con la presencia y desinteresada participación de Miriam Seghiri, quien, al frente de un salón colmado, dictó una charla sobre corpus.

| Por la Trad. Públ. **Silvia Focanti**, integrante de la Comisión de Terminología |

**S**i ya empezaron a leer este artículo, entonces habrán notado que dijimos todo, pero no dijimos nada, ¿cierto? Pues allá vamos, que por algo hubo tantos asistentes.

¿Cuándo? El lunes 13 de abril, a las 18.30.

¿Dónde? En el auditorio Tsugimaru Tanoue del CTPCBA. A sala llena, dicho sea de paso.

¿Quién? Acá se pone más interesante... Cuando la Comisión recibió el currículum de la expositora, se encontró con páginas y más páginas que detallaban la formación y la impresionante labor profesional y académica de Miriam Seghiri. Personalmente, llamó mi atención la sencillez de su presentación, que leía «Miriam Seghiri, Universidad de Málaga». Así de simple, o así de compleja; porque Miriam no solo es traductora, sino también terminóloga, docente e investigadora, y podríamos agregar más títulos y cargos. Que ella se presente a sí misma como «Miriam Seghiri, Universidad de Málaga» nos habla de su sencillez y humildad. Ahora bien, a no confundir estas nobles cualidades con timidez. Miriam es lo que yo suelo definir —en sentido figurado y con respeto, obviamente— como una «topadora».

¿Cómo? Durante un breve encuentro previo con la consejera a cargo y la coordinadora de la Comisión, Miriam accedió gustosa y generosamente a dictar en el CTPCBA la misma charla que venía de dictar en una universidad marplatense. ¡Y sabemos que Beatriz consigue lo que se propone!

¿Qué? La charla que ofreció Miriam Seghiri se titulaba «Cómo traducir con corpus y no morir en el intento».

Estela Lalanne, coordinadora de la Comisión, abrió la presentación leyendo el currículum —abreviadísimo, por supuesto— de Miriam y luego le cedió la palabra. Miriam cautivó al auditorio inmediatamente, por su calidez, su excelente predisposición a responder preguntas, su sentido del humor y simpatía.

A continuación, les contaremos «de qué va» (parafraseando a Miriam) este asunto del corpus y cómo puede ayudarnos en nuestra labor diaria, no solo a los traductores y estudiantes de traducción, sino a los profesionales de otras disciplinas relacionadas con las lenguas, como la terminología. Comencemos por una necesaria introducción.

## >> Cómo traducir con corpus y no morir en el intento

El término *corpus* es latino y significa 'cuerpo'. En el campo de la lingüística, se llama así al conjunto de textos auténticos representativos de la lengua, agrupados en formato electrónico según criterios específicos. Y como los traductores bien sabemos, el texto, es decir, la documentación —cualquiera sea su formato— es una herramienta fundamental para el trabajo diario.

Las fuentes documentales preferidas por los traductores son las siguientes:

- Diccionarios bilingües (el favorito de todos).
- Diccionarios especializados (aunque a veces no tanto); para las tecnologías de punta, no hay diccionarios especializados que cubran todas las áreas.
- Internet, pero sabemos que no suele ser confiable ni segura.

Sin embargo, no disponemos de diccionarios que vayan más allá de la definición y que incluyan todas las combinaciones posibles de palabras con sus respectivos significados.

El recurso más importante para el traductor son los términos *in vivo*. Por lo tanto, el recurso ideal está constituido por los corpus virtuales. Pero para obtenerlos «sin morir en el intento», necesitamos de una metodología protocolizada de compilación de dichos corpus. Para ello, Miriam vino en nuestro rescate con un ejemplo práctico, que fue mechando con teoría sobre corpus.

### Corpus en la práctica y didáctica traductora

**Ejemplo:** corpus bilingüe de textos científico-técnicos, con subcorpus en inglés (u otra lengua) y subcorpus en español.

El corpus presenta las siguientes características:

1. Colección de textos.
2. Texto auténtico.
3. Representa un ámbito o parte de él.
4. Selección en función de criterios específicos.
5. Almacenados en un ordenador.
6. Se utiliza para un fin lingüístico específico.

Tipos de corpus:

- Según el medio:
  - Impreso/electrónico/virtual
  - Escrito/audio/video/multimedia
- Según el diseño:
  - Abierto/cerrado
  - Completo/parcial
  - Oportunista/equilibrado/ReCor

— Según la lengua:

- Monolingüe/multilingüe
- Comparable (original)/paralelo (original + traducción)

Campos de aplicación del corpus:

- Lingüística
- Lexicografía
- Traducción
- Estudios de traducción
- Terminología
- Lengua
- PLN

