>>> revista tradumàtica
Traducció i Tecnologies de la Informació i la Comunicació
Número 07: L'aplicació dels corpus lingüístics a la traducció : desembre 2009 :
http://www.fti.uab.cat/tradumatica/revista : ISSN: 1578-7559

Corpus-based Interpreting Studies: Early Work and Future Prospects

Claudio Bendazzoli* & Annalisa Sandrelli[1]**
***Department of Interdisciplinary Studies in Translation, Languages and Cultures, University of Bologna at Forlì**
****Faculty of Interpreting and Translation LUSPIO, Rome**

Introduction

Corpus-based Interpreting Studies (CIS) is a new branch of Interpreting Studies that has begun to take shape in recent years, following in the footsteps of what has already been done in Translation Studies and in corpus linguistics. The present paper briefly discusses the main methodological issues to be taken into account when creating an interpreting corpus (§1, 2.1), reviews the first pioneering attempts (§2.2, 2.3), and concludes by outlining two projects currently being developed at the University of Bologna and LUSPIO university in Rome (§3.1 and 3.2, respectively).

1. Corpus-based Translation Studies (CTS) and Corpus-based Interpreting Studies (CIS): the challenge

The idea of applying corpus linguistics techniques and methods to Translation Studies was put forward for the first time by Mona Baker, who predicted that "[t]he availability of large corpora of both original and translated text, together with the development of a corpus-driven methodology will enable scholars to uncover the nature of translated texts as a mediated communicative event" (Baker, 1993: 243). This intuition has had positive and fruitful implications amongst Translations scholars (Laviosa, 2004), as it opened up new research lines and methodologies (Baker, 1995) that could be used to "study translation as a variety of language behaviour that merits attention in its own right" (Baker, 1996: 175). Just a few years later, the "corpus-based approach" could already be considered a fully-fledged "new paradigm in Translation Studies", as testified by the many contributions gathered in the 1998 special issue of the Translation journal *Meta* (edited by Laviosa). Those contributions gave insight into some of the key aspects of CTS, such as the possible types of corpora that can be created by using source and target texts (Baker, 1998), the problem of representativeness of one's data sample (Halverson, 1998), advantages and disadvantages of parallel corpora (Malmkjær, 1998), the potential benefits of using corpora in translator training (Zanettin, 1998; Bowker, 1998) and the need to correlate results to the context (and corpus with its specific features) from which the data are taken (Tymoczko, 1998). All but one of the remaining papers in the same issue concerned studies conducted on corpora made of written source and target texts. The only exception was Miriam Shlesinger's paper, which addressed the application of a corpus-driven methodology to the study of interpreting, i.e. translation of oral discourse. A number of challenges and opportunities were mentioned in that paper, and these are considered in more detail in the following section. What is interesting to point out at this stage is that, clearly, the development of CTS (corpus-based studies on written translation) has been more advanced than the development of CIS (corpus-based interpreting studies) since the very beginning of this scholarly venture. There is still a considerable gap between the two, both in terms of corpus size and availability and in terms of number of studies and pedagogical applications (see, among others, Kenny, 2001; Laviosa, 2002; Zanettin et al. 2003; Kruger 2004; Aston et al. 2004). This is probably due to the greater challenges and obstacles involved in setting up interpreting corpora, i.e. electronic corpora of transcribed speech events, which include an original (source language, hereafter SL) speech and its parallel (target language, hereafter TL) version into one or more foreign languages.

2. CIS: an overview

2.1 General obstacles

In corpus linguistics, corpus-based studies on spoken language are less advanced than studies on written language, mostly because of the time-consuming nature of data collection and transcription: "The recording and transcription of unscripted speech events is highly labour intensive in comparison to the work involved in collecting quantities of written text for analysis" (Thompson, 2005: 254).

