Terminology Preparation for Simultaneous Interpreters

Ran Xu

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

The University of Leeds

School of Languages, Cultures & Societies

Centre for Translation Studies

September 2015

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given with the thesis where reference has been made to the work of others.

"Chapter 5 Evaluation of automatic term extraction" of the thesis includes work which has been published in the jointly authored publication: Xu, R. and Sharoff, S. 2014. Evaluating term extraction methods for interpreters. In: *Proceedings of the 4th International Workshop on Computational Terminology*. Dublin, Ireland, pp. 86-93.

The candidate, as the first author, has undertaken the research work. She was responsible for the design of the work, collecting and analysing data, preparing the manuscript and handling the revisions. The corresponding author is her supervisor, who provided the intellectual input, contributed in the design of the work, supervised and verified data collection and data analysis. He was also responsible for the manuscript correction and proof reading.

Acknowledgements

There are a number of people behind this piece of work who deserve to be both acknowledged and thanked.

First, many special thanks go to my principal supervisor, Dr. Serge Sharoff and my second supervisor, Dr. Gracie Peng. They took me under their wing, encouraged my research and allowed me to grow as a researcher. Their unwavering enthusiasm for my project has been a constant source of motivation. At the end of my PhD, this thesis was examined by Professor Franz Pöchhacker from the University of Vienna and Dr. Dragoș Ciobanu from the University of Leeds. I would like to thank both of them for the many valuable suggestions that improved this manuscript.

I wish to thank my MA supervisor Dr. Svetlana Carsten and all the interpreting colleagues at Centre for Translation Studies, the University of Leeds for their great support, feedback and encouragement throughout my PhD. It has been a real pleasure working with them in the past few years.

I show extensive gratitude to those MAITS students who gave generously their time and effort to participate in this study. Without their willingness to participate, the research would not have even been possible.

I have also been lucky enough to meet many friends in Leeds. My affection and gratitude go to Mika Takewa, Marilena Di Bari, Angeline Liu, and all my fellow PhD students, who have accompanied me on this journey. David Yuan and his family deserve my particular appreciation for providing me with accommodation in the last stage of my study.

Words cannot express how grateful I am to my husband and my parents for all the sacrifices that they have made on my behalf. Special thanks go to my mother and my father for giving me their relentless support. Their unconditional love helped me sustain the PhD process. I want to dedicate this thesis to my family.

Abstract

Simultaneous interpreting requires efficient use of highly domain-specific terminology in the working languages of an interpreter. By necessity, interpreters often work in a wide range of domains and have limited time to prepare for new topics. To ensure the best possible simultaneous interpreting of specialised conferences where a great number of domain-specific terms are used, interpreters need preparation, usually under considerable time pressure. They need to familiarise themselves with concepts, technical terms, and proper names in the interpreters' working languages.

There is little research into the use of modern terminology extraction tools and pipelines for the task of simultaneous interpreting. A few previous studies mentioned the application of corpora as potential electronic tools for interpreters. For instance, Fantinuoli (2006) and Gorjanc (2009) discussed the functions of specific online crawling tools and explored ways to extract specialised terminology from disposable web corpora for interpreters. However, there has not been any empirical study to test how term extraction tools and the use of corpora can help interpreters increase their preparation efficiency and how these technologies and practices influence interpreters' simultaneous interpreting performance.

This study investigates a corpus-based terminology preparation pipeline integrating building small comparable corpora, using automatic term extractors and concordancers. We compared and evaluated several term extraction and concordance tools for Chinese and English, and a single term extractor and a concordancer with comparatively better performance were selected to be used in the empirical study of this research. With training on how to use the tools for interpreting preparation, interpreters are expected to develop the skills to build their own terminology resources and activate relevant terms for specialised simultaneous interpreting tasks.

This study also investigates the effect of using the tools on trainee interpreters' performances by looking at the quality of their simultaneous interpreting outputs. For this purpose, we ran two experiments with MA trainee interpreters at the University of Leeds using different preparation procedures (and tools) to prepare for simultaneous interpreting tasks (English and Chinese, both directions) on two specialised topics: Seabed Minerals (SM) and Fast Breeder Reactors (FR). I also collected data from focus groups to investigate the trainee interpreters' views on the use of different procedures (and tools).

Our results suggest that the preparation procedure using both the term extractor (Syllabs Tools) and the concordancer (Sketch Engine) yielded better preparation results compared with a traditional preparation procedure. It helped improve the trainee interpreters' terminological performance during simultaneous interpreting by significantly increasing term accuracy scores by 7.5% and reducing the number of omission errors by 9.3%. On the other hand, terminology preparation (through using both the term extractor and the concordancer) is not a "magical cure" for all errors. Our data shows that the preparation procedure (and the tools) only helped to improve the students' holistic SI scores by 2.8% (but not yielding any statistical significance).

This thesis demonstrates that training on terminology preparation for technical meetings could be a useful supplement to the already existing professional interpreting training. It is important for both students and trainers to be aware that electronic tools, when used properly, can assist the interpreters' terminology preparation and achieve an enhanced performance. It also offers directions for further research in the application of modern term extraction technology for conference interpreters.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

LSP	Language for Special Purposes
POS tagging	Part-of-speech tagging
SWT/MWT	Single-word term/multi-word term
KWIC	Key Word in Context
Auto-list	Automatically-generated termlist
SI	Simultaneous interpreting
SL	Source language
TL	Target language
ST	Source text
TT	Target text

Chapter 1 - Introduction

1.1 Problem statement

Since simultaneous interpreting in specialised conferences requires remarkably efficient access to and retrieval of domain-specific terminologies between the two working languages in order to ensure smooth delivery, simultaneous interpreters must acquire both lexical information and extra-linguistic information on the topic to a large extent before the beginning of the conference under considerable time pressure. By necessity, simultaneous interpreters often work in a wide range of domains and have limited time to prepare for and activate domain-specific terminologies before interpreting.