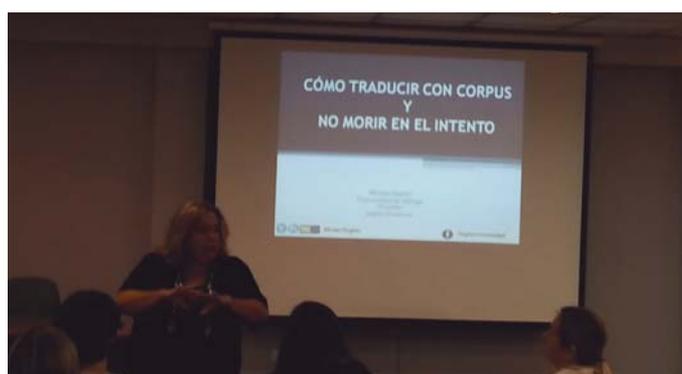
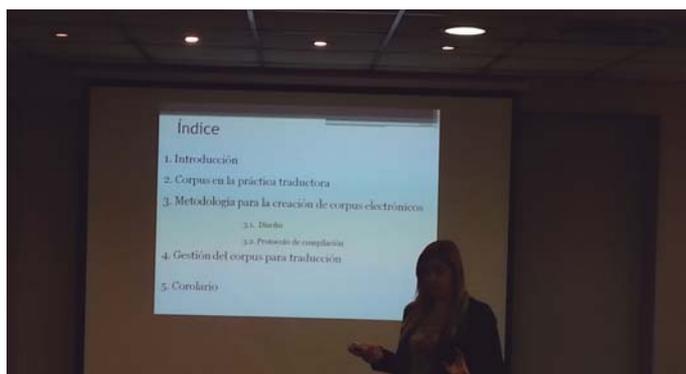
Internet como corpus presenta las siguientes desventajas:

- Textos no compilados para criterios específicos.
- Gran cantidad de documentos.
- La red no se encuentra limpia en términos de formato y lenguas.
- Se desconoce la población específica representada.
- Calidad dudosa.

### Metodología para la creación de corpus virtuales

La norma EN 15038:2006 establece las siguientes pautas:

- ❖ Competencias propias de los traductores: la única competencia que permite cumplir con todas las competencias es un corpus (Varantola, 2003).
  - a. Competencia traslativa
  - b. Competencia lingüística y textual tanto en LO como en LM
  - c. Competencia de investigación, adquisición y procesamiento de información
  - d. Competencia cultural
  - e. Competencia técnica
- ❖ Criterios de diseño: los criterios de diseño del corpus deben ser claros y seguir el protocolo de compilación establecido.
  - a. Comparable: el corpus incluirá documentos originales en inglés y en español.
  - b. Tipo textual: folletos, manuales, guías de usuario, artículos de investigación, etcétera.
  - c. Lenguas.
  - d. Restricciones diatópicas.
  - e. Original/traducción.
  - f. Texto completo/parcial.



#### ❖ Protocolo de compilación:

1. Búsqueda:
  - a. Institucional (organismos oficiales, empresas privadas, etcétera).
  - b. Por palabra clave con motores de búsqueda: Clusty, Yippi, etcétera.
2. Descarga: aplicaciones como BootCat o E-corpora (*proprietary*) ayudan a descargar gran cantidad de texto.
3. Formato: existen varias aplicaciones para convertir texto, como, por ejemplo, ABBY FineReader, ReadIris, etcétera.
4. Almacenamiento: se propone el armado de una carpeta con el título del corpus y subcarpetas con los documentos en formato original (TXT) y convertido, con etiquetas codificadas:

01ESTORF  
 ↓ ↓ ↓  
 N.º ↓  
 Español ↓  
 Texto original

La aplicación LupasRename etiqueta automáticamente los archivos.

- #### ❖ Gestión del corpus para la traducción: lo primero que se necesita es un programa de concordancias (KWIC) para el análisis de patrones lingüísticos más

frecuentes en el ámbito de trabajo. Algunas herramientas útiles son las siguientes: WordSmithTools, Concordance y AntConc (gratuita, multiplataforma). Y además, contar con ReCor (aplicación patentada por Miriam Seghiri), lo cual facilitará notablemente la tarea.

Finalmente, pudimos ver una demostración práctica de la compilación de corpus virtual con las herramientas mencionadas, para la traducción de una hoja técnica. El resultado fue asombroso y permitió constatar no solo la utilidad, sino la conveniencia de implementar esta metodología en las disciplinas asociadas a la lengua. Ahora, solo nos resta poner manos a la obra y practicar.

La charla finalizó pasadas las 21.00. Quedaron olvidados en la recepción los termos con café, pero las casi tres horas corridas del encuentro bien valieron la pena.

La Comisión de Terminología agradece la presencia y el interés de todos los participantes; pero, por sobre todo, agradece infinitamente la generosidad y la buena predisposición de Miriam, quien se ha ganado un espacio de reconocimiento en nuestro Colegio y, como no podía ser de otra manera, una invitación a participar del próximo VI Congreso Latinoamericano de Traducción e Interpretación, «El traductor después del mañana», que tendrá lugar del 21 al 24 de abril de 2016, en la ciudad de Buenos Aires. ■