

---

## Potencialidad del corpus de interpretación EPIC para el análisis léxico, morfosintáctico y semántico

By Mariachiara Russo (Università di Bologna, Italy)

### Abstract & Keywords

#### English:

Corpus-based Interpreting Studies have a fairly recent development compared to Corpus-based Translation Studies. By corpus we do not mean mere compilations of texts in electronic format but tools according to corpus linguistics criteria with an interface for semi-automated queries. This delay is mainly due to the 'object' of interpretation: oral speeches whose transcriptions are extremely time-consuming, and the limited access to recordings of original speeches and their corresponding interpretations. Faced with the need to rely on a sufficiently large corpus for the study of interpretation, the EPIC (European Parliament Interpreting Corpus) corpus was created in 2004 at the Department of Interpreting and Translation of Forlì: an online corpus that can be freely consulted at <https://corpora.dipintra.it>. This article briefly explains the development phases and features of EPIC and focuses on some contributions to the knowledge of interpreting processes, both semantic and morphosyntactic, made possible by this new approach. Firstly, the results of a comparative study on lexical variety and density between Spanish originals and corresponding Italian interpretations are presented: the study showed that the linguistic output of the interpreter is of greater density and variety than that of the speaker. Secondly, the results of study on morphosyntactic asymmetries and ambiguous Spanish structures that may give rise to interferences during simultaneous interpreting are presented. Finally, the results of another comparative study on lexical, morphosyntactic and semantic processes when interpreting simultaneously from the same Spanish speech into another Romance language (Italian) and a Germanic language (English) are presented: transcripts of speakers' and their interpreters' speeches have shown that the texts of the latter are generally shorter and slower.

#### Spanish:

Los estudios de interpretación basados en corpus (Corpus-based Interpreting Studies) tienen un desarrollo bastante reciente frente a los estudios de traducción basados en corpus (Corpus-based Translation Studies). Por corpus no se entienden meras recopilaciones de textos en formato electrónico sino herramientas según los criterios de la lingüística de corpus con una interfaz para su interrogación semiautomatizada. Este retraso se debe principalmente al 'objeto' de la interpretación: los discursos orales cuya transcripción requiere una dedicación larga e intensa y el limitado acceso a grabaciones de discursos originales y sus correspondientes interpretaciones. Ante la necesidad de contar con un corpus suficientemente amplio para el estudio de la interpretación, en 2004 se realizó en el Departamento de Interpretación y Traducción de Forlì el corpus EPIC (European Parliament Interpreting Corpus): un corpus en línea que se puede consultar libremente en <https://corpora.dipintra.it>. Este artículo expone brevemente las fases de desarrollo y las características de EPIC para centrarse luego en algunas de las aportaciones al conocimiento de los procesos de transferencia lingüístico-semánticos que el estudio de la interpretación mediante ese nuevo instrumento ha hecho posible. En primer lugar, se presentan los resultados de un estudio sobre variedad y densidad léxica comparando la lengua fuente del orador (español) con la lengua meta del intérprete (italiano), que demuestran que la lengua del intérprete resulta de densidad y variedad mayor que la del orador. Seguidamente, se exponen los resultados de un nuevo estudio sobre disimetrías morfosintácticas y estructuras ambiguas españolas que pueden dar lugar a interferencias en la prestación del intérprete simultáneo. Por último, se analizan los procesos y transformaciones léxicas, morfosintácticas y semánticas en interpretación simultánea partiendo del mismo texto en español hacia otra lengua romance (el italiano) y una lengua germánica (el inglés): las transcripciones de los discursos de los oradores y de sus intérpretes han puesto de relieve el hecho de que los textos de estos últimos son generalmente más breves y lentos.

**Keywords:** European Parliament Interpreting Corpus, simultaneous interpreting, morphosyntactic asymmetries, interpretación simultánea, asimetrías morfosintácticas

---

### Introducción

Los estudios de interpretación basados en corpus (*Corpus-based Interpreting Studies*, CIS) tienen un desarrollo bastante reciente frente a los correspondientes estudios de traducción (*Corpus-based Translation Studies*) (Baker 1993, Laviosa 1998a). La necesidad de realizar un corpus para un estudio sistemático de la producción lingüística de los intérpretes de conferencias fue planteada hace tiempo por Shlesinger (1998), pero los primeros corpus aparecen más tarde, como atestigua la panorámica ofrecida por Setton (2011): *CIAIR* (*Center for Integrated Acoustic Information Research*), *Simultaneous Interpretation Corpus* entre inglés y japonés (Tohyama *et al.* 2005, Tohyama, Matsubara 2006), *EPIC* (*European Parliament Interpreting Corpus*) entre inglés, español e italiano (Monti *et al.* 2005, Sandrelli *et al.* 2010), *CoSi* (*Consecutive and Simultaneous Interpreting*) entre alemán y portugués (Meyer 2008), *DIRSI-C* (*Directionality in Simultaneous Interpreting Corpus*) entre inglés e italiano (Bendazzoli 2010a, 2010b). Hoy contamos con las primeras publicaciones exclusivamente dedicadas a los CIS que presentan más corpus todavía (Straniero Sergio, Falbo 2012 y Russo *et al.* 2018).

