

Abstract

The paper will describe the structure and the main features of the European Parliament Interpreting Corpus (EPIC) (Monti et al. 2005, Russo et al. 2012), a corpus containing original speeches delivered in Spanish, Italian and English at the European Parliament and their corresponding simultaneous interpretations in these three languages. EPIC is, therefore, a parallel and comparable trilingual corpus. It was developed, in 2004, by an interdisciplinary team of researchers in interpreting, teachers of interpreters, corpus linguists and IT technicians from the Department of Interpreting and Translation of the University of Bologna. This first machine-readable on-line interpreting corpus inspired other Interpreting and Translation scholars who developed EPIC-G at the University of Gent (Defrancq 2015) or the intermodal European Parliament Interpreting and Translation Corpus (EPTIC) (Bernardini et al. 2016). Given its specific characteristics, EPIC lends itself to both research and interpreter training applications. The paper will provide an overview of the corpus-based evidence provided by the studies of EPIC both by scholars and by students of interpreting for their graduation theses. It will also show its pedagogic applications both in interpreting classes and in student self-learning sessions. Finally, it will highlight its untapped potential in pedagogic settings by fostering an awareness of the interpreting process and the product for trainees but also in pragma-linguistic studies on time-constrained and semantically-predetermined performance by professional interpreters.

Résumé français

Cet article décrira la structure et les caractéristiques principales du European Parliament Interpreting Corpus (EPIC) (Monti et al. 2005, Russo et al. 2012), un corpus contenant des discours originaux prononcés au Parlement européen en espagnol, en italien et en anglais, ainsi que leurs interprétations correspondantes en traduction simultanée dans ces trois langues. Ainsi, EPIC constitue un corpus trilingue tant parallèle que comparable. Il a été développé, en 2004, par une équipe interdisciplinaire composée de chercheurs en interprétation, de formateurs d'interprètes, de linguistes de corpus et de techniciens IT du Département d'Interprétation et Traduction de l'Université de Bologne. Ce premier corpus d'interprétation électronique lisible en ligne a inspiré d'autres chercheurs en Interprétation et Traduction qui ont développé EPIC-G à l'Université de Gent (Defrancq 2015) ou le corpus intermodal European Parliament Interpreting and Translation Corpus (EPTIC) (Bernardini et al. 2016). Grâce à ses spécificités, EPIC peut être utilisé tant pour la recherche que pour la formation en interprétation. Cet article donnera un aperçu sur les résultats atteints par les études basées sur le corpus EPIC et menées par des chercheurs ou par des étudiants en interprétation dans le cadre de leurs mémoires de maîtrise. Il montrera ensuite ses applications pédagogiques, aussi bien dans les cours d'interprétation que dans les sessions autodidactiques des étudiants. Enfin, cet article mettra en évidence le potentiel inexploité d'EPIC, non seulement dans des contextes pédagogiques en favorisant la connaissance et la conscience du procès et du produit de l'interprétation chez les étudiants interprètes, mais aussi dans des études pragma-linguistiques sur la performance des interprètes professionnels en tant que prestation limitée par le temps disponible et sémantiquement prédéterminée.

Bernd Meyer, *The CoSi-Corpus: Qualitative and Quantitative Aspects of a Parallel Corpus of Consecutive and Simultaneous Interpreting*

Abstract

CoSi is the Corpus of Consecutive and Simultaneous Interpreting. It was designed to compare consecutive and simultaneous interpreting in a semi-experimental setting. Professional interpreters (German-Portuguese) were hired to work under authentic conditions for a German-Brazilian Non-Governmental Organisation of environmental activists. The speaker and the speech delivered were the same in the three different venues while the interpreters and the mode of interpretation were varied systematically. This resulted in four recordings: two of consecutive interpreting by a single interpreter at two different events, and two of simultaneous interpreting by teams of two interpreters at a third event. The corpus design thus enables systematic comparisons between interpreting modes, individual interpreters and different interpretations of the same speech event. The corpus was created with the help of EXMARaLDA, a state-of-the-art system for working with spoken language corpora. It is available to the research community through the Hamburg Centre for Language Corpora (HZSK). In this paper, we will present the corpus and discuss its different options for further qualitative and quantitative research.

