

## CORPUS-BASED STUDIES IN CONFERENCE INTERPRETING

*M. Russo*<sup>1</sup>

<sup>1</sup>Bologna University  
136 Course of the Republic, Forlì, 47100, Italy  
Submitted on September 15, 2018  
doi: 10.5922/2225-5346-2019-1-6

*Corpus-based interpreting studies (CIS) are a relatively recent “Off-shoot of Corpus-based Translation Studies” to quote the seminal paper (1998) by the late Miriam Shlesinger, a constant source of inspiration for the T&I community. This line of research is now gaining ground in both conference interpreting and community interpreting. The present paper focuses on conference interpreting and covers the evolution of the concept of interpreting corpus by providing an overview of the most representative examples, from the early collections of transcribed source and target speeches to full-fledged machine-readable corpora based on corpus linguistic standards and tools. Furthermore, methodological issues and original results from a variety of recent CIS are presented.*

**Keywords:** *parallel corpus, comparable corpus, multimodal corpus, intermodal corpus, transcription, metadata.*

### Introduction

Over the decades conference interpreting has been studied through a variety of paradigms: cognitive, psycholinguistic, neurolinguistic, sociolinguistic, linguistic, pragmatic (Pöchhacker 2015). It was not until recently, however, that the prescriptive or anecdotal approaches to professional interpreters' performances, mainly based on the observation of a very limited number of interpreters during a handful of communicative events, have been enriched by descriptive approaches made possible by the implementation of new methodologies developed in the field of corpus linguistics (Bernardini and Russo 2018). This approach had already been embraced by translation studies scholars thus enabling 'a major leap from prescriptive to descriptive statements, from methodologizing to proper theorizing, and from individual and fragmented pieces of research to powerful generalizations' (Baker 1993, 248). An early milestone is the special issue of the journal *Meta*, published in 1998 and edited by Sara Laviosa, which established the corpus-based approach as a new paradigm in translation studies. That issue contained the seminal work "Corpus-based interpreting studies as an offshoot of corpus-based translation studies" by Miriam Shlesinger who was the first scholar to highlight the relevance and potential of the corpus-based approach for research into interpreting. She suggested that the corpus linguistics (CL) methodology could be extended to interpreting, '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' (Shlesinger 1998, 486). Interpreting corpora would add a new dimension to interpreting studies because they would overcome anecdotal observations and also provide information typical of CL, i.e. word frequencies, type-token ratios (lexical variety), co-occurrences, lexical density, grammatical constructions, textual operations, discourse patterns, etc. (ib.).

Shlesinger's call was first put into practice several years later by a multidisciplinary team made up of interpreting scholars/trainers, corpus linguists and IT technicians of the University of Bologna. They developed the first online machine-readable interpreting corpus, the European Parliament Interpreting Corpus (EPIC) (Monti et al. 2005; Russo et al. 2012), a trilingual corpus of source and target speeches delivered during EP sessions (see further on).

An interpreting corpus is a systematized, machine-readable collection of a mass of interpreters' performances, which lends itself to both quantitative and qualitative analyses. Interpreting corpora are insightful for many reasons. They are key resources for the observation and analysis of the surface structure organization of interpreting data of different natures. Rather than attempt to read the interpreter's mind, interpreting corpora provide an insight into textual operations: many of them, by multiple interpreters, in multiple settings (conference, institutional assemblies, community, court, and media), modes (sign-language, dialogue, simultaneous, consecutive, remote) levels of proficiency (professional, trainee, ad hoc interpreter) and conditions (real-life, simulated, experimental). They also allow for the observation of interpreters' translational behaviour. Indeed, the quantitative and qualitative analysis of parallel corpora can yield insights about interpreters' language transfer skills. These can profitably be contrasted with translators' language transfer skills through intermodal interpreting/translation corpora, an example of which is the European Parliament Translation and Interpreting Corpus (EPTIC), a multilingual corpus derived from EPIC (Bernardini et al. 2016). EPTIC is a bidirectional (English<>Italian) intermodal corpus of interpreted and translated EU Parliament proceedings, featuring the parallel outputs of interpreting and translation processes, aligned to each other and to the corresponding source texts.

Basic features to be included in interpreting corpora are: metadata (information concerning the ethnographic dimension of the study or 'situatedness', i.e. data on the speaker; date, speed and mode of delivery; subject; number of words, timing; location), linguistic features (information on morphosyntactic and lexical features), paralinguistic features (information on disfluencies, prosody, etc.). Depending on the corpus typology, proxemics, gestural and pragmatic features could also be included, e.g. for signed language.

Guidelines on the methodology to build interpreting corpora can be found in Sandrelli et al. (2010), Setton (2011) and Bernardini et al. (2018). The delay in the creation of interpreting corpora and, consequently, in the publi-



cation of corpus-based interpreting studies vs corpus-based translation studies is mainly due to two factors. First: the accessibility to conference interpreting events, including both originals and interpreted versions. This obstacle has now been partially overcome by the advent of the Internet which offers many live streaming or archived conferences and parliamentary debates with interpretations, for instance on the European Parliament (EP) website which is still the main source of materials for interpreting corpora with simultaneous interpretations.

Another issue linked to accessibility is the need for authorizations, which may create difficulties to compile conference interpreting corpora with genuine field data.

