

## **32. Corpus studies in conference interpreting**

*Claudio Bendazzoli*

*Orcid.org/0000-0003-4869-0146*

## **Abstract**

Research into conference interpreting has taken advantage of corpus methods since the early developments of the corpus-based approach. Overviews of major projects in corpus-based interpreting studies clearly show that interpreting corpora have been increasing in number and size over the last three decades, covering a variety of language combinations along with multiple modes and settings. Overall, conference interpreting corpora range from small data sets used for qualitative “manual” analysis to larger, machine-readable corpora, some of which are also accessible online. The systematic observations afforded by corpus methods have allowed interpreting scholars to study different features of conference interpreting with respect to source speeches and other Translation modes, combining a variety of methods including statistical testing, comparable and parallel inquiries, and discourse analysis. Some pedagogical applications have also begun to find their way in conference interpreting education, while there is still very limited evidence of the role of corpora in conference interpreters’ professional practice.

## **Keywords**

Corpus linguistics, parallel corpus, comparable corpus, intermodal corpus, data collection, corpus design, transcription, annotation, accessibility, frequency lists, collocations, reference corpus, spoken corpus, sign language corpus, occurrences

## **Introduction**

This chapter gives an overview of corpus-based methods and studies in conference interpreting. The content is organized into four main sections: methodology; overview of conference interpreting corpora; research deliverables; corpus applications to interpreting education and professional practice; future prospects. In particular, the methodological section introduces the corpus-based approach and explains how to create conference interpreting corpora, highlighting the main challenges in building a simultaneous interpreting corpus (including whispered interpreting) and a consecutive interpreting corpus. The fundamental steps of corpus design, data collection, transcription, annotation, access and distribution are illustrated with reference to major conference interpreting corpus projects, which may serve as inspiration for the development of new resources.

The second section gives an overview of conference interpreting corpora. The corpora presented in this (inevitably partial) overview are grouped according to types of setting and communicative situation, since these cut across language combinations and may entail different data collection strategies. In fact, some particular settings and communicative situations have lent themselves to scholarly attention more than others due to the availability of data. These range from international institutions (above all the European Parliament with the EPIC project along with EPIC-G and EPTIC, as well as other EU institutions and international organizations such as the UN) to governmental press conferences in China (with consecutive interpreting) and local markets with a variety of communicative situations such as thematic conferences, press conferences, festivals and so on.

The next section reports on the main results obtained from and observations based on conference interpreting corpora. Again, the selected studies obviously provide a non-exhaustive picture of what is known about conference interpreting thanks to corpora. Nevertheless, they are functional to pinpoint the research potential of these language resources and clarify the position of this research approach with respect to other approaches in interpreting research.

The fourth section examines the role of corpora (and corpus-based interpreting studies) in conference interpreting education and professional practice. The two areas are still underexplored and deserve greater attention on the part of the research community. On the one hand, since interpreting corpora are based on multimedia archives including videos or

audio recordings, they can be fruitfully exploited in training sessions. Moreover, trainees can take advantage of corpus projects to complete their final dissertations and reflect on their own interpreting practice. The potential of corpora, along with the results obtained in corpus-based research projects can also inform interpreting education programs. On the other hand, the use of corpus resources in professional practice seems to be almost inexistent. Although some tools have been developed specifically for the preparation stage of a conference interpreter, these are still limited in use and may take more time to be fully incorporated in the conference interpreting community of practice.

Finally, last section outlines possible future developments of this line of inquiry in terms of corpus features, research areas, and applications.

### **1 Developing conference interpreting corpora**

The corpus-based approach was introduced in Interpreting Studies towards the end of the Nineties (Shlesinger 1998) in the wake of the pioneering work done in translation research (Baker 1993). The availability of large datasets of written texts in electronic form made it possible to use computer programs to extract and count occurrences of various kinds of phenomena, accounting for typical and untypical features of translated texts. In Interpreting Studies, the same methodology was proposed as an attempt to overcome the limitations posed by “sparse, often anecdotal data” (Shlesinger 1998: 2) and to scrutinise larger samples of authentic data.

A corpus can be defined as “a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria” (Bowker and Pearson 2002: 9). Considering the procedure to create an interpreting corpus, the above definition can be integrated with reference to a number of methodological steps, namely corpus design, data collection, transcription, annotation, alignment, data access and distribution (Thompson 2004, Bendazzoli 2010, Bernardini et al. 2018). While the first three steps are essential (as there can be no corpus without transcribed data), the remaining steps are optional but contribute dramatically to increasing the research potential of an interpreting corpus (as well as any other spoken or sign language corpus). However, looking at the development of corpus-based interpreting studies, it is possible to find two major distinguishing features of interpreting corpora (Bendazzoli and Sandrelli 2009). On one hand, some interpreting corpora are not electronic, i.e. they are not machine-readable. The term *corpus* is used for data samples that

are analysed manually, without the aid of corpus linguistics methods, and with limited size. On the other hand, not all interpreting corpora are available to the research community at large, e.g. on internet-based platforms. Overall, the early studies conducted with the corpus-based approach concerned only conference interpreting, particularly in the simultaneous mode. Over time, the approach has been adopted to investigate other modes and multiple settings, including non-conference interpreting ones (Bendazzoli 2018; Setton 2011).

The limitations mentioned above are due to some special challenges in the development of interpreting corpora (be they spoken or sign language corpora). For conference interpreting corpora in particular, each stage of corpus development implies a number of hindrances. In *corpus design* it is necessary to consider what kind of interpreting mode will be represented in the corpus, whether simultaneous (with or without booth, whispered interpreting, sight translation), consecutive or both, how many working languages are involved and the target level of representativeness (Halverson 1998). It is possible to organise the data into *parallel corpora*, including source speeches and the resulting target speeches; *comparable corpora*, including source and target speeches in the same language; and *intermodal corpora*, including multiple target texts obtained from the same source but with different Translation modes, e.g. simultaneous interpreting and written translation (Bernardini et al. 2018; Shlesinger 2008).

Interpreting corpora normally rely on transcripts of recorded data, i.e. spoken and/or sign language data. Technological advancements have made it increasingly easier to video record communicative situations, though several projects have been carried out based on audio recordings only. In either case, *data collection* can be accomplished directly through fieldwork (Bendazzoli 2016; Metzger and Roy 2011), or indirectly, e.g. downloading recorded data from internet-based platforms and websites (e.g. the European Parliament video library) or recording interpreter-mediated events broadcast on TV (Falbo 2012, Straniero Sergio 2012). Fieldwork-based data collection may entail different challenges in conference interpreting, owing to the specific features of each interpreting mode and the variety of settings under consideration. For instance, in consecutive interpreting it would be interesting to keep track of the interpreter's notes along with the production of source and target speeches. This is now possible by means of digital pen technology (Kellet Bidoli 2016; Orlando 2010). In simultaneous interpreting without booth and in whispered interpreting, special care must be paid to the recording equipment to be able to obtain high quality recording of source and target speeches (for instance by means of lavalier microphones). In

simultaneous interpreting with booth it is sometimes necessary to coordinate data collection from different locations, as the booth may not be installed in the same room where the conference proceedings take place. Furthermore, in simultaneous interpreting there are (at least) two simultaneous audio inputs to be managed at the same time, and the different tracks should be kept separate to be able to transcribe them conveniently. All these logistic and technical variables have a strong bearing on how data collection can be organized most effectively.

The next step involves *transcription*, which can be managed by means of different software programs (see e.g. Niemants 2012) and according to different conventions, depending on the objective of one's analysis and the interaction format (i.e. monologic vs. dialogic). Transcripts may undergo *annotation* of a number of attributes, again depending on the kind of analysis to be carried out. Generally, metadata annotations are recorded to keep track of the kind of text/speech one is dealing with, including features such as the total number of words, the speed of delivery, its duration, speaker-related details and so on. In addition, further (optional) levels of annotation range from part-of-speech (POS) tagging, i.e. the annotation of grammatical categories for each word (or token) to specific verbal or nonverbal features (e.g. mispronounced words). Annotations can be applied automatically (as in the case of POS-tagging) or manually. In sign language interpreting, both the transcription and the annotation processes undergo further specific challenges, owing to the need to use glosses of signs, the variation resulting from the three-dimensional nature of signs themselves, and the impossibility of using transcripts as separate files as they are embedded in a tier of ELAN, i.e. the transcription program largely used in sign language research (Wehrmeyer 2015).

