The International Journal for Translation & Interpreting Research
trans-int.org

Computer-assisted preparation in conference interpreting

Claudio Fantinuoli
Johannes Gutenberg Universität Mainz/Germersheim, Germany
fantinuoli@uni-mainz.de

DOI: 10.12807/ti.109202.2017.a02

Abstract: Preparation has been proposed in the literature as one of the most important phases of an interpreting assignment, especially if the subject is highly specialised. Preparing an assignment in advance aims at bridging the linguistic and extra-linguistic gap between conference participants and interpreters and at reducing the cognitive load during interpretation. For these reasons it is considered crucial in ensuring higher interpreting quality. Yet, preparation is generally time-consuming and interpreters may often experience the feeling of not knowing exactly how to perform this task efficiently. Information technology could change this. Even though the first computer-assisted interpreting software has entered the profession in recent years, no tool has been specifically developed to satisfy the needs of interpreters during the preparatory phase. After analysing different theoretical frameworks of interpreting preparation, this paper aims at presenting a tool that implements a corpus-driven approach to preparation. According to this approach, the process of knowledge and language acquisition needed to perform well as an interpreter is optimized by making it corpus-driven: browsing the terminology of the domain in a specialised corpus, interpreters are able to reconstruct its conceptual structure, prepare subject-related glossaries and rationalise the preparatory work.

Keywords: interpreting, computer-based interpreter preparation, terminology, CAI tools

1. Introduction

Conference interpreters are language professionals who convey the meaning of an oral text from one language to another and do this simultaneously, i.e. producing the target text while a previously unknown original is orally delivered by the speaker[1]. Simultaneous interpreters typically work at highly specialised international conferences or meetings and have to translate a wide variety of subjects. Due to evolving market requirements, assignments are of a far more varied nature than in the past. This poses several challenges to the interpreting quality, an issue which is becoming increasingly important among interpreters, trainers, conference participants and scholars (cf. Kalina, 2006).

As interpreters are called to interpret many different topics for which they are not expert or do not have any specific qualification, conference preparation has been proposed in the literature as one of the most important phases of an interpreting assignment, especially if the subject is highly specialised (cf. Gile, 2009; Díaz-Galaz, 2015). The role of preparation is central for at least two reasons: it aims at bridging the linguistic and extra-linguistic gap between conference participants and interpreters (Will, 2009) and helps to reduce the cognitive load during the interpreting task as it anticipates parts of it in the preparatory phase (Stoll, 2009). Having more free cognitive capacities during an interpreting assignment, interpreters are able to manage the interpreting process more efficiently. Accordingly, preparing an assignment in advance supports interpretation quality, for example, by ensuring greater accuracy (Díaz-Galaz, 2015). Yet, preparation is generally time-consuming and interpreters may often experience the feeling of not knowing exactly how to perform this task efficiently. To cope with this, we propose a computer-assisted approach to conference preparation designed to help interpreters to rationalise the process.

The use of computer tools is not new in the language industry. Although information technology did not have the same impact on interpreting as it did on translation, during the last decade, the way interpreters work has been influenced by advances in informatics: the World Wide Web with its abundance of data, for example, has changed the way they access and elaborate knowledge (cf. Kalina, 2009; Fantinuoli, 2012) and the use of laptops and tablets has allowed interpreters to look up their reference material and terminology directly in the booth (cf. Fantinuoli, 2016b; Tripepi Winteringham, 2010; Costa, Corpas Pastor & Durán Muñoz, 2014). Yet, at the moment, no software has been specifically developed to satisfy the needs of interpreters in the preparatory phase. Considering the fact that information technology has played a central role and has changed the way many professionals work in the last decades, it is reasonable to assume that a process-oriented computer-assisted interpreting (CAI) tool (cf. Fantinuoli, 2017) specifically designed to address the preparatory phase of interpreters could contribute to enhance this task.

The rest of this paper is organised as follows. Firstly, the preparatory needs of conference interpreters are analysed with respect to the domain and the lexical knowledge needed to perform well at a conference. Subsequently, a corpus-based approach to preparation is proposed. Finally, the tool developed to implement the above mentioned approach is discussed.

2. Interpreter's preparation

2.1. Preliminary thoughts

At a typical conference, interpreters are called to work for specialists sharing knowledge totally or partially unknown to people who are not expert in the particular subject of the conference. Communication is therefore characterised by a linguistic and extra-linguistic gap between the interpreter and the participants (cf. Gile, 2009; Will, 2009; Kucharska, 2009). To fill this gap, interpreters have to prepare for the conference topic days or hours prior to the assignment. The preparation phase, and in particular the role of specialised terminology and the strategies to define, extract, organise and manage it, has been considered crucial to better cope with the difficulties arising during

interpreting and which may be the cause of problems and deficiencies (cf. Pöchhacker, 2000; Fantinuoli, 2006; Rütten, 2007; Will, 2009; Stoll, 2009).