The large-size reference term banks administrated by companies, governmental and international agencies, such as TERMIUM® and the United Nations' Multilingual Terminology Database (UNTerm) provide reliable terminology references for interpreters. However they may not be necessarily specific enough for the interpreters' individual preparation work. The simultaneous interpreters' terminology preparation nowadays is still very traditional, i.e. interpreters often have to spend a lot of time reading through meeting documents, and the actual collection of terms is still largely done manually.

A few previous studies mentioned the application of corpora as potential electronic tools for interpreters. For example, Fantinuoli (2006) and Gorjanc (2009) discussed the functions of specific online crawling tools and explored ways to extract specialised terminology from disposable web corpora for interpreters. However, there has not been any empirical study to reflect interpreters' perceptions of using corpora and corpus tools, nor any study to test how the tools can help interpreters increase their preparation efficiency and how they influence the interpreters' SI performance.

It seems that producing a relevant termlist is an important part of terminology preparation for simultaneous interpreters. However, interpreters sometimes may still find relevant terms not activated enough for their simultaneous interpreting tasks, and simply taking the termlist into the booth alone cannot guarantee spontaneous lexical access and retrieval of the terms. In fact, during SI, if the incoming terms in the SL are not familiar enough to be understood by the interpreter spontaneously, or the required term or concept in the TL does not surface fast enough, the interpreting process may break down due to the loss of valuable processing capacity and time. Therefore, specific steps need to be taken as part of advance preparation to ensure that interpreters increase the readiness of relevant terminology for both comprehension and production during interpreting. Concordance tools have been proven to benefit language learners in vocabulary acquisition. Using concordancers is therefore potentially helpful to consolidate the learning of specialised terminology for interpreters.

It is evident that the concept of using corpora and corpus tools is not familiar to average practitioner interpreters. Furthermore, the concept has not been well integrated in interpreting training so far. That is to say our future interpreters are not aware of or not familiar with these tools, either.

In summary, terminology preparation is important for simultaneous interpreters; furthermore, the use of corpora and corpus tools offers potential benefits to interpreters, yet so far it has not received enough attention from the interpreting academia. This study will investigate using comparable corpora and corpus tools for the simultaneous interpreters' terminology preparation, and demonstrate how a corpus-based terminology preparation pipeline might be useful for simultaneous interpreters. This research aims to contribute to a better understanding of the simultaneous interpreters' terminology preparation. It will also offer approaches to train interpreting students on how to use corpus tools to form and manage their own tailor-made terminology resources in their future work environments. It is expected to be helpful for interpreters' career development and for improving the training of interpreters.

1.2 Overview of chapters

Chapter 2 reviews literature relevant to the fundamental issues in this study. This chapter starts with introducing and clarifying several key concepts in this study, including terminology and specialised communication, users' expectations regarding terminology performance and terminology-driven vs. knowledge-driven preparation. It then provides summaries of major preparation models for interpreters and the approaches to increase terminology readiness, which provide a basis for assumptions about problems and interpreters' needs that will go into the design of this research. This chapter also provides a broad overview of the studies on using IT tools to assist interpreters' preparation. This chapter concludes with a rationale for this study, including the originality of the research design, the research goals and research questions that the study aims to address.

Chapter 3 aims to explore how to integrate the use of corpus tools into interpreters' preparation. The chapter focuses on investigating a corpus-based terminology preparation pipeline and the tools that can assist interpreters' preparation.

Chapter 4 introduces the main methodological approach of this study. It describes the groups of participants and the process of selection and design of experimental speeches and preparation materials. It then explains how an independent variable (preparation time) is controlled in this study. This chapter also presents the marking criteria applied to measure the dependent variables (i.e. participants' terminological and simultaneous interpreting performance). Then the tasks and procedures of the experiments are explained.

Chapter 5 presents the results of a numeric evaluation of three term extractors, and discusses technical challenges in term extraction in both English and Chinese. A single term extraction tool would then be selected to be used with the test groups of the experiments.

Chapter 6 presents the results of two experiments with the trainee interpreters using mainly two kinds of tools: an automatic term extractor and a concordancer. The main objective is to investigate whether the use of the two tools can influence the trainee interpreters' SI performance. This chapter also looks into the results from focus group discussions with the participants on the use of the tools.

Chapter 7 discusses the general patterns and implications of the experimental data and focus group discussions. This chapter aims to synthesise the key issues and the results obtained in the study by placing the findings in the context of theoretical and empirical frameworks reviewed in **Chapter 2** and **Chapter 4**. This chapter begins with a discussion of what auto-lists were like, what has been done with auto-lists and how shortlists were used in the experiments. It then focuses on term activation by different preparation procedures. This chapter also elaborates on the impact of challenges in rendering the source speeches (e.g. specialised terms, high density of information, and working into the B language) on the participants' interpreting performance. Based on the above discussions, this chapter concludes by addressing the pedagogical implication of the findings.

Chapter 8 reviews all the previous chapters. It summarises the findings and identifies contributions. The thesis ends by identifying limitations of the study and pointing out directions for future research.