Transcripts may also be processed for *alignment*, which in interpreting corpora can be performed on two levels. Alignment can be done so that source speech transcripts are displayed next to target speech transcripts for immediate comparison (Volk 2019). Depending on the interpreting mode involved and on corpus design (intermodal corpora entails more complex structure also in terms of alignment, see Bernardini et al. 2016), aligned transcripts can be displayed in tabular form or following a musical score layout. On the other hand, alignment can also be performed between a transcript and its video and/or audio recording. Finally, the last step in corpus development concerns its *accessibility and distribution*. Some corpora are made available to the entire research community with no particular restrictions, whereas others are only accessible to the researchers who developed them (e.g. owing to

confidentiality issues or copyright). Ready-made corpora are extremely useful language resources for replications of past studies or for new inquiries. They can also be used to develop further corpora with additional language combinations or specific purposes in mind (Bendazzoli et al. 2021). Conversely, DIY (i.e. do-it-yourself) corpora require considerable effort and are more likely to be created within larger team or PhD projects.

## **2 A partial overview of conference interpreting corpora**

Corpus research has been especially productive in conference interpreting settings since the early developments of this research paradigm. This is probably due to the fact that there is a longer research tradition concerning conference interpreting and that some specific settings have been far more accessible than others involving other types of interpreting such as community interpreting. The European Parliament (EP), along with other EU and international institutions, is a case in point. Owing to the principle of transparency applied by many international institutions and to the extensive use of professional interpreting services to preserve multilingualism (especially in the EU), EP plenary sittings have lent themselves to the creation of several corpus projects. Some of these investigations are based on transcripts of source and target speeches and may even feature text-video alignment, but are still analysed manually, i.e. without using any particular tool to look for occurrences (e.g. Vuorikoski 2004). Other studies take advantage of automatic queries to detect occurrences in verbatim reports, which are then checked against delivery and discussed more qualitatively (Bartłomiejczyk 2017, 2020). The first project to deliver a fully accessible, machine-readable corpus of EP interpreting was carried out at the University of Bologna at Forlì (Italy) and resulted in the European Parliament Interpreting Corpus (EPIC) (Monti et al. 2005). The EPIC consists of nine sub-corpora of transcribed and annotated speeches: three sub-corpora of source speeches by MEPs delivered in Italian, English, and Spanish, and six sub-corpora of simultaneously interpreted speeches covering all the possible combinations among the three languages involved. The particular structure of the EPIC makes it possible to use it both as a parallel and as a comparable corpus. The EPIC project has also been the springboard for other similar corpora including different language combinations, such as English, French, Dutch (and Spanish but only as a source language) in the EPIC-G (EPIC-Ghent). Given the unique features of the EP, the transcripts of oral speeches and simultaneous interpretations have also been analysed with respect to the relevant written proceedings and their translation, thus taking an intermodal perspective (e.g. see Defrancq et al. 2015). Along the same lines, the EPIC corpus has also been turned intermodal, covering other language combinations (i.e.

English, French, Italian, along with some sub-corpora of Polish, Slovene and Finnish) and data from different EP sessions, making up the European Parliament Translation and Interpreting Corpus (EPTIC) (Bernardini et al. 2016; Ferraresi and Bernardini 2019). On the other hand, taking a comparable perspective, the Translation and Interpreting Corpus (TIC) comprises EP source speeches and texts in English along with simultaneously interpreted speeches and translated texts into English from French, Spanish, German and Dutch (Kajzer-Wietrzny 2012). Another example of the successful exploitation of a ready-made corpus such as the EPIC is a purpose-specific corpus called Anglintrad (Bertozzi 2018), which was developed to look at how Anglicisms are managed in interpreting and in translation from Italian into Spanish. The aim of this corpus is primarily to inform translator and interpreter education (see next section below). Other than the EP, there are fewer corpora based on data from other EU institutions. For instance, the IMITES corpus (Spinolo 2018) comprises data from the European Commission to study figurative language in Italian and Spanish source speeches and the simultaneous interpretations into Spanish and Italian respectively.

Similar favourable conditions to data collection can be found in the context of the United Nations, which provide access to the meetings of the General Assembly and other interpreter-mediated events through its web TV online. However, the development of conference interpreting corpora based on this international institution is still quite limited. The SIREN corpus (Dayter 2018) is seemingly the first attempt to build an electronic corpus of simultaneous interpretations between Russian and English. In fact, the components of SIREN also include simultaneous interpretations “drawn from a variety of press conferences, briefings, and interviews by Russian, American and British politicians and public figures that were broadcast with SI on channels belonging to the video news agency Ruptly” (Dayter 2018: 245). Thanks to the annotations available in the SIREN corpus, Dayter could investigate linguistic variation in source and target speeches, contrasting her results with those obtained in other corpus-based translation and interpreting studies.

TV streaming online or even traditional TV broadcasts are another fruitful source of data for conference interpreting corpora. Fifty years of televised interpreter-mediated events have been collected in the Italian Television Interpreting Corpus (CorIT) (Straniero Sergio 2007), which includes consecutive and simultaneous interpretations into Italian performed in a variety of events, from moon landing to US presidential debates, broadcast on Italian TV channels (Dal Fovo 2013). Not all the target speeches contained in this corpus are paired with the relevant



source speeches, but the description of this particular medium has proved useful to pinpoint special features when it comes to interpreting on television (Falbo 2012) and interpreters' style (Straniero Sergio 2012).

A more targeted set of data obtained from TV broadcasts makes up the bilingual parallel corpus of Chinese-English Interpreting for Premier Press Conferences or CEIPPC, which was used to investigate consecutive interpreting norms (Wang 2012) and language-specific features such as syntactic asymmetry between Chinese and English (Wang and Zou 2018). China's premier press conferences have also been considered in other corpora, such as the Chinese-English conference interpreting corpus or CECIC (Hu and Tao 2013) with the addition of written reports and a comparable dataset from Web-sourced CNN press conferences, and another corpus including simultaneous interpreting and translated texts (Fu 2016). Further investigations have taken advantage of the availability of interpreting data (on TV or online) relating to political speeches in high-profile events held in China and other parts of Asia. For instance, a CDA (critical discourse analysis) approach was adopted to investigate ideology and self-referentiality in the Chinese-English Political Discourses Corpus (CE-PolitDisCorp) (Gu and Tipton 2020). Pragmatic markers were analysed in the Corpus of Interpreted Political Speeches from Chinese to English (CIPSCE) (Pan and Wong 2019), including both consecutive and simultaneous interpreting. This kind of data, along with translated texts have been gathered to create a new intermodal corpus similar to the EPTIC, i.e. the Chinese/English Political Interpreting Corpus (CEPIC) (Pan 2019). CEPIC is a large, machine-readable corpus of conference interpreting and translation of political speeches. It is also accessible online through a dedicated platform. At the time of writing, this unique resource is seemingly the largest ever created conference interpreting corpus of professional Chinese (Cantonese and Putonghua)-English interpreting and translation.

Conference interpreting on television can also provide other research opportunities, especially in case of special events with global coverage and media services. For instance, taking an intermodal perspective Taehyung (2011) analysed English-Korean simultaneous interpreting vs. live captions in three different editions of the Academy Awards Ceremony. Recently, the presence of sign language interpreters have become increasingly common in many daily news services and live addresses during natural disasters and pandemics, therefore favouring the creation of sign language interpreting corpora (Wehrmeyer 2019).

Direct access to interpreting data has been managed in various ways for conference interpreting corpora projects. The CoSi corpus (Meyer 2008) includes Portuguese-German interpretations of the same talk held in three different venues, two with consecutive and one with (two twin booths of) simultaneous interpreting. The researchers had full access to this real life event as they supported the organizers of the conference. The CoSi corpus is freely available from the EXMARaLDA portal (House et al. 2012). Another data collection strategy entails the direct involvement of a practisearcher, i.e. a professional interpreter (on duty) who is also a scholar engaged in research. Examples range from international medical conferences open to doctors and patients as in the DIRSI Corpus (Bendazzoli 2012) to pre- and post-football matches press conferences in the FOOTIE Corpus (Sandrelli 2012). Finally, there are conference interpreting corpora compiled by researchers who apparently acted as external observers or obtained recordings from the organizers, as is the case with the WAW corpus, which includes English-Arabic / Arabic-English simultaneous interpretations of conference presentations from three different events held in Qatar, for a total of 521 sessions (Abdelali et al. 2018; Temnikova et al. 2019).