Résumé français

CoSi, "Corpus of Consecutive and Simultaneous Interpreting", a pour objectif de comparer l'interprétation consécutive à l'interprétation simultanée dans un cadre semi-expérimental. Des interprètes professionnels (allemand-portugais) ont été engagés pour travailler, dans des conditions réelles, pour une organisation non gouvernementale germano-brésilienne d'activistes écologistes. Alors que l'intervenant et le discours tenu étaient les mêmes au cours des trois événements, les interprètes et le mode d'interprétation ont été systématiquement variés. Il en résulte au total quatre enregistrements, dont deux interprétations consécutives par un seul interprète lors de deux événements différents, et deux interprétations simultanées par des équipes de deux interprètes lors d'un troisième événement. La conception

du corpus permet donc des comparaisons systématiques entre les modes d'interprétation, entre les interprètes individuels, et entre différentes interprétations du même discours. Le corpus a été réalisé à l'aide d'EXMARaLDA, un système de traitement des corpus oraux prenant en compte toutes les avancées de la recherche, et il est diffusé au public scientifique par le Centre de Corpus Linguistique (HZSK) de Hambourg. Dans notre communication, nous présenterons le corpus et analyserons différentes options de recherche – qualitatives ou quantitatives – qui en résultent.

Bart Defrancq, *The European Parliament Discourse Community: Discourse Routines of MEPs and Interpreters Speaking the Same Language*

Abstract

There is quite significant evidence to support the idea that interpreters use discourse markers in a largely autonomous way and are only marginally driven by source text features. In simultaneous (Shlesinger & Ordan 2012; Defrancq et al. 2015; Defrancq 2016) as well as consecutive interpreting (in a dialogue context: Hale 2004, Mason 2008, Jacobsen 2010, Blakemore & Gallai 2014), interpreters have been shown to fail to render source text discourse markers, and add discourse markers that have no equivalent in the source text. This paper will concentrate on the addition of discourse markers.

Traditionnellement, l'ajout de marqueurs de discours en interprétation est expliqué comme une tentative de refléter plus adéquatement les intentions de l'orateur (le "modèle mental" de l'orateur, cf. Jacobsen 2002). Cependant, il a également été démontré que les marqueurs ajoutés reflètent des besoins de l'interprète, notamment des besoins liés à la face (Mason 2008; Defrancq et al. 2015).

Cette étude explore un nouveau facteur lié à l'ajout de marqueurs de discours, à savoir le développement de routines discursives dans un contexte de travail particulier. Il est en effet frappant de constater que dans les corpus consultés pour la présente étude (EPIC et EPICG), les fréquences des marqueurs de discours ajoutés par les interprètes sont presque identiques à celles des mêmes marqueurs utilisés par des orateurs de même langue travaillant dans le même contexte de travail, à savoir les membres du Parlement européen. Le développement de communautés de discours ou de communautés de pratiques n'a en soi rien de nouveau. Il a été montré que les traducteurs du PE constituent une communauté de discours à certains égards (Calzada Perez 2012). Or le fait que les interprètes et les membres du PE appartiennent à la même communauté et que cette communauté se reflète à travers l'usage de marqueurs de discours semble nouveau.

Résumé français

L'autonomie de l'interprète en termes d'utilisation des marqueurs de discours est bien documentée. Le texte source ne semble que partiellement déterminer leur utilisation. En interprétation simultanée (Shlesinger & Ordan 2012; Defrancq et al. 2015; Defrancq 2016) et en interprétation consécutive (dans un contexte de liaison: Hale 2004, Mason 2008, Jacobsen 2010, Blakemore & Gallai 2014), il a été démontré que les interprètes ne rendent pas les marqueurs du discours du texte source et ajoutent des marqueurs qui n'ont pas d'équivalent dans le texte source. La présente étude se concentrera sur les ajouts.

Traditionnellement, l'ajout de marqueurs de discours est expliqué comme une tentative de rendre plus adéquatement les intentions de l'orateur (le "modèle mental" de l'orateur, cf. Jacobsen 2002). Cependant, il a également été démontré que les marqueurs ajoutés reflètent des besoins de l'interprète, notamment des besoins liés à la face (Mason 2008; Defrancq et al. 2015).