Chapter 2 - Literature Review

This chapter will review the literature relevant to the fundamental issues in this study. I will start by introducing and clarifying the following topics: terminology and specialised communication (**Section 2.1**), the interpreting professional requirements and users' expectations of terminology performance (**Section 2.2**), and terminology-driven vs. knowledge-driven preparation (**Section 2.3**). I will then provide summaries of major preparation models for interpreters (**Section 2.4)** and the approaches to increase terminology readiness (**Section 2.5**). A further section discusses the basic assumptions about problems and interpreters' needs which were incorporated into the design of this research (**Section 2.6**). I will also give a broad overview of the studies on using IT tools to assist the interpreters' preparation (**Section 2.7**). This chapter concludes with a rationale for this study, including the originality of the research design, the research goals and research questions that the study aims to address (**Section 2.8**).

2.1 Terminology & specialised communication

Terminology is the study of the concepts and terms belonging to specialised languages. Terminology has both significant representative and communicative functions.

Terms fulfil a function of representation. The underlining theoretical model of terminology is the semiotic triangle (Ogden and Richards, 1923) which consists of an object, a concept and a term. A concept is the interpretation of a physical or an abstract object, and a term is the representation of the concept. As defined in ISO 1087-1:2000, a term is the verbal designation of a general concept in a specific subject field. The relationship between a term, a concept and knowledge could be summarised as this: a term is the formal representation of a specialised concept, which reflects specific or technical knowledge within a given subject field. Each knowledge structure consists of various interlinked concepts.

In addition, terms fulfil a communicative function. For a specific subject field, terminology is the set of units of expression and communication which allow specialised knowledge transfer. Terminology is a way of transferring and communicating (Cabré, 1996: 16-23). “Without terminology there is no professional communication and without professional communication there is no transfer of knowledge” (Zauberga, 2005: 107). The goal of terminology is to meet social and academic needs of the specialists, professionals, and the general public interested in specialised fields for various reasons (Mohammadi, 2013).

Terms do not come out of the blue, and they are not fixed, either. Terminology evolves over time, not only because of the new scientific discoveries and new artefacts, but also because of the need to make distinctions important for a particular theory. This development happens through the medium of language (Sharoff and Hartley, 2012: 319).

2.1.1 The users and creators of terminology

Terminology is “the set of terms of a subject field”. The direct users of terminology are specialists in each subject field. In specialised communication, specialists communicate with each other presupposing that they share a certain amount of information about the area of knowledge. They use terms to express themselves, exchange thoughts, and organise the structure of their disciplines. For them, terminology is a necessary tool for communication and an important element for conceptualising their own subject matter (Sager et al., 1980; Varantola, 1986; Cabré, 1998).

The other group of people who use terminology is professional communication mediators, for example, translators and interpreters, who facilitate communication for the specialised users (Cabré, 1998). As far as oral communication is concerned, when specialists participate in specialised international meetings, they may not understand or speak each other’s language, therefore conference interpreters are invited to facilitate the cross-language communication within specialised communities.

Translators and interpreters are not only end-users of the terminology products, but also creators of new terms in a language, too. They are influential in the development of terminology. At times, it happens to translators that they don’t find an equivalent term in specialised dictionaries and data banks; consequently, they create their own terminology mostly by an automatic transfer (e.g. transcription, semi-calque, and calque) which is different from terminologists’ methods who believe in semantic transformation and native original coinages, especially in small languages (Zauberga, 2005).

On the other hand, the interpreters' choice of terminology is based on the context and what their clients require. If there is a discrepancy between the terminology provided by terminologists and that of experts, interpreters opt for that of experts because experts are one of the user groups of interpreting services. Therefore, it can be claimed that interpreters' choice of terminology is more dynamic, unless the context and the audience are pre-determined (ibid).

2.1.2 The characteristics of LSP

Specialised language is also called LSP (Language for Special Purposes), which is defined as "a formalised and codified variety of language, used for special purposes with the function of communicating information of a specialist nature at any level in the most economic, precise and unambiguous terms possible" (Picht & Draskau, 1985:3). According to Berruto (1974), an LSP has a specialised lexicon, and this makes the language less accessible for those who do not have adequate knowledge of the field.

Cabré (1998: 68-77) discussed the characteristics of special languages for scientific and technical communication in terms of their users, communicative situations, and their main functions.

She specifies that the **primary users** of the special languages are professionals, while the recipients can be either experts or the general public, who passively receive special communication while acquiring knowledge. The **communicative situation** is usually formal and occurs in situations of a professional nature. The **basic purpose** of special languages is to inform and exchange objective information on a specialised topic. The **text types** generated in scientific and technical communication are primarily informative and descriptive in nature, yet may also attempt to persuade, but rather indirectly or implicitly by providing arguments, citing data, providing examples and explaining, etc. The special languages used in scientific and technical communication are precise, concise and impersonal. They tend to avoid ambiguities and redundancies, and they are not emotive. In addition, different from specialised scientific and technical texts in written articles or conference papers, the oral communication is more spontaneous.

She also discussed **the role of terminology** in an LSP. According to Cabré (1998: 47, 80-81), the use of terminology helps make communication between specialists more efficient. The use of terminology is the most important characteristic of specialised communication because terms differentiate special languages from the general language and also the various special languages from one another. Terminology contributes to the basic features of specialised texts: precision, concision, and suitability to the participants. In summary, special languages and the use of terminologies allow objective, precise and unambiguous exchange of information particularly between subject field experts and professionals.