The non-exhaustive overview of conference interpreting corpora provided above highlights that there are various data sources with some settings being more accessible than others. It also seems that conference interpreting corpora are turning larger and increasingly machine-readable, though smaller datasets can still be scrutinised with traditional methods without the aid of computer programs. In fact, small corpora are ideal for qualitative approaches (e.g. Petite 2005), while quantitative results can set the basis for in-depth, qualitative investigations (e.g. Bendazzoli 2019).

### **3 What we know about conference interpreting thanks to corpora**

Considering the relatively limited size of most interpreting corpora (especially in comparison with monolingual, reference corpora), the results reported so far can hardly be generalized. In addition, the plentiful variables at stake further narrows down the scope of many studies, as they focus on different kinds of language combinations, language directions, interpreting modes, directionality, and communicative situations. It can be argued that EP plenary sittings are the most scrutinised conference interpreter-mediated situation, followed by political discourse in China's governmental press conferences. In terms of languages, English is the most represented language, both as a source and target language. This is not surprising given the extensive use of English, both as native and non-native language (see Albl-Mikasa in this

volume) in most settings, and to the early development of advanced and reliable corpus linguistics tools that are able to process transcripts in that medium (e.g. automatic taggers).

As mentioned above, corpus research into conference interpreting has been carried out both manually, thus looking at the data without the aid of any particular computer software, and automatically, i.e. extracting occurrences by means of IT tools and, in some cases, applying statistical testing and measurements. The latter approach might be perceived with scepticism by the more sociologically-minded scholar and as an excessively dry take just based on mere number crunching, thus losing sight of situatedness. However, quantitative studies have also opened the way to more in-depth, qualitative investigations, which are in turn corroborated by well-founded results and not simply impressionistic evidence.

Overall, attention has been focused on a variety of areas of interest, such as verbal and nonverbal features of source speeches, features of interpreters' output (and so-called *interpretese*), manifestations of cognitive load and interpreting strategies, differences between interpreting and translation of the same source text, the role of gender, and ideology.

Regarding EP discourse, various corpus projects have provided a fuller picture of source speeches (albeit with obvious differences depending on the languages considered), featuring short duration in most speech events, particularly high speed of delivery, and read or semi-prepared presentation style (Bendazzoli 2010; de Manuel 2003b; Russo et al. 2012; Vuorikoski 2004). On the other hand, target speeches, i.e. interpreters' output, have been analysed from different perspectives. Looking at the EPIC as a comparable corpus, two studies on lexical patterns (Russo et al. 2006, Sandrelli et al. 2010) revealed that lexical density and lexical variety is higher in interpreted speeches than original speeches (in the same language). This result disconfirms findings obtained in translation (Laviosa 1998) but are similar to those obtained in the English component of the SIREN corpus (Dayter 2018). Conversely, lexical variety was found lower in interpreted speeches, in line with Laviosa's findings, with the only exception of the Italian booth, and with the Russian component of SIREN. The same result was also confirmed in the EPTIC, where target texts show less lexical variety than source texts, with interpreted texts being simpler than translated texts (Bernardini et al. 2016).

A comparable perspective was also used by Kajzer-Wietrzny (2018) to analyse linguistic patterns in interpreted English in comparison with native and non-native English. Focusing on complementizer *that*, a higher degree of explicitation is found in interpreted English and greater use of optional *that* is found in both interpreted English and non-native English source speeches vs. native ones. Still from a comparable perspective, pragmatic markers were analysed in the Corpus of Interpreted Political Speeches from Chinese to English (CIPSCE) (Pan and Wong 2019) and were found to be underused in interpreted speeches with the exception of contrastive and elaborative markers, leading the authors to consider the pedagogical implications of these results.

Moving to nonverbal features, the presence of disfluencies in interpreters' output and in source speakers' output was investigated in the EPIC (Bendazzoli et al. 2011) and also in the SIREN corpus (Dayter 2020). Despite considering different sets of features (only mispronounced and unfinished words in the EPIC; mispronounced words, hesitations, self-repairs, silent pauses, repetitions and fillers in SIREN), the two studies found similar patterns only to a certain extent, with more disfluencies in target speeches than original speeches and fewer repairs in interpreters' output. However, unfinished words were more frequent in English source speeches than interpreted English in the EPIC, pointing to the important role played by individuals' speaking style (which can also be true for interpreters, see Kajzer-Wietrzny 2013).

Interpreters' speaking style has been the object of analysis in corpus studies other than the EP setting, namely in TV interpreting through the CORIT corpus (Falbo 2012). Again, the systematic observation of the data included in CORIT has helped outlining the main characteristics of talk show interpreting (Straniero Sergio 2007), as well as showing that individual interpreters tend to make recurrent use of a limited number of set-phrases and expressions, including additional discourse markers that do not correspond to equivalent units of meaning in the source speech. The occurrence of interpreter-generated discourse markers was also confirmed in studies based on other conference interpreting corpora. For instance, in the DIRSI corpus, 30% of all the occurrences of discourse marker *so* found in target speeches are added by the interpreters (Bendazzoli 2019); in the EPIC-G (Defrancq et al. 2015), discourse markers are omitted in interpretations but also added and to a greater extent than is the case in translation. This intermodal perspective has shed light on different patterns found in translated texts from those found in interpreting. For example, in the EPTIC, interpreted

speeches show a higher degree of simplification (Bernardini et al. 2016; Ferraresi et al. 2018) and a lower degree of phraseological patterns along with greater use of non-standard word combinations with respect to translated texts and also to original texts (Ferraresi and Miličević 2017). Rather than contrasting interpreting with translation, the CoSi corpus was utilised intermodally to analyse interpreters' intervention in the renditions of proper names in simultaneous and consecutive interpreting (Meyer 2008). While there is ample evidence of variation in the way proper names are conveyed in target speeches, again pointing to individual interpreters' style, no particular difference has emerged between simultaneous and consecutive interpreting.

The analysis of textual and interactional features in source speeches and/or interpreters' output has been extended to account for interpreting strategies, norms, and cognitive load. Simultaneous interpreting corpora of football press conferences (Sandrelli 2017, 2018) and presidential debates (Dal Fovo 2018) have yielded insightful results about interpreters' management of interactional patterns, highlighting the occurrence of coordination activities even among simultaneous interpreters, along with differences in professional practices depending on team composition and working conditions (e.g. on-site vs. off-site interpreters).

Moving to consecutive interpreting, in a corpus of Chinese-English governmental press conferences (Wang 2012) interpreting shifts are identified in terms of addition, reduction, and correction, with addition being far more frequent than the others across the five interpreters represented in the corpus. These shifts are then investigated qualitatively and in relation to the relevant source speeches as evidence of adherence to four particular norms: adequacy, explicitation in logic relations, specificity in information content and explicitness in meaning. Another example still on consecutive interpreting tactics is the study of language-specific features such as syntactic asymmetry between Chinese and English (Wang and Zou 2018), which is likely to impact on interpreters' cognitive load as they transform long and complex front-loaded attributive modifying structures into back-loaded ones. Similar results have been obtained in another study of markers and predictors of cognitive load in simultaneous interpreting into Dutch from a comparable perspective (Defrancq and Plevoets 2018; Plevoets and Defrancq 2018). Greater frequency of filled pauses in interpreted Dutch is detected, especially when interpreters deal with linguistic asymmetry, compound nouns, and higher lexical density, whereas formulaic patterns are inversely correlated to this manifestation of cognitive load.

Two further areas of scrutiny concern gender and ideology. These lines of research have given way to interesting expansions in the scope of corpus-based interpreting studies, clearly showing that quantitative data along with the systematic retrieval of occurrences can then be considered for qualitative, more in-depth investigations. For instance, Russo (2018) found some statistically significant speaking patterns in a sub-section of the EPIC, with female interpreters showing faster speed of delivery and longer target speeches than male interpreters. Similarly, Magnifico and Defrancq (2017) found evidence of gender-based patterns in the use of hedges in interpreting (in a sub-section of the EPIC-G), with women interpreters wedging more than men (on account of more additions). Opposite findings emerged in the management of politeness, with male interpreters downtoning face threatening acts more than female interpreters (Magnifico and Defrancq 2016).

