

## **EPIC 2.0: ampliamento del corpus e allineamento testo-audio-video**

### **Obiettivi e compiti del/la borsista**

#### **Breve stato dell'arte**

Gli studi di interpretazione basati su corpora vengono per la prima volta menzionati da Miriam Shlesinger in “Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies” (1998) ove auspica l’adozione di questa metodologia, già sviluppata negli studi di traduzione, “through (1) the creation of parallel and comparable corpora comprising discourse which is relevant to interpreting; and (2) the use of existing monolingual corpora as sources of materials for testing hypotheses about interpreting’ (1998: 486). Il primo corpus di interpretazione online a disposizione della comunità scientifica è stato creato nel 2004 da un’equipe interdisciplinare proprio nel Dipartimento di Interpretazione e Traduzione dell’Università di Bologna a Forlì: l’European Parliament Interpreting Corpus (EPIC), un corpus trilingue (spagnolo-inglese-italiano), lemmatizzato, POS-tagged, indicizzato, di circa 180.000 parole (Monti et al. 2005, Russo et al 2012) con discorsi originali e corrispondenti interpretazioni simultanee. Numerosi sono stati gli studi condotti su EPIC (tra cui Sandrelli et al. 2010, Russo et al. 2012), anche a livello di tesi laurea magistrale (Russo 2010). Nel tempo, EPIC ha ispirato la creazione di altri corpora di interpretazione con nuove combinazioni linguistiche: EPIC-G (Università di Ghent, DeFrancq 2015), EPIC-UdS (Univ. des Saarlandes, Przybyl et al. 2022), PINC (Korzinek and Chmiel, 2021) e i corpora intermodali (testi originali, interpretati e tradotti) EPTIC (Università di Bologna, Bernardini et al. 2016) e TIC (University of Poznań, Kajzer-Wietrzny 2018).

La creazione di corpora di interpretazione che raccolgono un ampio numero di prestazioni di interpreti, e non studi di casi singoli su cui prevalentemente vertevano gli studi di interpretazione fino all’affermarsi di questa nuova metodologia, ha consentito di mettere in luce tendenze significative nei comportamenti linguistici, cognitivi e strategici di interpreti simultaneisti, con importanti ricadute teoriche e pedagogiche. Un’ampia panoramica dei principali risultati degli studi su corpora di interpretazione si trovano in Straniero Sergio e Falbo (2012), Russo et al (2018) e Bendazzoli et al. (2018). I diversi corpora di interpretazione creati finora non sempre condividono le medesime caratteristiche di progettazione e composizione, o identiche convenzioni di trascrizione, e ciò non permette di aggregare i dati e giungere così a risultati ancor più solidi. Pertanto è sorta l’esigenza di creare un mega-corpus che unifichi vari corpora di interpretazione esistenti per ottenere risultati e conclusioni generalizzabili e di cui EPIC 2.0 sarà una componente fondamentale.

## **Obiettivo del progetto**

L'obiettivo del progetto EPIC 2.0 è quello di ampliare il corpus EPIC che attualmente si può consultare liberamente online (<https://bellatrix.sslmit.unibo.it/noske/epic/#open>). EPIC è un corpus composto da 81 discorsi originali inglesi (org-en), 21 discorsi originali spagnoli (org-es) e 17 discorsi originali italiani (org-it) e le rispettive interpretazioni simultanee di ciascuno di essi (int-en-it, int-en-es, int-es-en, int-es-it, int-it-en, int-it-es) per un totale di 357 discorsi. I discorsi sono stati pronunciati nelle sessioni plenarie da membri del Parlamento europeo tra febbraio e luglio 2004 e sono stati registrati tramite il canale satellitare Europe by Satellite. L'attuale versione online del corpus EPIC (nella sola versione trascritta) costituisce solo una parte dell'ingente materiale salvato nell'Archivio multimediale EPIC (180 VHS con 280 ore di materiale utile), ove sono presenti anche gli audio-video originali degli oratori e gli audio delle interpretazioni simultanee registrate dalle cabine italiana, inglese e spagnola.

La persona selezionata avrà il ruolo di realizzare l'espansione di EPIC, il riequilibrio tra le 3 componenti linguistiche e l'allineamento audio-video-trascrizione discorso originale-trascrizione discorso interpretato. Inoltre sarà direttamente coinvolta nelle fasi di creazione di un mega-corpus di interpretazione unificato e nei meeting con il team di ricerca per essere pienamente partecipe delle attività e degli sviluppi del progetto.