Since interpreters work for specialists who share knowledge totally or partially unknown to outsiders, it is reasonable to assume that the resulting knowledge gap manifests itself at least at two levels, which can be defined as the level of *domain knowledge* and of *linguistic knowledge* of the specialised subject. Although there is consensus among scholars and practitioners on the crucial role of preparation and on some basic principles relating to it, particularly the fact that interpreters need an overall *thematic* knowledge into which *terminology* is embedded (Will, 2007), the approaches to preparation may diverge. Some believe that knowledge acquisition performed in advance should focus on extra-linguistic information (how things work, etc.) while others give priority to linguistic preparation, in particular to its terminological component (cf. Gile, 2009). Some authors claim that interpreters should be constantly up-to-date in all relevant topics (Feldweg, 1996), while others stress the importance of the specific meeting preparation based on reference materials (Seleskovitch & Lederer, 1989) or conference papers (Gile, 2009).

In recent years, scholars have stressed the need for a more holistic position which combines linguistic and extra-linguistic knowledge and describes knowledge as a combination of language, content and situational expertise, moving from simple and sparse data to the establishment of a complex knowledge system (cf. Kalina, 1998; Fantinuoli, 2006; Rütten, 2007; Gorjanc, 2009; Will, 2009). Accordingly, interpreters need to master both levels to a certain degree in order to provide a quality rendition of the original discourse. This is why both levels must be considered in any preparatory activity.

All approaches to interpreters' preparation are based on a more or less detailed and articulated division of the interpreting process (cf. Gile, 2009; Kalina, 2007; Will, 2009). They all share the basic idea that an assignment can be divided at least into three parts: *before*, *during* and *after* the interpreting task. Given the spontaneity and the time limitations of the interpreting process, knowledge acquisition occurs primarily prior to the conference. This is the phase in which preparatory work has to be performed (cf. Thrane, 2005; Gile, 2009; Stoll, 2009; Will, 2009). This poses a major challenge: different to translators, who can constitute their knowledge on an ad-hoc basis while translating (for example when comprehension or terminological problems arise), interpreters have to do it in advance and without knowing exactly which problems may arise while interpreting (comprehension, terminology, etc.)[ii]. In fact, during the conference itself it is only possible to integrate the knowledge acquired in the preparation phase, for example reading new documents handed in at the event, listening to the speeches and interacting with the participants (cf. Kalina, 2007).

In the next sections the two areas of preparation identified above are introduced and discussed briefly.

2.2. Domain knowledge

The domain or topic-specific knowledge concerns the expertise in a specific topic, information about the speaker, the situational context, etc. Communication among conference participants is based on knowledge which is shared by discourse producers and discourse receivers and which is indispensable to successful communication. This knowledge has been identified as important in enhancing interpreters' performance because it has a

major impact on the comprehension phase, as indicated by most cognitive models of translation and interpreting (cf. Gerver & Sinaiko, 1978; Gile, 2009; Setton, 1999). It is generally accepted that, in order to produce an acceptable rendition of the discourse, not only the lexical and semantic equivalence must be established during interpreting, but also a functional equivalence, and this requires a profound understanding of the domain and the communicative setting.

Because comprehension is essential in interpreting and the knowledge needed to facilitate comprehension is not always explicit in a text, interpreters need to acquire a sufficient working knowledge of the respective topic, i.e. a good level of familiarity with the underlying concepts in order to quickly comprehend the ideas (not the words) uttered by the speaker and to contextualise them into the specialised knowledge system which is shared by all communication participants. Although experimental studies in interpreting have come to divergent conclusions on the effect of studying the related materials prior to interpreting (e.g. Díaz-Galaz et al., 2015), modern comprehension models recognise the role of prior topic-specific knowledge in the processing of general and specialised discourse (e.g. McNamara & O'Reilly, 2009). If this process of meaning constitution is correctly performed by interpreters, it is more likely that meaning will be correctly transferred from one language into the other in the reformulation phase. Furthermore, a sound working knowledge of the conference topic and the communicative situation helps to anticipate and predict information (De Groot, 1999). This has obvious consequences on the cognitive load in the reception phase as the more interpreters know, the more they can predict, and the better the knowledge about everything, the faster it can be predicted (Stoll, 2009).

Traditionally, interpreters acquire the needed domain knowledge by reading topic-related texts preferably in both source and target language. The amount of preparatory work and the degree of text specialisation varies and depends on the interpreter's background knowledge and the level of specialisation of the conference itself. Interpreter associations generally require conference organisers to provide the interpreters with conference documents (program, minutes of the previous meeting, reports, etc.) in order to enable them to prepare for the assignment. This material is generally considered to be the most appropriate for preparation, as it gives interpreters the possibility to use only highly relevant, conference-related texts. Nevertheless, there are several reasons why preparatory material is not always made available to interpreters in advance: papers are not ready until the moment of speech delivery, speakers are not aware of interpreters' needs, they do not want to disclose the content of their speech in advance or the documents are confidential, etc. (Gile, 2009). All these cases require alternative approaches to information retrieval and knowledge acquisition, as is discussed in Section 3.