Cette étude propose une exploration d'un troisième facteur dans l'ajout de marqueurs de discours, à savoir le développement de routines discursives dans un contexte de travail particulier. Il est en effet frappant de constater que dans les corpus consultés pour la présente étude (EPIC et EPICG), les fréquences des marqueurs de discours ajoutés par les interprètes sont presque identiques à celles des mêmes marqueurs utilisés par des orateurs de même langue travaillant dans le même contexte de travail, à savoir les membres du Parlement européen. Le développement de communautés de discours ou de communautés de pratiques n'a en soi rien de nouveau. Il a été montré que les traducteurs du PE constituent une communauté de discours à certains égards (Calzada Perez 2012). Or le fait que les interprètes et les membres du PE appartiennent à la même communauté et que cette communauté se reflète à travers l'usage de marqueurs de discours semble nouveau.

Camille Collard, Bart Defrancq et Heike Prybyl, *Position of the Verb in Simultaneous Interpreting: a Corpus-Based Comparative Study*

Abstract

In Dutch and German subordinate clauses, the verb is generally placed after the clausal constituents (Subject-Object-Verb structure) thereby creating a middle field (also called a verbal brace) with the subordinate conjunction. This field makes interpreting from SOV into SVO languages particularly challenging as it requires further processing and feats of memory. It often requires the interpreter to use specific strategies (e.g. anticipation) (Lederer, 1981; Van Besien 1999; Setton, 1999; Liontjou, 2016).

However, few studies have tackled this issue from the point of view of interpreting into SOV languages. Producing SOV structures requires some specific cognitive effort as, for instance, subject properties need to be kept in mind in order to ensure the correct subject-verb agreement across a span of 10 or 20 words. Speakers therefore often opt for a strategy called extraposition placing specific elements after the verb in order to shorten the brace (Bevilacqua, 2009; Hawkins, 1994). It seems that Dutch speakers do so to a larger extent than German speakers (Haeseryn, 1990). Given the additional cognitive load generated by the interpreting process (Gile, 1999), it may be assumed that interpreters will shorten the verbal brace to a larger extent than speakers.

The present study is based on a corpus of original speeches and interpreting at the European Parliament and compares middle field lengths in Dutch and German interpreted and non-interpreted versions. Preliminary results from sets of 1273 German and 1826 Dutch subordinate clauses confirm that interpreters shorten the middle field more than speakers, and that Dutch interpreters shorten the middle field more frequently than German interpreters.

Résumé français

En allemand et en néerlandais, le verbe se place après tous les éléments d'une proposition subordonnée (structure sujet-objet-verbe) et forme un champ médial (également appelé parenthèse verbale) avec la conjonction de subordination. Ce champ rend l'interprétation à partir d'une langue SOV vers une langue SVO particulièrement difficile puisqu'il augmente l'effort de traitement et de mémoire et exige de l'interprète qu'il recoure à des stratégies (l'anticipation par exemple) (Lederer, 1981 ; Van Besien 1999; Setton, 1999; Liontou, 2016).

Toutefois peu d'études se penchent sur la problématique du point de vue des interprètes travaillant vers une langue SOV. Les structures SOV demandent un effort cognitif particulier pour le locuteur puisque, par exemple, les propriétés du sujet doivent être maintenues en mémoire pendant 10 à 20 mots pour permettre l'accord du verbe. Les locuteurs ont souvent recours à l'extraposition, une stratégie qui consiste à placer certains éléments après le verbe pour raccourcir la parenthèse (Bevilacqua, 2009 ; Hawkins, 1994). Les néerlandophones y auraient d'ailleurs recours plus fréquemment que les germanophones (Haeseryn, 1990). L'interprétation générant un effort cognitif supplémentaire (Gile, 1999), il est probable que les interprètes néerlandophones et germanophones raccourcissent la parenthèse verbale plus fréquemment que les orateurs.

Dans cette étude, un corpus de discours originaux et d'interprétations au Parlement européen est analysé afin de comparer la longueur des parenthèses verbales dans des subordonnées allemandes et néerlandaises. Les premiers résultats pour 1273 subordonnées allemandes et 1826 néerlandaises confirment la tendance des interprètes à raccourcir le champ plus que les orateurs, et que les interprètes néerlandais le font plus que les interprètes allemands.

