

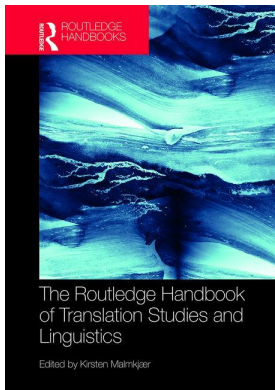
This article was downloaded by: 10.2.97.136

On: 31 May 2023

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



The Routledge Handbook of Translation Studies and Linguistics

Kirsten Malmkjær

Corpus linguistics, translation and interpreting

Publication details

<https://test.routledgehandbooks.com/doi/10.4324/9781315692845-23>

Silvia Bernardini, Mariachiara Russo

Published online on: 18 Dec 2017

How to cite :- Silvia Bernardini, Mariachiara Russo. 18 Dec 2017, *Corpus linguistics, translation and interpreting from*: The Routledge Handbook of Translation Studies and Linguistics Routledge
Accessed on: 31 May 2023

<https://test.routledgehandbooks.com/doi/10.4324/9781315692845-23>

PLEASE SCROLL DOWN FOR DOCUMENT

Full terms and conditions of use: <https://test.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Corpus linguistics, translation and interpreting

Silvia Bernardini and Mariachiara Russo

Introduction and definitions

During the first decade of the 21st century the corpus methodology established itself as one of the major paradigms in linguistics. Its fundamental assumption is that language should be studied by looking at genuine text samples stored electronically, rather than by relying on introspection and decontextualised, artificial examples. This view of linguistics as the study of language performance (or *E-language* to use Chomsky's (1986) term) rather than language competence (or *I-language*) is compatible with a product-oriented approach to the study of translation and interpreting. In this approach, the focus of attention is on the products delivered by translators and interpreters, rather than on their mental processes. While the latter can be studied through questionnaires, interviews, think-aloud protocols, key-logging, eye-tracking, and so forth, corpus-based translation and interpreting studies (hereinafter CTS and CIS) draw the bulk of their evidence from translated and interpreted texts assembled in *corpora*.

A corpus is a collection of texts, including transcriptions of spoken discourse, selected according to pre-defined criteria to be representative of a language variety, and stored in electronic format for consultation through a corpus query tool. In its simplest form, a corpus can consist of a few dozen text files stored in a local folder and searched through a stand-alone concordancer such as *AntConc* (Anthony 2014) or *Wordsmith Tools* (Scott 2016). However, corpora can also be very large and enriched with contextual metadata (about authors, publication details, intended audience, etc.) and structural and/or linguistic information (about textual subdivisions, graphical emphasis, pauses, hesitations, parts-of-speech, lemmas, etc.). The former, sometimes referred to as "DIY" or "disposable" corpora, are often constructed by single users (students, language professionals, linguists) for a specific task while the latter, requiring both linguistic and computational expertise and substantial efforts, are constructed by teams of corpus linguists and made available to the research community through client/server systems (see Baroni and Bernardini 2013 for further details on corpus preparation and corpus query systems).

Due to the nature of the object of study, positioned at the boundaries of two or more languages, corpora for translation and interpreting research tend to be more complex than those

used in other corpus linguistics (CL) fields, such as discourse studies or (monolingual) lexicography. Two main corpus typologies are used in CTS/CIS. The first, *monolingual comparable corpora*, include a minimum of two subcorpora, i.e. two collections of texts (“text” here subsumes oral language transcripts) in the same language, similar in all respects but for the existence vs. absence of a constraining source text (henceforth ST). The second, (bilingual) *parallel corpora*, include (transcripts of) STs and corresponding target texts (henceforth TTs) in one or more languages or by one or more translators/interpreters, aligned to each other, usually at the sentence level. Alongside ST–TT alignment in parallel corpora, interpreting corpora and corpora used in audiovisual translation or sign-language research may also include text-to-sound/video alignment, in which case they may be referred to as *multi-modal corpora*. These should not be confused with *intermodal corpora*, containing interpreted and translated language and/or samples from different interpreting modalities (see further “Current debates and future directions in CL, CTS and CIS”).

Historical perspectives

Originating from the so-called British school of linguistics and heavily influenced by its best-known exponent, J. R. Firth, CL became mainstream in the early 1990s, initially thanks to the work of John Sinclair (starting with the influential 1991 volume *Corpus, Concordance, Collocation*). Firth rejected the idea of an ideal native speaker competence, and believed that “we do not make up our lines as we go along, rather, to a large extent, they are already there for us, stereotyped and narrowly conditioned by our particular type of culture” (Firth 1935, 69). Sinclair further believed that “human intuition about language is highly specific, and not at all a good guide to what actually happens when the same people actually use the language” (Sinclair 1991, 4). The study of large repositories of written and spoken language samples could instead provide new and more revealing data. Corpus-derived insights and analytical methods have since been employed in a large variety of fields, for both descriptive/theoretical and applied purposes. These include, among many others, terminology/lexicography, analysis of register and style, discourse studies, language pedagogy and research on translation/interpreting.

The application of corpora and corpus methodologies to Translation Studies (TS) originates with Mona Baker’s (1993) paper in which she famously argued that “the most important task that awaits the application of corpus techniques in translation studies [. . .] is the elucidation of the nature of translated text as a mediated communicative event” (*ibid.* 243). Baker’s proposal was successful in establishing a strong connection between a powerful new methodology for the empirical study of language, and contemporary theoretical reflections within TS. Here, attention was shifting from the implicitly prescriptive, source-oriented and equivalence-focused search for translation shifts, to the descriptive, target-oriented and norm-focused account of typical traits of translated language (Toury 1995).