2.3. Linguistic knowledge

Given the fact that professional interpreters are language experts with a high command of their L1 and L2, the linguistic knowledge that needs to be acquired foremost concerns the terminology of a specific field as well as the subject-specific phrases and stylistic expressions used by a delimited group of people to exchange specialised information. This lexical and phraseological level is also referred to as *in-house jargon* (Kalina, 2006). As precise and complete communication can only be achieved by using the correct

terminology (Arntz, Picht & Mayer, 2009), there is a general consensus regarding the fact that the correct and appropriate use of specialised terminology is a major quality issue in scientific and technical conferences (cf. Gile, 2009; Will, 2009). As a matter of fact, audiences expect interpreters to use correct terminology to a much greater extent than in the past (Kalina, 2007).

For many practitioners the task of identifying the typical terms of a specialised domain is one of the main activities of the preparatory phase (Moser-Mercer, 1992). The preparatory work performed at the terminological level starts with the reading of the material provided by the conference organisers or autonomously collected from the web, underlining relevant, usually unfamiliar terms and phrases and searching for equivalents in the other working languages. This mainly results in bilingual or multilingual glossaries, i.e. lists of terms and their translation in one or more languages, or in terminological annotations on the preparatory documents handed out (Moser-Mercer, 1992).

In a communicative setting where time constraints play a crucial role, a good command of the domain terminology in both languages is important, at least because a) it is essential during the comprehension phase to understand the original discourse, and b) it helps formulating short and precise sentences in the target language, avoiding the abuse of alternative strategies – such as explanations, hypernyms – which are typically used to cope with terminological problems. In fact, although these strategies may be helpful if used in specific situations, their abuse can have negative consequences on the interpreting performance: it can lead to a cognitive overload in the interpreting process (cf. Gile, 2009); it can take away the time needed for other operations (for example listening); last but not least, it can be the cause of imprecise communication, which in turn may lead the audience to think that the interpreters are not experienced in the domain.

When preparing a new assignment, interpreters may need to acquire not only specialised terms but also the general terms which are typical of a specific domain (Rütten, 2007), depending on whether interpreters are working into or out of their foreign language or whether or not they are used to interpreting in that specific domain (Fantinuoli, 2006). In most general terms, the terminology used in a technical or scientific conference can be divided – as far as the interpreter perspective is concerned – into three main categories:

- General terms typically used in the specialised domain
- High-frequency terms of the specialised domain
- Low-frequency terms of the specialised domain

Category 1 contains terms which are typically used in a given domain, even if they are not highly specialised. These are basic terms shared with other disciplines or which are used in all sub-domains of a specific domain. In a technical meeting about *clutches*, for example, terms belonging to this category are *brake*, *pedal*, *torque*, etc. These terms should be at the interpreter's disposal at any time and without major efforts.

Category 2 contains terms which are typical of and frequent in the specialised domain. In our example (clutches), they could be *damper*, *centrifugal clutch*, *friction*, etc. As these terms are statistically very frequent in the domain of interest, they should be at the interpreter's disposal at any time

and without major efforts. Conference participants would expect the interpreter to use them correctly as they make up the core of the terminology of the sector in which they are experts.

Category 3 contains low-frequency, highly specialised terms of the conference domain. The probability that they will be used in the course of the conference is low. These terms generally make the bulk of the conference's terminology. As the probability of encountering these terms is low, it is reasonable to think that there is no need for them to be immediately at the interpreter's disposal, i.e. memorised, but could be accessed by means of a terminology look-up tool when needed (Costa et al., 2014; Fantinuoli, 2009, 2012, 2016b). Examples for these terms are *conical bellhousing*, *dog clutch*, *wrap-spring*, etc.

In practice, it can be difficult to objectively assign terms to a particular category. Notwithstanding, the proposed categorization is intended to help to better differentiate the terminological needs in the context of a conference, for example to guide the choice of whether a term should be memorised or rather saved in a glossary for look-up during the interpreting process.

Strictly related to terminology, phraseological knowledge plays a central role in the linguistic preparation of interpreters. It concerns the subject-specific phrases and stylistic expressions used by the experts of a particular domain, company, etc. Such specialised phrases are understood to be the connection of at least two linguistic elements that express a specialised content and are considered fixed expression in a specific context (cf. Rossenbeck, 1989). The appropriate use of phraseological items, most frequent collocations, plays a major role in interpreter-mediated communication. In the comprehension phase, the knowledge of collocations improves the quality of interpreting by supporting the anticipation process (Stoll, 2009). In the production phase, the correct use of typical phraseological units – even if other lexical alternatives could be perfectly acceptable – can increase the acceptance of the interpreted text and ultimately the perceived professionalism of the interpreter.