## **Compiti del borsista:**

- EPIC: trascrizione con software ASR delle videoclip EN-IT-ES (circa 20 ore)
- EPIC: controllo delle trascrizioni già effettuate dagli studenti EN-IT-ES
- EPIC: unificazione del corpus esistente con le nuove trascrizioni
- EPIC: allineamento testo-video-audio
- EPIC: predisposizione della nuova interfaccia online per l'interrogazione del corpus

## **Profilo del borsista**

Il/la candidato/a, di lingua madre italiana, deve possedere competenze di linguistica dei corpora e competenze di interpretazione di conferenza. Deve possedere una conoscenza approfondita principalmente di inglese (attivo e passivo) e, come seconda lingua, di spagnolo (anche solo passiva). Inoltre, il/la candidato/a deve essere in grado di utilizzare software di riconoscimento automatico della voce.

## Bibliografia citata

- Bendazzoli, C., M. Russo and B. Defrancq (2018) “Corpus-based interpreting studies: The state of the art. A booming research field: an Introduction by the guest editors”, *Special Issue: New Findings in Corpus-based Interpreting Studies, Intralinea* (20) [http://www.intralinea.org/specials/article/corpus\\_based\\_interpreting\\_studies\\_a\\_booming\\_research\\_field](http://www.intralinea.org/specials/article/corpus_based_interpreting_studies_a_booming_research_field)
- Bernardini, S., Ferraresi, A. and Miličević, M. (2016) “From EPIC to EPTIC – Exploring simplification in interpreting and translation from an intermodal perspective”, *Target*, 28, (1), 58–83.
- Defrancq, B. (2015) “Corpus-based research into the presumed effects of short EVS”, *Interpreting*, 17, (1), 26–45.
- Kajzer-Wietrzny, M. (2018) “Interpretese vs. Non-native Language Use: The Case of Optional That”. In Russo, M., Bendazzoli, C. and Defrancq B. (eds.), *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer Nature, 97-114.
- Korzinek, D. and A. Chmiel (2021) “Interpreter identification in the Polish Interpreting Corpus”, *Revista Tradumàtica*, 19: 276–288.
- Monti, C., Bendazzoli, C., Sandrelli, A. and Russo, M. (2005) “Studying Directionality in Simultaneous Interpreting through an Electronic Corpus: EPIC (European Parliament Interpreting Corpus)”. *Meta*, 50 (4). <https://doi.org/10.7202/019850ar>
- Przybyl, H., E. Lapshinova-Koltunski, K. Menzel, S. Fischer, and E. Teich (2022) “EPIC UdS - Creation and Applications of a Simultaneous Interpreting Corpus”, *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, pages 1193–1200 Marseille, 20-25 June 2022 <https://aclanthology.org/2022.lrec-1.127.pdf>
- Russo, M. (2010) “Reflecting on interpreting practice: Graduation theses based on the European Parliament Interpreting Corpus (EPIC)”. In L. Zybatow (ed.) *Translationswissenschaft – Stand und Perspektiven. Innsbrucker Ringvorlesungen zur Translationswissenschaft VI* (Forum Translationswissenschaft, Band 12), Frankfurt am Main, Peter Lang, 35-50.
- Russo, M., Bendazzoli, C., Sandrelli A. and Spinolo N. (2012) “The European Parliament Interpreting Corpus (EPIC): Implementation and developments”. In F. Straniero Sergio e C. Falbo (eds.) *Breaking Ground in Corpus-Based Interpreting Studies*, Bern, Peter Lang, 35-90.
- Russo, M., Bendazzoli, C. and Defrancq, B. (eds) (2018) *Making Way in Corpus-based Interpreting Studies*. Series: New Trends in translation Studies. Singapore: Springer.
- Sandrelli, A., Bendazzoli, C. and Russo, M. (2010) “European Parliament Interpreting Corpus (EPIC): Methodological issues and preliminary results on lexical patterns in SI”, *International*

*Journal of Translation* 22 (1-2): 165-203.

Shlesinger, M. (1998) “Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies” Special Issue of *Meta*, 43 (4), 486–493.

Straniero Sergio, F. and Falbo, C. (eds.) (2012) *Breaking Ground in Corpus-Based Interpreting Studies*, Bern: Peter Lang