Baker programmatically suggested that the nature of translated text could best be elucidated through the construction and analysis of monolingual comparable corpora. By suggesting that translated language should be compared monolingually against the benchmark of non-translated language, Baker was construing it as any other language variety. The whole expanse of inherently monolingual methods being developed and applied within CL could thus be brought to bear on translated language as an object of study. Corpus-driven observations such as type–token ratio, ratio of lexical to function words, sentence length (Laviosa 1998; Xiao 2010), as well as corpus-based observations of e.g. relative frequencies of specific words, word classes, collocations and lexico-syntactic structures (Dayrell 2007;

Mauranen 2002; Olohan and Baker 2000) could be employed to identify patterning “which is specific to translated texts, irrespective of the source or target languages involved” (Baker 1995, 234); in other words, patterning that is *universal*.

Similar concerns also emerged among interpreting studies (IS) scholars, who advocated a descriptive approach to replace anecdotal observations based on case studies or limited samples, to inform theorisations on interpreters’ linguistic output and cognitive processes. The scope of IS could thus be expanded via CL to include the observation of *textual operations* of various kinds – many of them, by multiple interpreters, in multiple settings (conference, institutional assemblies, community, court, media), modes (sign-language, dialogue, simultaneous, consecutive, remote), levels of proficiency (professional, trainee, ad hoc interpreter) and conditions (real-life, simulated, experimental) – and of interpreters behaviour, with insights into their language transfer skills.

The first scholar to highlight the relevance and potential of the corpus-based approach for research into interpreting was Shlesinger (1998, 486), who suggested that the 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”. Interpreting corpora would add a new dimension to IS because they would overcome anecdotal observations and also provide information typical of CL, i.e. about word frequencies, grammatical constructions, discourse patterns, co-occurrences, lexical density, type–token ratios, etc. The CL framework would provide for IS what Baker (1993, 248) had envisaged for TS, namely: “a major leap from prescriptive to descriptive statements, from methodologizing to proper theorizing, and from individual and fragmented pieces of research to powerful generalizations”.

Core issues, topics and methods

Introduction

As an approach to linguistics that aims to unearth the patterned nature of language performance, and to highlight co-occurrence regularities that escape the traditional categories of lexis and grammar, CL has traditionally focused on phraseological constructs, i.e. those frequently used phrases whose status is intermediate between that of fully lexicalised idioms and that of free word combinations, subject only to the rules of grammar. In particular, the frequency-based notion of *collocation* and the use of lexical association measures to search for collocations in corpora have been central to CL from the start (Sinclair, Jones and Daley 1970). Corpus linguists have also targeted repeated word sequences (for which a variety of terms exist, e.g. clusters, n-grams, lexical bundles), and co-occurrence preferences at the semantic level (semantic preferences), at the pragmatic level (semantic, or discourse, prosodies) and at the lexico-grammatical level (colligations).

At the lexical end of the cline, CL research has developed methods for extracting keywords from general and specialised corpora, i.e. words that are significantly more typical of a given language variety than of the language as a whole. Together with phraseological regularities, these have found application in lexicography (starting with the well-known COBUILD dictionaries) and language learning (e.g. for the development of vocabulary lists (Coxhead 2000) and corpus-informed syllabi (Lewis 2000)).

At the grammatical end of the cline, attempts have been made at developing new approaches to grammar based on patterns found in corpora (Sinclair and Mauranen 2006).

These are consistent with the general shift in theoretical linguistics away from rule-based approaches and towards usage-based accounts of language, in which the idiom principle plays a central role (e.g. frame semantics, construction grammar or cognitive linguistics).

For historical, methodological and practical reasons, and despite the close relationship between corpus-based research into translation and interpreting, their corpus-based investigation has developed separately, with work in TS leading the way, and IS following. For the sake of an orderly account, in the following subsections we follow the chronological order and present the main developments in CTS before those in CIS, pointing out convergences and divergences as appropriate.

Core issues, topics and research methods in CTS

From its early days, CTS research has accompanied the target-oriented turn in TS, focusing on the search for hypothesised typical features, or universals of translation. These include simplification, explicitation, normalisation or conservatism, levelling out or convergence. The search for evidence about such typical features has provided a wealth of new insights about translated language, construing it as a legitimate object of empirical linguistic research, beyond the particularistic perspective of previous theoretical approaches. An early milestone in this sense 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 TS.

Implicit in the target-oriented turn was the downplaying of source-focused approaches, such as those involving the use of parallel corpora. While some studies based on parallel corpora have been conducted in the past two decades, methodologically these have tended to replicate the text-based comparison of STs and TTs, mainly relying on the direct observation of parallel concordances and focusing on translation shifts, the small changes “that build up cumulatively over a whole text as a result of the choices taken by or imposed on the translator” (Munday 1998, 542). An example of this approach, going beyond single text comparisons, is Shih (2012), who studies the rendering of prepositions in translation from English to Chinese in a composite parallel corpus collection including contemporary fiction, instruction manuals, speeches and advertisements.

A more complex parallel design is used by Øverås (1998), who searches for shifts in cohesion/coherence in the English–Norwegian Parallel Corpus, a bidirectional corpus including STs in English and their Norwegian TTs and (comparable) STs in Norwegian and their English TTs. By carrying out the analysis in both translation directions, Øverås can factor out language-specific effects (the same aim is pursued in monolingual comparable approaches through the inclusion of translations from multiple source languages). She concludes that, in both translation directions, explicating shifts are more common than implicating ones, thus providing support for the explicitation hypothesis. A similar corpus (a bidirectional parallel corpus of French and Dutch) is used by Vanderbauwhede, Desmet and Lauwers (2011) in their study of demonstrative determiners, which suggests that around 30% of translation shifts occurring in either direction are due to omissions, additions or reformulations motivated by translator preferences.