Por corpus no se entienden meras recopilaciones de textos almacenados en archivos electrónicos, sino herramientas desarrolladas conforme a los criterios de la lingüística de corpus, es decir: una colección de discursos originales y de sus interpretaciones transcritos según determinadas convenciones para formar un corpus

electrónico indexado, *tokenizado*, lematizado y etiquetado para su interrogación semiautomatizada a través de una interfaz (Bernardini *et al.* 2018).

Antes de diseñar y realizar un corpus hay que plantearse una serie de asuntos preliminares referentes a la accesibilidad, cantidad y representatividad de los datos a almacenar. En el caso de la interpretación de conferencias, hay un acceso limitado a grabaciones de discursos originales y a sus respectivas interpretaciones realizadas por profesionales, porque los organizadores de eventos o los mismos intérpretes no siempre lo permiten; por otra parte, la cantidad de datos homogéneos suele ser escasa por la diversidad y unicidad de los eventos. Por consiguiente, el nivel de representatividad de los datos es reducido y no permite extraer conclusiones generalizables sobre esta modalidad de transferencia lingüística. A esto se añade otra peculiaridad de los corpus de interpretación, lo cual explica el retraso en los estudios de la misma: se trata de datos “orales”, los discursos son “evanescentes” por naturaleza y su transcripción requiere una dedicación larga e intensa.

Ante la necesidad de contar con un corpus suficientemente amplio y representativo para el estudio de la interpretación, en 2004 se creó en el Departamento de Interpretación y Traducción de la Universidad de Bolonia en Forlì un grupo multidisciplinar[1], integrado por intérpretes (docentes-investigadores-profesionales), lingüistas computacionales expertos en corpus e informáticos, que realizó el Proyecto *EPIC*. A continuación, se describirán los principales rasgos de *EPIC*, sus aplicaciones y desarrollo futuro para afianzar su potencialidad.

### 1. El Proyecto *EPIC*: el Corpus y el Archivo Multimedia

El corpus contiene discursos pronunciados en inglés, italiano y español en el Parlamento Europeo (PE), así como las interpretaciones simultáneas de cada discurso en dos de los tres idiomas implicados, cubriendo así todas las direcciones y combinaciones lingüísticas posibles. La elección de discursos y sus interpretaciones en el PE nos ha permitido solucionar los obstáculos mencionados anteriormente, es decir, accesibilidad, cantidad y representatividad: se ha grabado a partir de las emisiones en directo de las sesiones plenarias del PE que tuvieron lugar a lo largo de 2004 y que fueron transmitidas por la cadena *Ebs* (*Europe by Satellite*); el material es abundante, ya que se trata de plenarias con los mismos discursos interpretados en 11 idiomas oficiales[2], y homogéneo porque los discursos fueron pronunciados por los eurodiputados en el mismo contexto comunicativo e interpretados hacia su lengua nativa por intérpretes profesionales seleccionados con criterios muy estrictos.

Las emisiones grabadas fueron digitalizadas y almacenadas posteriormente. Se han grabado 280 horas de materiales útiles en 140 cintas VHS; hasta la fecha han sido digitalizadas 55 cintas VHS de las que 37 ya están editadas. Los discursos originales fueron digitalizados como videoclip (.MPEG1), las interpretaciones simultáneas, en cambio, fueron almacenadas como *audioclip* (.wav) en el Archivo Multimedia *EPIC* que contiene también documentación relativa a las sesiones del Pleno. Sucesivamente, los videoclips y los *audioclips* fueron transcritos según las convenciones establecidas a tal fin con tres niveles, a saber: lingüístico (transcripción ortográfica, en base al *Manuale interistituzionale di convenzioni redazionali* de la Unión Europea), paralingüístico (disfluencias: palabras interrumpidas, palabras mal pronunciadas, pausas llenas y pausas vacías) y extralingüístico (cabecera o *header* con información sobre orador y texto).

