



## Miriam Seghiri:

# «Armar un corpus lingüístico es uno de los mejores métodos para conocer una lengua»

**Miriam Seghiri, doctora en Traducción e Interpretación y profesora de la Universidad de Málaga, adelanta en esta entrevista los puntos principales de su exposición en el próximo Congreso Universitario de Formación en Traducción e Interpretación, titulada «Traducir con corpus: ¿una misión IMPOSIBLE?». Para Seghiri, es fundamental que todo traductor sepa armar sus corpus, para así generar herramientas necesarias en la traducción de un texto, especialmente con temáticas no tan conocidas.**

| Por Héctor Pavón

### ¿Por qué se sostiene que el corpus posee una importancia clave en la formación de un traductor?

El traductor tiene que estar al tanto de las herramientas que lo pueden ayudar: las de traducción automática (TA), las de traducción asistida por computadora (TAC) y cualquier herramienta tecnológica útil. Las principales están basadas en corpus. Por ejemplo, los diccionarios de la Real Academia Española (RAE). Los miembros de la RAE no se reúnen y dicen «uy, ¿qué palabra vamos a incluir en el diccionario?». No. Tienen muchos corpus: el Corpus del Español del Siglo XXI (CORPES XXI), el Corpus de Referencia del Español Actual (CREA), entre otros. La base de datos multilingüe de la Unión Europea, Interactive Terminology for Europe (IATE), se basa en corpus de la Unión Europea. Un vasto número de las herramientas que emplea hoy el traductor, desde un diccionario impreso o digital hasta una base de datos, se componen de grandes cantidades de corpus, textos que se utilizan para buscar términos, definiciones, colocaciones, equivalencias. Es crucial que ellos formen sus corpus para crear las herramientas específicas que necesiten.

### ¿Por qué se afirma que un buen corpus optimiza el trabajo del traductor?

Los corpus son textos reales de uso. Si un traductor tiene que traducir un texto sobre el *hallux valgus*, que

es una patología del pie que se llama vulgarmente *juanete*, no hay un diccionario sobre esto. Y, si tiene que traducir un texto sobre el último dispositivo tecnológico para ordenadores, no va a encontrar un diccionario específico. Entonces, ¿qué puede hacer? Crearse su propia herramienta. Le puede tocar traducir temas sobre los que hoy no hay diccionarios y sobre los que, por su especialización, tan específicos, tan concretos, seguramente no los haya nunca. Si nadie hizo nada sobre ese tema, no le queda otra alternativa que juntar textos del género y de la temática que tiene que traducir o estudiar, en la lengua original y en la lengua meta, con *software* como AntConc o ParaConc. Se pueden descargar los textos en segundos y con ellos se pueden cubrir todas las necesidades documentales. Puedo buscar la macroestructura, por ejemplo: las condiciones generales de un contrato en español van al principio y en alemán van al final. Puedo ver terminología, buscar equivalentes y conceptos, y no cli-quear aquí y allá en un diccionario. Además, al poder buscar en muchos textos, es posible ver la frecuencia de las palabras en contexto; el traductor necesita ver la terminología en vivo, contextualizada. Por ejemplo, *house: casa, hogar, lar, vivienda*; ¿y ahora cuál de todas ellas escojo en función de un género y una temática determinados (pongamos el caso de un contrato de arrendamiento)? Con un corpus es muy fácil ponerlo en contexto rápidamente.



Miriam Seghiri: «Armar un corpus lingüístico es uno de los mejores métodos para conocer una lengua»



### ¿Qué ocurre cuando el traductor, ya formado, se lanza al mercado laboral?

Va a depender mucho de dónde haya recibido su formación y de si el plan de estudios verdaderamente se adecúa al mercado actual de la traducción. Muchas veces, las universidades no van al ritmo del mercado; entonces, es importante que el alumno piense en seguir formándose, ver qué pide el mercado y no quedarse atrás. Debe ser curioso de todo lo que llega a sus manos. Cuanto más se forme y utilice herramientas que optimicen el trabajo, en menos tiempo podrá hacer su labor, cubrirá más trabajo y ganará más dinero.

### Básicamente, ¿cómo debe componerse el corpus de trabajo?

Normalmente, será de acuerdo con el encargo que haya que cumplir. Imaginemos que me llega un texto sobre patologías del pie, malformaciones. ¿Cómo compondré el corpus? Por ejemplo, con artículos científicos sobre malformaciones del pie, en inglés y en español. En español me servirán para buscar conceptos y en inglés para buscar equivalentes. Y, mediante su explotación, puedo buscar los términos y sus equivalentes al no ser yo, en este caso, traumatóloga. También puedo ver expresiones que no conozco por

no ser especialista en el tema y buscar en conceptos que no conozco por no ser médica.

### Esa búsqueda, ¿la hace pensando en el armado de un campo de especialización o con base en un amplio universo temático?

Los corpus para traductores se hacen en función del encargo que se debe realizar. Me llega un texto de medicina, como el citado, busco textos de la misma temática en las lenguas que tengo que traducir e incluso en la mía propia por ser un concepto que no entiendo. Una vez que los tengo, los junto. Pero no se trata solo de juntarlos, hay que seguir un diseño y un protocolo de compilación que garanticen la calidad de la muestra. Y luego hay un programa que he creado y patentado, llamado ReCor, que permite confirmar que con esos textos se ha cubierto la terminología de ese campo (es el programa con mayor solicitud de licencias de uso en la historia de la Universidad de Málaga). El corpus también es representativo en lo cuantitativo, en el número de términos existentes. Si el corpus es de textos originales, no traducidos, se interrogará con programas como AntConc, WordSmith Tools o Concordance (y es lo que denominamos corpus comparable). Cuando tenemos originales con sus

traducciones, es muy interesante ver los textos en paralelo. Eso se gestiona con *software* de corpus paralelos, como ParaConc o AntPConc. También hay programas que permiten gestionar ambos tipos de corpus, comparables y paralelos, como Sketch Engine.