This study will observe the trainee interpreters' terminology preparation and their terminological performance during the simultaneous interpreting of technical speeches. The above characteristics of oral scientific discourse will be taken into consideration in the process of selection and creation of source speeches used in the SI experiments of this study.

2.2 Interpreting quality criteria and user expectations regarding terminology use during SI

Having discussed the role of terminology in specialised communication, it is necessary to look into how the professional interpreters and the end-users of interpreting services actually perceive "terminology use" in simultaneous interpreting services. In other words, does correct terminology use really matter in the quality assurance of simultaneous interpreting? For this purpose, I focus my review of literature on one particular line of research, namely questionnaire-based surveys on interpreting quality criteria from both the interpreting practitioners' perspective and the users' perspective.

2.2.1 Interpreting quality from the practitioners' perspective

The first survey study of this kind is Bühler (1986). She asked 47 AIIC members which degree of importance they attributed to 16 linguistic (semantic) and extralinguistic (pragmatic) criteria on a four-point ordinal scale ranging from very important to irrelevant when sponsoring new applicants for AIIC membership. The result shows *that "sense consistency with the original"*, *"logical cohesion"*, *"correct use of terminology"* and *"fluency of delivery"* were the four top-rated criteria.

The latest replication of Bühler's (1986) study on quality criteria is a large-scale web-based survey among AIIC members conducted in 2008 by the Centre for Translation Studies of the University of Vienna (Pöchhacker and Zwishenberger, 2010). The relative importance of quality criteria in the 2008 survey was presented in the same order as in Bühler's (1986), but AIIC members responding to the 2008 survey seemed to be more demanding regarding the importance of form-based criteria such as *"correct terminology"*, *"correct grammar"* and *"appropriate style"*. The percentages of rating on the three items (in the category "very important") were noticeably higher than in Bühler (1986). In addition, nearly half of the respondents (43.3%) supported the idea that the importance of the quality criteria varied depending on the type of meeting or assignment, and correct terminology was considered the top priority for seminars/workshops.

2.2.2 Interpreting quality from the users' perspective

Some researchers have indicated that interpretation should be judged from the perspective of the audience (Séleskovitch, 1986:236; Déjean Le Féal, 1990:155). Since 1989, quite a number of survey studies have been done on user expectations of the conference interpreter's service (Kurz 1989, 1993, 1994, 1996; Gile 1990, Meak 1990; Ng 1992; Vuorikoski 1993, 1998; Kopczynski, 1994; Mark and Cattaruzza, 1995; Moser 1995, 1996, etc.). Correct use of terminology has been viewed as one of the important quality parameters from the users' perspective.

Kurz (1993, 1994, 2001) compared three different user groups (the participants in a medical conference, in a meeting of engineers on quality control and a Council of Europe meeting on equivalences) and there were high agreements between all groups on the importance of the following criteria: *"sense consistency"*, *"logical cohesion"*, and *"correct terminology"*.

Kopczynski (1994) conducted a survey among Polish users of interpreting services to determine their attitudes and expectations. All groups considered content more important than form, listing detailed content and terminological precision as their two top priorities. Wrong terminology was considered as the most irritating.

Mark and Cattaruzza (1995) conducted a survey among participants of five meetings where the simultaneous interpreting service was used. It was found that the ideal performance should, above all, be terminologically correct and informed, accurate and easy to follow. Pleasant speech rhythm and fluency were considered less important. Experienced users expected more, particularly with regard to the criteria *"informed"* and *"correct terminology"*.

Moser's survey study (1996) was funded by AIIC. 201 standardised interviews (using a questionnaire with both open-ended and specific questions) at 84 different meetings were carried out and this research found that terminological accuracy was considered more important in technical meetings than in general meetings, and was ranked higher by women than by men.

In the above studies, correct terminology use has been considered as one of the most important parameters for judging the quality of an interpreter's service by both professional interpreters and the users of the interpreter's service. The results indicate that the use of terminology in interpretation can influence the client's perception of interpreters. Inadequate or inconsistent terminology use in the target language may jeopardise the original message and produce a negative effect on the credibility of the interpreters.

In this study, the users of the students' simultaneous interpretation in the experiments are the three judges (interpreting trainers, who are also practitioners) rather than domain specialists. The terms used in the experimental speeches and their proper translations in the target language were discussed with domain experts before the experiments.

2.2.3 Error typologies in literature

As demonstrated so far, correct terminology use is important in interpreting quality assurance. The next question is how to assess terminological performance in SI. For the purpose of assessing terminological accuracy in the SI experiments, I reviewed the literature on frameworks used in interpreting and translation evaluation, with special focus on error classification schemes including interpreting error taxonomies, such as Barik (1975, 1994), Altman (1994), Napier (2002, 2004), and translation quality metrics and evaluation tools, such as SAE J2450 (2001), BlackJack in Eckersley (2002), MeLLANGE in Secară (2005), etc. Unfortunately, none of the existing typologies reviewed in the literature could alone fully cater to the specific needs of this study. The adaptation of the existing typologies seemed therefore necessary.

Some models (e.g. Barik and BlackJack) are based on the evaluation of every aspect of translation/interpreting, including general language use, terminology, accuracy, register and style, etc. However, not all the error types in the existing models are relevant to this study. What is needed in this study are quality metrics to be used to examine terminological accuracy rather than general accuracy in interpretation.