El corpus *EPIC* resultante[3] está vertebrado en 9 *subcorpus* (véase fig. 1), tres de los cuales contienen los discursos originales, mientras que los otros seis contienen los discursos interpretados. Los tipos de discursos originales contenidos en *EPIC* son monólogos, generalmente rápidos, pronunciados según tres modalidades distintas (leídos, improvisados, mixtos) y generalmente muy técnicos. Los discursos interpretados simultáneamente son una reconstrucción en directo del discurso original, que se lleva a cabo en el momento en que se escucha el discurso mismo. Son reconstrucciones cohesivas y coherentes, y las unidades de significado se interpretan una a una según la capacidad mnemónica del intérprete, del discurso fuente en discurso meta con posibles problemas de planificación, articulación y titubeos.

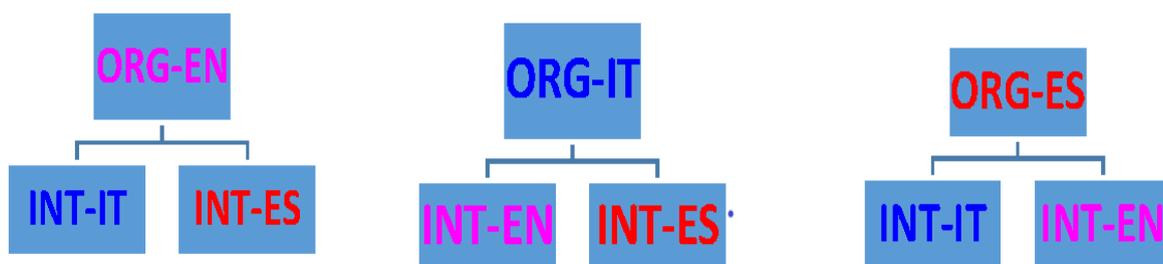


Fig. 1. Arquitectura del Corpus

ORG = discurso original o texto fuente; INT = discurso interpretado o texto meta; EN = inglés; IT = italiano; ES = español

Debido a su estructura, *EPIC* se puede analizar como corpus comparable y como corpus paralelo. Desde una perspectiva comparable, se pueden analizar los discursos originales en un idioma (pongamos el italiano) y las dos interpretaciones hacia el mismo idioma (es decir, del inglés al italiano y del español al italiano). Este enfoque es interesante para quienes quieran estudiar, por ejemplo, las características del mismo idioma hablado por los ponentes originales y por los intérpretes (desde dos idiomas fuente distintos). Por otra parte, *EPIC* se puede analizar como corpus paralelo, es decir, considerando los textos fuente y sus interpretaciones en los otros dos idiomas (por ejemplo, del inglés al italiano y al español). Las dimensiones actuales del corpus rondan las 178.000 palabras. Por ahora, los distintos *subcorpus* son de tamaño diferente (véase tabla 1), pero *EPIC* es un corpus abierto y se están realizando otras transcripciones que se añadirán.

| <i>Subcorpus</i> | N.º de discursos | Número de palabras | % en <i>EPIC</i> |
|------------------|------------------|--------------------|------------------|
| ORG-EN           | 81               | 42.705             | 25               |
| ORG-IT           | 17               | 6.765              | 4                |
| ORG-ES           | 21               | 14.406             | 8                |

|              |            |                |            |
|--------------|------------|----------------|------------|
| INT-IT-EN    | 17         | 6.708          | 4          |
| INT-ES-EN    | 21         | 12.995         | 7          |
| INT-EN-IT    | 81         | 35.765         | 20         |
| INT-ES-IT    | 21         | 12.833         | 7          |
| INT-EN-ES    | 81         | 38.066         | 21         |
| INT-IT-ES    | 17         | 7.052          | 4          |
| <b>TOTAL</b> | <b>357</b> | <b>177.295</b> | <b>100</b> |

Tabla 1: Composición del corpus EPIC

El actual corpus EPIC es un recurso en línea de libre acceso que desde 2006 se encuentra en la página web de nuestro Departamento de Interpretación y Traducción (<https://corpora.dipintra.it>). Además, para fines académicos el corpus EPIC se puede descargar gratuitamente desde de la *European Language Resource Association (ELRA)* ([http://catalog.elra.info/product\\_info.php?products\\_id=1145](http://catalog.elra.info/product_info.php?products_id=1145)) también en formato audio y vídeo con transcripciones.