Sabine Braun, *From Quantity to Quality: Using Corpus-Based Methods to Analyse Interpreting Performance*

Abstract

This presentation will discuss the development and use of a corpus of interpreter-mediated police interviews to investigate differences between on-site and remote interpreting in legal settings. The study was prompted by the increasing use of remote interpreting in legal proceedings worldwide and the questions of viability and reliability that this method of interpreting raises. This required a comparative approach and led to the development of a corpus consisting of (simulated) bilingual police interviews and their renditions by multiple interpreters using on-site and remote interpreting. The corpus includes video recordings along with time-aligned transcripts and has been annotated with regard to interpreting problems. After briefly outlining the corpus design, the presentation will focus on different ways of exploiting the corpus to analyse and compare interpreting performance in onsite and remote interpreting. It will be argued that quantitative corpus-based methods can be fruitfully applied to dialogue interpreting corpora, especially when combined with qualitative approaches. For example, it will be shown how a frequency analysis of different types of interpreting problems was used to identify important differences between on-site and remote interpreting. It will then be shown how a complementary qualitative analysis was used to corroborate, enrich and – equally important – challenge the quantitative findings; to identify mitigation strategies employed by the interpreters; and ultimately to paint a more comprehensive picture of the interpreters' performance than a quantitative analysis alone would have revealed. The presentation will conclude with a reflection on the main benefits and limitations of the above-mentioned approach.

Résumé français

Cette présentation traitera de la mise au point et de l'utilisation d'un corpus d'entretiens policiers interprétés afin d'étudier les différences entre l'interprétation sur place et l'interprétation à distance dans les milieux juridiques. L'étude a été motivée par l'utilisation croissante de l'interprétation à distance dans les procédures judiciaires et les questions de la viabilité de cette méthode d'interprétation. Cela nécessitait une approche comparative et a conduit à l'élaboration d'un corpus composé d'entretiens policiers bilingues (simulés) et de leurs interprétations par de multiples interprètes utilisant l'interprétation sur place et à distance. Le corpus comprend des enregistrements vidéo accompagnés de transcriptions et a été annoté en ce qui concerne l'interprétation des problèmes. Après avoir brièvement esquissé la conception du corpus, la présentation se concentrera sur les différentes façons d'exploiter le corpus pour analyser et comparer les performances dans l'interprétation sur place et à distance. On soutiendra que les

méthodes quantitatives peuvent être appliquées avec succès aux corpus d'interprétation du dialogue, surtout lorsqu'elles sont combinées avec des approches qualitatives. A titre d'exemple, on montrera comment une analyse en fréquence de différents types de problèmes d'interprétation a été utilisée pour identifier des différences entre l'interprétation sur place et à distance. On montrera alors comment une analyse qualitative complémentaire a été utilisée pour corroborer, enrichir et – tout aussi important – contester les résultats quantitatifs; identifier les stratégies d'atténuation employées par les interprètes; et finalement, présenter une analyse plus complète de la performance des interprètes qu'une seule analyse quantitative aurait révélé. La présentation se terminera par une réflexion sur les principaux avantages et limites de l'approche signalée.

Natacha Niemants et Laura Gavioli, *Some Notes on Electronic Collections of Dialogue Interpreting Data: Corpora or Archives?*

Abstract

Corpora are increasingly used in translation studies and practice in order to better understand equivalent lexical patterns and to see how items of phraseology are used in comparable text-types and genres (e.g. Zanettin 2012). The use of corpora for glossary making (e.g. Williams 2008), moreover, may be considered a useful resource even for conference interpreting practice, and attempts at modelling corpora have been carried out (Cencini & Aston 2002) for research into interpreting.

As for dialogue interpreting (DI), the use(fulness) of corpora is controversial. On the one hand, DI research is increasingly moving towards the study of authentic interaction with a consequent growth of recorded and transcribed material which might well be explored as corpora. On the other hand, DI interest in "interaction" rather than "text" (Wadensjö's 1998) has made traditional methods of corpus linguistics not so attractive for DI studies.

In this paper, we will report on our experience with quite a large DI collection of data (the AIM corpus, 500 encounters, over 100 hours recording), which we are trying to implement for corpus use. In particular, we have tested the ELAN software on a sub-set of 40 encounters and we have aligned the transcript with the audio version. So far, our experiments have taken two directions which we will illustrate in our presentation: 1) exploring potentially interesting search items for the analysis of interaction structures; 2) extracting audio-aligned segments for use in specific training or learning activities.