Some translation quality metrics/tools incorporate sub-categories for terminology errors. For example, **the BlackJack translation evaluation tool** specifies terminological errors as *"non-application of glossary term"*, *"inappropriate technical term in TT"*, *"inconsistent term in TT"* and *"wrong treatment of acronym/proper noun"*. Similarly, **the MeLLANGE error annotation scheme** defines terminological and lexical errors as *"incorrect (meaning inconsistent with ST)"*, *"false cognate"*, *"inconsistent with glossary"*, *"inconsistent with TT"* and *"user-defined error"*. In the above two metrics, terminological error types having similar impact on the TT are covered in great details, but terminological errors relevant to poor use of language (e.g. wrong collocation use and grammatical error) are not considered as terminological errors. In this study, in order to balance the total number of error types at a manageable level, the existing terminological error types having similar impact on the TT need to be incorporated into a more general category. Moreover, specific terminological error types relevant to poor use of language should be included as well to examine the students' terminological accuracy in the SI experiments.

In addition, some of the error categories in the existing translation quality matrix are written communication-specific, which is not relevant to interpreting – for example, "misspelling" and "punctuation errors" in **SAE J2450 Translation Quality Metric** (2001).

"Omission" has received attention in interpreting studies (e.g. Barik, 1975, 1994; Kopczynski, 1980; Cokely, 1992; Wadensjö, 1998; Napier, 2002, 2004). Although there is still no agreement as to how omission should be evaluated, it is generally agreed that an omission is "when information transmitted in the source language with one or more lexical items does not appear in the target language, which therefore potentially alters the meaning" (Napier, 2002:121). Some researchers highlighted that omission could be used as a strategy to achieve effective interpretation, and omission of terms is not necessarily an indication of poor translation/interpreting (Winston, 1989; Livingston et al., 1994). In this study, whether omission of a term is judged as an error does depend on whether the omission affects/alters the original meaning in the ST. If the omission of a term does not affect the original meaning, it is not counted as an error.

Chapter 4 - Methodology (Section 4.5.3) will further discuss six terminological error categories defined in this study.

2.2.4 The meeting as a genre

"A genre established within a particular community serves as an institutionalised template for social interaction, an organizing structure that influences the ongoing communicative action of members through their use of it within and across their community." Genres carry expectations about the purpose, content, participants, form, etc. of social interactions (Yates and Orlikowski, 2002).

The meeting is a genre of oral communication. Meetings that involve simultaneous interpreting services could be ranging from large to small-scale (e.g. international conference, summit, seminar, round table meeting, etc.), and their topics could range from general to specific (e.g. administrative, non-technical, technical, market research focus group discussions, etc.). These are just a few examples of meeting sub-genres.

This research focuses on specialised communication in technical meetings. However, the perceived technical meetings may include sessions only addressing general issues, and therefore the requirements for terminology precision in the TT within one specific meeting may differ. I want to find out what kind of technical meetings and which sessions in technical meetings have more specialised content, and therefore possibly require higher terminology accuracy. Such a discussion will point to a clearer direction for this research in terms of choice of themes and speeches to be used in a series of experiments, which are designed to test the impact of terminology preparation on trainee interpreters' SI performance.

2.2.4.1 The level of specialisation

Specialisation is an important factor. The level of specialisation of a scientific/technical discourse depends not only on the subject matter in question, but also on the recipients and the sender's communicative purpose. As discussed in Picht & Draskau (1985), specialised discourses can be divided into different levels of specialisation, with the highest corresponding to communication between experts, and the lowest to general-purpose information meant for laymen. For instance, a technical seminar on *climate change* involving mainly climatology experts would have more specialised content and denser terminology use than a world summit on *climate change* addressing world leaders and policy makers, even if their subject matter is, broadly speaking, the same. Similarly, a technical seminar among domain experts generally requires more expert knowledge or technical language and terminology use than a world summit.

2.2.4.2 Different sessions within a meeting

Moreover, each meeting is also a genre system. Yates & Orlikowski (2002) defined the meeting genre system as composed of the following genres: logistics, agenda, meeting itself, and report.

Logistics covers information exchange about time, place and who participates in the meeting. **Agenda** represents information stating the meeting objectives, which is normally covered in the *welcome and opening remarks*. **The meeting itself** is the interactions among meeting participants necessary to accomplish the meeting objectives. It is implemented in the form of *keynote address, presentation and panel discussion*, etc. **Report** is the meeting outcomes, serving two purposes: as meeting summary, and as a trigger for subsequent work. It can be included in *closing remarks and/or conference proceedings, etc*.

Apparently, even in a technical meeting, the opening remarks by a chairperson, or the sessions addressing logistics and agenda generally contain limited technical content, while the meeting itself (*keynote address, presentation and panel discussion*) contains more specialised content and requires more expert knowledge and specialised language, and is therefore more relevant to this research.

Based on the discussion above, this study will focus on keynote speeches and presentations in technical meetings among specialists. In **Chapter 3 - Methodology**, I will further discuss the choice of themes and speeches to be used in a series of experiments.

2.3 Preparation is indispensable

Professional conference interpreters work with LSP, of which technical conferences form a large part. According to Jiang's survey study (2013), technical conferences are seen as the most challenging among the different types of conference by professional interpreters: "Nearly 30% of respondents, while not preparing glossaries for other conferences, would do so for technical or unfamiliar ones". In technical meetings, interpreters are called to work for groups of specialists who do not share a common language, yet share knowledge and terminologies that are totally or partially unknown to laypersons or outsiders (e.g. interpreters). Therefore, in order to interpret specialised texts, interpreters must acquire sufficient knowledge of terminology and conceptual content.