Clearly, the same applies to CIS, since observational studies on interpreting necessarily involve the recording and transcribing of a communicative event (Gile, 1994, 2000; Shlesinger, 1998). However, an added difficulty is the limited access to authentic data, as was recently emphasised by Pöchhacker (2008): it is difficult to obtain collaboration from conference organisers and speakers (for confidentiality reasons) and even harder to obtain consent from professional interpreters, who tend to perceive scientific research as attempts to evaluate the quality of their work (Cencini, 2002; Gile, 1997; Kalina, 1994). This hurdle clearly makes it harder to collect sufficiently large samples of representative and homogeneous interpreting data.
Shlesinger (1998) points out that, given the complexity of the interpreting process, as many variables as possible must be controlled to obtain reliable results. The first variable is the type of interpreter-mediated event, since, clearly, interpreting in a medical conference is not the same as in court. The type of event determines participants and their roles, as well as the interpreter's role, and therefore has an impact on the interpreting service. Interpreting mode is also an issue, since consecutive, simultaneous and liaison interpreting all have their own specific characteristics. There is also high variability concerning speakers (their public speaking experience, their language skills, their accent, style, etc.) and speeches (topic, length, speed, degree of technicality, position along the written-to-spoken continuum, use of accompanying visual information, and so on). The target audience is also important, since the expectations of a small gathering of experts are different from those of the general public in a popular science lecture. Likewise, it is not possible to compare different interpreters, unless one tries to control certain factors, such as their training, expertise, language combination, preparation for a specific assignment, working conditions (equipment), and so on.
If, notwithstanding the above-mentioned practical and methodological hurdles, researchers manage to collect sufficiently homogeneous data which can be considered representative of a specific communicative situation, they are then faced with the problem of deciding how to transcribe the data. In other words, depending on the aims of the study, they have to decide what to transcribe and which conventions to use, and how to encode the data to make automatic or semi-automated analysis possible (see Armstrong, 1997; Cencini, 2002; Monti et al. 2005; Bendazzoli & Sandrelli, 2005-2007; Bendazzoli, *forthcoming(a)* for a fuller discussion of these issues).

2.2 'Manual' corpora and early machine-readable corpora

Despite this overall background, there have been a growing number of observational and experimental studies based on corpus data. However, until not long time ago most of these studies have been based on 'traditional' or 'manual' analyses, since they do not take advantage of computational linguistics or corpus linguistics methods; moreover, generally speaking, these studies are still based on relatively small samples, which are not available in electronic form. They also use different transcription conventions and the audio/video recordings and transcripts are not directly available to the scientific community (see Setton *forthcoming* for a comprehensive overview). For example, Vourikoski (2004) compiled a corpus of 122 speeches in four languages (English, Finnish, Swedish and German) recorded at the European Parliament. The transcripts of these speeches (together with their target versions) are in electronic form, and some of them feature a link to the relevant audio file. However, these transcripts would probably need further processing before they can be studied using corpus linguistics computer programs. Similarly, Straniero Sergio (2007) recorded a great number of interpreter-mediated events on Italian TV in order to study talk show interpreting from a Conversation Analysis perspective. Here again, the transcripts and the video

recordings are there, but it is up to IT technicians, computational linguists and interpreting scholars to pool their expertise and work together to compile a machine-readable corpus.