The last area in this overview of corpus-based research on conference interpreting encompasses ideology and (Critical) Discourse Analysis. Examples in this respect are still based on EP data (Vuorikoski 2004; Bartłomiejczyk 2020; Beaton 2007, 2010, 2013) or Chinese governmental press conferences (Gu and Tipton 2020), which are analysed either manually or automatically. These studies demonstrate the importance for source speakers to be aware of the interpreting process through which the sense of their message will be conveyed to target listeners. They also reveal that multilingual environments, especially high-level conferences, pose special challenges to interpreters, from rhetorical strategies (such as self-referentiality) to irony, from racist speech to conflicting views. Still, there seems to be a lack of commonly shared interpreting standards or protocols to deal with them.

#### **4 Corpora in conference interpreting education and professional practice**

The constant development of the corpus-based approach and the availability of multiple corpora are having a slow, yet growing impact on conference interpreting education and, to a much lesser extent, professional practice.

Easier access to conference interpreting data and the implementation of extensive conference interpreting projects (e.g. EPIC and CorIT) have given many conference interpreting students the opportunity to take advantage of conference interpreting corpora projects for their undergraduate and postgraduate theses (Dal Fovo 2011; Russo 2010).

EU institutions, especially the European Parliament, are once again the main data source for corpus-related pedagogical applications. Going back in time, Dollerup and Ceelen (1996) provide possibly one of the earliest examples of pedagogical corpora for consecutive interpreting training. Their “corpus of consecutive interpreting in Danish, Dutch, English, French, German and Italian” is in fact a collection of audiotaped recordings and transcripts, not a machine-readable corpus proper, of lectures and trainees’ output within a training course held in 1976. A more advanced, digital development of this kind is the Marius database, i.e. a collection of speeches from the EP and the World Social Forum set up by de Manuel (2003a, 2003b) and enriched with metadata annotations. Annotations make it possible to search for transcripts and the relevant multimedia in a targeted way, though these transcripts are not searchable as in a standard electronic corpus. These early examples resemble other resources such as speech repositories including a variety of annotated speeches that can be used for interpreter training. Their annotations range from duration, speed of delivery, and topic to the kind of interpreting mode and level of difficulty involved. In fact, these databases and repositories may well be considered proto-corpora and could set the basis for more sophisticated language resources.

More recently, (still EU-related) conference interpreting corpora have led to the development of pedagogical resources. For simultaneous interpreting, Sandrelli (2010) designed a pedagogical activity taking advantage of corpus methods to extract a wide range of examples of verb tenses in Italian, English and Spanish from the EPIC corpus and see how these are used by EP interpreters. In the same vein, Aston (2018) argues that the analysis of concordances from a corpus of transcripts of EP interpreters’ output into English can be useful to enhance formulaic phraseologies among interpreters in autonomous learning activities. Spinolo (2018) organised training sessions to deal with figurative language utilizing the IMITES corpus, while Bertozzi (2018) developed the Anglintrad corpus and platform with a wealth of terminological information concerning the treatment of Anglicisms in interpreting and translating from Italian into Spanish. Cresswell (2018) combined different SI corpora (including EP data but also medical conferences) to study the frequency of phrasal verbs in English target speeches and found a higher frequency in the output of native English interpreters in comparison with the output of interpreters working into English as their B language (i.e. active foreign language). This finding, along with the retrieval of formulaic language from the corpora taken into account, prompted the development of targeted teaching material for conference interpreters, particularly in language enhancement modules. The

analysis of collocations from an intermodal perspective, i.e. as found in translated and interpreted target texts from the same source text in the EPTIC, led to the design of task-based activities even in the translation classroom: conference interpreting data are contrasted with translation data and stimulate reflection on the differences between the two processes (Ferraresi 2016).

Non-interpreting corpora have also been exploited for pedagogical purposes in interpreter education and training. For example, Bale (2013) took advantage of the BACKBONE corpus (Kohn 2012) for improving trainees' lexical knowledge in (English/German) consecutive interpreting. BACKBONE is a corpus of interviews in several languages (i.e. French, German, Polish, Spanish, Turkish, native and non-native English) but not relating to translation or interpreting *per se*. The usefulness of this corpus in interpreter training stems from its pedagogically-oriented annotations, which make it possible to select transcripts and videos according to language, topic (e.g. business and education), and particular language features (e.g. verb tenses and colloquial expressions). Interpreter trainees can thus use these materials as source speeches for interpreting practice and as parallel texts for terminology building and checking. A similar implementation was proposed by Leeson (2008) for sign language interpreter training with the Signs of Ireland (SOI) corpus. This corpus can be used not only as a linguistic resource for retrieving source speeches, but also as a reference corpus to check and learn collocations, and to perform self-analysis of one's interpreting performance. The last two examples clearly unveil the educational potential of corpora even when these do not represent mediated communication but simply gather instances of (direct) communication that have been put together according to certain criteria and, most importantly, have been enriched with pedagogically-oriented annotations.

As regards professional practice, it is fair to assume that corpus methods have not yet broken through. While it is hard to think of any use of corpora *during* an interpreting assignment, they may find applications prior to assignments for preparation and, more generally, to further terminological competence and working language skills (e.g. with collocation searches and phraseology). In this respect, potential solutions may be found in web-crawled corpora (Baroni and Ueyama 2006) to build ad-hoc comparable corpora and term banks (Fantinuoli 2018), and software programs such as InterpretBank (Fantinuoli 2016) which was specifically designed as a kind of workstation for interpreters, including tools for automatically extracting texts and terms, managing glossaries, and accessing terminology on the fly. The real impact



of interpreting corpora development and research on conference interpreting practice has yet to be investigated in detail, especially among the younger generations of conference interpreters who may have become familiar with corpus linguistics methods in their educational programmes.

## **5 Future prospects**

Considering the development and use of conference interpreting corpora over the last three decades, further advancements can be expected in terms of corpus features, availability, research, and application.

Modern technology and collaborative efforts may lead to an increase in the number and size of conference interpreting corpora (Bendazzoli 2018). Some data sources, in particular, have already lent themselves to the establishment of large, machine-readable corpora, and these will continue to expand. In this respect, a shared effort is necessary in swapping methodological choices, especially with respect to transcribing and annotating interpreting data. For data sources such as the European Parliament this would make it possible to have “a massively multilingual interpreting and intermodal corpus” (Bernardini et al. 2018: 23).

Each approach to data collection entails both advantages and disadvantages for corpus compilers. While practisearchers may need to deal with issues relating to reflexivity or expose themselves to possible conflicts with other interpreters not as keen on being the subject/object of academic research, external observations and indirect access inevitably offer more limited insight into the many variables at stake (e.g. information about the interpreters’ background). After all, source and target speeches in the form of transcripts remain a partial representation of spoken/signed communication and are the result of a selective process. This unavoidable limitation may be softened with greater inclusion of multimodal data in corpora with direct linking to the relevant textual data. It is also hoped that technology may streamline the transcription, annotation, and alignment processes, so that even the less tech-savvy can easily exploit corpus methods. As for annotation in particular, depending on the aims behind the creation of a conference interpreting corpus, different types of annotation may be added along with extralinguistic details of the communicative situation, the role of participants and the speech events or text types involved. Whether a corpus is machine-readable or not, sharing (if not standardising) annotation systems would be a welcome improvement to allow for replications.

In addition to offering the possibility to carry out systematic observations and to point out similarities and differences of interpreted speeches with respect to other modes of delivery and translation, conference interpreting corpora may also expand in terms of research scope following the recent call made for “a new multifactorial, multi-methodological and interdisciplinary research agenda for empirical translation studies” (De Sutter and Lefer 2020: 18). This is already being the case with quantitative studies that are setting the basis for more in-depth, qualitative investigations based on other methodologies (e.g. thematic or critical discourse analysis). In fact, this is well in line with the notion of ‘corpus’ in interpreting research, as it has been loosely attached to both large and small, machine-readable and non-computerised data sets.