As Seleskovitch (1998:58) pointed out in order to analyse what is said and to understand it, the interpreter must raise his/her level of understanding of the subject to a level which is distinctly higher than that of an ordinary educated person. Although it is not necessary to have the same depth of knowledge as an expert in the field, there is a minimum threshold that must be met. If the gap between the interpreter's knowledge of the subject and that of the expert is too great or has not been sufficiently reduced by knowledge acquisition before the meeting, it is impossible for the interpreter to grasp the rationale behind the speakers' words, and consequently the interpreter fails to communicate instantly the speaker's intended messages as accurately, faithfully, and completely as possible.

2.3.1 SI as working mode

Interpreting simultaneously means that the interpreter works in a soundproofed booth, conveying the speakers' ideas from source to target language almost simultaneously; the audience in the meeting room listens through headsets. As conference interpreting is a professional communication service, the simultaneous interpreters' job is to communicate instantly the speaker's intended messages as accurately, faithfully and completely as possible (SCIC[1]).

This is different from translation, which could be done over periods of hours, days and weeks; simultaneous interpretation is immediate, and the speakers' ideas must be conveyed from source to target language almost simultaneously (within seconds). Simultaneous interpreters have to perform their mental operations under severe time pressure. Unlike translators who can devote all their attention to comprehension at one moment and to reformulation at another, the interpreters' processing capacity is always shared at a given moment, and terminology availability requirements are noticeably higher in interpretation than in translation, both in comprehension and production (see Gile, 1995:132-141).

It is therefore generally accepted that (simultaneous) interpreters need to acquire additional (linguistic and specialised) knowledge to fill the gaps that they may have largely prior to the interpreting process, as this knowledge will need to be used live during interpretation and the interpreting process cannot be interrupted (not the same as in written translation). However, the acquisition process does not stop there: further information is added and new terms are acquired also at the conference venue. The interpreter's acquisition of information can thus be viewed as a continuous process (Moser-Mercer, 1992:509). In fact, preparation takes place not only before their interpreting tasks, but also during and after the specific assignments (Gile, 1995:147; Kalina. 2005:257).

Due to its limited scope, this study however only focuses on advance preparation for technical meetings which requires specialised language use in interpreting.

2.3.2 Terminology-driven preparation

As we have mentioned earlier (in **Section 2.22**), domain specialists at international meetings have a high expectation of terminology accuracy in the interpreting service. They may on the other hand have a wrong impression that

---

1 SCIC: Service Commun Interprétation-Conférences, is the European Commission's interpreting service and conference organizer who provides interpreters for around 11,000 meetings every year, thus being the largest interpreting service in the world. http://ec.europa.eu/dgs/scic/what-is-conference-interpreting/simultaneous/index_en.html

interpretation is simple word-for-word translation, and the only requirement for interpreters is the knowledge of specialised terminology. In other words, they may believe that merely knowing the technical terms is sufficient for interpreters to work. As a matter of fact, only knowing technical terms is generally of limited help for interpreters if they do not understand the concepts involved.

2.3.2.1 Terminology-driven vs. knowledge-driven preparation

Ideally, interpreters should have an in-depth understanding of the domain knowledge approaching that of a specialist. Some authors claim that in preparing for a specific meeting, interpreters should gain specialised knowledge through reading (systematically organised) reference series or introductory handbooks on the subject matter and digest the fundamentals (Séleskovitch and Lederer, 1989:87; Séleskovitch, 1998:56). Séleskovitch (1998) also suggested that specialised knowledge should be acquired in a rational, logical and coherent manner which can stand the interpreter better than rote learning.

However, in real practice, as it would be impossible for interpreters to acquire the similar amount of knowledge as their specialised speakers within limited period of preparation time, conference interpreters thus “have to be able to use individual texts (conference papers) as the principal source for preparation” before the conference takes place (Gile, 1995:147; Moser-Mercer, 1992:507) to acquire important concepts and ideas more effectively.

As discussed at the beginning of this chapter (in **Section 2.1**), terms are the formal representations of specialised concepts, which reflect specific or technical knowledge within a given subject field. One particular term is also related to other terms depending on its meaning within a concept system (Will, 2007:2). Therefore, terminology is the basis for the structure of thematically-specialised knowledge (Cabré, 1998:43).

The theoretical model of terminology (the relationship between term, concept and knowledge) allows us to further assume that through learning the most relevant terminology and concepts behind them, the knowledge system of a specialised field could also be generally formed for a learner. We need to be clear that interpreters do not have to be trained as nuclear physicist to interpret for a technical meeting on nuclear reactors. For interpreters, there is no 'royal road' to knowledge of a domain other than by starting with the terminology of that domain. That is the reason why this research supports terminology-driven preparation approach for interpreters.

2.3.2.2 Specialised professionals vs. interpreters in terms of knowledge acquisition

Different from specialised professionals, who acquire terminology naturally as their knowledge of a certain field advances, an interpreter's knowledge acquisition is performed primarily through learning terms in individual texts (conference papers). Terms/concepts are identified through reading the individual texts, and the knowledge systems are constituted according to their relevance to the specific interpreting task. In other words, the interpreters' specialised knowledge acquisition involves deliberate constitution of term-specific and superordinate knowledge structures (Will, 2007). Then, by putting together various bits and pieces of information, fragments of knowledge gradually blend into a more coherent picture for interpreters.