The second major obstacle is the requirement to transcribe both the source speeches and the interpreters' linguistic output. This explains the scarcity of large machine-readable interpreting corpora. As is well known, transcription is an extremely time-consuming task and, at the same time, the first level of data selection for subsequent analyses. The lack of user-friendly and shared conventions for transcribing linguistic and paralinguistic features of orality in conference interpreting further adds to the problem (Cencini 2002; Hu and Tao 2013; Niemants 2015). A possible course of action implemented by some authors has been to keep corpus transcription and annotation to basic features, thus striking the best possible balance between user-friendliness in both coding and using corpus data. This makes it possible to share corpora to be used on different platforms. This was the case with the EPIC corpus that could also be exploited by Shlesinger's team in Bar Ilan University (Russo et al. 2012).

As to transcription, speech recognition software, often combined with shadowing (the transcriber repeats aloud what s/he hears), may speed up the process, even though transcripts still need double-checking and editing before creating/integrating a corpus.

Despite the use of software or methods to streamline the transcription procedure, the production of source and target text transcripts remains a major challenge for a major interpreting corpus project. That is why interpreting corpora are still considered a "cottage industry" by some scholars (Setton 2011) or, more audaciously, a "cottage (wired) industry" by others (Bendazzoli 2018).

Yet, as also reported in detail by the above-mentioned authors, since 2004 several electronic interpreting corpora were created. These display different designs:

- parallel corpora include transcripts of source texts and corresponding target texts with or without text-to-sound / video alignment;
- comparable corpora include source texts and c target texts as monolingual productions, i.e. English source texts and English interpreted target texts;
- multimodal corpora include several interpreting modalities or input / output channels (video, audio, transcripts);
- intermodal corpora include source texts and the corresponding interpreted and translated target texts.



The source-target text / sound / video alignment is a very important feature, which is difficult to obtain, due to the laborious manual encoding. The alignment software generally used in corpus-based studies are: CLAN, ELAN, EXMARaLDA, syncWRITER, TRANSCRIBER, TRANSANA, WINPITCH (Niemants 2015).

In the following sections, the development of interpreting corpora from collections of speeches to electronic corpora will be briefly described (section 1), then a review of the available conference interpreting corpora will be provided with some significant research results (section 2) and some concluding remarks (section 3).

### 1. From collections of transcribed speeches to electronic interpreting corpora

Conference interpreting, both simultaneous and consecutive, entails the interlinguistic transfer of an oral message, which, by its very nature, is evanescent, and, therefore, any attempt to study the product and process of interpreting for didactic or research purposes requires the fixation on a material support (transcription) of the interpreter's linguistic output, usually coupled with that of the speaker's. Interpreting corpora, that is a collection of transcribed source and target speeches, were created and their development went through a series of stages leading up to the present availability of full-fledged electronic corpora. Both Setton (2011) and Bendazzoli (2018) provide a detailed account of the main features of interpreting corpora appeared in the literature so far, providing updated information on their language composition, size, availability (or lack of) etc.

Here, we shall provide an overview of the characteristics of the interpreting corpora developed at each stage.

At first, collections of transcripts of moderate size and generally involving only a few interpreters were taken as a basis for theorizing on interpreting processes and products. Despite their limits, these studies exerted a great influence on interpreting theories and interpreter education: a notable example is Seleskovitch's *Langage, langues et mémoire. Etude de la prise de notes en interprétation consecutive* (1975), where interpreters' notes were collected and analysed.

In a second phase, scholars started collecting larger quantities of real-life interpreting data from specific professional settings. They carried out qualitative analyses of their data sets with manually aligned STs and TTs. Given their vast amount of field data and the extended recording periods (from several months to several years), these can be considered the first genuine descriptive studies (in the sense of Toury 1995), thus providing insights into interpreters' operational norms, styles, strategies, skills and field challenges.

Examples of these corpora are those developed by Vuorikoski and Straniero Sergio.

Vuorikoski (2004) evaluated the quality of 30 interpreters' linguistic outputs, in a corpus of 120 original speeches in English, Finnish, German and



Swedish delivered at the European Parliament and their simultaneous interpretation into these languages. Her focus was 'accuracy' and 'faithfulness'. In a subsequent publication on the same corpus (2012), she concentrated on speech acts containing modals in English EP speeches and concluded that interpreters were not always aware of the several roles of speech acts, an issue that she recommended should be incorporated into interpreter training.

Straniero Sergio developed the Italian Television Interpreting Corpus (CorIT), featuring 1200 consecutive and simultaneous interpretations broadcast by public and private TV networks. His aim was 'to respond to the pressing need for authentic data on SI' (2003, 136), tracing the history of media interpreting and highlighting differences with conference interpreting and other forms of dialogue interpreting. Since 1999, numerous CorIT-based studies have appeared (Straniero and Falbo 2012). CorIT does not contain performances in traditional conference settings, but it is nevertheless a unique and invaluable interpreting corpus of reference for the massive quantity of consecutive and interpreting performances.

Before full-fledged electronic corpora, a third phase can be identified. This includes large sets of real-life interpreting data, collected and stored with criteria inspired by corpus linguistics, in that they envisage the use of tools to retrieve features of source texts and target texts, albeit still manually aligned (Wallmach 2000), or of tools to allow for multiple visualizations of the texts stored (Collados et al. 2004). Wallmach (2000) recorded 110 hours of simultaneous interpretations by 16 professional interpreters working between English, Afrikaans, Zulu and Sepedi to investigate the effect of speed on interpreters' performance and to highlight interpreters' strategies and language-specific norms in a South African legislative context. In her pilot study (8 hours, approximately 40.000 tokens), using the parallel concordancing programme, ParaConc for Windows, she identified language-specific difficulties and strategies influenced by text complexity and lack of source text-target text equivalents.