Corpus research may continue to pursue the objective of identifying interpreting norms and universals, though the many variables at stake, even in large corpora, impinge on the extent to which results can be generalised (e.g. a recurrent observation is that there can be a language-specific effect). The description of typical and untypical patterns is nevertheless useful to have a better understanding of what it means to communicate through an interpreter and possibly to develop high quality standards.

Conference interpreting education should continue to benefit from interpreting corpora. It would be interesting to support the dissemination of pedagogical activities designed by individual researchers with a view to fine-tuning relevant side-modules (e.g. targeted language enhancement, communication, linguistics, terminology) other than interpreting classes *stricto sensu*. On the other hand, it is not clear when the impact of corpora on professional practice might start gaining ground – perhaps this may be seen more in the future, especially among younger generations of conference interpreters and in forthcoming working conditions.

## References

- Abdelali, Ahmed, Temnikova, Irina, Hedaya, Samy & Vogel, Stephan 2018. The WAW Corpus: The first corpus of interpreted speeches and their translations for English and Arabic. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis & T. Tokunaga (eds) *Proceedings of LREC 2018, Eleventh International Conference on Language Resources and Evaluation*. Miyazaki: ELRA, 2135–2140.
- Aston, Guy 2018. Acquiring the language of interpreters: A corpus-based approach. In M. Russo, C. Bendazzoli & B. Defrancq. (eds) *Making way in Corpus-based Interpreting Studies*. Singapore: Springer, 83–96.
- Baker, Mona 1993. Corpus Linguistics and Translation Studies: Implications and applications. In M. Baker, G. Francis & E. Tognini-Bonelli (eds) *Text and Technology: In Honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, 233–250.
- Bale, Richard 2013. Undergraduate consecutive interpreting and lexical knowledge. The role of spoken corpora. *The Interpreter and Translator Trainer* 7 (1), 27–50.
- Baroni, Marco & Ueyama Motoko 2006. Building general- and special-purpose corpora by Web crawling. In *Proceedings of the 13th NIJL International Symposium. Language Corpora: Their Compilation and Application*, 31–40.
- Bartłomiejczyk, Magdalena 2017. The interpreter's visibility in the European Parliament. *Interpreting* 19 (2), 159–185.
- Bartłomiejczyk, Magdalena 2020. How much noise can you make through an interpreter? A case study on racist discourse in the European Parliament. *Interpreting* 22 (2), 238–261.
- Beaton, Morven 2007. Interpreted ideologies in institutional discourse: The case of the European Parliament. *The Translator* 13 (2), 271–296.
- Beaton-Thome, Morven 2010. Negotiating identities in the European Parliament: The role of simultaneous interpreting. In M. Baker, M. Olohan, & M. C. Perez (eds), *Text and Context: Essays on Translation and Interpreting in Honour of Ian Mason*. Manchester: St. Jerome, 117–138.
- Beaton-Thome, Morven 2013. What's in a word? Your enemy combatant is my refugee: The role of simultaneous interpreters in negotiating the lexis of Guantánamo in the European Parliament. *Journal of Language and Politics* 12 (3), 378–399.
- Bendazzoli, Claudio & Sandrelli, Annalisa 2009. Corpus-based Interpreting Studies: Early work and future prospects. *Tradumatica* 7. *L'aplicació dels corpus linguistics a la*

- traducció*. <http://webs2002.uab.es/tradumatica/revista/num7/articles/08/08art.htm> (accessed 12 March 2020).
- Bendazzoli, Claudio 2010. *Corpora e interpretazione simultanea*. Bologna: Asterisco.
- Bendazzoli, Claudio 2012. From international conferences to machine-readable corpora and back: An ethnographic approach to simultaneous interpreter-mediated communicative events. In F. Straniero Sergio & C. Falbo (eds) *Breaking Ground in Corpus-Based Interpreting Studies*. Frankfurt am Main [etc.]: Peter Lang, 91–117.
- Bendazzoli, Claudio 2016. The ethnography of interpreter-mediated communication: Methodological challenges in fieldwork. In C. Bendazzoli & C. Monacelli (eds) *Addressing Methodological Challenges in Interpreting Studies Research*. Newcastle upon Tyne: Cambridge Scholars Publishing, 3–30.
- Bendazzoli, Claudio 2018. Corpus-based Interpreting Studies: Past, present and future developments of a (wired) cottage industry. In M. Russo, C. Bendazzoli & B. Defrancq (eds) *Making way in Corpus-based Interpreting Studies*. Singapore: Springer, 1–19.
- Bendazzoli, Claudio 2019. Discourse markers in English as a target language: The use of *so* by simultaneous interpreters. *Textus* 32 (1), 183–201.
- Bendazzoli, Claudio, Bertozzi, Michela & Russo, Mariachiara 2021. Du texte aux ressources multimodales : faire avancer la recherche en interprétation à partir d'un corpus déjà existant. *Meta* 65 (1), 211–236.
- Bendazzoli, Claudio, Sandrelli, Annalisa & Russo, Mariachiara 2011. Disfluencies in simultaneous interpreting: A corpus-based analysis. In A. Kruger, K. Wallmach & J. Munday (eds) *Corpus-based Translation Studies: Research and Applications*. London/New York: Continuum, 282–306.
- Bernardini, Silvia, Ferraresi, Adriano & Miličević, Maja 2016. From EPIC to EPTIC – Exploring simplification in interpreting and translation from an intermodal perspective. *Target* 28 (1), 61–86.
- Bernardini, Silvia, Ferraresi, Adriano, Russo, Mariachiara, Collard, Camille, & Defrancq, Bart 2018. Building interpreting and intermodal corpora: A how to for a formidable task. In M. Russo, C. Bendazzoli & B. Defrancq (eds) *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer, 21–42.
- Bernd, Meyer 2008. Interpreting proper names: Different interventions in simultaneous and consecutive Interpreting? *trans-kom* 1 (1), 105–122.
- Bertozzi, Michela 2018. ANGLINTRAD: Towards a purpose specific interpreting corpus. In C. Bendazzoli, M. Russo & B. Defrancq (eds) *New Findings in Corpus-based*

- Interpreting Studies*. Special issue of *inTRAlinea*.  
<http://www.intralinea.org/specials/article/2317> (accessed 16 January 2020).
- Bowker, Lynne & Pearson, Jennifer 2002. *Working with Specialized Language. A Practical Guide to Using Corpora*. London/New York: Routledge.
- Cresswell, Andy 2018. Looking up phrasal verbs in small corpora of interpreting: An attempt to draw out aspects of interpreted language. In C. Bendazzoli, M. Russo & B. Defrancq (eds) *New Findings in Corpus-based Interpreting Studies*. Special issue of *inTRAlinea*. (accessed 21 September 2020).
- Dal Fovo, Eugenia 2011. Through the CorIT looking-glass and what MA students found there. *The Interpreters' Newsletter* 16, 1–20.
- Dal Fovo, Eugenia 2013. Il progetto CorIT: corpus e prospettive di ricerca. *Rivista internazionale di tecnica della traduzione* 15, 45–62.
- Dal Fovo, Eugenia 2018. European Union politics interpreted on screen: A corpus-based investigation on the interpretation of the third 2014 EU presidential debate. In M. Russo, C. Bendazzoli & B. Defrancq (eds) *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer, 157–184.
- Dayter, Daria 2018. Describing lexical patterns in simultaneously interpreted discourse in a parallel aligned corpus of Russian-English interpreting (SIREN). *Forum* 16 (2), 241–264.
- Dayter, Daria 2020. Variation in non-fluencies in a corpus of simultaneous interpreting vs. non-interpreted English. *Perspectives*, 1–18.
- de Manuel Jerez, Jesús (ed.) 2003a. *Nuevas Tecnologías y Formación de Intérpretes*. Granada: Comares.
- de Manuel Jerez, Jesús 2003b. Medios audiovisuales e informáticos en la formación de intérpretes. In M. A. Vega Cernuda (coord.) *Una mirada al taller de San Jerónimo: bibliografías, técnicas y reflexiones en torno a la traducción*. Madrid: Universidad Complutense, 53–62.
- De Sutter, Gert & Lefer, Marie-Aude 2020. On the need for a new research agenda for corpus-based translation studies: A multi-methodological, multifactorial and interdisciplinary approach. *Perspectives* 28 (1), 1–23.
- Defrancq, Bart & Plevoets, Koen 2018. Over-uh-load, filled pauses in compounds as a signal of cognitive load. In M. Russo, C. Bendazzoli & B. Defrancq (eds) *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer, 43–64.

