Applications of Simultaneous-Interpreting Corpora
in Terminology Research

Tamara Cabrera Castro
University of North Carolina at Chapel Hill, USA

ABSTRACT

Simultaneous-interpreting corpora are unique research tools that have been created with the purpose of exploring interpreting and interpreting-related phenomena. Like translation corpora, they consist of a set of source speeches in a source language, and a set of target speeches in one or more target languages. Unlike translation corpora however, simultaneous-interpreting corpora reflect the unique features of simultaneous interpreting in terms of orality, simultaneity and contextualization. Furthermore, because the setting in which simultaneous interpreting develops is usually a specialized conference or an international organization, another feature of these corpora is their specialization. To date, simultaneous-interpreting corpora have been used to explore simultaneous interpretations as samples of oral textual productions, and to delve into aspects that concern the way in which interpreters interpret. Therefore, their use has been put to the service of Interpreting Studies almost exclusively. However, Interpreting Studies may not be the only research field to which simultaneous-interpreting corpora can make a contribution. This article will argue that these tools have research applications beyond the disciplinary field of Interpreting Studies, more specifically in Terminology.

KEYWORDS: corpus-based Interpreting Studies, machine-readable corpora, simultaneous interpreting, terminology

Introduction

Simultaneous-interpreting corpora are unique research tools designed to further advance research on simultaneous interpreting, a complex oral activity that involves the immediate transfer of

knowledge from a source language into a target language. Simultaneous-interpreting corpora are similar to translation corpora in that they consist of a set of source speeches in a source language, and a set of target speeches in one or more target languages. Because these corpora have been compiled with the purpose of conducting research on simultaneous interpreting, they also have a series of characteristics that set them apart from other corpora. For instance, simultaneous-interpreting corpora capture the way that knowledge is grasped and conveyed in this mode: orally, contextualized and in meaning units or chunks, as it is being produced by a speaker. Furthermore, because simultaneous interpreting is the preferred modality in international institutions and conference settings (Pöchhacker 2004: 16), and because these settings usually involve the transfer of knowledge among specialists in a given domain of expertise, these corpora are oftentimes further enriched by a specialized dimension.

Currently, simultaneous-interpreting corpora are being used as research tools to observe the unique features of interpreted speeches, either in isolation or as compared to other forms of textual productions (e.g., translations), and also as a means to explore topics related to interpreting performance, such as the effect of experience and expertise on the use of strategies (Monti et al. 2005). They are also being used as key elements for constructing real-time simultaneous translation systems (Shimizu et al. 2014). With the exception of the use of simultaneous-interpreting corpora to build automatic translation systems, to date they have been used to study interpreting-related phenomena almost exclusively. However, their research applications may not be limited to these two areas.

Because simultaneous-interpreting corpora are often enriched by a specialized dimension, they capture the specialization of the source speeches that are being interpreted and the way that interpreters understand specialized meaning units in a source language and render them in the target language. Likewise, because specialized conferences are a fertile ground for discussing innovative ideas, and for introducing new discoveries and technological advances, simultaneous-interpreting corpora can become springboards for the exploration of interpreter-induced neology and neology-related phenomena. However, to date there are little or no empirical studies that explore these topics, or that put forward theoretical proposals that suggest ways in which simultaneous-interpreting corpora could be used beyond Interpreting Studies.

In response to this research gap, the goal of this article is to initiate a discussion on the feasibility of using machine-readable simultaneous-interpreting corpora in the field of Terminology. To achieve this goal, this article will provide a general overview of simultaneous-interpreting corpora, summarize trajectories of research in Interpreting Studies, and further reflect on feasible applications of these corpora in Terminology research. Because different modes of interpreting have their own specific characteristics, for consistency purposes, this paper will focus exclusively on machine-readable, oral, simultaneous-interpreting corpora.

Simultaneous-interpreting corpora

If a corpus is a collection of naturally occurring examples of language, a definition of (oral) simultaneous-interpreting corpora would then be that of collections of examples of naturally-occurring simultaneously interpreted (oral) speeches, produced online and onsite, by the interpreters working in their respective booths.

The idea of exploring samples of authentic conference interpretations from real events recorded and/or partially transcribed and analyzed can be traced back to the first fifty years of interpreting research (Setton 2011: 38) and especially to the work of Danica Seleskovitch and the Paris School (Seleskovitch 1975; Lederer 1981). Most corpora from this period however were based on single conferences, often too small to draw conclusions that could be extrapolated to other interpreting scenarios, or not adapted to conduct statistical analysis (for an extensive review see Setton 2011: 38-44).

It was only after the year 2000 that large, machine-readable simultaneous-interpreting corpora were developed, and Corpus Linguistics methods and tools for data analysis were implemented in the study of oral, simultaneously translated output. The development of these corpora was to a large extent influenced by the empirical turn in Translation Studies prompted by the work of Baker (1995, 1996) and Laviosa (1998, 2002), and the influential article published by Shlesinger in 1998 titled *Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies*. Since then, interpreting scholars have witnessed the creation of large machine-readable corpora (e.g., EPIC, ECIS, ARCHINT, to name but a few), the development of Corpus-based

Interpreting Studies; the increase of specialized publications (Bendazzoli 2010ab; Bendazzoli and Sandrelli 2005, 2009; Sandrelli et al. 2010; Iglesias Fernández 2010, and others); the launch of specialized monographs (e.g., Straniero and Falbo 2012); the organization of thematic workshops (e.g., *First Forlí International Workshop*, 2015) and the consolidation of the use of corpora and Corpus Linguistics methods and techniques to explore interpreting and interpreting-related phenomena.