There are very few studies which actually make use of corpus linguistics techniques and these are either student dissertations (which did not continue after the author's graduation) or small-scale projects carried out by individual researchers. For example Cencini (2000) created the Television Interpreting Corpus (TIC), a 36,000 word corpus with transcripts of interpreter-mediated TV programmes (the languages involved are English and Italian). The corpus is based on the TEI standard (see Web References) to transcribe, annotate and index the material, which can thus be queried using computer programs to automatically retrieve occurrences. As was commented in the previous section (§2.1), this research work highlighted many critical issues in CIS. Regrettably, TIC is not available online and it is a closed project. Similarly, Fumagalli (1999-2000) created a parallel corpus of 18 English source speeches on international current affairs and corresponding Italian target speeches interpreted consecutively by interpreting trainees, and a comparable corpus of 15 Italian speeches. The aim of the study was to verify whether the main trends of *translationese* (explicitation, simplification, normalisation and levelling; see Baker, 1996) could be identified in interpreted output too. The study was conducted by transcribing and aligning the source and target speeches by means of the *MultiConcord-Parallel Concordancer* application. Although the size of the corpora involved makes it impossible to generalize her conclusions, it is interesting to note that her starting hypothesis was confirmed: Fumagalli did find evidence of explicitation, simplification, normalisation and levelling in the interpreted speeches she analyzed. Unfortunately, like the other corpora mentioned in this section, the corpus is not openly available to the scientific community. The same applies to another innovative corpus-based study (Shlesinger, 2008), which compared the output of different Translation modes (written translation and simultaneous interpreting), using the same SL text and the same group of subjects. By using corpus linguistics methods, it was possible to calculate type-token ratio and study a number of lexical and grammatical features with the aim of isolating any features of 'interpreted language' or *interpretese*. Therefore, as well as parallel and comparable corpora, CIS can also include intermodal corpora, i.e. corpora "consisting solely of translations, in different modalities or in different modes" (Shlesinger, 2008: 240). Although this is surely a promising development to be further explored, to our knowledge there are no such corpora currently available to the scientific community at large.
It can be concluded that early attempts to develop CIS were first based on 'manual' corpora, i.e. sample data and transcripts that could not be studied using corpus linguistics methods. Then, more steps were made towards fully-fledged machine-readable corpora, including easier (local) access to recordings. However, general accessibility to these electronic corpora was limited and most projects have remained isolated attempts.

2.3 Machine-readable corpora

The current sub-section provides a brief overview of current projects based on machine-readable corpora that are already available to the scientific community. To our knowledge, the first corpus of this kind was EPIC, the European Parliament Interpreting Corpus (Monti et al. 2005). In 2004 the Directionality Research Group in Forlì decided to create an interpreting corpus with speeches taken from the European Parliament plenary sessions, in order to bypass the practical and methodological problems mentioned in §2.1. From a practical point of view, the EP plenary sessions are in the public domain and can be used for research and educational purposes. Moreover, homogeneity of the data is ensured by the institutionalized setting (the plenary debates, with their procedures and routines) and the interpreters' selection process and working environment. EPIC is a trilingual (English, Italian and Spanish) open corpus, with transcripts of SL speeches and corresponding TL versions in all the possible combinations and directions of the three languages involved (a total of 9 sub-corpora). In other words, it can be used both as a parallel and as a comparable corpus. The transcripts have been POS-tagged and indexed and the corpus can be queried online by means of a dedicated web interface. Currently, EPIC is the largest electronic corpus available in CIS, standing at almost 180,000 words in total. It has already been used to carry out research on lexical density and variety (Russo et al. 2006, Sandrelli et al. *forthcoming*) and on disfluencies in simultaneous interpreting (Bendazzoli

et al. *forthcoming*), as well as for a number of graduation dissertations on various issues (Russo, *forthcoming*).
Another example of machine-readable interpreting corpus with its own freely available web interface is the K6 corpus compiled by Meyer (2008). This includes recordings (5 hours) and transcripts (35,000 words in total) of lectures originally given in Brazilian Portuguese and interpreted into German, using both simultaneous and consecutive interpreters. The speaker was invited to Germany by an environmentalist NGO and toured the country to give lectures on the Amazon in three different German cities. The talks were recorded and transcribed using the *EXMARaLDA* software; Meyer used the corpus to compare the treatment of proper names in consecutive and simultaneous interpreting.

The same researcher also created the K2 corpus for the "Interpreting in Hospitals (DiK)" project. This is a 160,000 word corpus including transcripts of monolingual and interpreted doctor-patient communication (in German, Turkish, Portuguese and Spanish). In total, about 25 hours of audio recordings and transcribed words are available from the same website as the K6 corpus.