In summary, the interpreters' knowledge acquisition of specialised topics is term-based and job-oriented. It is geared towards the anticipated needs of the ensuing conference that they are going to interpret for. The interpreters' knowledge acquisition is mostly done before the interpreting assignment, and also updated and revised both during and after the assignment.

2.4 Models and procedures for the simultaneous interpreters' terminology preparation

Gile (2009:132) defined "terminology work" in the context of translation and interpretation as "the quest for information for the purpose of gaining better understanding of specialised terms and finding acceptable equivalents in the target language". Systematic studies have been made on the interpreters' terminology preparation, and the research findings were drawn not only from within interpreting studies itself, but also neighbouring disciplines, e.g. terminology, knowledge and information management, etc. This section will provide summaries of interpreters' preparation models and procedures discussed in the literature.

Gile (1995 & 2009) distinguished three different phases of terminology preparation for an interpreting assignment: advance, last-minute, and in-conference preparation. Kalina (2005) approached interpreters' preparation from the perspective of quality assurance. She proposed a model of interpreting conditions and processes covering the whole interpreting workflow (including pre-process, peri-process, in-process and post-process), as all these phases may have a significant impact on the interpreting output.

The focus and the depth of preparation in different phases/processes varies. “Advance or pre-process preparation is geared to the acquisition of subject knowledge, whereas last minute preparation often takes the form of a terminology search or merely the marking of manuscripts or presentation slides where available. As interpreters may be called, or assigned, at short notice, they often have to rely on last-minute preparation” (Kalina, 2015:319). “In-conference (in-process) preparation is necessary when manuscripts are not made available before a meeting, but only just before the speech is delivered” (ibid.) However, “due to situational constraints and the high cognitive load, opportunities for in-process terminology work are mostly limited to the occasional search for a specific term. It is therefore more essential that in-depth preparation, taking into account the conceptual background and specific context, takes place pre-process as well as peri- and post-process in order to ensure correct understanding and efficient retrieval and production” (Rütten, 2015).

In order to identify the key elements of terminology-driven preparation for this study, in the following sections (**2.41-2.43**), I will mainly compare three terminology preparation models focusing on “advance preparation” or “pre-process”, namely 1) Moser-Mercer’s terminology workflow, 2) Will’s knowledge management model and 3) Rütten’s information and knowledge management model.

2.4.1 Moser-Mercer’s terminology workflow (1992)

Barbara Moser-Mercer (1992) described an interpreter’s terminology workflow, which could be summarised into the following six steps:

Step 1: Request and receive conference documents from the clients.
Step 2: Read through all the conference documents provided and underline unfamiliar terms.
Step 3: Search for the equivalents in the other working language.
Step 4: Establish a bilingual termlist.
Step 5: Study the terms and the essential subject knowledge.
Step 6: Polish and update the termlist before, during and after the conference.

Table 1: Moser-Mercer’s terminology workflow (1992:508-509)

This workflow reflects a general process of terminology preparation by professional interpreters. However, the description of some key steps is still too simple and too vague. For example, **Step 5** suggests terms should be studied together with essential subject preparation, but it is still not clear how terms and subject preparation is carried out and whether there is a more detailed structural process within this step.

Moreover, this workflow also has scope for improvement. For example, in **Step 2, "**reading through all the documents to search for relevant terms" and copying them from the texts onto a list are quite time-consuming. We could possibly find a way to automatise the term extraction process in order to make the whole workflow more efficient.

In summary, the nature of terminology and knowledge acquisition is left largely untouched by this workflow. Moreover, the workflow also needs to evolve by keeping up with the development of technology, so that the tedious preparation task could possibly be accomplished with a higher level of efficiency. This, in turn, may enhance the interpreters' performance, resulting in better interpreting quality.

2.4.2 Will's knowledge management model (2007)

Martin Will (2007) approaches the interpreters' terminology preparation from the perspective of knowledge management. His model describes a more detailed structural process within the interpreters' terminology preparation. He suggested that knowledge management within an interpreting assignment include three stages, - reception, transfer and reproduction -, which could be further divided into ten steps. They are summarised in the following table.

Reception stage
1. A specific term is identified in a text (conference document).
2. The specific corresponding concept is identified.
3. The textterm is related to reference term system to form a *Terminological Knowledge Entity (TKE)*.
4. All the *TKEs* are pulled together into an unstructured *Terminological Knowledge Constellation (TKC)*.
5. A knowledge system is constituted. It consists of functionally interrelated and hierarchically organised holemes and sub-holemes.
6. Different holemes and subholemes in the knowledge system are referred back to the

corresponding textterm in order to understand them. This marks the end of the reception phase.
Transfer phase
7. A corresponding knowledge system in the target language is constituted along the same principles as in the source language.
8. Holemes and subholemes of both knowledge systems are compared with respect to their conceptual and functional content.
9. Adaptations have to be made in the target language in case of differences between the two languages. This marks the end of transfer phase.
Reproduction stage
10. The interpreting process itself represents the reproduction stage.

Table 2: Knowledge management within an interpreting assignment (Will, 2007:7-8)

This model for the first time explains how terminology is acquired and the knowledge system is constituted within the interpreters' preparation. From the model above we could see that the interpreters' preparation is a complex knowledge-intensive process. This "self-organised learning" starts with the identification of terms and concepts, followed by the forming of a hypotheses on a *Terminological Knowledge Entity (TKE)* and *Terminological Knowledge Constellation (TKC)*; then, through reference, logical analysis and inferring, a hierarchically-organised knowledge system is formed and structured. It is then ready to be applied to understand the terms in the original texts better. This acquisition result also needs to be transferred and updated in the other language before the acquired knowledge is reproduced in the interpreting setting.