Less central to CTS than monolingual comparable corpora, parallel corpora have been used extensively in research carried out at the crossroads of CTS and corpus-based contrastive linguistics (see Johansson 2007; Dom and Declerck 2015). Parallel corpora include “simple” ones, bidirectional ones consisting of STs and TTs in two directions, multi-target ones with one source and many targets, and so forth. Where the contrastive focus predominates, the purpose is to highlight systemic differences between languages. This is the aspect that

translation research attempts to factor out in order to highlight translation-inherent shifts occurring regardless of the languages involved. Apart from this fundamental difference, the two fields share a substantial common ground in terms of methods, tools and resources. More importantly, awareness of each other's findings may prove indispensable for the progress of both disciplines, as

lack of familiarity with [translation studies] findings may lead [contrastive linguistics] researchers to interpret their data in terms of differences between language systems when they result from translation norms or strategies, while [translation studies] researchers may similarly misinterpret their data because of a lack of awareness of a systematic difference between the two language systems established by [contrastive linguistics].
(Granger 2003, 25–26)

The corpus methodology offers a shared framework in which this cross-fertilisation may occur.

The focus on translated language, construed as a mediated variety of the target language, resulted in scholarly attention being drawn not only to universal features shared by all translated texts, but also to radically local patterns, typical of individual translators. Borrowing insights from literary stylistics, studies of translators' style that adopt a monolingual comparable corpus approach attempt to identify the stylistic fingerprints of translators, or their "characteristic use of language, [their] individual profile of linguistic habits, compared to other translators" (Baker 2000, 245). Objections have been raised, however, against the total neglect in such studies of the ST, since "many important questions about writer motivation [...] may not arise in the case of translated texts unless the texts are seen in the context of their source texts" (Malmkjær 2004, 22). Several ingenious corpus designs have been used to zoom in on such a slippery research object. For instance, Ji (2010) and Marco (2004) adopt a multi-target parallel structure, i.e. a ST and two TTs. More specifically, Ji focuses on phraseology and translators' profiles in two translations of *Don Quixote* into Chinese, while Marco (2004) analyses two translations of Poe's short story, "The Fall of the House of Usher" into Catalan, also factoring in aspects of the style displayed by the two Catalan translators in their own writing. More complex still is the corpus used by Dirdal (2014), namely the fiction component of the English–Norwegian Multiple Translation Corpus (Johansson 2007). Dirdal investigates variation in the use of clause building and clause reduction in 10 different commissioned translations of the same short story (A.S. Byatt's "A Lamia in the Cevennes"), finding that translators differ in their tendency to preserve or modify the syntactic structure of the original, and that this tendency is not to be explained in terms of straightforward "fidelity", since syntactic conservatism is not matched by lexical conservatism. Approaching style in translation between English and Chinese through a series of case studies and multiple perspectives, Huang (2015) concludes that the complexity of the object requires triangulation based on a composite set of corpus resources, both intralingual (or monolingual) and interlingual (or parallel).

Summing up, three main research objectives have attracted the attention of CTS researchers from the beginning: first, the search for textual patterning supporting hypotheses about the existence of norms or universals of translation; second, the analysis of translation shifts; and, third, the identification of stylistic fingerprints left by translators in their work. Even though parallel corpora have also been used for these purposes, particularly for the search for shifts, there is no denying that the corpus approach has become associated with (various kinds of) monolingual comparable corpora. After initial

enthusiasm, however, it is now generally agreed that ST-specific effects should not be ignored by design (Pym 2008).

Core issues, topics and research methods in CIS

Corpora for IS are characterised by a number of specific features requiring the close attention of the researcher. First and foremost, the requirement to transcribe both the source discourse and the interpreters' output is arguably the main cause behind the scarcity of large machine-readable interpreting corpora (one notable exception being the *CLAIR* Simultaneous Interpreting Corpus, with its 182 hours of recordings, amounting to one million words of simulated lectures interpreted by professionals (Tohyama *et al.* 2004)). The lack of user-friendly and shared conventions for transcribing linguistic and paralinguistic features of orality further adds to the problem (Niemants 2015; Cencini 2002; Hu and Tao 2013). All in all, and despite the use of speech recognition software or methods to streamline the transcription procedure (for instance shadowing), transcription remains a major challenge for interpreting corpus projects.

Further challenges to the development of interpreting corpora are multilingualism (source and target texts, language pair and directionality) and situated synchronicity (Setton 2011, 36), as well as the accessibility of interpreting events including both originals and interpreted versions, the need for authorisations and the representativeness of events/speakers/interpreters. In terms of corpus processing, the ST–TT/sound/video alignment is another very demanding feature, due to the laborious manual encoding (alignment software used in CIS includes: CLAN, ELAN, EXMARaLDA, syncWRITER, TRANSCRIBER, TRANSANA, WINPITCH, Niemants 2015). Lastly, at the corpus annotation stage, further work is needed to include metadata concerning the ethnographic dimension of the data (speaker; date, speed and mode of delivery; subject; number of words, timing; location), linguistic features (morpho-syntactic and lexical features), paralinguistic features (disfluencies, prosody, etc.) and, depending on the corpus typology, proxemics, gestural and pragmatic features e.g. for signed language.

In terms of topics and methods, research in CIS before the availability of machine-readable corpora can be divided into several phases. At first, collections of transcripts of moderate size and generally involving only a few interpreters were taken as a basis for theorising on interpreting processes and products. Despite their limits, these studies exerted a great influence on interpreting theories and interpreter education (see e.g. Seleskovitch 1975).

In a second phase, scholars started collecting larger quantities of real-life interpreting data from specific professional settings and carried out qualitative analyses aligning STs and TTs manually. Given their vast amount of field data and the extended recording periods (from months to years), these can be considered as the first genuinely descriptive studies (Touy 1995), providing insights into interpreters' operational norms, styles, strategies, skills and field challenges. For instance, Vuorikoski (2004) evaluated the quality of 30 interpreters' linguistic outputs, in a corpus of 120 original speeches in English, Finnish, German and Swedish and their simultaneous interpretation (SI) into these languages. Her focus was on "accuracy" and "faithfulness". In a subsequent publication (Vuorikoski 2012), she concentrated on speech acts containing modals in English-language European Parliament speeches and concluded that interpreters were not always aware of the functions of speech acts, an issue that she recommended should be incorporated into interpreter training.