- Defrancq, Bart, Plevoets, Koen & Magnifico, Cédric 2015. Connective items in interpreting and translation: Where do they come from? In J. Romero-Trillo (ed.) *Yearbook of Corpus Linguistics and Pragmatics 2015*. Singapore: Springer, 195–222.
- Dollerup, Cay & Ceelen, Leo 1996. *A Corpus of Consecutive Interpreting, Comprising Danish, Dutch, English, French and Italian*. Copenhagen: University of Copenhagen.
- Falbo, Caterina 2012. CorIT (Italian Television Interpreting Corpus): Classification criteria. In F. Straniero Sergio & C. Falbo (eds) *Breaking Ground in Corpus-Based Interpreting Studies*. Frankfurt am Main [etc.]: Peter Lang, 155–185.
- Fantinuoli, Claudio 2016. InterpretBank. Redefining computer-assisted interpreting tools. *Proceedings of the 38<sup>th</sup> Conference Translating and the Computer, London, UK, November 17-18*, 42–54.
- Fantinuoli, Claudio 2018. The use of comparable corpora in interpreting practice and training. *The Interpreters' Newsletter* 23, 133–149.
- Ferraresi, Adriano & Bernardini, Silvia 2019. Building EPTIC: A many-sided, multi-purpose corpus of EU parliament proceedings. In I. Doval & M. T. Sánchez Nieto (eds) *Parallel Corpora for Contrastive and Translation Studies. New Resources and Applications*. Amsterdam/Philadelphia: John Benjamins, 23–139.
- Ferraresi, Adriano 2016. Intermodal corpora and the translation classroom: What can translation trainers and trainees learn from interpreting? *Linguaculture* 7 (2), 27–51.
- Ferraresi, Adriano and Miličević, Maja 2017. Phraseological patterns in interpreting and translation: Similar or different? In G. De Sutter & M.-A. Lefer (eds) *New Ways of Analysing Translational Behaviour*. Berlin: Mouton de Gruyter, 157–182.
- Ferraresi, Adriano, Bernardini, Silvia, Miličević Petrović, Maja & Lefer, Marie-Aude 2018. Simplified or not simplified? The different guises of mediated English at the European Parliament. *Meta* 63 (3), 717–738.
- Fu, Rongbo. 2016. Comparing modal patterns in Chinese-English interpreted and translated discourses in diplomatic setting. A systemic functional approach. *Babel* 62 (1): 104–121.
- Gu, Chonglong & Tipton, Rebecca 2020. (Re-)voicing Beijing's discourse through self-referentiality: A corpus-based CDA analysis of government interpreters' discursive mediation at China's political press conferences (1998–2017). *Perspectives* 28 (3), 406–423.

- Halverson, Sandra 1998. Translation Studies and representative corpora: Establishing links between translation corpora, theoretical/descriptive categories and a conception of the object of study. *Meta* 43 (4), 494–514.
- House, Juliane, Bernd, Meyer & Schmidt, Thomas 2012. CoSi – A Corpus of consecutive and simultaneous interpreting. In T. Schmidt & K. Wörner (eds) *Multilingual Corpora and Multilingual Corpus Analysis*. Amsterdam/Philadelphia: John Benjamins, 295–304.
- Hu, Kaibao & Tao, Qing 2013. The Chinese-English conference interpreting corpus: Uses and limitations. *Meta* 58 (3): 626–642.
- Kajzer-Wietrzny, Marta. 2012. *Interpreting Universals and Interpreting Style*. Unpublished PhD dissertation, Adam Mickiewicz University.
- Kajzer-Wietrzny, Marta. 2013. Idiosyncratic features of interpreting style. *New Voices in Translation Studies* 9, 38–52.
- Kajzer-Wietrzny, Marta. 2018. Interpretese vs. Non-native language use: The case of optional *that*. In M. Russo, C. Bendazzoli & B. Defrancq (eds) *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer, 97–113.
- Kellett Bidoli, Cynthia J. 2016. Methodological challenges in consecutive interpreting research: Corpus analysis of notes. In C. Bendazzoli & C. Monacelli (eds.) *Addressing Methodological Challenges in Interpreting Research*. Newcastle upon Tyne: Cambridge Scholars Publishing, 143–171.
- Kohn, Kurt 2012. Pedagogic corpora for content and language integrated learning. Insights from the BACKBONE project. *The EUROCALL Review* 20 (2), 3–22.
- Laviosa, Sara 1998. Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta* 43 (4), 557–570.
- Leeson, Lorraine 2008. Quantum leap: Leveraging the Signs of Ireland Digital Corpus in Irish sign language/English interpreter training. *The Sign Language Translator & Interpreter* 2 (2), 149–176.
- Magnifico, Cédric, & Defrancq, Bart 2016. Impoliteness in interpreting: A question of gender? *The International Journal of Translation and Interpreting* 8 (2), 26–25.
- Magnifico, Cédric, & Defrancq, Bart 2017. Hedges in conference interpreting. The role of gender. *Interpreting* 19 (1), 21–46.
- Metzger, Melanie & Roy, Cynthia 2011. The first three years of a three-year grant. When a research plan doesn't go as planned. In B. Nicodemus & L. Swabey (eds) *Advances in Interpreting Research: Inquiry in Action*. Amsterdam/Philadelphia: John Benjamins, 59–84.

- Meyer, Bernd 2008. Interpreting proper names: Different interventions in simultaneous and consecutive interpreting? *trans-kom* 1 (1). [http://www.trans-kom.eu/ihv\\_01\\_01\\_2008.html](http://www.trans-kom.eu/ihv_01_01_2008.html) (accessed 18 February 2020).
- Monti, Cristina, Bendazzoli, Claudio, Sandrelli, Annalisa & Russo, Mariachiara 2005. Studying directionality in simultaneous interpreting through an electronic corpus: EPIC (European Parliament Interpreting Corpus). *Meta* 50 (4). <http://id.erudit.org/iderudit/019850ar> (accessed 12 March 2020).
- Niemants, Natacha S. A. 2012. The transcription of interpreting data. *Interpreting* 14 (2), 165–191.
- Orlando, Marc 2010. Digital pen technology and consecutive interpreting: Another dimension in note-taking training and assessment. *The Interpreters' Newsletter* 15, 71–86.
- Pan, Jun & Wong, Billy T. M. 2019. Pragmatic competence in Chinese-English retour interpreting of political speeches. A corpus-driven exploratory study of pragmatic markers. In P. Pietrzak (ed.) *New Insights into Translator Training*. Special issue of *inTRAlinea*. <http://www.intralea.org/specials/article/2415> (accessed 9 March 2020).
- Pan, Jun 2019. The Chinese/English Political Interpreting Corpus (CEPIC): A new electronic resource for translators and interpreters. In I. Temnikova, C. Orasan, G. Corpas Pastor & R. Mitkov (eds) *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019), Varna, Bulgaria, September 5 – 6*. Shoumen: Incoma, 82–88.
- Petite, Christelle 2005. Evidence of repair mechanisms in simultaneous interpreting. A corpus-based analysis. *Interpreting* 7 (1), 27–49.
- Plevoets Koen & Defrancq, Bart 2018. The cognitive load of interpreters in the European Parliament. A corpus-based study of predictors for the disfluency uh(m). *Interpreting* 20 (1), 1–29.
- Russo, Mariachiara 2010. Reflecting on interpreting practice: Graduation theses based on the European Parliament Interpreting Corpus (EPIC). In L. N. Zybatow (ed.) *Translationswissenschaft – Stand und Perspektiven. Innsbrucker Ringvorlesungen zur Translationswissenschaft VI (Forum Translationswissenschaft, Band 12)*. Frankfurt am Main [etc.]: Peter Lang, 35–50.
- Russo, Mariachiara 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 & B. Defrancq (eds) *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer, 115–131.