Interpreting corpora available to date vary to a certain extent in size and language combination, and are similar in the type of material they include (audiovisual recordings and transcripts of source and target speeches) and in the specialized communicative situations they portray. One of the first large machine-readable simultaneous-interpreting corpora to date is the *European Parliament Interpreting* or EPIC corpus, which was developed by a group of researchers at the University of Bologna (Italy) (Bendazzoli and Sandrelli 2005, 2009). The corpus is formed by speeches in Italian, English and Spanish and their corresponding simultaneous interpretations, and includes the plenary debates of the European Parliament recorded from the EbS news channel (Sandrelli et al. 2010). Along the same lines, ECIS (*Quality Evaluation in Simultaneous Interpreting*) is another large machine-readable specialized simultaneous-interpreting corpora formed by European Parliament debates and their simultaneous interpretations. It was developed by researchers at the University of Granada, Spain. The speeches are in English, Spanish and French and, like the EPIC corpus, the material was also recorded from the Ebs news channel (Collados Aís 2009).

Reflecting the reality of highly specialized thematic conferences, DIRSI *(Directionality in Simultaneous Interpreting Corpus)* is a bilingual corpus of simultaneously interpreted medical conferences in English and Italian. The corpus was developed as part of a doctoral dissertation at the University of Bologna (Bendazzoli 2010b). It is formed by texts in Italian and English and includes various speech events ranging from papers to lectures related to the institutional frame of each conference (ibid). FOOTIE *(Football in Europe Corpus)* is a multilingual corpus of simultaneously interpreted football press conferences that were held during the 2008 European football championship. It was created in the Faculty for Interpreting and Translation at LUSPIO University in Rome, and is formed by individual speech events in Italian and English (Sandrelli

2012). The *Interpreting Architecture* or ARCHINT corpus is a parallel, specialized corpus formed by the video-recordings and transcripts of eleven specialized conferences on Architecture and their simultaneous interpretations into Spanish (Cabrera 2016). The speeches are highly technical, and the corpus was developed at the University of Granada as part of a doctoral dissertation.

Taking advantage of the benefits of new technologies, online news distribution and virtual platforms, the CLAIR (*Center for Integrated Acoustic Information Research*) corpus is an experimental, parallel and bilingual corpus developed by researchers at the University of Nagoya in Japan. It contains speeches in English and Japanese, and includes simultaneous interpretations of recorded material from several talk and news channels (TED, CNN, CSJ and NHK) performed by interpreters with varying levels of expertise (Shimizu et al. 2014).

These corpora serve to illustrate the diversity and richness of readily available interpreting corpora. With the exception of the CLAIR corpus, in which the simultaneous interpretations were produced in a controlled setting, these corpora are formed by speeches that have been live-recorded in a conference setting and later transcribed. They share the features of being parallel and multilingual, with each corpus providing access to textual variables (e.g., lexical information, type-token ratios, collocations, etc.), and to paralinguistic and contextual features (e.g., setting, participants, speech rate, audiovisual materials used by the speakers, technical problems, contextual features, etc.). The corpora are available for research purposes on internet platforms (EPIC, CLAIR) or by contacting the principal researcher. The next section will provide an overview of specific applications of these corpora in interpreting research.

Trajectories of Current Research

The development of simultaneous-interpreting corpora has facilitated the study of product-related aspects of simultaneous interpreting (e.g., orality features, effect, source-target correspondence, etc.) and revolutionized the way simultaneous interpreting research is conducted. Such is the extent of this influence that some researchers have come to believe that it is corpus-based lexical approaches, with a strong semantic and pragmatic component, that better

meet the needs of Interpreting Studies (Pöchhacker 1993; Diriker 2004; García de Quesada 2011). In addition to specific applications of these corpora to the study of product-related aspects of simultaneous interpreting, there are at least five other major trends that result from the use of these tools in interpreting research (see also Setton 2011: 45-47).

Trajectory 1: Interpreter's Speech

The first research trajectory addresses questions such as "Are there, and if so what are the unique features of interpreters' output?", and the goal is to learn more about simultaneous interpreting as a textual production. Mainstream topics are the identification of patterns specific to interpreted speeches (as pieces of oral discourse) regardless of their source language, and in relation to comparable speeches in the same language. Also, the identification of patterns which stem from interpreted (oral or signed) speeches in a given language, regardless of their source languages, and their comparison with translated (written) products.