The theoretical framework and the methods applied in other studies of this second phase are drawn from sociolinguistics and the ethnography of communication, with interpreting

data triangulated with observation notes and interviews with the participants involved (Berk-Seligson 1990/2002; Angelelli 2004), or investigated by means of conversation analysis (Straniero Sergio 2003, 2007). Berk-Seligson investigated how Spanish–English interpreters faced the challenges of legal discourse in 114 hours of interaction in US courtrooms, highlighting the influence on the receivers’ perceptions of the way in which people spoke and were interpreted. Angelelli (2004) compiled the *California Hope Corpus*, based on data collected over 22 months, during almost 400 interpreter-mediated hospital encounters including both face-to-face and telephone interpreting. Her objective was to investigate, in particular, the interpreters’ perceptions of their role and their actual practice. Straniero Sergio studied quality, interpreters’ styles and interpreting norms based on the world’s largest multilingual media corpus, his Italian Television Interpreting Corpus (*CorIT*), currently featuring over 2,700 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” (Straniero Sergio 2003, 136), tracing the history of media interpreting and highlighting differences between conference interpreting and other forms of dialogue interpreting. Since 1999, numerous *CorIT*-based studies have appeared (Straniero Sergio and Falbo 2012).

A third phase includes large sets of real-life interpreting data, collected and stored according to criteria inspired by CL, in that they envisage the use of tools to retrieve features of STs and TTs, albeit still manually aligned (Wallmach 2000), or of tools to allow for multiple visualisations of the stored texts (Collados Aís *et al.* 2004). Wallmach recorded 110 hours of SIs by 16 professional interpreters working between English, Afrikaans, Zulu and Sepedi to investigate the effect of speed, complexity and lack of ST–TT equivalents on interpreter strategies and language-specific norms in a South African legislative context. Collados Aís *et al.* (2004) developed the multilingual *ECIS* (*Evaluación de la Calidad en Interpretación Simultánea*) corpus to explore 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 the users.

The turn from collections of computer-stored speeches to the use of CL tools and methodologies has allowed for numerous new perspectives on the investigation of interpreting (hence, fully fledged CIS), thanks to the addition of new features and the speed of information retrieval. What follows is an overview of the most prominent lines of investigation in the interpreting corpora available and their contributions to our understanding of interpreting processes and products.

Between 2004 and 2006, the first free, open, machine-readable, on-line corpus was developed: the European Parliament Interpreting Corpus (*EPIC*), containing English, Spanish and Italian speeches and corresponding simultaneous interpretations. The *EPIC* parallel and comparable design allows for a variety of investigations (Monti *et al.* 2005; Russo *et al.* 2012). Lexical patterns were investigated to ascertain whether the results obtained by Laviosa (1998) for translation held true also for interpreting. Laviosa had 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 in terms of lexical density, but generally not in terms of lexical variety (Russo, Bendazzoli and Sandrelli 2006). A similar result was obtained by Shlesinger (2009), who applied a different method, calculating the ratio of types to tokens, to identify linguistic richness in her intermodal corpus. Other topics investigated are disfluencies and repairs (Bendazzoli, Sandrelli and Russo 2011), and text-processing strategies (Russo 2011; see also Russo 2010 for an overview of qualitative studies carried out for Master’s theses). Building on the expertise gained through *EPIC*,

Bendazzoli (2012) created the Directionality in Simultaneous Interpreting (*DIRSI*) corpus, an English–Italian text-to-sound and ST–TT aligned corpus of medical conferences with a dedicated web interface. A further spin-off of *EPIC* is the European Parliament Interpreting Corpus (at) Ghent (*EPICG*) which is an open, multilingual (French > Dutch and English), partly aligned (time–ST–TT) corpus. Several topics have been explored so far, such as connective markers (Defrancq, Plevoets and Magnifico 2015), and ear–voice span (Defrancq 2015).

Press conference data from a variety of 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, a multimedia, multilingual (French, English, Spanish, Italian) corpus of original speeches and SIs (Sandrelli 2012); the Chinese–English Interpreting Corpus of the Chinese Premier’s annual press conferences (*CEIPPC*) recorded over 14 years (Wang 2012); and the Chinese–English Conference Interpreting Corpus (*CECIC*), compiled by Hu and Tao (2013), who found that interpreted texts in the corpus exhibit greater normalisation and explicitation than corresponding translated texts.

Research on interpreter language, or “interpretese”, has spurred the creation of small comparable, PoS-tagged, annotated corpora designed to identify lexical and morphosyntactic features. Shlesinger (2009) and Shlesinger and Ordan (2012) developed an English > Hebrew intermodal corpus of ST-interpreted and -translated TTs, and Aston (2018) detected typical lexical patterns in his small corpus of English interpreted speeches at the European Parliament.

An example of a multimodal (audio and video) corpus is the open-source consecutive and simultaneous corpus *CoSi* (House, Meyer and Schmidt 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 data exchange in CIS.

Finally, studies of dialogue interpreting have gained ground, especially in public service, health and court interpreting. Two large corpora have been developed: the Analysis of Mediated Interaction (*AIM*) multilingual corpus (Baraldi and Gavioli 2012), a sound-to-text and ST–TT aligned corpus of 528 interpreter-mediated medical encounters; and the Community Interpreting Database (*ComInDat*, Angermeyer, Meyer and Schmidt 2012), a corpus of interpreting data from a variety of settings and language dyads, designed to develop common standards for annotating multilingual interpreter-mediated interactions.

Current debates and future directions in CL, CTS and CIS

Consistently with the general trend within CL, the second decade of the 21st century has witnessed three main developments concerning corpus methods in the study of translation, which may continue shaping the field. First, more composite corpus designs counteract the downplaying of the ST, often through a combination of “different components of multilingual corpora as well as of reference corpora not originally created for translation-oriented purposes” (Zanettin 2012, 12). Second, quantitative methods and more sophisticated techniques for data analysis are borrowed from neighbouring research fields, e.g. those employed for authorship attribution and stylometry (Oakes and Ji 2012). Third, new hypotheses and research paradigms are emerging, extending the initial focus on translation universals and combining corpus techniques with other data sources. As concerns CIS, the priorities appear to be the triangulation of qualitative and quantitative approaches, the enrichment of corpus data with information about cognitive and pragmatic constraints and the pooling of resources within and beyond the interpreting community.