- 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). *Forum* 4 (1), 221–254.
- Russo, Mariachiara, Bendazzoli, Claudio, Sandrelli, Annalisa & Spinolo, Nicoletta 2012. The European Parliament Interpreting Corpus (EPIC): Implementation and developments. In F. Straniero Sergio & C. Falbo (eds) *Breaking Ground in Corpus-Based Interpreting Studies*. Frankfurt am Main [etc.]: Peter Lang, 53–90.
- Sandrelli, Annalisa 2010. Corpus-based Interpreting Studies and interpreter training: A modest proposal. In N. L. Zybatow (ed.) *Translationswissenschaft – Stand und Perspektiven. Innsbrucker Ringvorlesungen zur Translationswissenschaft VI (Forum Translationswissenschaft, Band 12)*. Frankfurt am Main [etc.]: Peter Lang, 69–90.
- Sandrelli, Annalisa 2012. Introducing FOOTIE (Football in Europe): Simultaneous interpreting in football press conferences. In F. Straniero Sergio & C. Falbo (eds) *Breaking Ground in Corpus-Based Interpreting Studies*. Frankfurt am Main [etc.]: Peter Lang, 119–153.
- Sandrelli, Annalisa 2018. Interpreter-Mediated Football Press Conferences: A Study on the Questioning and Answering Strategies. In M. Russo, C. Bendazzoli, B. Defrancq (eds) *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer, 185–204.
- Sandrelli, Annalisa, Bendazzoli, Claudio & Russo, Mariachiara 2010. European Parliament Interpreting Corpus (EPIC): Methodological issues and preliminary results on lexical patterns in simultaneous interpreting. *International Journal of Translation* 22 (1-2), 165–203.
- Sandrelli, Annalisa. 2017. Simultaneous dialogue interpreting: Coordinating interaction in interpreter-mediated football press conferences. *Journal of Pragmatics* 107, 178–194.
- Setton, Robin 2011. Corpus-Based Interpreting Studies (CIS): Overview and prospects. In A. Kruger, K. Wallmach, & J. Munday (eds) *Corpus-based Translation Studies: Research and Applications*. London/New York: Continuum, 33–75.
- Shlesinger, Miriam 1998. Corpus-based Interpreting Studies as an offshoot of Corpus-based Translation Studies. *Meta* 43 (4), 486–493.
- Shlesinger, Miriam 2008. Towards a definition of interpretese. An intermodal, corpus-based study. In G. Hansen, A. Chesterman & H. Gerzymisch-Arbogast (eds) *Efforts and Models in Interpreting and Translation Research. A Tribute to Daniel Gile*. Amsterdam/Philadelphia: John Benjamins, 237–253.

- Spinolo, Nicoletta 2018. Studying figurative language in simultaneous interpreting: The IMITES (Interpretación de la Metáfora entre ITaliano y ESpañol) corpus. In M. Russo, C. Bendazzoli, B. Defrancq (eds) *Making Way in Corpus-bases Interpreting Studies*. Singapore: Springer, 133–155.
- Straniero Sergio, Francesco 2007. *Talkshow Interpreting. La mediazione linguistica nella conversazione spettacolo*. Trieste: Edizioni Universitarie Trieste.
- Straniero Sergio, Francesco 2012. Using corpus evidence to discover style in interpreters' performances. In F. Straniero Sergio & C. Falbo (eds) *Breaking Ground in Corpus-Based Interpreting Studies*. Frankfurt am Main [etc.]: Peter Lang, 211–230.
- Taehyung, Lee 2011. English into Korean simultaneous interpretation of Academy Awards Ceremony through open captions on TV. *Meta* 56 (1), 145–161.
- Temnikova, Irina, Abdelali, Ahmed, Djabri, Souhila & Hedaya, Samy 2019. Human-informed speakers and interpreters analysis in the WAW Corpus and an automatic method for calculating interpreters' décalage. In I. Temnikova, C. Orăsan, G. Corpas Pastor & R. Mitkov (eds) *Proceedings of the 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019), Varna, Bulgaria, September 5 – 6*. Shoumen: Incoma, 105–115.
- Thompson, Paul 2004. Spoken language corpora. In Martin, Wynne (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*.  
<http://users.ox.ac.uk/~martinw/dlc/index.htm> (accessed 28 September 2020).
- Volk, Martin 2019. Innovations in parallel corpus alignment and retrieval. In I. Doval & M. T. Sánchez Nieto (eds) *Parallel Corpora for Contrastive and Translation Studies: New resources and applications*. Amsterdam/Philadelphia: John Benjamins, 79–90.
- Vuorikoski, Anna-Riitta 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.
- Wang, Binhua & Zou, Bing 2018. Exploring language specificity as a variable in Chinese-English interpreting. A corpus-based investigation. In M. Russo, C. Bendazzoli & B. Defrancq (eds.) *Making Way in Corpus-bases Interpreting Studies*. Singapore: Springer, 65–82.
- Wang, Binhua 2012. A descriptive study of norms in interpreting: Based on the Chinese-English Consecutive Interpreting Corpus of Chinese Premier Press Conferences. *Meta* 57 (1), 198–212.

Wehrmeyer, Ella 2015. An annotation system for signed language interpreting corpora.

*Hermeneus* 17, 79–318.

Wehrmeyer, Ella 2019. A corpus for signed language interpreting research. *Interpreting* 21

(1), 62–90.

### **Biographical Note**

Claudio Bendazzoli is Assistant Professor of English Language and Translation at the Department of Economic, Social, Mathematical and Statistical Sciences of the University of Turin (Italy). Previously (2004-2011), he worked at the Department of Interpreting and Translation of the University of Bologna at Forli, where he obtained an MA in Conference Interpreting (Italian, English, Spanish) and a PhD in Interpreting Studies. His main research interests are Corpus-based Interpreting Studies, Theatre and Interpreter Training, Ethnography of Speaking, English as a Lingua Franca, Business English, and English Medium Instruction. He also works as a freelance translator and interpreter.



ROUTLEDGE



ROUTLEDGE  
HANDBOOKS

# The Routledge Handbook of Conference Interpreting

Edited by Michaela Abl-Mikasa and Elisabet Tiselius

# The Routledge Handbook of Conference Interpreting

Providing comprehensive coverage of both current research and practice in conference interpreting, *The Routledge Handbook of Conference Interpreting* covers core areas and cutting-edge developments, which have sprung up due to the spread of modern technologies and global English.

Consisting of 40 chapters divided into seven parts—Fundamentals, Settings, Regions, Professional issues, Training and education, Research perspectives and Recent developments—the *Handbook* focuses on the key areas of conference interpreting. This volume is unique in its approach to the field of conference interpreting as it covers not only research and teaching practice but also practical issues of the profession on all continents.

Bringing together over 70 researchers in the field from all over the world and with an introduction by the editors, this is essential reading for all researchers, trainers, students and professionals of conference interpreting.

**Michaela Albl-Mikasa** is Professor of Interpreting Studies at ZHAW Zurich University of Applied Sciences in Switzerland. She is currently a member of the Executive Council of the International Association for Translation and Intercultural Studies (IATIS) and of the Board of the European Network of Public Service Interpreting (ENPSIT). She is Principal Investigator of the SNSF Sinergia project CLINT.

**Elisabet Tiselius** is Associate Professor of Interpreting Studies at Stockholm University, Sweden, where she teaches public service, conference and signed language interpreting. Tiselius is a member of AIIC, serving on its research committee, is accredited to the EU institutions and is a state-authorized public service interpreter. She is a board member of the European Society for Translation Studies (EST).



---

# Routledge Handbooks in Translation and Interpreting Studies

Routledge Handbooks in Translation and Interpreting Studies provide comprehensive overviews of the key topics in translation and interpreting studies. All entries for the handbooks are specially commissioned and written by leading scholars in the field. Clear, accessible and carefully edited, Routledge Handbooks in Translation and Interpreting Studies are the ideal resource for both advanced undergraduates and postgraduate students.