An illustrative example of this trend is a corpus-based study developed by Shlesinger (2008) in which she compared the output of different translation modes (written translation and simultaneous interpreting), using the same source language speech and the same group of subjects. More specifically, the researcher developed a computerized analysis of the interpreted outputs of six professional translator-interpreters, interpreting and later translating the same (non-domain) specific source text from their second language (English) into their first language (Hebrew). The researcher found a set of marked differences between them in terms of richness (as defined by the type-token ratio) and a set of lexico-grammatical features, and provided evidence supporting the fact that interpreted discourse displays features of *interpretese* which set it apart from other types of discourse (Shlesinger 2008: 250; Setton 2011: 45-46).

Trajectory 2: Language and Directionality

The second research trajectory covers questions such as "Are there any language-specific and direction-specific features of the interpreted output?" A key goal of this trajectory is to test hypotheses about the strategies involved in specific language pairs. Mainstream topics are the identification of language-specific and direction-specific features of interpreters' output, the observation of interactions between translational patterns and personal variables such as gender,

experience, language background and so forth, that is, beyond language-pair specific factors. Some work in this respect has been developed by the Forli group (Monti et al. 2005; Russo et al. 2006). The work developed so far has focused on topics such as interpreters' strategies, recurrent lexical patterns and morpho-syntactical structures across all the possible language combinations and directions, as well as on identifying and analyzing interpreting strategies when interpreting from a Germanic language into a Romance language and between more than one Romance language.

Trajectory 3: Process

The third research trajectory explores process-related aspects of simultaneous interpreting and addresses questions such as "Can any insight be gained about the human process in simultaneous interpreting by exploring corpus data?" As in any other process-oriented research path, the goal is to learn more about the workings behind the unusual task of simultaneously understanding a speech in one language and producing another speech in a different language. Mainstream research topics in this trajectory focus on the factors conducive to (or that allow for) anticipation, strategies, segmentation, etc. Intensive work has been developed by the Nagoya group on the CLAIR corpus as part of their ongoing research towards analyzing differences in interpretation styles and segmentation patterns (Shimizu et al. 2014; Setton 2011: 44-45). Also in tune with this process-oriented research trajectory is the work developed on the ARCHINT corpus and its future applications (Cabrera 2016). To date, the corpus has provided the foundations for a corpus-based study on terminological accuracy with further applications in terminology management and advance preparation for simultaneous interpreting. The study focused on the analysis of a sample of terms selected from simultaneous interpretations into Spanish. The objective was to identify whether the interpreter was using terminologies accurately and, if not, to explore the causes for terminological inaccuracies (Cabrera 2015). The analysis followed the premises of Frame-based Terminology (Faber 2012), a cognitive approach to the descriptive study of terms and the creation of knowledge bases and terminological resources for translators and interpreters. The study highlighted examples of terminological inaccuracies affecting the knowledge representations of lexical items and suggested at least four causes for such inaccuracies (Cabrera and Faber, in progress).

Trajectory 4: Quality/Performance

The fourth research trajectory deals with quality-related aspects of simultaneous interpreting, and covers questions such as "Are there, and if so what are the factors affecting quality and performance?" Studies in line with this research trajectory are mostly concerned with objectively identifying the intrinsic and extrinsic criteria that lead towards a feasible definition of quality for simultaneous interpreting, and to explore the specific weight of each criteria in assessment. A representative example of quality research is the work developed by the ECIS research team in Granada (Collados Aís 1998; Collados Aís et al. 2007: 201; Iglesias Fernández 2010, Pazos et al. 2010). As inferred from their earliest publications, the corpus is a later addition to their studies, and is being used as an observational tool for their experimental research (Setton 2011).

Trajectory 5: Automatic Interpreting Systems

The fifth research trajectory targets the mechanization of simultaneous interpreting and addresses questions such as "Can simultaneous-interpreting corpora provide the foundations for the construction of real-time simultaneous interpreting systems?" Studies in line with this research trajectory are mostly concerned with improving the translation performance of automatic speech translation by observing real-time performances, and with constructing simultaneous interpretation databases that provide the foundations for constructing these systems (Shimizu et al. 2014). The CLAIR corpus has been created with this research trajectory in mind.

Trajectory 6: Pedagogical Applications

The sixth research trajectory focuses on identifying the pedagogical uses of simultaneous-interpreting corpora in interpreter training, their benefits and limitations. Because interpreting corpora typically contain video clips and transcripts of the source and target speeches, they can be used as a self-learning tool by students not only to familiarize themselves with the terminologies of specific domains of expertise, but also for building the required term bank of a specific field of expertise or for the actual interpreter training. Furthermore, they can be implemented in a CAIT (Computer Assisted Interpreter Training) tool, which can in turn be integrated into a traditional interpreting course and enhance classroom-based training (Sandrelli 2015: 117). Illustrative examples of CAIT tools available to date are Black Box (Melissi Ltd), a

fully-fledged commercial software developed by Sandrelli (2003a, 2003b, 2007; 2015) and the European Union's Speech Repository, an online version of a speech bank developed by Directorate-General for Interpretation, the European Commission's interpreting service. Interpreting corpora as a didactic resource are useful in designing activities which simulate the translator/interpreter practice in professional contexts and reflect the multiple challenges of real-life interpreting scenarios (e.g., specialized speeches, technological problems, speaker's orality features, speech rhythm, accent and so forth). By engaging in activities that are based on authentic audiovisual content, interpreting learners are able to gain a better overall impression of a real interpreting event. Such activities expose learners to actual problems and strategies, and encourage them to come up with their own solutions in the safe environment of the classroom.