Although the problem of collocational competence is traditionally confined to non-native speakers, it has major relevance also for native speakers when it comes to specialised languages (LSP), as the use of the appropriate collocational items is not intuitive, but is based on frequency of use. This is the reason why it is reasonable to suggest that, when preparing for a new assignment, interpreters should also learn the correct collocates of specialised terms in order to master the in-house jargon used by conference participants.

3. Corpus-driven Interpreter Preparation

Corpus-driven Interpreter Preparation (CDIP) aims at solving the challenges and problems introduced in the previous sections by means of a computer-based, corpus-driven approach to preparation. The basic idea of CDIP is to turn the preparatory phase into a discovery-oriented task for terminology and knowledge acquisition (Fantinuoli, 2006). Adapting the corpus-based approach originally developed for L2 acquisition (Carter, McCarthy & O'Keeffe, 2007), CDIP aims at resolving the dichotomy between terminology-oriented and content-oriented preparation introduced in Section 2.1 and described by Gile (2009, p. 149) with the following words:

> […] interpreters experience very concretely the deleterious effects of insufficient familiarity with technical terms that are used in conferences. Since very little time is available for advanced preparation, they generally have to choose between primarily extralinguistic preparation and primarily terminological preparation. Most of them give preference to terminology […]

CDIP is based on the idea that corpora, and in particular specialised monolingual corpora, can be the source of a potentially endless "serendipity process" (Johns, 1988), as one term can lead to another, depending on the interpreter's intuition and needs. In this approach, interpreters *explore* the corpus starting from a list of specialised terms and learn them in real context, understanding their meaning and usage and, at the same time, getting a grasp of the subject.

This turns interpreters into a kind of *special learner* who needs to acquire as much linguistic and extra-linguistic knowledge as possible in an autonomous way. The use of corpora for conference preparation is in line with the idea of placing learners in the centre of the learning process with their needs, cognitive processes and learning strategies (Kiraly, 2000). The approach is based on Data Driven Learning (DDL), as introduced by Boulton (2009, p. 82):

DDL typically involves exposing learners to large quantities of authentic data – the electronic corpus – so that they can play an active role in exploring the language and detecting patterns in it. They are at the centre of the process, taking increased responsibility for their own learning rather than being taught rules in a more passive mode.

DDL can be operationalised by means of computer tools: the learner can gain insights into the language and the domain by using a concordance program to locate authentic examples of the language in use (cf. Johns, 1988). Experimenting with corpora offers "virtually unlimited opportunities for learning by discovery, as learners embark on challenging journeys whose outcomes are unpredictable and usually rewarding" (Bernardini, 2001, p. 246). Thanks to the interactivity of concordancers, the approach provides the amount of flexibility and active interaction typical of the interpreter's profession.

The ideas mentioned above seem to apply well to the challenges posed by the terminology-oriented interpreter preparation discussed by Will (2009). Describing the complexity of the knowledge systems that must be mastered by interpreters, he applies the context-related term model of Gerzymisch-Arbogast (1996), which considers possible deviations from the classic unique correlation between concept and designation, as traditionally advocated by terminologists: knowledge always manifests itself within real texts and as part of a knowledge system; terminology is embedded in texts and therefore can be *contaminated* by the knowledge system itself. In order to take account of the *variability* of terms which manifest itself in real texts, Will (2009) pleads for a preparation that resembles *detective work*, allowing interpreters to constitute and represent knowledge in context: from the term and the term definition to the specific knowledge system. This kind of knowledge acquisition allows the interpreter to gain a systematic overview of the knowledge systems involved in the conference as well as their ranking in terms of importance and priority. The structured knowledge systems emerging from this approach can ultimately be recorded in a database (glossary) and used during interpretation.

Corpus-driven Interpreter Preparation aims at optimizing the preparation process of conference interpreters by making use of the discovery attitude of corpus-driven analyses. In order to discover their meaning and usage in context, the interpreter explores the corpus using a list of specialised terms.

The process of "knowledge/language learning" needed by interpreters in order to prepare themselves for a conference can be optimized if "terminology driven", i.e., "bottom-up": from the terminology to the conceptual structure of a particular domain (Fantinuoli, 2006, p. 174).

The terms to start exploring the domain can be obtained from an automatic extraction method based on corpora collected from the web. In his experiment, Fantinuoli (2006) uses BootCaT to bootstrap text from the web and implements a series of scripts to extract the specialised terminology from the corpus. The evaluation of the terminology extraction quality, based on the categorization of the terms according to their level of specialisation and well-formedness, confirms that the results of the procedure are suitable for CDIP.