3. CIS: work-in-progress

In this last section, two ongoing CIS projects are briefly presented, namely DIRSI and FOOTIE. These two corpora share a number of features but, at the same time, they also have several differences. Both DIRSI and FOOTIE concern the same language pair (English and Italian) and the same interpreting mode, i.e. simultaneous interpreting (provided by means of technical equipment and a sound-proof booth). Permission to use the material (i.e. source and target speeches) for research and teaching purposes was obtained by the principal investigators, who were also directly involved as interpreters (in Gile's terms, *practisearchers*; see Gile, 2000) in most recorded assignments. However, the communicative situations under study are different in the two corpora, with DIRSI being based on international conferences about health-related subjects, and FOOTIE concerning football press conferences.

3.1 DIRSI

The first corpus is named DIRSI, i.e. Directionality in Simultaneous Interpreting (Bendazzoli, *forthcoming(b)*), since it includes interpreters' output into both their native language and their foreign working language. Notwithstanding the criticism which the latter language direction (A to B) has always received by interpreting scholars and professionals working in international institutions in the West, 'working into B' is common practice in most domestic private markets and used to be the norm in Eastern Europe.
DIRSI is being created by using audio recordings from international conferences held in Italy over the last three years. These are always structured into different sessions (i.e. opening, presentation, debate and closing sessions) and involve more than one participant giving paper presentations or lectures. Five professional interpreters collaborated in this project by granting permission to be recorded (one English and four Italian native speakers). Their collaboration also had a positive influence in obtaining consent from conference organizers and other participants.

Debates and Q&A sessions are actually excluded from the corpus, owing to their high degree of interactivity in communication (dialogue), which strongly differentiates them from the 'monologic' speech events delivered in all the other working sessions. At the time of writing, three conferences have been fully transcribed, POS-tagged, lemmatized and indexed, totalling more than 130,000 words. In particular, the SL sub-corpus includes approximately 70,000 words and the TL sub-corpus includes nearly 60,000 words from 20 hours of selected recordings overall. For this project, further material was collected from other conferences and this will be added as transcripts are completed. Text-to-sound alignment is also envisaged as a next step and the resulting corpus will be made accessible via a dedicated online web-interface. The creation of this new corpus has been possible thanks to the experience previously gained with the EPIC project.

3.2 FOOTIE

FOOTIE is much more restricted in scope than either DIRSI or EPIC. The latter corpora include texts on various topics from different plenaries or conference sessions; by contrast, all the texts in FOOTIE come from the same setting and the same type of communicative event, namely the press conferences scheduled before and after every game played by Italy's national team during the 2008 European football championships (UEFA EURO 2008) held in Switzerland and Austria. This new project has just been started by Annalisa Sandrelli at LUSPIO University in Rome, where a number of undergraduate students are collaborating by transcribing portions of the available data for their dissertations on various aspects of interpreting.

As happened with data collection for DIRSI, once again it was possible to obtain the relevant recordings because the principal investigator was recruited to work as Italy's interpreter during EURO 2008; after the championship, UEFA granted permission to use the video and audio materials for research purposes. Clearly, this procedure may involve a degree of researcher bias, but, as was pointed out in § 2.1, it is very hard to obtain authentic recordings of interpreter-mediated events and therefore the analysis of one's work is, unfortunately, rather common among *practisearchers*. In order to reduce the effect, permission is being sought from the other interpreters at work in the relevant games to use their recordings too.

During EURO 2008, Italy played Holland, Romania and France in the first round, and went out to Spain in the quarter finals. For each game there was one pre-match and one post-match press conference for Italy and the same for their opponents. Interpreters were 'assigned' to a specific team and worked in all of the press conferences involving that particular team. Two interpreters were always at work in all the press conferences, each of them working in both translation directions (A to B and B to A) in his/her own booth. The official languages always included English (for the international press) and the two languages spoken in the countries of the two teams. English was also used by the interpreters as a *pivot* language whenever the foreign language used by the speakers was not one of their working languages. Overall, a total of 16 press conferences involving Italian, English, French and Spanish as SLs and TLs are available for transcription and analysis. However, in order to begin this project with an analysis of the more homogeneous part of the corpus, it has been decided to start by transcribing all of Italy's press conferences (in Italian) and corresponding English target versions. This part of the corpus corresponds to over two and a half hours of SL material and matching interpreted version produced by the same interpreter working from her A language (Italian) into her B language (English).