Opportunities beyond Interpreting Studies

Machine-readable simultaneous-interpreting corpora have made their contribution to Interpreting Studies, and now it remains to be seen if these tools have specific applications in fields such as Terminology. Since the 1990s, Terminology has consistently relied on the use of corpora to draw insights about the behavior of terms in large running collections of text of various types, fields and levels of specialization. The use of corpora in terminology research has had positive and fruitful research outcomes, as evidenced by the number of publications specific to this topic since the 1990s (Bowker 1998; Bowker and Pearson 2002; Meyer 2001; Meyer and Mackintosh 1996; Pearson 1998, 1999, Corpas Pastor and Seghiri 2016, etc.). The questions that arise are: Could simultaneous-interpreting corpora be useful to terminologists? Could they have any application at all in terminology research?

The following subsections will introduce a first and modest attempt to illustrate how simultaneous-interpreting corpora could be used for research purposes in the specific areas of terminology research.

Oral Discourse

Specialized simultaneous-interpreting corpora are oral corpora and, as such, they allow for the study of oral discourse, regardless of the level of specialization of the discourse. To date, oral discourse, and oral communication in general is a much less explored area than written

discourse, or written communication. To a large extent, this is due to the difficulties associated with accessing and compiling oral corpora (Gumperz et al. 1993). Recordings of original speeches are difficult to obtain because consent may be denied by the parties involved for reasons of confidentiality (Pöchhacker 2008), or due to a lack of understanding of the purpose of research (Bendazzoli and Sandrelli 2005). Furthermore, if the recordings are obtained, they might not be adequately representative and/or might not be of the desired quality. Even if these are of the desired sound and video quality, the oral material still needs to be transcribed, which can be a difficult and time-consuming task. These are just some of the difficulties involved in compiling an oral corpus. However, because simultaneous-interpreting corpora are readily compiled, easily accessible, and typically include transcripts of the oral speeches, they represent a valuable and time-saving source of data for oral discourse researchers.

These corpora can be used to explore textual aspects such as word frequency, discourse patterns, co-occurrences, lexical usages, type-token ratios and so forth. They can also be used in contrastive studies between oral and written discourses to unveil unique features of oral (specialized) discourse and lead to a better understanding of such features. In addition, these corpora may have pedagogical applications in areas such as Languages for Specific Purposes (LSP), as a better understanding of specialized discourses in oral form could enhance language training programs.

Specialized Communication

Simultaneous-interpreting corpora portray two types of communicative event, namely, those in which the source speaker addresses the audience that can understand the source speech directly, and those in which the interpreter, in his or her role as a second speaker, addresses the audience that cannot understand the source speech directly because of the language barrier. Although these two communicative events are different in nature (e.g., different languages, produced under different conditions), they also share many features in common. For instance, they have been produced and recorded at the same time, in the same setting and involve the same topic, and the same participants. Thus, they can provide the foundations to reception studies which may address questions such as: "How is a simultaneous interpreting of a specialized speech perceived by different user profiles (i.e. expert and non-expert users)? Are there any differences? If so,

what are these differences, and why do they arise? Are there effective mechanisms that contribute towards improving mediated specialized communication in more than one language?" In addition, because simultaneous-interpreting corpora—especially those where video recordings are available—provide access to the contextual factors in which the source and target speeches were produced, as well as to the support material that was used in the conference in question (e.g., Power Points, graphs, charts, images and so forth), they can also provide the foundations to studies exploring the role of visual elements in specialized communication. Overall, this type of study may contribute to research on the role of prior knowledge in language comprehension, and report on the role of support material when it comes to the conceptualization of a specialized speech.

Terminological Variation

Because interpreters are not usually specialists in the domain of the conference in which they are interpreting, they use the domain-specific terminology according to their own understanding of the specialized concepts, purpose and ideas that the source speaker conveys or wishes to convey. According to Faber (2012: 234), the success of a specialized communication act is, to a great extent, conditioned by the terms selected by the sender and by the receivers' ability to make the right inferences and correctly interpret this information. If simultaneous interpreters share the same knowledge of the topic as the audience for whom they interpret, then they can make the right inferences and correctly interpret this information. However, if they do not share the same knowledge, their abilities to perform these tasks are limited to their own understanding of the content that they need to convey. This reality is a breeding ground for the proliferation of variation caused by different communicative registers (Freixa 2006: 52). In fact, findings from the ARCHINT corpus showed that the interpreters (non-specialists), when transferring specialized knowledge associated to the domain of Architecture, labeled specialized concepts such as 'roof' or 'floor' by using words that are frequently used in a day-to-day conversation, e.g., *tejado* (roof, tile roof), *suelo* (floor, ground), *piso* (floor, flat). The study involved different expert subjects, questionnaires and assessment videos showing fragments of one of the simultaneous interpretations included in the corpus. In one of the questionnaires, subjects with expertise on Architecture reported that the use of those terms contravened their expectations. In fact, they were expecting the use of the terms *cubierta* (roof) and *forjado* (floor) instead (Cabrera

2015). This led to the identification of instances of variation in the interpreters' rendering of the source speech content. The results of this study illustrate how simultaneous-interpreting corpora can be used to explore functional variation, a core research topic in Terminology, and further deepen our understanding of this phenomenon.