## 2. Tipos de investigaciones con EPIC

EPIC es un corpus electrónico, es decir, se puede explotar utilizando las herramientas de la lingüística de corpus para extraer automáticamente o semiautomáticamente datos analizables. Esto es posible gracias a que cada palabra de las transcripciones lleva etiquetas que indican su categoría gramatical (POS, *Part-Of-Speech tag*) y el lema. Este proceso se realizó por medio de etiquetadores distintos según el idioma: *TreeTagger* (Schmid 1994) para el inglés y el italiano, y *Freeling* (Carreras *et al.* 2004) para el español. La indexación del corpus se ha realizado mediante los recursos del *IMS Corpus Work Bench – CWB* (Christ 1994), lo que permite llevar a cabo búsquedas simples y avanzadas a través de una interfaz especial en la que se pueden introducir distintos parámetros (por ejemplo, el modo de presentación del texto, el tema, el tipo de ponente, etc.) para filtrar las búsquedas. Todos los parámetros disponibles están reunidos en la cabecera o *header* inicial, que proporciona información acerca del archivo de texto, del ponente (nombre, género, cargo, etc.) y del discurso (tema, número de palabras, duración, etc.). Las búsquedas simples incluyen palabras individuales o secuencias de palabras (véase fig. 2):

The screenshot shows the SSLMIT Dev Online web application in a Microsoft Internet Explorer browser. The page title is 'SSLMIT Dev Online' and the URL is 'http://sslmittdev-online.sslmit.unibo.it/corpora/results.php?mode=simple&path=E.P.I.C./Source%20Texts&name=org-es'. The interface includes a navigation menu on the left with options like 'Corpora Home', 'Simple Query', and 'Advanced Query'. The main content area displays 'ORG-ES SIMPLE QUERY RESULTS' for a query 'su señoría'. It shows 'Query cached (QRY1221640385)', 'Results (Random set): 3 out of 3 found', and a list of search results with highlighted words and their POS tags. The search results include text snippets such as 'crecimiento se contempla cuál es el nivel de cobertura su señoría en varios países de la Unión se están llevando a cabo reformas para ehm garantizar de durabilidad en el tiempo a estos sistemas de pensiones' and 'pero bueno qué le vamos a hacer // no se trata de su señoría // decir que el que Afganistán no aparezca específicamente contemplado no quiere decir que no vamos a mantener su la acción en Afganistán // en cuanto a la Agenda de Lisboa Señorías su señoría que lo que hace falta es cumplir con lo que se acordó en Lisboa cumplir llevar a cabo y desarrollar en todos los países de'.

Fig. 2: Búsqueda simple: su señoría

Las búsquedas avanzadas incluyen POS, lemas, palabras interrumpidas o mal pronunciadas (las únicas disfluencias anotadas), combinación de palabra+POS, palabra+LEM, etc. (véase fig. 3):



Fig. 3: Búsqueda avanzada: *ir* + gerundio

Todas las aplicaciones y los estudios que se exponen en el sucesivo apartado 3 fueron realizados con la interfaz arriba descrita y activa desde 2006 hasta 2018. Ahora EPIC y los demás corpus del Departamento (<https://corpora.dipintra.it>) tienen una nueva única interfaz con el mismo método de búsqueda e interrogación: NoSketch Engine.

### 3. Aplicaciones

Los materiales de *EPIC* han sido explotados principalmente para tres tipos de aplicaciones: en la didáctica para prácticas en clase, como material fuente para tesis de licenciatura y para la investigación.

En cuanto a la didáctica, la cantidad y variedad de interpretaciones simultáneas del mismo discurso en otros dos idiomas ofrece al estudiante de interpretación un abanico realmente único de estilos profesionales sobre los que reflexionar y forjar el modelo de prestación profesional al que desea aspirar. Además, al contar con grabaciones de discursos originales y con sus interpretaciones, los estudiantes pueden practicar en clase y cotejar su prestación con las de los profesionales (Sandrelli 2010). Ambas aplicaciones se han demostrado verdaderos catalizadores del aprendizaje para el estudiante de interpretación.

En cuanto a las tesis de licenciatura (Russo 2010), se han realizado numerosos estudios que han investigado aspectos gramaticales (elaboración de nombres propios, prefijos, anglicismos, verbos en posición final, etc.), pragmáticos (mitigación, cortesía, rasgos interpersonales, etc.) y rasgos estratégicos frente a problemas específicos (metáforas, alta velocidad de habla, etc.).

En cuanto a las investigaciones, los análisis cuantitativos y cualitativos de los procesos de transferencia lingüístico-semánticos en *EPIC* han proporcionado resultados originales y han abierto caminos novedosos en el estudio de la interpretación.