Given the time-consuming aspect of the typical preparation workflow, which comprises collecting parallel texts and manually extracting the relevant terminology, computer-based CDIP, if properly implemented, seems particularly suitable for conference preparation. It allows interpreters to obtain within minutes a list of relevant terms and a collection of specialised texts that can be used as reference material for consultation.

CDIP's workflow can be summarized as follows:

1. Topic identification through selection of a set of highly specialised terms
2. Collection of monolingual specialised texts dealing with the topic
3. Automatic extraction of statistically relevant terminology according to the categories introduced in Section 2.3
4. Dynamic exploration of textual material starting from the extracted terminology, extraction of collocational patterns for the terms of interest, etc.

The feasibility of this approach has recently been the focus of experimental studies. Xu (2015), for example, has experimentally investigated how corpus-based terminology preparation, which integrates the building of small comparable corpora as well as the use of automatic term extractors and concordance tools, can improve the performance of trainee interpreters. The results show that the experimental groups had consistently better terminology performance during simultaneous interpreting: they correctly interpreted more terms, had higher terminology accuracy scores and made less term omissions[iii]. Furthermore, they also had higher holistic simultaneous interpreting performance scores than the control groups. These results seem to suggest that the CDIP approach can help interpreters to improve their performance on specialised topics. As the experiment was performed with a series of tools not specifically developed for interpreters, it is reasonable to think that the use of a tool specifically developed for this target group may further improve the above-mentioned scores.

In the next section, such a tool is briefly discussed.

4. CorpusMode for CDIP: Architectural design

CorpusMode[iv] is a documentation software designed for translators and interpreters. It comprises a tool to build specialised corpora from the web, a terminology and collocation extraction module and an easy-to-use concordancer to explore the texts in an exploration-oriented way. The tool bundles a set of topic-related information such as:

- a corpus of specialised texts automatically collected from the web
- a list of statistically relevant terms for the conference topic
- a search engine-like tool to dynamically explore the corpus
- candidate translations for the extracted terms
- a definition for each extracted term.

The tool has been developed in the framework of InterpretBank (Fantinuoli, 2012; 2016b), a comprehensive terminology and knowledge program for conference interpreters, adapting the tool TranslatorBank (Fantinuoli, 2016a), a corpus analysis tool developed at the University of Mainz in Germersheim, to the needs of interpreters.

In the next sections, the main parts of the tool are briefly described.

4.1. Corpus creation

The corpus creation utility is designed to automatically build on-the-fly specialised corpora, i.e. collections of electronic texts dealing with the conference subject, using the web as a text repository.

It is typical of the profession that interpreters are handed out only a limited amount of preparatory material (Stoll, 2009) and that they are expected to be autonomous in retrieving the information they need (Kalina, 2007). In all these cases, an automatically generated corpus can be used as a source of comparable texts in order to acquire as much information and specialised knowledge as possible and to extract the terminology typical of the domain under investigation, as described in Section 3.

The nearly unsupervised corpus creation procedure shows some similarities with the one proposed by Baroni & Bernardini (2004). In the past, scholars have successfully used this procedure to create corpora from the web for translation (Bernardini & Castagnoli, 2008) and interpreting tasks (Fantinuoli, 2006). The workflow is straightforward: the process requires a small set of terms that are expected to be representative of the conference's domain. To prevent the software from collecting unrelated texts, the searching terms should be unambiguous, highly specialised and possibly used only within the domain of interest. These terms are used as a query string in a search engine[v] and the top pages (PDF and/or HTML) returned for each query are downloaded and saved as XML together with meta-information, such as original URL, source and date. The user can influence the corpus building procedure by means of the following parameters: the number of documents to be collected (size of the corpus); the language of the documents; the format (PDF/HTML); the possibility to restrict the query to a specific domain or Internet address (for example to create a company-related corpus).

The relatedness and quality of the collected documents can be assessed manually by the user and texts not suitable for inclusion in the corpus can be discarded. The selected documents are loaded in the concordancer (4.3) and

are ready to be looked up and used for terminology and collocation extraction (4.2).

4.2. Terminological and collocational extraction

The purpose of the terminology extraction utility is to identify a list of monolingual specialised terms and phrases from the collected corpus that can be used by the interpreter to create a conference glossary as well as to start the learning process described in Section 3.