Specialized Language Processing

Simultaneous-interpreting corpora of specialized texts reflect, on the one hand, the way that experts convey specialized knowledge orally through the source speeches. On the other hand, they also reflect the way in which interpreters understand meaning units in a source language, and search for equivalents in the target language. Because these are intrinsic aspects to any simultaneous-interpreting corpora, they can also be used as an observational complement to studies on topics such as entrenchment, collocations and even word choice.

According to Langacker (1999: 93), entrenchment is a phenomenon that has to do with the formation of a habit or routine that is easily elicited and reliably executed. In his view, when a complex language structure is susceptible to being manipulated as a 'pre-packaged' assembly, that is, it no longer requires conscious attention to its parts or their arrangement, then it has the status of an entrenched unit. Schmid (2007: 118) agrees that the activation of familiar concepts (e.g., dog, poodle) is highly automated because they are deeply entrenched in our memory. In fact, when language users are faced with unfamiliar concepts (e.g., specialized denominations that fall beyond their field of expertise), they are likely to need more time to classify the unfamiliar concept. Interpreters at work are in a constant quest for equivalents, that is, they are constantly faced with concepts that need to be recognized and labeled. Obviously, factors such as the influence of the source speech on the target language will, very likely, need to be accounted for, but a priori, simultaneous-interpreting corpora seem to be a rich source of information when it comes to observing this phenomenon, or when it is used as an observational complement to studies exploring entrenched terminological usages across linguistic and cultural boundaries. Such studies could further deepen our knowledge of the mechanics of specialized languages in contact.

Another contribution of simultaneous-interpreting corpora to specialized language processing concerns collocations. It is well known that, more often than not, specialized terms come in the form of phrasemes and complex nominal forms, which have specific syntactic structures. For example, in the domain of Architecture, the term 'retaining wall' is used to refer to a vertical structure which is built to prevent the earth behind it from moving by resisting the lateral pressure of soil, while 'reinforced concrete' is used to refer to concrete in which steel bars are embedded in such a manner that the two materials come together when resisting a force. To experts of Architecture, these collocations come almost naturally. But what about users of specialized terminology who are not experts on the topic (e.g., interpreters)? Do these collocations also come naturally to them in the target language? Specific examples taken from ARCHINT suggest that this does not seem to be the case. Typically, in a Peninsular Spanish-speaking context, experts would refer to 'retaining wall' and 'reinforced concrete' as *muro de contención* and *hormigón armado* respectively. In contrast, in ARCHINT, the false cognates *muro de retención* (retaining wall) and *hormigón reforzado* (reinforced concrete) were found instead. Why is there a terminological mismatch between experts and interpreters? Some possible reasons could be the saturation of the interpreters' limited cognitive resources, the time constraints imposed by the interpreting task and poor terminology preparation. However, a more important factor could be the failure to identify and process these noun compounds as a whole semantic unit and not as the sum of their parts; that is processing each item within the collocation individually. Therefore, there are grounds to believe that simultaneous-interpreting corpora can provide a fertile ground for the study of collocations in different ways. For instance, contrastive studies could be developed between simultaneous-interpreting corpora and *ad-hoc* oral corpora in search of the similarities and differences of collocational usages. Results from these studies may contribute to the research on the processing of noun compounds in specialized discourse and support the need to further include information about specialized collocations when creating terminological resources and specialized databases.

The third topic that falls within the broad area of language processing and to which simultaneous-interpreting corpora could make a contribution is that of word choice in general. According to Faber (2012: 59), the choice of a word or a grammatical structure profiles aspects of a particular situation and provides a particular construction or construal of reality. In fact, the

terminology used in a source speech is a reflection of the speaker's decision to prompt a specific knowledge representation in the mind of the target audience. The terminology used in a target speech of an interpreting corpus symbolizes the interpreter's word choice to convey the specialized content that lies beneath the linguistic cover of the source speech. Because of the wealth of contextual and linguistic information to which simultaneous-interpreting corpora give access, these corpora could be used to explore the topic of word choice from a comparative, intra-lingual (i.e. within the same speech) and inter-lingual perspective (i.e. across different languages), and as an observational complement to studies exploring the motivations that lie behind those word choices (e.g., Is it a well-delivered reasoning process? Is it a semi-automatic activity? Is it an entrenchment process that matches more or less defined knowledge representations?).

Neology

Another research area to which simultaneous-interpreting corpora could make a strong contribution is that of neology. When new knowledge is introduced at a conference, interpreters are often faced with the problem of determining the type of denomination that should be used in the target language. Since standard denominations have yet to be coined in the target language, interpreters are faced with a number of options including borrowing source language terms, borrowing source language terms with some degree of modification, translating source language terms literally, coining entirely new terms or attributing specialized senses to commonly used words. Interpreters might also introduce new terms not due to the absence of standard equivalents in the target language, but to convey information more effectively by using the terminology that is being used by the target audience in the conference.