En primer lugar, se estudiaron las pautas léxicas comparando discursos originales y discursos interpretados desde una perspectiva comparable trilingüe inglés-español-italiano (Russo *et al.* 2006). El punto de partida fue un estudio de Laviosa (1998b), el cual había descubierto que la lengua del traductor en el *Translational English Corpus* resulta ser de menor densidad (hay menos palabras-contenido que palabras-función) y variedad (un número reducido de palabras aparecen con mayor frecuencia) frente a la lengua del original. Utilizando la misma metodología, nuestros resultados fueron distintos: las tendencias que se destacan en *EPIC* indican que la densidad léxica es generalmente superior en los discursos interpretados que en los originales en la misma lengua (solo hay dos excepciones: interpretación del español al italiano y del español al inglés, frente a sus respectivos originales en la misma lengua), mientras que la variedad léxica es generalmente inferior en los discursos interpretados que en los originales en la misma lengua, con excepción del italiano. El resultado de la cabina italiana puede interpretarse como una muestra de estilo más rebuscado, pero también como un intento por parte de los intérpretes de alejarse todo lo posible de la forma del texto fuente supuestamente por miedo a producir interferencias y calcos (véase también un breve ejemplo en la tabla 2).

En un segundo estudio (Bendazzoli *et al.* 2011) analizamos la incidencia de las palabras mal pronunciadas y de las palabras interrumpidas entre discursos fuente y discursos meta, suponiendo que por la sobrecarga cognitiva de la interpretación simultánea estas disfluencias serían más frecuentes en las actuaciones de los intérpretes. Descubrimos que, entre estos dos tipos de disfluencias, las palabras interrumpidas son generalmente más frecuentes que las mal pronunciadas. Además, se confirma la hipótesis inicial, es decir, los oradores tienden a cometer menos errores de producción que los intérpretes, con algunas tendencias contrarias destacadas: los oradores ingleses dejan más palabras sin acabar que sus intérpretes españoles e italianos y los intérpretes ingleses cometen menos errores de pronunciación con respecto a sus oradores italianos y españoles.

Otro estudio (Russo 2011) se centró en las interferencias durante la interpretación simultánea del español al italiano. La modalidad de búsqueda avanzada que nuestro corpus anotado *EPIC* contempla permite encontrar automáticamente en el texto fuente estas unidades sintácticas compuestas y agilizar el análisis de su efecto en el texto meta del intérprete. Concretamente nuestro estudio se focalizó en el efecto de disimetrías morfosintácticas (Russo 1990) y de ambigüedades léxico-sintácticas (Morelli 2010). En nuestro corpus solo aparecen las siguientes:

*ir* + gerundio, *quando* + subjuntivo, *de que, seguir* + gerundio, *dejar de* + V; *su / su señoría, por, hablamos* – tiempo presente/pretérito. Como se desprende de nuestros resultados, estas estructuras plantean escasos problemas para los intérpretes del PE que trabajan del español al italiano, a diferencia de lo que se observó en otros contextos (Russo 1990, Morelli 2010). A veces estas estructuras solo ocasionan leves pérdidas de cohesión y/o coherencia. Cabe señalar que el corpus *EPIC* todavía cuenta con una cantidad de discursos españoles reducida (21) y sería preciso obtener más datos para poder confirmar nuestras observaciones.

Los estudios más recientes sobre *EPIC* han investigado las pautas morfosintácticas y la elaboración semántica en el discurso del intérprete simultáneo que trabaja del español al italiano (Russo 2014) y, en la totalidad del corpus trilingüe, la perspectiva de género que ha revelado algunos efectos estadísticamente significativos en la actuación de mujeres intérpretes frente a sus colegas varones (Russo 2016a, 2016b, 2018).

Las transcripciones de los discursos de los oradores y de sus intérpretes en *EPIC* han puesto de manifiesto interesantes rasgos recurrentes, entre estos estriba el hecho de que los textos de los intérpretes son en la gran mayoría de los casos más breves y lentos. Detectar mecanismos relevantes en la producción lingüística del intérprete desde un enfoque comparativo parece sumamente interesante tanto desde un prisma psicolingüístico como pedagógico. *EPIC* ofrece además la posibilidad de llevar a cabo pormenorizados análisis de los procesos de transformaciones léxicas, morfosintácticas y semánticas partiendo del mismo texto hacia dos lenguas afines (pongamos del inglés al español y al italiano), o bien diferentes (del español hacia otra lengua romance como el italiano y hacia una lengua germánica como el inglés). Como *EPIC* cuenta con una anotación básica, pueden añadirse otras según el enfoque de la investigación.