The extraction algorithm used by CorpusMode is described in detail in Fantinuoli (2016a). The implemented method is hybrid as it combines linguistic knowledge and statistical measures. To improve the usability of the software for interpreters, the focus is on precision rather than recall. This means that the majority of extracted terms should be potentially useful for the user while the number of malformed terms should be kept to a minimum, even at the risk of missing some eligible candidates. The level of importance of terms is determined by means of frequency (see Section 2.3). Both single-word as well as multi-word terms are extracted. The extraction is based on the assumption that single-word and multi-word terms have a certain fixed set of linguistic structures, for example *Noun + Preposition + Noun* are likely to be candidate terms in Italian (*cima di recupero*, *barca da riporto*, etc.). The tool assigns a part-of-speech tag to each word and extracts all candidate terms that adhere to predefined patterns. The resulting term list is then filtered by means of statistical measures in order to rank the candidate terms and select the most appropriate. For example, common words can be excluded from the final list. This allows the list to be trimmed depending on the interpreter's profile. Novice interpreters, or interpreters not accustomed to working with a particular subject, may also need the general terms used in a particular field (see Section 2.3 for term categorization), especially if they are working into their foreign language in order to activate such terms before the beginning of the conference.

For each term a list of collocates is automatically retrieved. This function aims at identifying those collocates which are the most frequent for the given term in the specific domain, leaving out rather atypical collocational patterns. This is in line with our assumption that interpreters predominately need the most typical and therefore most frequent linguistic information for a given term. Collocates are identified statistically by counting the number of occurrences of all tokens conforming to the part of speech pattern of interest which occurs in a defined window span. The most frequent collocates are finally presented to the user as a list of collocates and their frequency or as a word cloud.

4.3. Concordancer

The texts collected can be analysed for learning or analysis purposes through a concordancer, a program whose function is to bring together passages of text and show how a word is used in its context by means of Key Words in Context (KWiC).

The concordancer (Figure 1 below) has been designed to offer a user-friendly and intuitive interface, giving priority to clarity and simplicity over a large number of options (typical of concordancers designed for linguists). The query system replicates as far as possible the behaviour of search engines, as they are considered to be the most familiar working environment for interpreters (cf. Zanettin, 2002). By default, queries are performed in a case-

insensitive way. If the input string is a single word, all sentences containing that word will be shown among the results. If the input string is made of two or more words, then the so-called *proximity search* is performed: all sentences containing the words inside a certain window span are displayed.

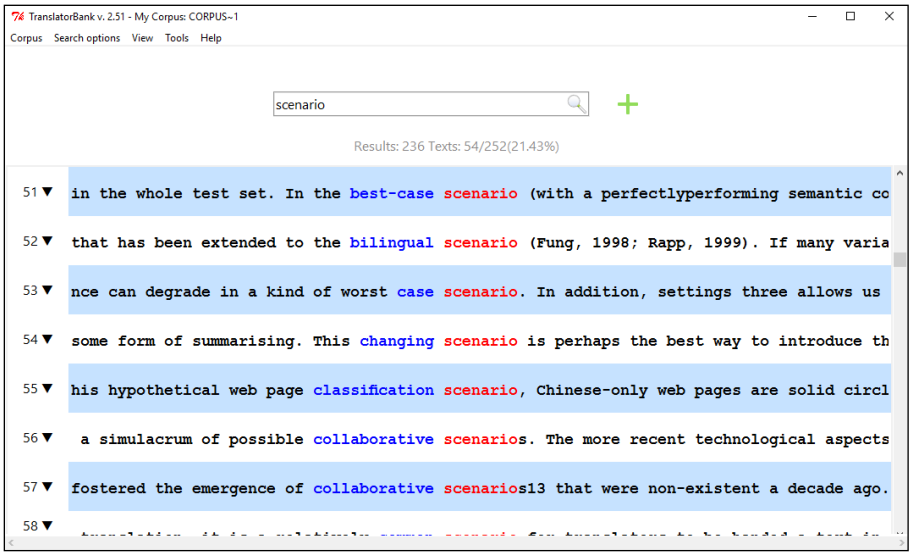

Figure 1: Graphical interface for the concordance

The proximity search is particularly useful in offering a flexible way to explore the corpus with the discovery approach introduced in the previous sections. Exact matching of two or more words is still possible by using double quotation marks (“) operator. In order to spot regularities in language use, results can be ordered alphabetically by the first, second or third element to the left or right of the query word. For every KWiC, the user can show the wider textual context in which the result occurs or directly access the original source (PDF or Webpage)[vi].

4.4. Candidate translation and definition of key terms

In order to extend the monolingual, corpus-based approach adopted by CDIP, users are offered a set of bilingual information, such as translation of terms and phrases or their definitions. They are retrieved from sources freely available on the web (dictionaries, lexica, encyclopaedias, glossaries, etc.) replicating the typical web searches done by interpreters. The number of data sources that can be potentially integrated in the software is very large and depends on the language combination and the user’s needs. By default, the typical sources used by interpreters, like IATE for terms and Wikipedia for definitions, are available.

When right-mouse-clicking on a term, a list of available sources is provided. By selecting the source of choice, a new window will pop up, showing the webpage containing the translation or definition.