Simultaneous Interpreting assignments which are very likely to involve the creation of new terms are those which involve disciplines with high creative components such as Art and Architecture. In such interpreting assignments, new designs and concepts are often introduced in the form of spontaneous speeches. Examples of this phenomenon are found in ARCHINT. A case in point is the creative (or metaphorical) usages that architect Zaha Hadid did with the words 'blister' and 'cone' when trying to explain the structure of one of her buildings. To convey these ideas, the

interpreter used the denominations *conos* (cones) and *ampollas* (blisters). Although relatively small in number, there is a chance that the conference participants who were also interpreting users heard these interpreter-made equivalents of 'cones' and 'blisters' for the first time, and ended up using them in their professional career as architects or architecture scholars. The question is: would interpreters become a catalyst for the creation of new terms, even if just temporarily? Does the use of these terminologies become widespread among experts? Terminologists may find this phenomenon appealing, and could well become interested in conducting studies involving both the domain-specific terminologies introduced by interpreters in specialized conferences, as well as the use and acceptability of these denominations within expert communities.

Conclusions

In the last thirty years, the number of machine-readable simultaneous-interpreting corpora has increased tremendously, and so have the studies that have been developed as a result. Interpreting corpora have become a valuable tool in simultaneous interpreting research and have contributed significantly to the development of Interpreting Studies.

The aim of this paper was to initiate a discussion on the feasibility of using machine-readable simultaneous-interpreting corpora as a research tool in Terminology. Undeniably, any attempt to use simultaneous-interpreting corpora beyond the realm of Interpreting Studies is bound to raise questions, paraphrasing Shlesinger's words, not only about the feasibility of the exercise but also about its point (1998: 3). Nevertheless, there may also be research opportunities that, unless explored, could well be disregarded. Given the specificity of the reality portrayed, there are some limitations and challenges of using simultaneous-interpreting corpora beyond the field of Interpreting Studies that need to be accounted for. One of these limitations is the need to use simultaneous-interpreting corpora as a research complement to other studies, or as part of a larger study, and not as the sole methodology. Ahead lie also the challenges of adapting the corpus material to the research needs of individual studies, and the training of individual researchers so that they become familiar with the object of the study that the corpus portrays. This can be achieved with increased collaborations and more interdisciplinary research projects.

While the real outcomes of this endeavor are still unknown, the overall balance seems to weigh more heavily on the positive side. In fact, the article has provided some modest examples suggesting how simultaneous-interpreting corpora could contribute to Terminology research in areas such as specialized communication, terminological variation, specialized language processing and neology. It is worth noting that although the article has only focused on the applications of simultaneous-interpreting corpora in the field of Terminology, they could also be used in any other area of language research. It seems that bright opportunities may lie ahead. The question that remains though, is whether these corpora are or will be appealing enough to terminologists and language researchers.

References

Baker, Mona (1995) 'Corpus Linguistics and Translation Studies. An Overview and Suggestions for Future Research', *Target*, 7:2.

Baker, Mona (1996) 'Corpus-based Translation Studies. The Challenges that Lie Ahead', in Harold Sommers (ed.) *Terminology, LSP and Translation*, Amsterdam/Philadelphia: John Benjamins.

Bendazzoli, Claudio and Annalisa Sandrelli (2005) 'An Approach to Corpus-based Interpreting Studies: Developing EPIC (European Parliament Interpreting Corpus)', in Sandra Nauert (ed.) *Proceedings of The Marie Curie Euroconferences MUTRA: Challenges of Multidimensional Translation*, *Saarbrücken 2-6 May 2005*, available online at [http://www.euroconferences.info/ proceedings/2005_proceedings/2005_bendazzoli_sandrelli.pdf] (accessed 08 august 2016).

Bendazzoli, Claudio and Annalisa Sandrelli (2009) 'Corpus- Based Interpreting Studies: Early Work and Future Prospects'*, Revista Tradumàtica. Traducció i Tecnologies de la Informació i la Comunicació* 7, available online at [http://www.raco.cat/index.php/tradumatica/ article/download/154835/206729] (accessed 08 august 2016).

Bendazzoli, Claudio (2010a) 'The European Parliament as a Source of Material for Research into Simultaneous Interpreting: Advantages and Limitations', in Lew N. Zybatow (ed.) *Translationwissenschaft: Stand und Perspektiven*, Frankfurt: Peter Lang, 51-68.

Bendazzoli, Claudio (2010b) *Il corpus dirsi: creazione e sviluppo di un corpus electrónico per lo studio della direzionalità in interpretazione simultanea [The Dirsi Corpus: creation and development of an electronic corpus for the study of directionality in Simultaneous Interpreting]*, Ph.D Thesis, University of Bologna.

Bowker, Lynne (1998) 'Using Specialized Monolingual Native-Language Corpora as a Translation Resource: a Pilot Study', *Meta*, 43:4, 631-651.

Bowker, Lynne and Jennifer Pearson (2002) *Working with Specialized Language—a Practical Guide to Using Corpora*, London: Routledge.