Por ahora los análisis cualitativos de tipo semántico se pueden realizar alineando manualmente el texto fuente y los dos textos meta como en el ejemplo siguiente (véase tabla 2; para un análisis más extenso, véase Russo 2014). Se trata de un fragmento de un discurso en español sobre la Política agrícola común (PAC):

| Original español   | Interpretación en italiano   | Interpretación en inglés  |
|--|--|---|
| muchas gracias Presidente // (1) <b>me sumo al elogio general</b> de este magnífico informe del señor Garot (1.1) que es tan oportuno como necesario // (2) señor Presidente (2.1) el riguroso análisis del señor Garot <b>demuestra con datos</b> nuestra inquietud ante el <b>peligro</b> de la reforma de la PAC (2.2) si no se le se neutraliza su impacto negativo sobre las rentas agrícolas // (3) lamento mucho (3.1) <b>la grave equivocación incoherencia</b> en la que incurren algunos colegas que por un lado se deshacen en elogios del informe porque es bueno pero a la vez de una manera muy incoherente anuncian (3.2) <b>que no van a apoyarlo</b> // | grazie Presidente grazie Presidente // (1) <b>a mia volta vorrei riprendere quanto hanno detto i colleghi che si sono congratulati</b> col collega Garot per questa relazione opportuna e necessaria // (2) Presidente (2.1) l'analisi del collega Garot <b>dimostra sulla base di elementi concreti che la nostra preoccupazione</b> di fronte alla riforma della politica agricola comune è <b>giustificata</b> (2.2) se l'impatto negativo sui redditi agricoli non verrà neutralizzato // (3) mi dispiace (3.1) <b>che sia stata data un'interpretazione errata cioè che non ci sia sufficiente coerenza</b> nelle considerazioni di alcuni colleghi che si rallegrano col relatore per l'ottima relazione e poi in modo del tutto incoherente alla fine del loro intervento annunciano (3.2) <b>che non voteranno a favore</b> // | thank you President // (1) <b>I'd like to support the general praise</b> for this excellent support by Monsieur Garot (1.1) which is both expedient and necessary // (2) President (2.1) the rigorous analysis carried out by Monsieur Garot <b>shows and proves with figures</b> our concern about <b>the danger</b> of reform of the CAP (2.2) unless its negative impact on farm income is offset // (3) I'm very sorry (3.1) <b>about the ambiguity and incoherence</b> being expressed by some colleagues ... who on the one hand praised the report because it's good but then very illogically say (3.2) <b>that they are not going to support it</b> // |

**Tabla 2: Fragmento de un discurso en español interpretado en italiano y en inglés en simultánea**

Comparando los dos textos meta se puede apreciar la diversidad de planteamiento entre el intérprete italiano y el intérprete inglés con más reelaboración semántica y sintáctica por parte del primero (pese al mayor paralelismo entre español e italiano, Russo 2012). La mayor reelaboración en italiano se manifiesta con expansiones léxico-sintácticas que hacen que el texto meta sea más explícito (segmentos 1, 2.2 y 3.2), pero también con una desviación semántica (segmento 3.1).

### Conclusiones

El corpus *EPIC* es un recurso que nos permite arrojar luz a los mecanismos de la interpretación simultánea y los rasgos de las lenguas en contacto en base a una gran cantidad de datos del habla oral procedentes de situaciones comunicativas profesionales reales. Las próximas etapas de este esfuerzo colectivo, “épico”, prevé la realización de una herramienta más potente aún, *EPIC 2.0*, de doble tamaño y con alineación audio, texto y vídeo para llevar a cabo análisis mucho más pormenorizados y rápidos (Russo et al. 2012). El campo de la investigación lingüística y de sus aplicaciones didácticas, con especial referencia a nuestro ámbito, que es el de la interpretación de conferencias, es amplísimo y esperamos que el novedoso paradigma de estudios basado en corpus pueda enriquecer el panorama de conocimientos a nuestro alcance.