In 2003, Collados Aís and collaborators (2004) developed the multilingual ECIS corpus (Evaluación de la Calidad en Interpretación Simultánea) which contains 43 EP speeches and 73 interpretations, with an interface for multivariate visualizations. They explored other important aspects of quality, namely non-verbal, paralinguistic and prosodic features, thus providing a more comprehensive evidence-based evaluative framework for the study of interpreters' performances and their effect on users.

The turn from collections of manually transcribed speeches to the use of corpus linguistic tools and methodologies in compiling interpreting corpora has allowed for numerous new perspectives on the investigation of interpreting from a corpus-based approach.

## 2. Interpreting corpora and study results

While corpus-based translation studies also tackled common topics in different corpora and approaches (see Bernardini and Russo 2018 for an overview), corpus-based interpreting studies do not seem to follow this pat-



tern. Therefore, what follows is an overview of the most prominent lines of investigation through the available interpreting corpora and their contributions to our understanding of interpreting processes and products in conference interpreting.

Between 2004 and 2006, the first free, open, machine-readable, on-line corpus was developed in the Forlì Campus of the University of Bologna: the European Parliament Interpreting Corpus (EPIC), a pos-tagged, lemmatised and indexed corpus enabling simple and advanced queries (<http://sslmitdev-online.sslmit.unibo.it/corpora/corporaproject.php?path=E.P.I.C>). EPIC is made up of nine sub-corpora (approx. 180,000 tokens), three sub-corpora of English, Spanish and Italian original speeches and six sub-corpora of the corresponding simultaneous interpretations in these three languages (for a detailed description of EPIC, its applications and developments, see Russo et al. 2012). The EPIC parallel and comparable design allows for a variety of study typologies. For instance, lexical patterns were investigated to ascertain whether the results obtained by Laviosa (1998) for translated versus non-translated texts held true also for original vs. interpreted speeches. Laviosa found that non-translated texts displayed higher lexical density (content vs. grammatical words) and lexical variety (proportion of high-frequency words vs. low-frequency words) compared to translated English texts. EPIC-based results differed from Laviosa's findings on lexical density, but generally not for lexical variety (Russo 2018). Shlesinger (2009), who applied a different method, calculating the ratio of types to tokens, to identify linguistic richness in her intermodal corpus, obtained a similar result. Other topics investigated in EPIC are disfluencies and repairs (Bendazzoli et al. 2011), text-processing strategies (Russo 2010), gender-based trends (Russo 2011, 2016), universals in interpreting (Lobascio 2017).

Building on the expertise gained through EPIC, another corpus was created in Forlì: the Directionality Simultaneous Interpreting Corpus (DIRSI), an English-Italian corpus of medical conferences (approx. 130.000 tokens) with a dedicated web interface to study the effect of directionality on interpreter's output (Bendazzoli 2012). DIRSI is text-to-sound and source text - target text aligned, indexed, pos - and time-tagged: this enables the contextual analysis of transcripts and sound.

A further development arising from EPIC is the European Parliament Interpreting Corpus (at) Ghent (EPICG) which is an open, multilingual (initially French>Dutch and English), partly aligned (time-ST-TT) and pos-tagged corpus of about 250.000 tokens, also containing metadata (speaker, speech and situational details). Several topics have been explored, such as connective markers (Defrancq et al. 2015), ear-voice-span (Defrancq 2015) gender-based trends (Magnifico and Defrancq 2016, 2017).

Press conference data from different cultural and professional settings are included in three corpora compiled to study communicative interactions and interpreters' strategies and norms: the Football in Europe (FOOTIE) corpus, the Chinese-English Interpreting Corpus of the Chinese Premier's annual press conferences (CEIPPC) and the Chinese-English Conference Interpreting Corpus (CECIC).



FOOTIE was developed by Sandrelli (2012) at UNINT University of Rome. It contains 16 interpreter-mediated press conferences held during the 2008 European Football Championship. It is a multimedia, multilingual (French, English, Spanish, Italian), closed, untagged corpus; the transcripts of the source texts and simultaneously interpreted target texts are organized as a table which also includes extra-linguistic information (word/turn, word/min, etc.).

CEIPPC was developed at Guangdong University of Foreign Studies, China. The corpus data span 14 years (1998–2011) and include transcripts (over 100,000 tokens) of video recordings of seven interpreters (Wang 2012; Wang and Zou 2018). The Chinese CECIC is an annotated corpus in TEI format, with head information mark-up, pos-tags and paralinguistic information tags compiled by Hu and Tao (2013), who, based on this corpus, found that interpreted texts exhibit greater normalization and explicitation than written translated texts.

One of the few common research topics is the interpreter's language or 'interpretese', which spurred the creation of small comparable, pos-tagged, annotated corpora designed to identify lexical and morphosyntactic features. At Bar Ilan University, Shlesinger (2009) and Shlesinger and Ordan (2012) developed an English>Hebrew intermodal corpus of source texts, interpreted target texts and translated target texts. At the University of Bologna in Forlì, Aston (2015, 2018) detected typical lexical patterns in his 2249i, a corpus of English interpreted speeches at the EP consisting of approx. 60,000 words. A more recent study on interpretese, Kajzer-Wietrzny (2018) from the University of Poznań (Poland) investigated the use of optional "that" in her TIC corpus.