**The Routledge Handbook of Translation and Education**

*Edited by Sara Laviosa and Maria González-Davies*

**The Routledge Handbook of Translation and Cognition**

*Edited by Fabio Alves and Arnt Lykke Jakobsen*

**The Routledge Handbook of Translation and Activism**

*Edited by Rebecca Ruth Gould and Kayvan Tahmasebian*

**The Routledge Handbook of Translation, Feminism and Gender**

*Edited by Luise von Flotow and Hala Kamal*

**The Routledge Handbook of Translation and Globalization**

*Edited by Esperança Bielsa and Dionysios Kapsaskis*

**The Routledge Handbook of Translation and Ethics**

*Edited by Kaisa Koskinen and Nike K. Pokorn*

**The Routledge Handbook of Translation and Health**

*Edited by Şebnem Susam-Saraeva and Eva Spišáková*

**The Routledge Handbook of Translation and the City**

*Edited by Tong King Lee*

**The Routledge Handbook of Translation History**

*Edited by Christopher Rundle*

**The Routledge Handbook of Translation and Media**

*Edited by Esperança Bielsa*

**The Routledge Handbook of Conference Interpreting**

*Edited by Michaela Albl-Mikasa and Elisabet Tiselius*

For a full list of titles in this series, please visit [www.routledge.com/Routledge-Handbooks-in-Translation-and-Interpreting-Studies/book-series/RHTI](http://www.routledge.com/Routledge-Handbooks-in-Translation-and-Interpreting-Studies/book-series/RHTI)

---

# The Routledge Handbook of Conference Interpreting

*Edited by Michaela Albl-Mikasa and Elisabet Tiselius*



Cover image: © Getty Images

First published 2022

by Routledge

2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge

605 Third Avenue, New York, NY 10158

*Routledge is an imprint of the Taylor & Francis Group, an informa business*

© 2022 selection and editorial matter, Michaela Albl-Mikasa and Elisabet Tiselius;  
individual chapters, the contributors

The right of Michaela Albl-Mikasa and Elisabet Tiselius to be identified as the authors  
of the editorial material, and of the authors for their individual chapters, has been asserted  
in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised  
in any form or by any electronic, mechanical, or other means, now known or  
hereafter invented, including photocopying and recording, or in any information  
storage or retrieval system, without permission in writing from the publishers.

*Trademark notice:* Product or corporate names may be trademarks or registered trademarks,  
and are used only for identification and explanation without intent to infringe.

*British Library Cataloguing-in-Publication Data*

A catalogue record for this book is available from the British Library

*Library of Congress Cataloging-in-Publication Data*

Names: Albl-Mikasa, Michaela, editor. | Tiselius, Elisabet, editor.

Title: The Routledge handbook of conference interpreting/edited by  
Michaela Albl-Mikasa and Elisabet Tiselius.

Description: Abingdon, Oxon; New York, NY: Taylor and Francis, 2022. |

Series: Routledge handbooks in translation and interpreting studies |

Includes bibliographical references and index.

Identifiers: LCCN 2021027164 | ISBN 9780367277895 (hardback) |

ISBN 9781032134642 (paperback) | ISBN 9780429297878 (ebook)

Subjects: LCSH: Congresses and conventions—Translating services. |  
Translating and interpreting.

Classification: LCC P306.2 .R663 2022 | DDC 418/.02—dc23

LC record available at <https://lcn.loc.gov/2021027164>

ISBN: 978-0-367-27789-5 (hbk)

ISBN: 978-1-032-13464-2 (pbk)

ISBN: 978-0-429-29787-8 (ebk)

DOI: 10.4324/9780429297878

Typeset in Times New Roman

by Newgen Publishing UK

# Contents

---

<i>Contributors</i>	<i>ix</i>
Introduction <i>Michaela Albl-Mikasa and Elisabet Tiselius</i>	1
<b>PART I</b> <b>Fundamentals</b>	<b>7</b>
1 Historical developments in conference interpreting: an overview <i>Jesús Baigorri-Jalón, María Manuela Fernández-Sánchez and Gertrudis Payàs</i>	9
2 Modes of conference interpreting: simultaneous and consecutive <i>Magdalena Bartłomiejczyk and Katarzyna Stachowiak-Szymczak</i>	19
3 Note-taking for consecutive conference interpreting <i>Barbara Ahrens and Marc Orlando</i>	34
4 Conference and community interpreting: commonalities and differences <i>Elisabet Tiselius</i>	49
<b>PART II</b> <b>Settings</b>	<b>65</b>
5 Diplomatic conference interpreting <i>Barry Slaughter Olsen, Henry Liu and Sergio Viaggio</i>	67
6 Conference interpreting at press conferences <i>Annalisa Sandrelli</i>	80
7 Media conference interpreting <i>Caterina Falbo</i>	90

## Contents

8	Conference interpreting in the European Union institutions <i>Alison Graves, Marina Pascual Olaguibel and Cathy Pearson</i>	104
9	Conference interpreting at the United Nations <i>Lucía Ruiz Rosendo and Marie Diur</i>	115
<b>PART III</b>		
<b>Regions</b>		<b>127</b>
10	Conference interpreting in the United States <i>Renée Jourdenais</i>	129
11	Conference interpreting in Russia <i>Igor Matyushin and Dmitry Buzadzhi</i>	140
12	Conference interpreting in Japan <i>Kayoko Takeda and Kayo Matsushita</i>	150
13	Conference interpreting in South Korea <i>Jieun Lee</i>	159
14	Conference interpreting in Australia <i>Marc Orlando</i>	169
15	Conference interpreting in China <i>Andrew C. Dawrant, Binhua Wang and Hong Jiang</i>	182
16	Conference interpreting in India <i>Chitra Harshvardhan and Anya Malhotra</i>	197
17	Conference interpreting in Sub-Saharan Africa <i>Kim Wallmach and Nina Okagbue</i>	216
18	Conference interpreting in Brazil <i>Reynaldo J. Pagura and Jayme Costa Pinto</i>	231
<b>PART IV</b>		
<b>Professional issues</b>		<b>241</b>
19	Quality and norms in conference interpreting <i>E. Macarena Pradas Macías and Cornelia Zwischenberger</i>	243

20	Testing for professional qualification in conference interpreting <i>Andrew C. Dawrant and Chao Han</i>	258
21	Status and profession(alization) of conference interpreters <i>Helle V. Dam and Paola Gentile</i>	275
22	Ethics and codes of ethics in conference interpreting <i>Ildikó Horváth and Małgorzata Tryuk</i>	290
<b>PART V</b>		
	<b>Training and education</b>	<b>305</b>
23	Aptitude for conference interpreting <i>Mariachiara Russo</i>	307
24	Learning and teaching conference interpreting <i>Sylvia Kalina and Rafael Barranco-Droege</i>	321
25	Theory and training in conference interpreting: initial explorations <i>Daniel Gile and Rafael Barranco-Droege</i>	337
<b>PART VI</b>		
	<b>Research perspectives: theoretical and empirical</b>	<b>355</b>
26	Working memory and cognitive processing in conference interpreting <i>Ena Hodzik and John N. Williams</i>	357
27	Strategies and capacity management in conference interpreting <i>Alessandra Riccardi</i>	371
28	Conference interpreting and expertise <i>Barbara Moser-Mercer</i>	386
29	Stress and emotion in conference interpreting <i>Pawel Korpala</i>	401
30	Sex and gender in conference interpreting <i>Bart Defrancq, Camille Collard, Cédric Magnifico and Emilia Iglesias Fernández</i>	414
31	Discourse analysis in conference interpreting <i>Alicja M. Okoniewska and Binhua Wang</i>	428

Contents

32	Corpus studies in conference interpreting <i>Claudio Bendazzoli</i>	443
33	Eye-tracking studies in conference interpreting <i>Agnieszka Chmiel</i>	457
34	Neuroimaging of simultaneous conference interpreters <i>Alexis Hervais-Adelman</i>	471
<b>PART VII</b>		
<b>Recent developments</b>		<b>489</b>
35	Distance conference interpreting <i>Kilian G. Seeber and Brian Fox</i>	491
36	Conference interpreting and new technologies <i>Claudio Fantinuoli</i>	508
37	Bridging the gap between conference interpreters and researchers with online media <i>Sarah Hickey, Jonathan Downie, Alexander Gansmeier and Alexander Drechsel</i>	523
38	Sign language conference interpreting <i>Graham Turner, Nadja Grbić, Christopher Stone, Christopher Tester and Maya de Wit</i>	531
39	Conference interpreting and English as a lingua franca <i>Michaela Albl-Mikasa</i>	546
40	Mindfulness training for conference interpreters <i>Julie E. Johnson</i>	564
	<i>Index</i>	<i>581</i>