Composite corpus designs and sophisticated techniques for data analysis are employed, for instance, by Delaere, De Sutter and Plevvoets (2012) in their study of growing standardisation. They rely on a corpus of non-translated Dutch, and Dutch translated from English and French, including six different text types (administrative, journalistic, instructive, fiction, non-fiction and external communication). The novel approach taken in this study consists in splitting the corpus into nine language varieties, six based on the text type and three based on translational status (non-translated, translated from English, translated from French). Using a statistical technique called profile-based correspondence analysis, the authors calculate how different the varieties are in terms of their level of standardisation (measured as the frequency of occurrence of non-standard Belgian Dutch terms for which a standard term exists). Non-translated Belgian Dutch turns out to be (slightly) less standard than Belgian Dutch translated from English and substantially less standard than Belgian Dutch translated from French. The method further reveals that translated fiction, external communication and administrative texts conform to the general trend, while journalistic texts and non-fiction texts do not. Rybicki (2012) adopts an even more complex corpus set-up, carrying out multiple stylometric analyses of translated and non-translated fiction texts from and into several languages, arranged into different groups. He employs the technique known as Burrows's Delta, which is borrowed from the authorship attribution field and known to reliably identify authors on the basis of words that they use frequently, but which proves less apt at identifying translators. If these results somewhat undermine claims about the existence of translators' style, a different picture emerges from a subsequent study by Rybicki and Heydel (2013). Applying the same method to the Polish translation of Virginia Woolf's *Night and Day*, which was carried out partly by one translator and partly by another (who was also responsible for editing the entire text), the authors show that it is possible to identify the point where the second translator took over.

Besides ingenious corpus designs and sophisticated analytical methods, advances in CTS have been pursued through triangulation with other methods and data sources. Alves and Vale (2011) collect and analyse five types of process data (key-logging, screen recording, eye-tracking, recordings/transcriptions of retrospective protocols and questionnaires), as well as the translated texts (both final and interim versions). This composite corpus, which crosses the traditional boundary between process and product approaches, is annotated and searched using specially designed software. The triangulation of product and process is also at the basis of Jiménez-Crespo's (2015) study of explicitation in which he compares translation choices made under experimental conditions with corpus data. This study is innovative in its attempt to isolate the effect of different working methods on the tendency to explicitate (i.e. "traditional" translation vs. selection from a precompiled list). Results suggest that when subjects select their preferred translation from a set of choices – as is becoming increasingly common when post-editing/leveraging machine-translation and translation-memory output – they tend to favour more explicit formulations. Triangulation of methods and data sources is not limited to the process/product dichotomy. Another example is the application of corpus methods to audiovisual translation, for which "an integrated approach is needed to account for the complex semiotic fabric of audiovisual texts, their hybrid nature and multiple codes" (Baños, Bruti and Zanotti 2013, 483). Balirano (2013), combines corpus and multimodal analysis to account for both verbal and non-verbal humour in the TV series *The Big Bang Theory*. On the basis of a parallel corpus containing 87 transcribed episodes in their English and Italian (dubbed) versions, he shows how the Italian dubbed version (which was not as successful as the English one) "lacks both linguistic and semiotic internal coherence" (2013: 573).

In terms of new hypotheses and research paradigms, it is worth mentioning the work pioneered by Shlesinger, at the interface of CTS and CIS. Shlesinger suggested that

[i]deally, the notion of comparable corpora in interpreting studies should be extended to cover setting up three separate collections of texts in the same language: interpreted texts, original oral discourses delivered in similar settings, and *written translations of such texts*. (1998, 488, emphasis added).

Bernardini, Ferraresi and Miličević (2016) reproduce Laviosa's (1998) study of simplification using the European Parliament Translation and Interpreting Corpus (*EPTIC*), a bidirectional (English < > Italian) intermodal corpus of interpreted and translated European Parliament proceedings, featuring the parallel outputs of interpreting and translation processes, aligned to each other and to the corresponding source texts. From the monolingual comparable point of view, both interpreted and translated outputs into English and Italian are simpler than the corresponding non-translated/non-interpreted speeches, while at the intermodal level, interpretations are simpler than translations.

Moving on to CIS, there is a widely felt need for commonly shared transcription conventions and information on corpus designs, as well as for the exchange of CIS results, despite concerns about data comparability (Straniero Sergio and Falbo 2012, 37). Several quantitative and qualitative studies have provided evidence of linguistic traits characteristic of interpreted language, thus framing interpreted speech as the result of interpreters' strategic behaviour subject to time, cognitive and situational constraints. This is particularly evident when the outputs of expert and novice interpreters are compared. However, even interpreters with comparable levels of expertise display diverse speech styles and skills. Much remains to be investigated, therefore, through a triangulation of quantitative and qualitative approaches, by zooming in on, among other phenomena, target speeches that are shorter than their sources, fast speech and consequent semantic loss, and mode of delivery (improvised vs. read) in relation to the interpreter's voice quality and comprehensibility.

Further issues in need of exploration within CIS concern cognitive, pragmatic, ethical, socio-cultural and ideological aspects of interpreting (Straniero Sergio 2012). While lexical and morphosyntactic features can be tagged automatically, they need to be transcribed and annotated manually, raising complex methodological questions in terms of objectivity and replicability.

Future directions in CIS also include expanding and aligning existing interpreting corpora and furthering multimodal, intermodal, multiple language/direction/setting corpora, ideally by cooperating and pooling resources within the CIS community. The wide range of research topics expounded in Russo, Bendazzoli and Defrancq (2018) and Bendazzoli, Russo and Defrancq (2017) testifies to the vitality of this research agenda.