An example of a multimodal corpus is the open-source consecutive and simultaneous corpus CoSi (House et al. 2012), compiled to study the effect of the interpreting mode on the processing of discourse markers, mitigators and proper nouns. Extensive information on the corpus design is provided in this work, to encourage corpus exchange in corpus-based interpreting studies.

The most recent publications providing further insights into corpus-based interpreting studies emerge from two events gathering the most active scholars in the field: Emerging Topics in Translation and Interpreting (Trieste, 16–18 June 2010) with one session devoted to corpus-based interpreting studies (Straniero and Falbo 2012) and the First Forlì International Workshop on Corpus-based Interpreting Studies: The State of the Art (Forlì, 7–8 May 2015). A selection of the papers describing several new corpora and insightful research results presented at Forlì appeared in Russo et al. (2018) and Bendazzoli et al. (2018).

### 3. Concluding remarks

The corpus-based approach in interpreting studies is opening unprecedented opportunities to investigate conference interpreters' linguistic output and their cognitive behaviour highlighted by text-processing strategies. The



quantitative approach, typical of corpus linguistics, serves the purpose to detect trends and peculiarities, which could be better understood through a qualitative approach, which allows for an in-depth analysis. As we have seen, however, compiling spoken corpora is beset with more difficulties than translation corpora, which explains the delay in their development in conference interpreting.

In order to test the hypothesis, theorize on interpreting products and processes, and exploit interpreting corpora for didactic purposes, massive and representative data are required. Therefore, it appears to be high time for the interpreting community to join efforts and harmonize methodologies to foster the sharing corpora and comparison of results.

### References

Aston, G., 2015. Learning phraseology from speech corpora. In: A. Lenko-Szymanska and A. Boulton (eds.). *Multiple affordances of language corpora for data-driven learning*. Amsterdam: Benjamins, pp. 65–84.

Aston, G., 2018. Acquiring the language of interpreters: A Corpus-based Approach. In: M. Russo, C. Bendazzoli and B. Defrancq (eds.). *Making Way in Corpus-based Interpreting Studies*. Springer Nature Switzerland, pp. 83–96.

Baker, M., 1993. Corpus linguistics and translation studies. Implications and applications. In: M. Baker, G. Francis and E. Tognini-Bonelli (eds.). *Text and Technology*. Amsterdam: John Benjamins, pp. 233–250.

Bendazzoli, C., 2012. From international conferences to machine-readable corpora and back: An ethnographic approach to simultaneous interpreter-mediated communicative events. In: F. Straniero Sergio and C. Falbo (eds.). *Breaking Ground in Corpus-based Interpreting Studies*. Peter Lang AG, pp. 91–117.

Bendazzoli, C., 2018. Corpus-based Interpreting Studies: Past, Present and Future Developments of a (Wired) Cottage Industry. In: M. Russo, C. Bendazzoli and B. Defrancq (eds.). *Making Way in Corpus-based Interpreting Studies*. Springer Nature Switzerland, pp. 1–19.

Bendazzoli, C., Sandrelli, A. and Russo, M., 2011. Disfluencies in simultaneous interpreting: A corpus-based analysis. In: A. Kruger, K. Wallmachand and J. Munday (eds.). *Corpus-based Translation Studies: Research and Applications*. New York, Continuum, pp. 282–306.

Bendazzoli, C., Russo, M. and Defrancq, B. (eds.), 2018. New Findings in Corpus-based Interpreting Studies. *InTRALinea*, 20, available at: <http://www.intraline.org/current> (accessed 23 September 2018).

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), pp. 58–83.

Bernardini, S. and Russo, M., 2018. Corpus Linguistics, Translation and Interpreting. In: K. Malmkjær (ed.). *Routledge Handbook of Translation Studies and Linguistics*. London, Routledge, pp. 342–356.

Bernardini, S., Ferraresi, A., Russo, M., Collard, C. and B. Defrancq., 2018. Building Interpreting and Intermodal Corpora: A How to for a Formidable Task. In: M. Russo, C. Bendazzoli and B. Defrancq (eds.). *Making Way in Corpus-based Interpreting Studies*. Springer Nature Switzerland, pp. 21–42.

Cencini, M., 2002. On the importance of an encoding standard for corpus-based interpreting studies: Extending the TEI scheme. *InTRALinea*. *CULT2K*, available at: [www.intraline.org/specials/article/1678](http://www.intraline.org/specials/article/1678) (accessed 23 September 2018).





Collados Aís, Á., Fernández Sánchez, M.M., Iglesias Fernández, E., Pérez-Luzardo, J. Pradas Macías, E.M., Stévaux, E., Blasco Mayor, M.J. and Jiménez Ivars A., 2004. Presentación de Proyecto de Investigación sobre Evaluación de la Calidad en Interpretación Simultánea (Bff2002-00579). *Actas del IX Seminario Hispano-Ruso de Traducción e Interpretación*. Moscú/Moscow: Universidad Estatal Lingüística de Moscú. MGLU, pp. 3–15.