Résumé français

Les corpus sont de plus en plus exploités dans les études et la pratique de la traduction, pour mieux comprendre les patrons lexicaux équivalents et observer l'utilisation des unités phraséologiques dans des types/genres de textes comparables (Zanettin 2012). En outre, l'emploi des corpus pour la création de glossaires (Williams 2008) pourrait être considéré utile pour la pratique de l'interprétation de conférence et il existe des tentatives de modélisation des corpus pour la recherche (Cencini & Aston 2002).

Pour ce qui est de l'Interprétation de Dialogue (ID), l'utilisation et l'utilité des corpus sont autant de sujets controversés. D'une part, la recherche est de plus en plus orientée vers l'étude de l'interaction authentique, ce qui entraîne une augmentation des données enregistrées et transcrites qui pourraient être explorées comme corpus ; de l'autre, l'intérêt de l'ID envers l'« interaction » plutôt que le « texte » (Wadensjö's 1998) diminue l'attrait des méthodes traditionnelles de la linguistique de corpus.

Dans cet essai, nous partagerons notre expérience avec une collection de données d'ID assez importante (corpus AIM, 500 rencontres, plus de 100h d'enregistrement), que nous cherchons maintenant à outiller. Nous avons testé le logiciel ELAN sur un sous-ensemble de 40 rencontres, dont la transcription est alignée avec l'audio, et nous avons jusqu'à présent entamé deux pistes de recherche que nous allons illustrer : 1. Explorer des requêtes d'unités lexicales potentiellement intéressantes pour l'analyse des structures de l'interaction ; 2. Extraire des segments alignés à l'audio que l'on puisse utiliser dans des activités d'enseignement et d'apprentissage

Caterina Falbo, *La collecte de corpus d'interprétation : un défi permanent*

Résumé français

Depuis l'article programmatique de Shlesinger (1998), beaucoup de corpus d'interprétation ont vu le jour. Il s'agit de corpus constitués d'interprétations effectuées dans un contexte de conférence (comme par exemple EPIC, DIRSI, FOOTIE), dans le domaine médical (DiK-corpus, CoSI), ou encore dans une situation institutionnelle comme la télévision (CorIT) (cf. Straniero Sergio/Falbo 2012). Malgré les différences portant sur le contexte situationnel, ils présentent des traits communs au niveau de leur dimension, qui est en général limitée, et de leur exploitation, qui est quant à elle conditionnée par la présence ou l'absence de l'alignement des transcriptions avec l'audio, de même que par le niveau d'étiquetage choisi. Nous nous proposons de discuter ces deux traits, qui marquent la plupart des corpus d'interprétation, en illustrant brièvement les caractéristiques de deux corpus collectés à l'Université de Trieste tels que CorIT et CorILex (Corpus d'Interprétation judiciaire) ainsi que le travail mené sur eux. Cette approche comparative/contrastive nous permettra de mettre au jour les aspects théoriques et pratiques qui ont eu une influence

sur les différentes étapes de leur constitution et sur les possibilités d'exploitation, comme par exemple la modalité d'interprétation (simultanée, consécutive...), le type d'interprétation (interprétation de dialogue, interprétation de conférence...), et le type de transcription (qui détermine la réalisation de spoken ou speech corpora).

Abstract

Since Shlesinger's (1998) programmatic article was published, a steadily increasing number of corpora for interpreting have been developed. They range from interpreted discourses delivered in conference interpreting situations (e.g. EPIC, DIRSI, FOOTIE), healthcare settings (e.g. DiK-corpus, CoSI, HCIQ.1415), or other institutional contexts, such as television (CorIT) (cf. Straniero Sergio/Falbo 2012). Notwithstanding the differences concerning the setting and the communicative context, all these corpora share a number of traits in terms of their scope, which is generally limited, as well as their use, which is heavily dependent on the presence or absence of transcript-audio alignment and the level of tagging. Our aim is to discuss these two aspects, which are common to the majority of the corpora for interpreting, drawing upon the characteristics of two corpora gathered at the University of Trieste, namely CorIT and CorILex (Legal Interpreting Corpus), and the analyses carried out on them. This comparative/contrastive approach should allow us to highlight aspects, both theoretical and practical, which can influence different phases of corpus development and their possible use, for example, *interpreting mode* (simultaneous or consecutive interpreting...), *type* (dialogue or conference interpreting), and *transcription method* (which has a direct impact on the very nature of spoken or speech corpora).