## Bibliografia

- Baker, Mona (1993) "Corpus Linguistics and Translation Studies: Implications and Applications", en Baker, Gill, Tognini-Bonelli (eds): 233-250.
- Baker, Mona; Gill, Francis; Tognini-Bonelli, Elena (eds) (1993) *Text and Technology: in Honour of John Sinclair*, Amsterdam/Philadelphia: John Benjamins.
- Bendazzoli, Claudio (2010a) *Il Corpus "DIRSI": creazione e sviluppo di un corpus elettronico per lo studio della direzionalità in interpretazione simultanea*, Tesi dottorale non pubblicata, Dipartimento SITLeC, Università di Bologna.
- Bendazzoli, Claudio (2010b) *Corpora e interpretazione simultanea*, Alma-DL, Università di Bologna. URL: <http://amsacta.unibo.it/2897/>
- Bendazzoli, Claudio, Sandrelli, Annalisa, Russo, Mariachiara (2011) "Disfluencies in simultaneous interpreting: a corpus-based analysis", en Kruger; Wallmach; Munday (eds): 282-306.
- Bernardini, Silvia; Ferraresi, Adriano; Russo, Mariachiara (2018) "Building Interpreting and Intermodal Corpora: A How to for a Formidable Task", en Russo; Bendazzoli; Defrancq (eds): 21-42.
- Calvo, Cesáreo; Spinolo, Nicoletta (eds) (2016) *Translating orality/La traducción de la oralidad*, «MonTI Special issue» III.
- Carreras, Xavier; Chao, Isaac; Padró, Lluís; Padró, Muntsa (2004) "Freeling: An open-source suite of language analyzers", en *Proceedings of the 4th International Conference on Language Resources and Evaluation* (Vol. I) Lisbona: ELRA: 239-242.
- Christ, Oliver (1994) "A modular and flexible architecture for an integrated corpus query system", en *Proceedings of COMPLEX'94, 3rd Conference on Computational Lexicography and Text Research, Budapest, Hungary, July 7-10*: 23-32. URL: <https://arxiv.org/abs/cmp-1g/9408005>
- Gatta, Francesca (ed) (2016) *Parlare insieme. Studi per Daniela Zorzi*, Bologna: BUP.
- Gran, Laura; Taylor, Christopher (eds) (1990) *Aspects of applied and experimental research on conference interpretation*, Udine: Campanotto Editore.
- Kruger, Alet; Wallmach, Kim; Munday, Jeremy (eds) (2011) *Corpus-based Translation Studies: Research and Applications*, London/New York: Continuum.
- Laviosa, Silvia (ed) (1998a) "The corpus-based approach", en *META* (LXIII/IV): 474-664. URL: <https://www.erudit.org/en/journals/meta/1998-v43-n4-meta169/003424ar.pdf>
- Laviosa, Silvia (1998b) "Core patterns of lexical use in a comparable corpus of English narrative prose", en *META* (LXIII: IV): 557-570.
- Meyer, Bernd (2008) "Interpreting Proper Names: Different Interventions in Simultaneous and Consecutive Interpreting?", en *trans-kom* (I/I): 105-122. URL: [http://www.trans-kom.eu/ihv\\_01\\_01\\_2008.html](http://www.trans-kom.eu/ihv_01_01_2008.html)
- Monti, Cristina; Bendazzoli, Claudio, Sandrelli, Annalisa, Russo, Mariachiara (2005) "Studying Directionality in Simultaneous Interpreting through an Electronic Corpus: EPIC (European Parliament Interpreting Corpus)", en *META* (LXX/IV) URL: <https://www.erudit.org/fr/revues/meta/2005-v50-n4-meta1024/019850ar/>
- Morelli, Mara (2010) *La interpretación español-italiano: planos de ambigüedad y estrategias*, Granada: Comares.
- Ohnheiser, Ingeborg; Pöckl, Wolfgang; Sandrini, Peter (eds) (2011) *Festschrift in Honour of Prof. Lew Zybatow Translation – Sprachvariation – Mehrsprachigkeit*, Frankfurt am Main: Peter Lang.
- Russo, Mariachiara (1990) "Disimetrías y actualización: un experimento de interpretación simultánea (español-italiano)", en Gran, Taylor (eds): 158-225.
- Russo, Mariachiara; Bendazzoli, Claudio; Sandrelli, Annalisa (2006) "Looking for lexical patterns in a trilingual corpus of source and interpreted speeches: extended analysis of EPIC (European Parliament Interpreting Corpus)", en *Forum. International journal of interpretation and translation* (IV/I): 221-254.
- Russo, Mariachiara (2010) "Reflecting on Interpreting Practice: Graduation Theses Based On the European Parliament Interpreting Corpus (EPIC)", en Zybatow (ed): 35-50.
- Russo, Mariachiara (2011) "Text Processing Patterns in Simultaneous Interpreting (Spanish-Italian): a Corpus-based Study", en Ohnheiser; Pöckl; Sandrini (eds): 83-103.
- Russo, Mariachiara (2012) *Interpretare lo spagnolo. L'effetto di dissimmetrie morfosintattiche nella simultanea*, Bologna: CLUEB.
- Russo, Mariachiara; Bendazzoli, Claudio; Sandrelli, Annalisa; Spinolo, Nicoletta (2012) "The European Parliament Interpreting Corpus (EPIC): implementation and developments", en Straniero Sergio, Falbo (eds): 53-90.
- Russo, Mariachiara (2014) "Fenomeni dell'oralità critici per l'interpretazione simultanea: un'analisi contrastiva spagnolo-italiano basata sul corpus EPIC", en *Cuadernos AISPI. Estudios de lenguas y literaturas hispánicas*, IV, 165-182.
- Russo, Mariachiara (2016a) *Orality and Gender: A corpus-based study on lexical patterns in simultaneous interpreting*, en Calvo; Spinolo (eds): 307-322.
- Russo, Mariachiara (2016b) "Cooperativismo retorico in una prospettiva di genere, ovvero il 'noi' in monologhi originali e interpretati", en Gatta (ed): 335-345.
- Russo, Mariachiara; Bendazzoli, Claudio; Defrancq, Bart (eds) (2018a) *Making Way in Corpus-based Interpreting Studies*, Series: New Trends in Translation Studies, Singapore: Springer Verlag.
- Russo, Mariachiara (2018b) "Speaking patterns and gender in the European Parliament Interpreting Corpus. A quantitative study as a premise for qualitative investigations", en Russo; Bendazzoli; Defrancq (eds): 115-131.
- Sandrelli, Annalisa; Bendazzoli, Claudio; Russo, Mariachiara (2010) "European Parliament Interpreting Corpus (EPIC): Methodological issues and preliminary results on lexical patterns in SI", en *International Journal of*