Defrancq, B., 2015. Corpus-based research into the presumed effects of short EVS. *Interpreting*, 17 (1), pp. 26–45.

Defrancq, B., Plevoets, K. and Magnifico, C., 2015. Connective markers in interpreting and translation: Where do they come from? *Yearbook of Corpus Linguistics and Pragmatics*, 3, pp. 195–222.

House, J., Meyer B. and Schmidt, T., 2012. CoSI-A Corpus of Consecutive and Simultaneous Interpreting. In: T. Schmidt and K. Wörner (eds.). *Multilingual Corpora and Multilingual Corpus Analysis*. Amsterdam: John Benjamins, pp. 295–304.

Hu, K. and Tao, Q., 2013. The Chinese-English conference interpreting corpus: Uses and limitations. *Meta*, 58 (3), pp. 626–642.

Kajzer-Wietrzny, M., 2018. Interpretese vs. Non-native Language Use: The Case of Optional That. In: M. Russo, C. Bendazzoli and B. Defrancq, eds. *Making Way in Corpus-based Interpreting Studies*. New Frontiers in Translation Studies. Springer, Singapore, pp. 97–114.

Laviosa, S., 1998. Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta*, 43 (4), pp. 557–570.

Lobascio, M., 2017. *Genitive variation and unique items hypothesis in simultaneous interpreting from Italian into English. An intermodal study based on EPIC*. MA Dissertation. University of Bologna.

Magnifico, C. and Defrancq, B., 2016. Impoliteness in interpreting: a question of gender? *Translation and Interpreting*, 8 (2), pp. 26–45.

Magnifico, C. and Defrancq B., 2017. Hedges in conference interpreting: The role of gender. *Interpreting*, 19 (1), pp. 21–46.

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), available at: <http://www.erudit.org/revue/meta/2005/v50/n4/019850ar.pdf> (accessed 23 September 2018).

Niemants, N., 2015. Transcription. In: F. Pöchhacker (ed.). *Routledge encyclopedia of interpreting studies*. London: Routledge, pp. 421–422.

Pöchhacker, F. (ed.), 2015. *Routledge Encyclopedia of Interpreting Studies*. London, Routledge.

Russo, M., Bendazzoli, C. and Sandrelli, A., 2006. Looking for lexical patterns in a trilingual corpus of source and interpreted speeches: Extended analysis of EPIC (European Parliament Interpreting Corpus). *Forum*, 4 (1), pp. 221–254.

Russo, M., 2010. Reflecting on interpreting practice: Graduation theses based on the European Parliament Interpreting Corpus (EPIC). In: L. Zybatow (ed.). *Translationswissenschaft – Standund Perspektiven. Innsbrucker Ringvorlesungen zur Translationswissenschaft VI (Forum Translationswissenschaft, Bd. 12)*. Frankfurt am Main: Peter Lang, pp. 35–50.

Russo, M., 2011. Text processing patterns in simultaneous interpreting (Spanish-Italian): A corpus-based study. In: I. Ohnheiser, W. Pöckl and P. Sandrini (eds.). *Festschrift in Honour of Prof. Lew Zybatow Translation – Sprachvariation – Mehrsprachigkeit*. Frankfurt am Main: Peter Lang, pp. 83–103.

Russo, M., 2016. Orality and Gender: A corpus-based study on lexical patterns in simultaneous interpreting. In: C. Calvo and N. Spinolo (eds.). *MonTI. Special iss. Translating orality / La traducción de la oralidad*, pp. 307–322.



Russo M., 2018. Speaking patterns and gender in the European Parliament Interpreting Corpus. A quantitative study as a premise for qualitative investigations. In: M. Russo, C. Bendazzoli and B. Defrancq (eds.). *Making Way in Corpus-based Interpreting Studies*. Springer Nature Switzerland, pp. 115–131.

Russo, M., Bendazzoli, C., Sandrelli A. and Spinolo N., 2012. The European Parliament Interpreting Corpus (EPIC): Implementation and developments. In: F. Straniero Sergio and C. Falbo (eds.). *Breaking ground in corpus-based Interpreting Studies*. Frankfurt am Main: Peter Lang, pp. 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., 2012. Introducing FOOTIE (Football in Europe): Simultaneous interpreting in football press conferences. In: F. Straniero Sergio and Falbo, C. (eds.). *Breaking Ground in Corpus-based Interpreting Studies*. Peter Lang AG, pp. 119–153.

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), pp. 165–203.

Seleskovitch, D., 1975. *Langage, Langues et Mémoire. Etude de la Prise de Notes en Interprétation Consécutive*. Paris: Minard.

Setton, R., 2011. Corpus-based Interpreting Studies (CIS): Overview and prospects. In: A. Kruger, K. Wallmach and J. Munday (eds.). *Corpus-Based Translation Studies. Research and Applications*. New York: Continuum, pp. 33–75.

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

Shlesinger, M., 2009. Towards a definition of Interpretese. An intermodal, corpus-based study. In: G. Hansen, A. Chesterman and H. Gerzymisch-Arbogast (eds.). *Efforts and Models in Interpreting and Translation Research. A tribute to Daniel Gile*. Amsterdam: John Benjamins, pp. 237–253.

Shlesinger, M. and Ordan, N., 2012. More spoken or more translated? Exploring a known unknown of simultaneous interpreting. *Target*, 24 (1), pp. 43–60.