5. Conclusions

In this paper, we have discussed the role of linguistic and extra-linguistic preparation in the interpreting profession. Different approaches proposed by scholars have been briefly analysed and a computer-assisted, corpus-driven approach has been introduced. In order to give interpreters a practical tool to optimize their information retrieval needs, the free tool CorpusMode has been presented and its main features briefly discussed. There is reason to believe this software will prove a useful addition to the traditional way interpreters prepare for a conference, yet more empirical studies are needed to test and possibly improve the way it can be integrated with current preparation workflows. It is our hope that this program could be of use for professional interpreters wishing to implement a more organic computer-based approach to interpreter preparation, and could stimulate other researchers to analyse the emerging needs of interpreters in a digitalised era.

References

Arntz, R., Picht, H., & Mayer, F. (2009). *Einführung in die Terminologiearbeit*. Hildesheims: Olms.

Baroni, M., & Bernardini, S. (2004). BootCaT: Bootstrapping corpora and terms from the web. In *Proceedings of the forth international conference on language resources and evaluation*, Paris. ELRA.

Bernardini, S. (2001). Spoilt for choice: A learner explores general language. In G. Aston (Ed.), *Learning with corpora* (pp. 220–249). Bologna: CLUEB.

Bernardini, S., & Castagnoli, S. (2008). Corpora for translator education and translation practice. In E. Yuste Rodrigo (Ed.), *Topics in language resources for translation and localisation* (pp. 39–55). Amsterdam: John Benjamin.

Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35(1): 81–106.

Carter, R., McCarthy, M. M., & O'Keeffe, A. (2007). *From corpus to classroom*. Cambridge: Cambridge University Press.

Costa, H., Corpas Pastor, G., & Durán Muñoz, I. (2014). A comparative user evaluation of terminology management tools for interpreters. In *Proceedings of the workshop on Computational Terminology (CompuTerm'14)* (pp. 68–76).

De Groot, G.-R. (1999). Zweisprachige juristiche Wörterbücher. In P. Sandrini (Ed.), *Übersetzen von Rechtstexten. Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache*, (pp. 203–227). Tübingen: Gunter Narr Verlag.

Díaz-Galaz, S. (2015). *La influencia del conocimiento previo en la interpretación simultánea de discursos especializados: Un estudio empírico*. PhD thesis, Universidad de Granada.

Díaz-Galaz, S., Padilla, P., & Bajo, M. T. (2015). The role of advance preparation in simultaneous interpreting: A comparison of professional interpreters and interpreting students. *Interpreting*, 17(1): 1–25.

Fantinuoli, C. (2006). Specialized corpora from the web for simultaneous interpreters. In M. Baroni & S. Bernardini (Eds.), *Wacky! Working papers on the web as corpus*, 173–190. Bologna: GEDIT.

Fantinuoli, C. (2009). InterpretBank: Ein Tool zum Wissens- und Terminologie-management für Simultandolmetscher. In *Tagungsband der internationalen Fachkonferenz des Bundesverbandes der Dolmetscher und Übersetzer e.V. (BDÜ)* (pp. 411–417), Berlin.

Fantinuoli, C. (2012). *InterpretBank - Design and implementation of a terminology and knowledge management software for conference interpreters*. PhD thesis, University of Mainz.

Fantinuoli, C. (2016a). Revisiting corpus creation and analysis tools for translation tasks. *Cadernos de Tradução*, 36(1): 62–87.

Fantinuoli, C. (2016b). InterpretBank. Redefining computer-assisted interpreting tools. *Proceedings of the Translating and the Computer 38 Conference in London*, 42-52. Geneva: Editions Tradulex.

Fantinuoli, C. (2017). Computer-assisted interpreting: challenges and future perspectives. In G. Corpas Pastor & I. Durán Muñoz (Eds.), *Trends in e-tools and resources for translators and interpreters*. Leiden: Brill.

Feldweg, E. (1996). *Der Konferenzdolmetscher im internationalen Kommunikations-prozeß*. Julius Groos, Heidelberg.

Gerzymisch-Arbogast, H. (1996). *Termini im Kontext: Verfahren zur Erschließung und Übersetzung der textspezifischen Bedeutung von fachlichen Ausdrücken*. Tübingen: Gunter Narr Verlag.

Gerver, D., & Sinaiko, H. W. (1978). *Language interpretation and communication*. Boston: Springer.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training: Revised edition*. Amsterdam: John Benjamins.

Gorjanc, V. (2009). Terminology resources and terminological data management for medical interpreters. In D. Andres, & S. Pöllabauer (Eds.), *Spürst Du, wie der Bauch rauf-runter? Fachdolmetschen im Gesundheitsbereich. Is everything all topsy turvy in your tummy? Healthcare Interpreting* (pp. 85–95). München: Meidenbauer.

Johns, T. (1988). Whence and whither classroom concordancing. In T. Bongaerts (Ed.), *Computer applications in language learning*. Dordrecht: Foris.

Kalina, S. (1998). *Strategische Prozesse beim Dolmetschen*. Tübingen: Gunter Narr Verlag.