*Translation* (XXII/I-II): 165-203.

Sandrelli, Annalisa (2012) *Corpus-based Interpreting Studies and Interpreter Training: A Modest Proposal*, en Zybatow (ed): 69-90.

Schmid, Helmut (1994) "Probabilistic Part-of-Speech tagging using decision trees", en *Proceedings of International Conference on New Methods in Language Processing*, Manchester: 44-49. URL: <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/data/tree-tagger1.pdf>

Setton, Robin (2011) "Corpus-based interpreting studies (CIS): reflections and prospects", en Kruger, Wallmach, Munday (eds): 33-75.

Shlesinger, Miriam (1998) "Corpus-based interpreting studies as an offshoot of corpus-based translation studies, en *META* (LXIII/IV): 486-493.

Straniero Sergio, Francesco; Falbo, Caterina (eds) (2012) *Breaking Ground in Corpus-Based Interpreting Studies*, Bern: Peter Lang.

Tohyama, Hitomi; Matsubara, Shigeki; Kawaguchi, Nobuo; Inagaki, Yasuyoshi (2005) "Construction and Utilization of Bilingual Speech Corpus for Simultaneous Machine Interpretation Research", en *Proceedings of 9th European Conference on Speech Communication and Technology (Interspeech-ICSLP 2005)* pp.1585-1588. URL: <http://hdl.handle.net/2237/93>

Tohyama, Hitomi; Matsubara, Shigeki (2006) "Collection of Simultaneous Interpreting Patterns by Using Bilingual Spoken Monologue Corpus", en *Proceedings of 5th International Conference on Language Resources and Evaluation, May 2006 (LREC-2006)*: 2564-2569.

Zybatow, Lew (ed) (2010) *Translationswissenschaft – Stand und Perspektiven. Innsbrucker Ringvorlesungen zur Translationswissenschaft VI (= Forum Translationswissenschaft, Band 12)*. Frankfurt am Main: Peter Lang.

## Notas

[1] El grupo estaba integrado por Elio Ballardini, Marco Baroni, Claudio Bendazzoli, Silvia Bernardini, Gabriele Mack, Peter Mead, Cristina Monti, Annalisa Sandrelli, Nicoletta Spinolo y Mariachiara Russo (coordinadora). La interfaz web fue realizada por Lorenzo Piccioni y Eros Zanchetta.

[2] Actualmente los idiomas oficiales son 24.

[3] Conseguimos el permiso de los Servicios Audiovisuales y Mediateca del Parlamento Europeo para realizar y compartir este corpus con fines académicos.

---

©inTRAlinea & Mariachiara Russo (2019).

"Potencialidad del corpus de interpretación EPIC para el análisis léxico, morfosintáctico y semántico", *inTRAlinea* Special Issue: Le ragioni del tradurre.

Stable URL: <http://www.intraline.org/specials/article/2371>