Cabrera, Tamara (2015) *La interpretación simultánea especializada y la precisión terminológica. La incidencia de las imprecisiones terminológicas en la transmisión de conocimiento y la calidad de la interpretación [Specialized Simultaneous Interpreting and Terminological Accuracy. The effect of Terminological Inaccuracies in the Transfer of Knowledge and the Quality of Interpreting]*. Ph.D Thesis. Granada: University of Granada.

Cabrera, Tamara (2016) 'Interpreting Architecture: The ARCHINT Corpus'. *Revista Tradumàtica. Tecnologies de la Traduciò*. Revistes Digitals- Universitat Autònoma de Barcelona, 14. 156-171.

Cabrera, Tamara and Pamela Faber (in progress) Accuracy in Simultaneous Interpreting.

Collados Aís, Ángela (1998) *La evaluación de la calidad en interpretación simultánea. La importancia de la comunicación no verbal [Quality Assessment in Simultaneous Interpreting. The Importance of Non-Verbal Communication]*, Granada: Comares.

Collados Aís, Ángela, Esperanza Macarena Pradas Macías, Elisabeth Stévaux and Olalla García Becerra (eds) (2007) *La evaluación de la calidad en interpretación simultánea: parámetros de incidencia [Quality Assessment in Simultaneous Intepreting: Quality Parameters]*, Granada: Comares.

Collados Aís, Ángela (2009) 'Marco evaluador de la interpretación simultánea', *Estudios de traducción: perspectivas. Zynaida Lvóskaya in memoriam*, Frankfurt: Peter Lang, 145-69.

Corpas Pastor, Gloria and Miriam Seghiri (2016) *Corpus-based Approaches to Translation and Interpreting*, Frankfurt: Peter Lang.

Tamara Cabrera Castro, Applications of Simultaneous-Interpreting Corpora in Terminology Research, 31 – 51.

Diriker, Ebru (2004) *De-/Re-Contextualizing Conference Interpreting: Interpreters in the Ivory Tower?*, Amsterdam/Philadelphia: John Benjamins.

Faber, Pamela (ed.) (2012) *A Cognitive Linguistics View of Terminology and Specialized Language*, Berlin/Boston: Mouton de Gruyter.

Freixa, Judit (2006) 'Causes of Denominative Variation in Terminology: A Typology Proposal' *Terminology* 12(1): 51-77.

García de Quesada, María Mercedes (2011) 'Terminologie', in Collados Aís, Ángela, Emilia Iglesias Fernández, Esperanza Macarena Pradas Macías y Elisabeth Stévaux (eds) *Qualitätsparameter Beim Simultandolmetschen. Interdisziplinäre Perspektiven,* Tübingen: Narr.

Gumperz, John Joseph and Norine Berenz (1993) 'Transcribing Conversational Exchanges', in Jane A. Edwards and Martin D. Lampert (eds) *Talking Data: Transcription and Coding in Discourse Research*, Hillsdale, N.J.: Lawrence Erlbaum, 91-121.

Iglesias Fernández, Emilia (2010) 'Speaker Fast Tempo and its Effect on Interpreter Performance: a Pilot Study of a Multilingual Interpreting Corpus', *International Journal of Translation*, 145-70.

Langacker, Ronald (1999) *Grammar and Conceptualization*. Berlin/New York: Mouton de Gruyter.

Laviosa, Sara (1998) 'The Corpus-based Approach: a New Paradigm in Translation Studies', *Meta,* 43:4.

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

Lederer, Marianne (1981) *La traduction simultanée*, Paris: Minard Lettres Modernes.

Meyer, Ingrid and Kristen Mackintosh (1996) 'The Corpus from a Terminographer's Viewpoint', *International Journal of Corpus Linguistics*, 1:2, 257-285.

Meyer, Ingrid (2001) 'Extracting Knowledge-Rich Contexts for Terminography: a Conceptual and Methodological Framework', in Didier Bourigault, Christian Jacquemin and Marie-Claude L' Homme (eds) *Recent Advances in computational Terminology*, Amsterdam/Philadelphia: John Benjamins, 279-302.

Monti, Cristina, Claudio Bendazzoli, Annalisa Sandrelli and Mariachiara Russo (2005) 'Studying Directionality in Simultaneous Interpreting through an Electronic Corpus: EPIC (European Parliament Interpreting Corpus)', *Meta*, 50:4.

Pazos-Bretaña, José-Manuel, Olalla García Becerra and Rafael Barranco-Droege (2010) 'Diseño y técnicas de explotación de un corpus oral para el análisis de parámetros de calidad en interpretación', in: Isabel Moskowich-Spiegel Fandiño, Begoña Crespo García, Inés Lareo Martín and Paula Lojo Sandino (eds) *Language windowing through corpora- Visualización del lenguaje a través de corpus*. A Coruña: Universidade A Coruña, 627–640.

Pearson, Jennifer (1998) *Terms in context*, Amsterdam/Philadelphia: John Benjamins.

Pearson, Jennifer (1999) 'Comment accéder aux elements définitoires dans les texts spécialisés?', *Terminologies Nouvelles*, 19, 21-28.