Finally, since interpreting corpora are invaluable resources for language learning and Natural Language Processing (development of speech recognition software and synchronous machine speech-to-speech translation), collaboration with these fields could also prove mutually beneficial.

CL, CTS and CIS: Implications for practice

CL has provided a wealth of descriptive data that have found application in a number of fields, from lexicography/terminology to language learning/teaching/testing. Concerning CTS and CIS specifically, translation-driven corpora are employed in applied fields such as machine

translation, translation practice and translator education. Statistical machine translation, currently the leading paradigm in machine-translation research, relies on massive parallel corpora for the generation of its translation models (Koehn 2005). In translation practice, some computer-assisted translation (CAT) tools (such as Kilgray's *MemoQ*) now include facilities for performing on-the-fly alignment and search of parallel corpora, alongside better-established resources like translation memories and term bases; more widespread use of corpora as documentation resources is likely to follow, thanks to this inclusion in the standard translation environment and workflow. Lastly it is worth mentioning, in connection with translator education, attempts at building translation learner corpora, i.e. collections of source texts and multiple student translations. These corpora, which can track the development of translation competence over the duration of a degree, typically feature error annotation alongside contextual metadata (Castagnoli *et al.* 2011; Espunya 2014; Kutuzov and Kunilovskaya 2014). They can help learners to develop self- and peer-evaluation skills, provide instructors with innovative teaching resources and offer researchers a vantage point for observing the development of translation competence.

Similar arguments concern practical applications to interpreter education. Access to large interpreting corpora would help trainees hone their analytical and evaluative skills, familiarise themselves with different speech styles, norms and lexical patterns, face real-life interpreting challenges and reflect on ways to meet them. In so doing they would gradually develop the metalinguistic knowledge that favours the shift towards implicit linguistic competence (Paradis 1994) required to perform the complex, time-constrained, multi-componential interpreting task. Professional interpreters too would benefit from exposure to a variety of interpreting styles. Moreover, multiple performances by the same interpreter (sometimes in different language combinations) offer opportunities for self-revision. Interpreters could monitor the pragmatic adequacy and the linguistic quality of their output, and become more aware of their potential.

Further reading

Kruger, A., Wallmach, K. and Munday, J., eds. 2011 *Corpus-Based Translation Studies: Research and Applications*. London: Continuum.

This edited volume is divided into three highly informative sections providing an overview of core concepts and topics in CTS and CIS, methods for the qualitative analysis of contrastive patterns in large translation corpora and studies in specific sub-fields, including an example of simultaneous interpreting.

Laviosa, S., ed. 1998. "L'approche basée sur le corpus/The Corpus-based Approach". *Meta* 43(4)
Available from: <http://www.erudit.org/revue/meta/1998/v/n4/>.

This special issue features 14 contributions addressing a wide range of topics central to CTS from a variety of perspectives, descriptive, theoretical and applied. Most of the studies have been highly influential and are still quoted today. The collection includes Shlesinger's contribution marking the birth of CIS.

McEnery, T. and Hardie, A. 2012. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

This volume provides an accessible and thorough introduction to CL, surveying its development, methods, findings and applications, and establishing connections with other approaches to empirical linguistics research.

Oakes, M. and Ji, M., eds. 2012. *Quantitative Research Methods in Corpus-based Translation Studies*. Amsterdam: John Benjamins.

This edited volume offers a description of several quantitative research techniques. It is of particular relevance to researchers in TS wanting to move beyond qualitative approaches.

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

As the title suggests, this is the first volume containing only CIS. The editors' introduction offers a full account of the development of CIS and of key theoretical and methodological issues. It includes the corpus designs of the first large interpreting corpora *CorIT*, *EPIC*, *DIRSI*, *FOOTIE* and accounts of studies based on them.

Related topics

Theories of linguistics and of translation and interpreting.

References

- Alves, F. and Vale, D. 2011. On drafting and revision in translation: a corpus linguistics oriented analysis of translation process data. *Translation: Computation, Corpora, Cognition* 1, pp. 105–122.
- Angelelli, C. 2004. *Medical Interpreting and Cross-Cultural Communication*. Cambridge: Cambridge University Press.
- Angermeyer, P., Meyer, B. and Schmidt, T. 2012. Sharing community interpreting corpora: a pilot study. *Multilingual Corpora and Multilingual Corpus Analysis* 14, pp. 275–294.
- Anthony, L. 2014. *AntConc (Version 3.4.3)*. Tokyo: Waseda University.
- Aston, G. 2018. “Acquiring the Language of Interpreters”. In *Making Way in Corpus-based Interpreting Studies*, edited by M. Russo, C. Bendazzoli and B. Defrancq. Singapore: Springer.
- Baker, M. 1993. “Corpus Linguistics and Translation Studies: Implications and Applications”. In *Text and Technology: In Honour of John Sinclair*, edited by M. Baker, G. Francis and E. Tognini-Bonelli, 233–250. Amsterdam: John Benjamins.
- Baker, M. 1995. Corpora in translation studies: an overview and some suggestions for future research. *Target* 7(2), pp. 223–243.
- Baker, M. 2000. Towards a methodology for investigating the style of a literary translator. *Target* 12(2), pp. 241–266.
- Balirano, G. 2013. The strange case of *The Big Bang Theory* and its extra-ordinary Italian audiovisual translation: a multimodal corpus-based analysis. *Perspectives* 21(4), pp. 563–576.
- Baños, R., Bruti, S. and Zanotti, S. 2013. Corpus linguistics and audiovisual translation: in search of an integrated approach. *Perspectives* 21(4), pp. 483–490.
- Baraldi, C. and Gavioli, L., eds. 2012. *Coordinating Participation in Dialogue Interpreting*. Amsterdam: John Benjamins.
- Baroni, M. and Bernardini, S. 2013. “Corpus Query Tools for Lexicography”. In *Dictionaries: An International Encyclopaedia of Lexicography*, edited by R. Gouws, U. Heid, W. Schweickard and H. Wiegand, 1395–1405. Berlin: Mouton de Gruyter.
- Bendazzoli, C. 2012. “From International Conferences to Machine-readable Corpora and Back: An Ethnographic Approach to Simultaneous Interpreter-mediated Communicative Events”. In *Breaking Ground in Corpus-based Interpreting Studies*, edited by F. Straniero Sergio, and C. Falbo, 91–117. Bern: Peter Lang.
- Bendazzoli, C., Russo, M. and Defrancq, B., eds. 2017. *Corpus-based Interpreting Studies: New Trends in Corpora Development and Exploitation*. Special issue of *InTRAlinea*.
- Bendazzoli, C., Sandrelli, A. and Russo, M. 2011. “Disfluencies in Simultaneous Interpreting: A Corpus-based Analysis”. In *Corpus-based Translation Studies: Research and Applications*, edited by A. Kruger, K. Wallmach and J. Munday, 282–306. London: Continuum.
- Berk-Seligson, S. 1990/2002. *The Bilingual Courtroom: Court Interpreters in the Judicial Process*. Chicago: The University of Chicago Press.