Kalina, S. (2006). Zur Dokumentation von Maßnahmen der Qualitätssicherung beim Konferenzdolmetschen'. In C., Heine, K., Schubert, & H. Gerzymisch-Arbogast (Eds.), *Translation theory and methodology* (pp. 253–268). Tübingen: Gunter Narr Verlag.

Kalina, S. (2007). "Microphone Off" – Application of the Process Model of Interpreting to the Classroom. *Kalbotyra*, 57(3): 111–121.

Kalina, S. (2009). Dolmetschen im Wandel - neue Technologien als Chance oder Risiko. In *Tagungsband der internationalen Fachkonferenz des Bundesverbandes der Dolmetscher und Übersetzer e.V. (BDÜ)* (pp. 393–401), Berlin.

Kiraly, D. (2000). *A social constructivist approach to translator education. Empowerment from theory to practice*. Manchester: St. Gerome.

Kucharska, A. (2009). *Simultandolmetschen in defizitären Situationen. Strategien der translatorischen Optimierung*. Leipzig: Frank & Timme.

McNamara, D., & O'Reilly, T. (2009). Theories of comprehension skill: Knowledge strategies versus capacity and suppression. In A. Columbus (Ed.), *Advances in psychology research* (pp. 113–136). Nova Science Publishers.

Moser-Mercer, B. (1992). Terminology documentation in conference interpretation. *Terminologie et traduction*, 2/3, Office des Publications des Communautés Européennes.

Pöchhacker, F. (2000). *Dolmetschen - Konzeptuelle Grundlagen und deskriptive Untersuchungen*. Tübingen: Stauffenburg Verlag.

Rossenbeck, K. (1989). Lexikologische und lexikographische Probleme fachsprachlicher Phraseologie aus kontrastiver Sicht. In M. Snell-Hornby & P. Pöhl (Eds.), *Translation and Lexicography* (pp. 197–211). Amsterdam: John Benjamins

Rütten, A. (2007). *Informations- und Wissensmanagement im Konferenzdolmetschen*. Frankfurt am Main: Peter Lang.

Seleskovitch, D., & Lederer, M. (1989). *Pédagogie raisonnée de L'interprétation*. Bruxelles: Didier Érudition Opoce.

Setton (1999). *Simultaneous interpretation. A cognitive-pragmatic analysis*. Amsterdam: John Benjamins.

Stoll, C. (2009). *Jenseits simultanfähiger Terminologiesysteme*. Trier: Wvt Wissenschaftlicher Verlag.

Thrane, T. (2005). 'Representing interpreters' knowledge: Why, what, and how? In V. Dam, H., Engberg, J., and Gerzymisch-Arbogast, H. (Eds.), *Systems and translation* (pp. 31–60). Berlin: de Gruyter.

Tripepi Winteringham, S. (2010). The usefulness of ICTs in interpreting practice. *The Interpreters' Newsletter*, 15, 87–99.

Will, M. (2007). Terminology work for simultaneous interpreters in LSP conferences: Model and method. In *Proceedings of the EU-High-Level Scientific Conference Series MuTra*, 65–99.

Will, M. (2009). *Dolmetschorientierte Terminologiearbeit. Modell und Methode*. Tübingen: Gunter Narr Verlag.

Zanettin, F. (2002). Corpora in translation practice. In *Proceedings of the Workshop Language Resources for Translation Work and Research*. Retrieved from http://www.lrec-conf.org/proceedings/lrec2002/pdf/ws8.pdf

Xu, R. (2015). *Terminology preparation for simultaneous interpreters*. Doctoral Thesis, University of Leeds.

---

[i] Although there are many similarities between different forms of interpreting (liaison, community, court interpreting, etc.), the present paper focuses on simultaneous interpreting (SI), in particular in the setting of technical and scientific conferences. Notwithstanding, many aspects dealt with in this paper can apply to other forms of oral mediated communication.

[ii]The relation between knowledge acquisition and the quality of the interpretation is analysed for example by Stoll (2009, p. 7). The author introduces the idea of "*kognitive Hypothek*": an insufficient preparation causes an increasing cognitive load during interpretation. This leads to a poor text analysis, memory activation and text production. As a consequence, interpreters need to apply "repairing strategies" with negative consequences on their performances. Efficient preparatory work can thus contribute to anticipate a part of the cognitive load from the interpreting phase to the preparatory phase (see Stoll, 2009, and Kalina, 1998).

[iii] For example, the author reports an improvement of term accuracy scores by 7.5% and reduction of the number of omission errors by 9.3%.

[iv] CorpusMode is released as freeware and is available at www.staff.uni-mainz.de/fantinuo

[v] CorpusMode uses Bing.

[vi] For a practical description of how to use concordancer in a discovery-oriented way, see for example Zanettin (2002).