Straniero Sergio, F., 2013. Norms and quality in media interpreting: The case of Formula One press-conferences. *The Interpreters' Newsletter*, 12, pp. 135–174.

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

Tohyama, H., Ryu, K., Mastubara, S., Kawaguchi, N. and Inagaki, Y., 2004. *Simultaneous Interpreting Corpus*. Proceedings of Oriental COCOSDA, available at: [http://slp.itc.nagoya-u.ac.jp/web/papers/2004/Oriental-COCOSDA2004\\_tohyama.pdf](http://slp.itc.nagoya-u.ac.jp/web/papers/2004/Oriental-COCOSDA2004_tohyama.pdf) (accessed 23 September 2018).

Vuorikoski, A.R., 2004. *A Voice of Its Citizens or a Modern Tower of Babel? The Quality of Interpreting as a Function of Political Rhetoric in the European Parliament*. Tampere: Tampere University Press, available at: <http://tampub.uta.fi/handle/10024/67348> (accessed 23 September 2018).

Vuorikoski, A.R., 2012. Fine-tuning SI quality criteria: Could speech act theory be of any use? In: C. J. Kellett Bidoli (ed.). *Interpreting across Genres: Multiple Research Perspectives*. Trieste: EUT, pp. 152–170.

Wallmach, K., 2000. Examining simultaneous interpreting norms and strategies in a South African legislative context: A pilot corpus analysis. *Language Matters*, 31 (1), pp. 198–221.

Wang, B., 2012. Interpreting strategies in real-life interpreting: Corpus-based description of seven professional interpreters' performance. *Translation Journal*, 16 (2), available at: <http://translationjournal.net/journal/60interpreting.htm> (accessed 23 September 2018).



Wang, B. and Zou, B., 2018. Exploring Language Specificity as a Variable in Chinese-English Interpreting. A Corpus-Based Investigation. In: M. Russo, C. Bendazzoli and B. Defrancq (eds.). *Making Way in Corpus-based Interpreting Studies*. Springer Nature Switzerland, pp. 65–82.

### The author

*Mariachiara Russo*, Professor, Department of Translation and Interpreting, Bologna University, Forli, Italy.

E-mail: mariachiara.russo@unibo.it

### To cite this article:

Russo, M. 2019, Corpus-based studies in conference interpreting, *Slovo.ru: baltijskij accent*, Vol. 10, no. 1, p. 87–100. doi: 10.5922/2225-5346-2019-1-6.

## КОРПУСНЫЕ ИССЛЕДОВАНИЯ В КОНФЕРЕНЦ-ПЕРЕВОДЕ

*М. Руссо*<sup>1</sup>

<sup>1</sup>Болонский университет  
47100, Италия, Форли, ул. Курс Республики, 136  
Поступила в редакцию 15.09.2018 г.  
doi: 10.5922/2225-5346-2019-1-6

*Корпусные исследования в рамках теории устного перевода появились относительно недавно как «следствие применения корпусов в изучении письменного перевода». Так их происхождение описывается в основополагающей статье Мириам Шлезингер (1998), работы которой являются постоянным источником вдохновения для переводоведов. В настоящее время переводческие корпусы все чаще используются для изучения конференц- и сопровождающего перевода. В статье основное внимание уделяется исследованию процесса конференц-перевода, рассматривается эволюция собственно концепта «корпус устных переводов», а также приводится обзор наиболее репрезентативных примеров из ранних собраний транскрибированных оригиналов речей и их переводов, которые были обработаны с учетом принятых лингвистических стандартов и инструментов цифровых корпусов. Используя аутентичные примеры, автор анализирует ряд методологических проблем, связанных с применением переводческих корпусов.*

**Ключевые слова:** параллельный корпус, сопоставительный корпус, мультимодальный корпус, интермодальный корпус, транскрипция, метаданные.

### Список литературы

Aston G. Learning phraseology from speech corpora // Multiple affordances of language corpora for data-driven learning / ed. by A. Lenko-Szymanska, A. Boulton. Amsterdam : Benjamins, 2015. P. 65–84.

Aston G. Acquiring the language of interpreters: A Corpus-based Approach // Making Way in Corpus-based Interpreting Studies. Series: New Trends in translation Studies / ed. by M. Russo, C. Bendazzoli, B. Defrancq. Singapore : Springer, 2018. P. 83–96.



*Baker M.* Corpus linguistics and translation studies. Implications and applications // Text and Technology. In Honour of John Sinclair / ed. by M. Baker, G. Francis, E. Tognini-Bonelli. Amsterdam : Benjamins, 1993. P. 233–250.

*Bendazzoli C.* From international conferences to machine-readable corpora and back: An ethnographic approach to simultaneous interpreter-mediated communicative events // Breaking Ground in Corpus-Based Interpreting Studies / ed. by F. Straniero Sergio, C. Falbo. Bern : Peter Lang, 2012. P. 91–117.

*Bendazzoli C.* Corpus-based Interpreting Studies: Past, Present and Future Developments of a (Wired) Cottage Industry // Making Way in Corpus-based Interpreting Studies. Series: New Trends in translation Studies / ed. by M. Russo, C. Bendazzoli, B. Defrancq. Singapore : Springer, 2018. P. 1–19.

*Bendazzoli C., Sandrelli A., Russo M.* Disfluencies in simultaneous interpreting: A corpus-based analysis // Corpus-based Translation Studies: Research and Applications / ed. by A. Kruger, K. Wallmach, J. Munday. N. Y. : Continuum, 2011. P. 282–306.