All the FOOTIE press conferences were interpreted simultaneously. Press conferences are an example of dialogic communication characterised by high interactivity, and in this sense the FOOTIE material can be said to resemble conference Q&A sessions. The type of dialogue is also specific to this setting, in that it features examples of one-to-one communication (interviewers posing questions to the interviewee/s, generally a football manager and sometimes a player) and of one-to-many communication (the interviewee replying to each question for the benefit of all the journalists present in the room). There is also a composite audience: there is a primary audience that is entirely made up of potential interviewers (only journalists were admitted to the press conference rooms) and a secondary audience that is not physically present, i.e. the football fans from all the countries involved (although the press conferences were not fully televised, excerpts were used by TV channels and information obtained during the press conferences was used by media people to write match reports and articles).

The project is still in its infancy and many aspects are still to be defined, including data encoding methods. However, the data certainly look interesting and it is hoped that their homogeneity will make it possible to carry out interesting studies on the corpus when it is ready for analysis.

4. Conclusions

Nearly two decades after the initial efforts to apply corpus linguistics to Translation Studies, Corpus-based Interpreting Studies are still at a less advanced stage of development than Corpus-based (written) Translation studies. This is probably due to the many obstacles involved in creating spoken corpora in general and to the many variables at stake in setting up Interpreting corpora.

Early attempts were based on small samples of data, which usually were not machine-readable. Over the last few years, technological advancements and greater collaboration between Translation scholars and IT experts have made it possible to obtain larger samples of data and store them in electronic form to create corpora. Unfortunately, once completed, many of these projects have not been made accessible to the scientific community at large. However, there are some exceptions, such as the European Parliament Interpreting Corpus (EPIC) and the K6 and K2 corpora, which are publicly available and can be accessed online.
The latest examples of ongoing CIS projects presented in this paper are called DIRSI and FOOTIE, two interpreting corpora based on health-related international conferences and football-related press conferences respectively. In particular, with these two corpora it will be possible to study the role played by directionality (i.e. whether interpreters work into their A or B language). Despite the principal investigators' direct involvement in data collection in both projects, we believe that *practisearchers* may have the key to accessing real life data. Although self-analysis is likely to be criticised, its limits are still to be demonstrated in most cases. If we really want CIS to catch up with CTS in terms of number of resources and contributions to research, closer collaboration between *practisearchers* and their colleagues in both academic and professional settings is vital and strongly called for.

References

Armstrong, S. (1997). "Corpus-based methods for NLP and translation studies", *Interpreting*, 2/1-2.

Aston, G., Bernardini, S. & D. Stewart (eds.) (2004). *Corpora and Language Learners*. Amsterdam / Philadelphia: John Benjamins.

Baker, M. (1993). "Corpus Linguistics and Translation Studies: implications and applications", in Baker, M., Francis, G. & E. Tognini-Bonelli (eds.) (1993), *Text and Technology: In Honour of John Sinclair*, Amsterdam / Philadelphia: John Benjamins.

Baker, M. (1995). "Corpora in Translation Studies. An overview and suggestions for future research", *Target*, 7/2.

Baker, M. (1996). "Corpus-based Translation Studies. The challenges that lie ahead", in Somers, H. (ed.) (1996), *Terminology, LSP and Translation*, Amsterdam / Philadelphia: John Benjamins.

Baker, M. (1998). "Réexplorer la langue de la traduction: une approche par corpus", *Meta*, 43/4.

Bowker, L. (1998). "Using specialized monolingual native-language corpora as a translation resource: a pilot study", *Meta*, 43/4.

Bendazzoli, C. (*forthcoming*(a)). "The European Parliament as a source of material for research into simultaneous interpreting: advantages and limitations", in Zybatow L. (ed.) (*forthcoming*), *Translation Studies: State of the Art and Perspectives*, series 'Forum Translationswissenschaft' (vol 12).