Pöchhacker, Franz (1993) 'From Knowledge to Text Coherence in Simultaneous Interpreting', in Yves Gambier and Jorma Tommola (eds) *Translation and Knowledge*: SottIV, Turku: University of Turku, Centre for Translation and Interpreting, 87-100.

Pöchhacker, Franz (2004) *Introducing Interpreting Studies*, London: Routledge.

Pöchhacker, Franz (2008) 'Inside the 'black box'. Can Interpreting Studies help the Profession if Access to Real Life Settings is Denied?', *The Linguist*, 48:2.

Russo, Mariachiara, Claudio Bendazzoli and Annalisa Sandrelli (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, Annalisa (2003a) 'Herramientas informáticas para la formación de intérpretes: Interpretations y The Black Box', in Jesús de Manuel Jerez (ed.) *Nuevas tecnologías y formación de intérpretes*, Granada: Atrio, 67-112.

Sandrelli, Annalisa (2003b) 'New Technologies in Interpreter Training: CAIT' in Heidrun Gerzymisch-Arbogast, Eva Hajicová, Peter Sgall, Zuzana Jettmarová, Annely Rothkegel and Dorothee Rothfuss-Bastian (eds) *Textologie und Translation*, Tübingen: Gunter Narr Verlag, 261-293.

Sandrelli, Annalisa (2007) 'Designing CAIT (Computer-Assisted Interpreter Training) Tools: Black Box', in Gerzymisch Arbogast, Heidrun and Sandra Nauert (eds) *Challenges of*

Multidimensional Translation. Proceedings of the Marie Curie Euroconferences MuTra: Challenges of Multidimensional Translation, Saarbrücken 2-6 May 2005.

Sandrelli, Annalisa, Claudio Bendazzoli and Mariachiara Russo (2010) 'European Parliament Interpreting Corpus (EPIC): Methodological Issues and Preliminary Results on Lexical Patterns in SI', *International Journal of Translation*, 22:1-2, 165-203.

Sandrelli, Annalisa (2012) 'Interpreting Football Press Conferences: the FOOTIE Corpus', in Jane C.Kellett, (ed.) *Interpreting Across Genres: Multiple Research Perspectives*, Trieste: E.U.T.: 78-101.

Sandrelli, Annalisa (2015) 'Becoming an Interpreter: the Role of Computer Technology' *MonTi,* 2, 111-138.

Schmid, Hans Jörg (2007) 'Entrenchment, Salience and Basic Levels', in Dick Geeraerts and Hubert Cuyckens (eds) *The Oxford Handbook of Cognitive Linguistics*, Oxford: Oxford University Press, 117-138.

Seleskovitch, Danica (1975) *Language, langues et memoir: études de la prise de notes en interpretation consécutive*, Paris: Minard Lettres Modernes.

Setton, Robin (2011) 'Corpus-based Interpreting Studies (CIS): Overview and Prospects', in Alet Kruger, Kim Wallmach and Jeremy Munday (eds) *Corpus-Based Translation Studies: Research and Applications,* London: Continuum International Publishing, 34-75.

Shimizu, Hiroaki, Graham Neubig, Sakriani Sakti, Tomoki Toda and Satoshi Nakamura (2014) 'Collection of a Simultaneous Translation Corpus for Comparative Analysis', *International Language Resources and Evaluation*, available online at [http://www.phontron.com/paper/shimizu14lrec.pdf] (accessed 08 august 2016).

Shlesinger, Miriam (1998) 'Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies', *Meta*, 43:4, 486-493.

Shlesinger, Miriam (2008) 'Towards a Definition of Interpretese. An Intermodal, Corpus-based Study', in Gyde Hansen, Andrew Chesterman and Heidrun Gerzymisch-Arbogast (eds) *Efforts and Models in Interpreting and Translation Research. A Tribute to Daniel Gile*, Amsterdam / Philadelphia: John Benjamins.

Straniero Sergio, Francesco and Caterina Falbo (eds) (2012) *Breaking Ground in Corpus-based Interpreting Studies*. Bern: Peter Lang.

Web pages

EbS, The European Commission Audiovisual Service, available online at [http://ec.

europa.eu/avservices/ebs/schedule.cfm] (accessed 08 august 2016).

ECIS Group, available online at [http://ecis.ugr.es/index-en.htm] (accessed 08 august 2016).

EPIC Corpus, available online at [http://sslmitdev-online.sslmit.unibo.it/corpora/corporaproject.

php?path= e.p.i.c] (accessed 08 august 2016).

Epic Transcription Conventions, available online at [http://dev.sslmit.unibo.it/corpora/

additionalpages.php?path=e.p.i.c.&source=project&content=transcription_conventions.des

c] (accessed 08 august 2016).

The European Union's Speech Repository, available online at [https://webgate.ec.europa.eu/sr/]

(accessed 08 august 2016).

Wordsmith Tools. *Windows software for finding patterns*, available online at

[http://www.lexically.net/wordsmith/] (accessed 08 august 2016).

First Forlí International Workshop 2015, [http://eventi.dipintra.it/cis1/] (accessed 08 august

2016).