- 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.
- Castagnoli, S., Ciobanu, D., Kunz, K., Volanschi, A. and Kübler, N. 2011. “Designing a Learner Translator Corpus for Training Purposes”. In *Corpora, Language, Teaching and Resources: From Theory to Practice*, edited by N. Kübler, 221–248. Bern: Peter Lang.
- Cencini, M. 2002. On the importance of an encoding standard for corpus-based interpreting studies: extending the TEI scheme. In *CULT2K*. Special issue of *InTRAlinea*. Available from: www.intraline.org/specials/article/1678 [Accessed 18 March 2016].
- Chomsky, N. 1986. *Knowledge of Language: Its Natural Origin and Use*. New York: Praeger.
- 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)”. In *Actas del IX Seminario Hispano-Ruso de Traducción e Interpretación*, 3–15. Moscú: Universidad Estatal Lingüística de Moscú, MGLU.
- Coxhead, A. 2000. A new academic word list. *TESOL Quarterly* 34(2), pp. 213–238.
- Dayrell, C. 2007. A quantitative approach to compare collocational patterns in translated and non-translated texts. *International Journal of Corpus Linguistics* 12(3), pp. 375–414.
- 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.
- Delaere, I., De Sutter, G. and Plevoets, K. 2012. Is translated language more standardized than non-translated language? *Target* 24(2), pp. 203–224.
- Dirdal, H. 2014. Individual variation between translators in the use of clause building and clause reduction. *Corpus-based Studies in Contrastive Linguistics. Oslo Studies in Language* 6(1), pp. 119–142.
- Dom, S. and Declerck, B. 2015. Translating English non-human subjects in agentive contexts: a closer look at Dutch. *Across Languages and Cultures* 16(2), pp. 285–310.
- Espunya, A. 2014. The UPF learner translation corpus as a resource for translator training. *Language Resources and Evaluation* 48(1), pp. 33–43.
- Firth, J. R. 1935. “On sociological linguistics”. Reprinted in *Language in Culture and Society*, edited by D. Hymes, 1964, 66–70. New York: Harper International.
- Granger, S. 2003. “The Corpus Approach: A Common Way Forward for Contrastive Linguistics and Translation Studies”. In *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, edited by S. Granger, J. Lerot and S. Petch-Tyson, 17–29. Amsterdam: Rodopi.
- House, J., Meyer, B. and Schmidt, T. 2012. “CoSI-A Corpus of Consecutive and Simultaneous Interpreting”. In *Multilingual Corpora and Multilingual Corpus Analysis*, edited by T. Schmidt and K. Wörner, 295–304. Amsterdam: John Benjamins.
- Hu, K. and Tao, Q. 2013. The Chinese-English conference interpreting corpus: uses and limitations. *Meta* 58(3), pp. 626–642.
- Huang, L. 2015. *Style in Translation: A Corpus-based Perspective*. Dordrecht: Springer.
- Ji, M. 2010. *Phraseology in Corpus-based Translation Studies*. Bern: Peter Lang.
- Jiménez-Crespo, M. A. 2015. Testing explicitation in translation: triangulating corpus and experimental studies. *Across Languages and Cultures* 16(2), pp. 257–283.
- Johansson, S. 2007. *Seeing through Multilingual Corpora*. Amsterdam: John Benjamins.
- Koehn, P. 2005 “Europarl: A Parallel Corpus for Statistical Machine Translation”. <http://homepages.inf.ed.ac.uk/pkoehn/publications/europarl-mtsummit05.pdf>.
- Kutuzov, A. and Kunilovskaya, M. 2014. “Russian Learner Translator Corpus: Design, Research Potential and Applications”. In *Text, Speech and Dialogue*, edited by P. Sojka, A. Horák, I. Kopeček and K. Pala, 315–323. Dordrecht: Springer.
- Laviosa, S. 1998. Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta* 43(4), pp. 557–570.