Bendazzoli, C. (*forthcoming*(b)). *Il corpus DIRSI: creazione e sviluppo di un corpus elettronico per lo studio della direzionalità in interpretazione simultanea*. PhD thesis, University of Bologna at Forlì.

Bendazzoli, C. & Sandrelli, A. (2005-2007). "An approach to corpus-based interpreting studies: developing EPIC (European Parliament Interpreting Corpus)", in Nauert S. (ed.) (2005-2007), *Proceedings of the Marie Curie Euroconferences MuTra: Challenges of Multidimensional Translation - Saarbrücken 2-6 May 2005*.
< http://www.euroconferences.info/proceedings/2005_Proceedings/2005_proceedings.html >. Page consulted on date: 23.10.09.

Bendazzoli, C., Russo, M. & Sandrelli, A. (*forthcoming*). "Disfluencies in simultaneous interpreting: a corpus-based analysis", to appear in Kruger, A. & K, Walmach (eds.) (*forthcoming*), *Corpus-based Translation Studies: Research and Applications* (provisional title; St. Jerome Publishing). Paper delivered at Second IATIS Conference, Intervention in Translation, Interpreting and Intercultural Encounters, University of the Western Cape, South Africa, 12 - 14 July 2006.

Cencini, M. (2000). *Il Television Interpreting Corpus (TIC). Proposta di codifica conforme alle norme TEI per trascrizioni di eventi di interpretazione in televisione*. Unpublished dissertation. Advanced School for Translators and Interpreters (SSLMIT), University of Bologna at Forlì.

Cencini, M. (2002). "On the importance of an encoding standard for corpus-based interpreting studies". *inTRAlinea*, Special Issue: CULT2K.
< http://www.intralinea.it/specials/cult2k/eng_open.php?id=P107 >. Page consulted on date: 23.10.09.

Fumagalli, D. (1999-2000). *Alla ricerca dell'interpretese. Uno studio sull'interpretazione consecutiva attraverso la corpus linguistics*. Unpublished dissertation, Advanced School for Translators and Interpreters (SSLMIT), University of Trieste.

Gile, D. (1994). "Methodological aspects of Interpretation and Translation research", in Lambert S. & B. Moser-Mercer (eds.) (1994), *Bridging the Gap: Empirical Research in Simultaneous Interpretation*, Amsterdam / Philadelphia: John Benjamins.

Gile, D. (1997). "Interpretation research: realistic expectations", in Klaudy K. & J. Kohn (eds.) (1997), *Transferre Necesse Est. Proceedings of the 2nd International Conference on Current Trends in Studies of Translation and Interpreting, 5-7 September 1996*, Budapest, Hungary: Scholastica.

Gile, D. (2000). "Issues in interdisciplinary research into Conference Interpreting", in Englund Dimitrova B. & K. Hyltenstam (eds.) (2000), *Language Processing and Simultaneous Interpreting: Interdisciplinary Perspectives*, Amsterdam / Philadelphia: John Benjamins.

Halverson, S. (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.

Kalina, S. (1994). "Analyzing interpreters' performance: methods and problems", in Dollerup C. & A. Loddegaard (eds.) (1994), *Teaching Translation and Interpreting 2: Insights, Aims, Visions*, Amsterdam / Philadelphia: John Benjamins.

Kenny, D. (2001). *Lexis and Creativity in Translation: A Corpus-based Study*. Manchester / Northampton: St. Jerome.

Kruger, A. (ed.) (2004). *Corpus-based Translation Studies: Research and Applications*. Special Issue of *Language Matters. Studies in the Languages of Africa* 35/1.

Laviosa, S. (1998). "The Corpus-based approach: a new paradigm in Translation Studies", *Meta*, 43/4.

Laviosa, S. (2002). *Corpus-based Translation Studies: Theory, Findings, Application*. Amsterdam / New York: Rodopi.

Laviosa, S. (2004). "Corpus-based translation studies: where does it come from? Where is it going?", in Kruger, A. (ed.) (2004), *Corpus-based Translation Studies: Research and Applications*. Special Issue of *Language Matters. Studies in the Languages of Africa*, 35/1.