### ¿Qué herramientas o conocimientos allanan el camino del traductor en el momento de armar un corpus?

Ante todo, tenemos que tener muy claro el diseño y cumplir el protocolo de compilación. Ese protocolo consta de cuatro pasos muy fáciles. En primer lugar, la búsqueda documental en páginas de reconocido prestigio (institucionales, empresariales, etc.), en aras de asegurar la calidad. El segundo paso consiste en la descarga de los documentos que compondrán el corpus. Esto se puede hacer uno a uno, pero también hay programas que permiten descargar lotes de páginas web (como BootCat). El tercer paso es dar el mismo formato a los documentos y convertirlos a formato de texto plano. Por último, se archivan en carpetas específicas. Todo ello sirve para garantizar la representatividad cualitativa de la muestra compilada. Para asegurar la representatividad del corpus desde el punto de vista de

la cantidad, aplicaremos ReCor. Este programa te ayuda a saber si tu corpus cubre el campo de la terminología de tu especialidad, y te aseguras de que están todos los términos que necesitas. Hay muchas herramientas que facilitan corpus ya compilados, como Sketch Engine, pero difícilmente cubrirán el género y la temática que requiere el traductor. Así, cuando son corpus de traducción, es conveniente que los compile el propio traductor.

### ¿En qué consiste el protocolo de compilación?

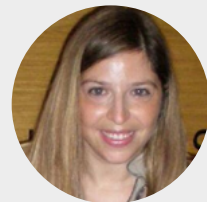
Se compone de cuatro fases: búsqueda de los textos; descarga; formato de esos textos; almacenamiento y codificación. Hay herramientas que permiten optimizar la búsqueda de corpus. Si quiero buscar textos de España o de la Argentina, puedo subrayar que quiero páginas .ar o .es. Luego hay webs, como Clusty, que permiten buscar por *clusters*, es decir, por temáticas. Por ejemplo, escribo *gato* y separa los textos que tienen que ver con el animal y otros sobre la herramienta para levantar los coches. Si no utilizo las palabras clave en la búsqueda, aparecerá el «ruido documental».

### ¿Qué variantes o criterios se deben tener en cuenta cuando componemos un corpus?

No me interesan los corpus enormes, sino que sean de calidad. La calidad viene determinada por la autoría. No es lo mismo que descargue un artículo de investigación de un médico que un chat de medicina. Además, si estás traduciendo como género textual artículos de investigación, no puedes poner revistas divulgativas, porque en lugar del término *hallux valgus* va a poner *juanete*. Siempre es mejor que sean artículos de publicaciones de primera línea, de organizaciones internacionales, de

## ↓ Biografía de Miriam Seghiri

Miriam Seghiri es licenciada en Traducción e Interpretación (español-inglés, francés, italiano) por la Universidad de Málaga (2002). Obtuvo el grado de doctora por la misma universidad (2006), con la calificación de sobresaliente *cum laude* por unanimidad (Premio Extraordinario de Doctorado). Actualmente, es profesora titular en el Departamento de Traducción e Interpretación de dicha casa de estudios.



empresas de reconocido prestigio, de autores de relevancia.

### ¿Qué ocurre cuando el traductor trabaja con dos idiomas distintos?

Nada. Los corpus pueden ser monolingües porque queremos estudiar algo en una lengua determinada. Por ejemplo, un corpus de español de España y de español de la Argentina. Tenemos dos variedades diatópicas del español y podemos ver sus diferencias. Es posible agregar «otros idiomas», como inglés, francés. Imaginemos que tenemos un contrato en inglés y su traducción al español, al francés, al italiano. Podemos interrogarlo en paralelo. ParaConc y Sketch Engine permiten interrogar varios idiomas simultáneamente. Una vez que tenemos el corpus, los textos se suben a AntConc o a ParaConc.

### ¿En qué medida los corpus lingüísticos nos permiten conocer mejor una lengua?

Son el mejor método para conocer una lengua. El traductor se puede servir de ellos para conocer un género o una temática actuales, de modo muy profundo, desde su macroestructura hasta el modo de manejo de la información. Son los llamados corpus sincrónicos. Así, un contrato de compraventa en español tiene las

condiciones generales al principio, y también su terminología, sus colocaciones, sus preposiciones. Es decir, con una herramienta de gestión de corpus adecuada, podemos conocer todo un género de una temática concreta. Asimismo, la RAE tiene millones y millones de textos y palabras, y podemos estudiar la evolución de la lengua (son los denominados corpus diacrónicos): cómo se decía una palabra hace trescientos años y cómo se dice hoy. Es maravilloso, los corpus son una de las mejores maneras que nos permiten conocer una lengua. Sean escritos u orales.

### ¿Lleva tiempo aprender a usar un corpus?

Es muy rápido, como montar en bicicleta: cuanto más pedaleas, más fácil y rápido es. Esto es igual; cuanto más la usas, la metodología es muy fácil. Un corpus yo lo hago en cinco minutos. El primer día que compiles uno te llevará un poquito más. Cuando doy un taller, después de una hora, a los alumnos les digo que armen un corpus en quince minutos y que en otros quince traduzcan un texto que les doy. Lo hacen rapidísimo, son pruebas reales que he hecho. Uno se puede interiorizar muy bien en ellos muy velozmente. ■