- Lewis, M. 2000. *Teaching Collocation: Further Developments in the Lexical Approach*. Hove: Language Teaching Publications.
- Malmkjær, K. 2004. Translational stylistics: Dulcken's translations of Hans Christian Andersen. *Language and Literature* 13(1), pp. 13–24.
- Marco, J. 2004. Translating style and styles of translating: Henry James and Edgar Allan Poe in Catalan. *Language and Literature* 13(1), 73–90.
- Mauranen, A. 2002. "Where's cultural adaptation?" *InTRAlinea*, Available from: <http://www.intraline.it/> [Accessed 12 March 2016].
- 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 from: <http://www.erudit.org/revue/meta/2005/v50/n4/019850ar.pdf> [Accessed 18 March 2016].
- Munday, J. 1998. A computer-assisted approach to the analysis of translation shifts. *Meta* 43(4), pp. 542–556.
- Niemants, N. 2015. "Transcription". In *The Routledge Encyclopedia of Interpreting Studies*, edited by F. Pöchhacker, 421–422, Routledge VitalSource Bookshelf Online.
- Oakes, M. and Ji, M., eds. 2012. *Quantitative Research Methods in Corpus-based Translation Studies*. Amsterdam: John Benjamins.
- Olohan, M. and Baker, M. 2000. Reporting that in translated English. *Across Languages and Cultures* 1(2), pp. 141–158.
- Øverås, L. 1998. In search of the third code: an investigation of norms in literary translation, *Meta* 43(4), pp. 571–588.
- Paradis, M. 1994. Towards a neurolinguistic theory of simultaneous translation: the framework. *International Journal of Psycholinguistics* 10(3), pp. 319–335.
- Pym, A. 2008. "On Toury's Laws of How Translators Translate". In *Beyond Descriptive Translation Studies*, edited by A. Pym, M. Shlesinger and D. Simeoni, 311–328. Amsterdam: John Benjamins.
- Russo, M. 2010. "Reflecting on Interpreting Practice: Graduation Theses based on the European Parliament Interpreting Corpus (EPIC)". In *Translationswissenschaft – Stand und Perspektiven. Innsbrucker Ringvorlesungen zur Translationswissenschaft VI* (Forum Translationswissenschaft, Vol. 12), edited by L. Zybatow, 35–50. Frankfurt am Main: Peter Lang.
- Russo, M. 2011. "Text Processing Patterns in Simultaneous Interpreting (Spanish-Italian): A Corpus-based Study". In *Translation – Sprachvariation – Mehrsprachigkeit: Festschrift in Honour of Prof. Lew Zybatow*, edited by I. Ohnheiser, W. Pöckl and P. Sandrini, 83–103. Frankfurt am Main: Peter Lang.
- 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., Bendazzoli, C., Sandrelli, A. and Spinolo, N. 2012. "The European Parliament Interpreting Corpus (EPIC): Implementation and Developments". In *Breaking Ground in Corpus-based Interpreting Studies*, edited by F. Straniero Sergio and C. Falbo, 35–90. Bern: Peter Lang.
- Russo, M., Bendazzoli, C. and Defrancq, B., eds. 2018. *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer.
- Rybicki, J. 2012. "The Great Mystery of the (almost) Invisible Translator: Stylometry in Translation". In *Quantitative Methods in Corpus-based Translation Studies*, edited by M. Oakes and M. Li, 231–248. Amsterdam: John Benjamins.
- Rybicki, J. and Heydel, M. 2013. The stylistics and stylometry of collaborative translation: Woolf's "Night and Day" in Polish. *Literary and Linguistic Computing* 28(4), pp. 708–717.
- Sandrelli, A. 2012. "Introducing FOOTIE (Football in Europe): Simultaneous Interpreting in Football Press Conferences". In *Breaking Ground in Corpus-based Interpreting Studies*, edited by F. Straniero Sergio and C. Falbo, 119–153. Bern: Peter Lang.
- Scott, M. 2016. *WordSmith Tools Version 7*. Stroud: Lexical Analysis Software.
- Seleskovitch, D. 1975. *Langage, Langues et Mémoire. Etude de la Prise de Notes en Interprétation Consecutive*. Paris: Minard.

- Setton, R. 2011. "Corpus-based Interpreting Studies (CIS): Overview and Prospects". In *Corpus-based Translation Studies: Research and Applications*, edited by A. Kruger, K. Wallmach and J. Munday, 33–75. London: Continuum.
- Shih, C.-I. 2012. A corpus-aided study of shifts in English-to-Chinese translation of prepositions. *International Journal of English Linguistics* 2(6), pp. 50–62.
- 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 *Efforts and Models in Interpreting and Translation Research: A Tribute to Daniel Gile*, edited by G. Hansen, A. Chesterman and H. Gerzymisch-Arbogast, 237–253. Amsterdam: John Benjamins.
- Shlesinger, M. and Ordan, N. 2012. More spoken or more translated? Exploring a known unknown of simultaneous interpreting. *Target* 24(1), pp. 43–60.
- Sinclair, J. McH. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. McH., Jones, S. and Daley, R. 1970. *English Collocation Studies: The Osti Report*. Reprinted 2004. London: Continuum.
- Sinclair, J. McH. and Mauranen, A. 2006. *Linear Unit Grammar*. Amsterdam: John Benjamins.
- Straniero Sergio, F. 2003. Norms and quality in media interpreting: the case of Formula One press-conferences. *Interpreters' Newsletter* 12, pp. 135–174.
- Straniero Sergio, F. 2007. *Talkshow Interpreting: La Mediazione Linguistica nella Conversazione Spettacolo*. Trieste: Edizioni Università di Trieste.
- 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. "CIAIR Simultaneous Interpreting Corpus". *Proceedings of Oriental COCOSDA 2004*. Available from: http://slp.itc.nagoya-u.ac.jp/web/papers/2004/Oriental-COCOSDA2004_tohyama.pdf [Accessed 18 March 2016].
- Toury, G. 1995. *Descriptive Translation Studies and Beyond*. Amsterdam: John Benjamins.
- Vanderbauwhede, G., Desmet, P. and Lauwers, P. 2011. The shifting of the demonstrative determiner in French and Dutch in parallel corpora. *Meta* 56(2), pp. 443–464.
- 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 from: <http://tampub.uta.fi/handle/10024/67348> [Accessed 18 March 2016].
- Vuorikoski, A. R. 2012. "Fine-tuning SI Quality Criteria: Could Speech Act Theory Be of Any Use?" In *Interpreting across Genres: Multiple Research Perspectives*, edited by C. J. Kellett Bidoli, 152–170. Trieste: EUT.
- 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 from: <http://translationjournal.net/journal/60interpreting.htm> [Accessed 18 March 2016].
- Xiao, Z. 2010. How different is translated Chinese from native Chinese? *International Journal of Corpus Linguistics* 15(1), pp. 5–35.
- Zanettin, F. 2012. *Translation-driven Corpora*. Manchester: St Jerome.