Malmkjær, K. (1998). "Love thy neighbour: will parallel corpora endear linguists to translators?", *Meta*, 43/4.
Meyer, B. (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 >. Page consulted on date: 23.10.09.

Monti, C., Bendazzoli, C., Sandrelli, A., Russo, M. (2005). "Studying directionality in simultaneous interpreting through an electronic corpus: EPIC (European Parliament Interpreting Corpus)", *Meta*, 50/4.
< http://www.erudit.org/revue/meta/2005/v50/n4/ >. Page consulted on date: 23.10.09.

Pöchhacker, F. (2008). "Inside the 'black box'. Can Interpreting Studies help the profession if access to real-life settings is denied?", *The Linguist*, 48/2.

Russo, M. (*forthcoming*). "Reflecting on interpreting practice: graduation theses based on the European Parliament Interpreting Corpus (EPIC)", in Zybatow, L. (ed.) (*forthcoming*), *Translation Studies: State of the Art and Perspectives*, series "Forum Translationswissenschaft" (vol 12).

Russo, M., Bendazzoli, C. & 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, International journal of interpretation and translation*, 4/1.

Sandrelli, A., Bendazzoli, C., Russo, M. (*forthcoming*). "European Parliament Interpreting Corpus (EPIC): methodological issues and preliminary results on lexical patterns in simultaneous interpreting", *Target. International Journal of Translation*, to appear in 2010 issue.

Setton, R. (*forthcoming*). "Corpus-based interpretation studies (CIS): reflections and prospects", to appear in Kruger, A. & K. Walmach (eds.) (*forthcoming*), *Corpus-based Translation Studies: Research and Applications* (provisional title; St. Jerome Publishing) (Paper delivered at Symposium on Corpus-based Translation Studies: Research and Applications, Pretoria, July 22-25, 2003).

Shlesinger, M. (1998). "Corpus-based Interpreting Studies as an offshoot of Corpus-based Translation Studies", *Meta*, 43/4.

Shlesinger, M. (2008). "Towards a definition of Interpretese. An intermodal, corpus-based study", in G. Hansen, Chesterman, A. & H. Gerzymisch-Arbogast (eds.) (2008), *Efforts and Models in Interpreting and Translation Research. A tribute to Daniel Gile*, Amsterdam / Philadelphia: John Benjamins.

Straniero Sergio, F. (2007). *Talkshow Interpreting. La mediazione linguistica nella conversazione spettacolo*. Trieste: Edizioni Universitarie Trieste.

Thompson, P. (2005). "Spoken language corpora", in Wynne, M. (ed.) (2005), *Developing Linguistic Corpora: A Guide to Good Practice*, Oxford: Oxbow Books.
< http://ahds.ac.uk/linguistic-corpora/ >. Page consulted on date: 23.10.09.

Tymoczko, M. (1998). "Computerized corpora and the future of Translation Studies", *Meta*, 43/4.

Vuorikoski, A.R. (2004). *A Voice of its Citizens or a Modern Tower of Babel?* Tampere: Tampere University Press.

Zanettin, F. (1998). "Bilingual comparable corpora and the training of translators", *Meta*, 43/4.

Zanettin, F., Bernardini S. & D. Stewart (eds.) 2003. *Corpora in Translator Education*. Manchester / Northampton: St Jerome.

Web references

EPIC website:
< http://sslmitdev-online.sslmit.unibo.it/corpora/corporaproject.php?path=E.P.I.C. >. Page consulted on date: 23.10.09.
EXMARaLDA project website:
< http://www.exmaralda.org >. Page consulted on date: 23.10.09.
Text Encoding Initiative TEI:
< http://www.tei-c.org >. Page consulted on date: 23.10.09.

[1] Although the present paper is the result of a joint effort, Claudio Bendazzoli can be identified as the author of §1, 2.2, 3, 3.1 and 4, while Annalisa Sandrelli wrote the Introduction, §2.1, 2.3 and 3.2.