*Bendazzoli C., Russo M., Defrancq B.* (eds.). New Findings in Corpus-based Interpreting Studies // InTRALinea. 2018. Special Iss. Vol. 20. URL: <http://www.intralinea.org/current> (дата обращения: 23.09.2018).

*Bernardini S., Ferraresi, A. and Miličević, M.* From EPIC to EPTIC – Exploring simplification in interpreting and translation from an intermodal perspective // Target. 2016. Vol. 28 (1). P. 58–83.

*Bernardini S., Russo M.* Corpus Linguistics, Translation and Interpreting // Routledge Handbook of Translation Studies and Linguistics / ed. by K. Malmkjær. L. : Routledge, 2018. P. 342–356.

*Bernardini S., Ferraresi A., Russo M. et al.* Building Interpreting and Intermodal Corpora: A How to for a Formidable Task // Making Way in Corpus-based Interpreting Studies. Series: New Trends in translation Studies / ed. by M. Russo, C. Bendazzoli, B. Defrancq. Singapore : Springer, 2018. P. 21–42.

*Cencini M.* On the importance of an encoding standard for corpus-based interpreting studies: Extending the TEI scheme // InTRALinea. 2002. Special Iss. : CULT2K. URL: [www.intralinea.org/specials/article/1678](http://www.intralinea.org/specials/article/1678) (дата обращения: 23.09.2018).

*Collados Aís Á., Fernández Sánchez M.M., Iglesias Fernández E. et al.* Presentación de Proyecto de Investigación sobre Evaluación de la Calidad en Interpretación Simultánea (Bff2002-00579) // Actas del IX Seminario Hispano-Ruso de Traducción e Interpretación. Moscú (Moscow) : Universidad Estatal Lingüística de Moscú (MGLU), 2004. P. 3–15.

*Defrancq B.* Corpus-based research into the presumed effects of short EVS // Interpreting. 2015. Vol. 17 (1). P. 26–45.

*Defrancq B., Plevoets K., Magnifico C.* Connective markers in interpreting and translation: Where do they come from? // Yearbook of Corpus Linguistics and Pragmatics. 2015. Vol. 3. P. 195–222.

*House J., Meyer B., Schmidt T.* CoSI-A Corpus of Consecutive and Simultaneous Interpreting // Multilingual Corpora and Multilingual Corpus Analysis / ed. by T. Schmidt, K. Wörner. Amsterdam : Benjamins, 2012. P. 295–304.

*Hu K., Tao Q.* The Chinese-English conference interpreting corpus: Uses and limitations // Meta. 2013. Vol. 58 (3). P. 626–642.

*Kajzer-Wietrzny M.* Interpretese vs. Non-native Language Use: The Case of Optional That // Making Way in Corpus-based Interpreting Studies. Series: New Trends in translation Studies / ed. by M. Russo, C. Bendazzoli, B. Defrancq. Singapore : Springer, 2018. P. 97–114.

*Laviosa S.* Core patterns of lexical use in a comparable corpus of English narrative prose // Meta. 1998. Vol. 43 (4). P. 557–570.



*Lobascio M.* Genitive variation and unique items hypothesis in simultaneous interpreting from Italian into English. An intermodal study based on EPIC : MA dissertation. Bologna : University of Bologna, 2017. URL: <http://amslaurea.unibo.it/12721/> (дата обращения: 23.09.2018).

*Magnifico C., Defrancq B.* Impoliteness in interpreting: a question of gender? // Translation and Interpreting. 2016. Vol. 8 (2). P. 26–45.

*Magnifico C., Defrancq B.* Hedges in conference interpreting: The role of gender // Interpreting. 2017. Vol. 19 (1). P. 21–46.

*Monti C., Bendazzoli C., Sandrelli A., Russo M.* Studying directionality in simultaneous interpreting through an electronic corpus: EPIC (European Parliament Interpreting Corpus) // Meta. 2005. Vol. 50 (4). URL: <http://www.erudit.org/revue/meta/2005/v50/n4/019850ar.pdf> (дата обращения: 23.09.2018).

*Niemants N.* Transcription // Routledge Encyclopedia of Interpreting Studies / ed. by F. Pöchhacker. L. : Routledge, 2015. P. 421–422.

*Pöchhacker F.* (ed.). Routledge Encyclopedia of Interpreting Studies. L. : Routledge, 2015.

*Russo M., Bendazzoli C., Sandrelli A.* Looking for lexical patterns in a trilingual corpus of source and interpreted speeches: Extended analysis of EPIC (European Parliament Interpreting Corpus) // Forum. 2006. Vol. 4 (1). P. 221–254.

*Russo M.* Reflecting on interpreting practice: Graduation theses based on the European Parliament Interpreting Corpus (EPIC) // Translationswissenschaft – Standund Perspektiven. Innsbrucker Ringvorlesungen zur Translationswissenschaft VI (Forum Translationswissenschaft, Bd. 12) / ed. by L. Zybatow. Frankfurt a/M: Peter Lang, 2010. P. 35–50.

*Russo M.* Text processing patterns in simultaneous interpreting (Spanish-Italian): A corpus-based study // Festschrift in Honour of Prof. Lew Zybatow Translation – Sprachvariation – Mehrsprachigkeit / ed. by I. Ohnheiser, W. Pöckl, P. Sandrini. Frankfurt a/M : Peter Lang, 2011. P. 83–103.

*Russo M.* Orality and Gender: A corpus-based study on lexical patterns in simultaneous interpreting // MonTI. 2016. Special iss. №3: Translating orality / La traducción de la oralidad / ed. by C. Calvo, N. Spinolo. P. 307–322.

*Russo M.* Speaking patterns and gender in the European Parliament Interpreting Corpus. A quantitative study as a premise for qualitative investigations // Making Way in Corpus-based Interpreting Studies. Series: New Trends in translation Studies / ed. by M. Russo, C. Bendazzoli, B. Defrancq. Singapore : Springer, 2018. P. 115–131.

*Russo M., Bendazzoli C., Sandrelli A., Spinolo N.* The European Parliament Interpreting Corpus (EPIC): Implementation and developments // Breaking Ground in Corpus-Based Interpreting Studies / ed. by F. Straniero Sergio, C. Falbo. Bern : Peter Lang, 2012. P. 35–90.

*Russo M., Bendazzoli C., Defrancq B.* (eds.). Making Way in Corpus-based Interpreting Studies. Series: New Trends in translation Studies. Singapore: Springer, 2018.

*Sandrelli A.* Introducing FOOTIE (Football in Europe): Simultaneous interpreting in football press conferences // Breaking Ground in Corpus-Based Interpreting Studies / ed. by F. Straniero Sergio, C. Falbo. Bern : Peter Lang, 2012. P. 119–153.

*Sandrelli A., Bendazzoli C., Russo M.* European Parliament Interpreting Corpus (EPIC): Methodological issues and preliminary results on lexical patterns in SI // International Journal of Translation. 2010. Vol. 22 (1–2). P. 165–203.

*Seleskovitch D.* Langage, Langues et Mémoire. Etude de la Prise de Notes en Interprétation Consécutive. P. : Minard, 1975.

*Setton R.* Corpus-based Interpreting Studies (CIS): Overview and prospects // Corpus-Based Translation Studies. Research and Applications / ed. by A. Kruger, K. Wallmach, J. Munday. N. Y. : Continuum, 2011. P. 33–75.



*Shlesinger M.* Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies // *Meta*. 1998. Vol. 43 (4). P. 486–493.

*Shlesinger M.* Towards a definition of Interpretese. An intermodal, corpus-based study // *Efforts and Models in Interpreting and Translation Research. A tribute to Daniel Gile* / ed. by G. Hansen, A. Chesterman, H. Gerzymisch-Arbogast. Amsterdam : Benjamins, 2009. P. 237–253.

*Shlesinger M., Ordan N.* More spoken or more translated? Exploring a known unknown of simultaneous interpreting // *Target*. 2012. Vol. 24 (1). P. 43–60.

*Straniero Sergio F.* Norms and quality in media interpreting: The case of Formula One press-conferences // *The Interpreters' Newsletter*. 2013. Vol. 12. P. 135–174.

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

*Tohyama H., Ryu K., Mastubara S. et al.* CIAIR Simultaneous Interpreting Corpus, Proceedings of Oriental COCOSDA [2004]. URL: [http://slp.itc.nagoya-u.ac.jp/web/papers/2004/Oriental-COCOSDA2004\\_tohyama.pdf](http://slp.itc.nagoya-u.ac.jp/web/papers/2004/Oriental-COCOSDA2004_tohyama.pdf) (дата обращения: 23.09.2018).

*Vuorikoski A. R.* A Voice of Its Citizens or a Modern Tower of Babel? The Quality of Interpreting as a Function of Political Rhetoric in the European Parliament. Tampere : Tampere University Press, 2004. URL: <http://tampub.uta.fi/handle/10024/67348> (дата обращения: 23.09.2018).

*Vuorikoski A. R.* Fine-tuning SI quality criteria: Could speech act theory be of any use? // *Interpreting across Genres: Multiple Research Perspectives* / ed. by C.J. Kellert Bidoli. Trieste : EUT, 2012. P. 152–170.

*Wallmach K.* Examining simultaneous interpreting norms and strategies in a South African legislative context: A pilot corpus analysis // *Language Matters*. 2000. Vol. 31 (1). P. 198–221.

*Wang B.* Interpreting strategies in real-life interpreting: Corpus-based description of seven professional interpreters' performance // *Translation Journal*. 2012. Vol. 16 (2). URL: <http://translationjournal.net/journal/60interpreting.htm> (дата обращения: 23.09.2018).

*Wang B., Zou B.* Exploring Language Specificity as a Variable in Chinese-English Interpreting. A Corpus-Based Investigation // *Making Way in Corpus-based Interpreting Studies. Series: New Trends in translation Studies* / ed. by M. Russo, C. Bendazzoli, B. Defrancq. Singapore : Springer, 2018. P. 65–82.

### Об авторе

*Мариякьяра Руссо*, профессор лингвистики и устного перевода, кафедры устного и письменного перевода, Болонский университет, Форли, Италия.

E-mail: [mariachiara.russo@unibo.it](mailto:mariachiara.russo@unibo.it)

### Для цитирования:

*Russo M.* Corpus-based studies in conference interpreting // *Слово.ру: балтийский акцент*. 2019. Т. 10, №1. С. 87–100. doi: 10.5922/2225-5346-2019-1-6.