Learning terms and constituting relevant knowledge systems will for sure help the interpreters' comprehension of the ST during interpreting. However, it does not guarantee that the terms are ready for production in the TL (especially in simultaneous interpreting mode). For example, during SI, if the required term or concept in the TL does not surface fast enough, it is likely that the interpreting process may break down due to the loss of valuable processing capacity and time. Therefore, this Model needs to add an extra step, focusing on increasing readiness of relevant terms for production in SI. (The concept of terminological readiness will be discussed in **Section 2.5**.)

2.4.3 Rütten's information and knowledge management model (2003 & 2015)

Like Will (2007), Anja Rütten also discussed the interpreters' terminology preparation in the context of information and knowledge management. She contributed to an entry "Terminology" in the forthcoming *Routledge Encyclopaedia of Interpreting Studies* (Pöchhacker, 2015). According to Rütten (2015:416), the interpreters' information and knowledge management involves "three levels of 'enrichment', from data to information to knowledge".

> "The first level involves rather mechanical retrieval of all sorts of data (manuscripts, presentations, glossaries, etc.). The second level consists of extracting from the 'raw material' the elements which are potentially relevant for the interpreting assignment (terms, definitions, context), thus turning data into information, and organising it to ensure that it is visible or retrievable when needed. The third level involves the interpreter's personal knowledge. It consists of checking which relevant items of information are already actively known by the interpreter (i.e. retrievable from memory even under cognitive load) and memorising the most relevant previously unknown information before the conference…" (ibid: 416-417).

Rütten (2003) also investigated the basis of optimum information and knowledge management for interpreters. She suggested a conceptual model, consisting of five modules: "online+offline research module, document management module, terminology extraction module, terminology management module and trainer module" (ibid). In her "terminology extraction module", termlists are expected to be extracted (semi-) automatically and then to be revised by their users, the interpreters, who can concentrate on those terms which are relevant and important to remember. This idea offers a solution to the time-consuming manual selection of terms from texts, the problem we have discussed about Moser-Mercer's workflow in **Section 2.4.1**.

However, Rütten's study (2003) only provided a conceptual model; it neither tested the functions of the term extraction tools, nor discussed the interpreters' perception of the usefulness of the automatically-generated lists in their preparation for interpreting tasks.

Apart from the "terminology extraction module", Rütten's model also contains a "trainer module", which is supposed to help systematic memorisation of terminology. It basically provides a testing environment in which “unknown or problematic terms are presented or tested automatically at regular intervals”. This module could be a useful supplement to Will (2007)’s Knowledge Management Model. As Rütten (2003) put it, “interpreters will never be machines spitting out word equities; however, in order to retain very technical terminology within (and for) a short period of time, some automation of word pairs may be necessary”. (There will be more discussions on the concepts of terminological readiness and automaticity in **Section 2.5.**)

2.4.4 Key elements of terminology-driven preparation

As we have discussed so far, although there is no universally-accepted mode of preparation, it is generally agreed that the interpreters’ preparation is indispensable, and that “terminology is a tangible vehicle for the construction of the conceptual knowledge that supports interpreting” (Jiang, 2013).

Based on the models we have reviewed in **Section 2.4**, I could therefore summarise that terminology preparation for a specific interpreting assignment (especially technical meetings) should include the following four key elements to ensure good interpreting performance and proper terminology use in interpreting:

a. Establishing the interpreter’s own termlist for a specific interpreting assignment, either through manual selection or automatic extraction of terms;
b. Checking information to gain a better understanding of the terms and relevant concepts, and building a hierarchically organised knowledge system of functionally interrelated terms and concepts;
c. Finding acceptable equivalents to the terms in the target language;
d. Enhancing the interpreter’s readiness of terminology access and retrieval for both comprehension and production.

2.5 Approaches to increase terminology readiness

Now I will focus on the last key element **(d)** of the interpreters’ terminology preparation to further explore ways to increase the interpreters’ readiness of terminology access and retrieval.

As far as interpreters are concerned, terminological readiness is for both comprehension and production. Gile (1995 & 2009) suggested the **concept of availability/readiness** in language comprehension and production, according to which low availability in speech production results mostly in pauses and hesitations which slow down the utterance. It can also lead to lack of accuracy in expressing ideas and various grammatical and other errors. During simultaneous interpreting, when comprehension availability is low, the speed of processing is slowed down and lag accumulates. As a result, the maximum storage capacity of the working memory can be exceeded rapidly and if it is saturated at the time the speaker utters the next speech segments, either previously-heard sounds or incoming sounds cannot be fully processed to yield meaning, and the corresponding speech segments are not understood (Gile, 2009:222-225). This has far-reaching implications for terminology preparation for simultaneous interpreting.

In this section, I will look into some neighbouring disciplines, e.g. psycholinguistics and vocabulary acquisition – especially cognitive psychological studies concerning vocabulary acquisition – to review the nature of lexical development and various ways to promote lexical fluency/readiness in language learning.

2.5.1 Learning in context

In both first and second language learning, vocabulary acquisition develops when learning in context, in circumstances that make possible linking the new vocabulary to other terms and prior knowledge (Nagy and Herman, 1987; Nation, 1993; Segalowitz and Gabonton, 1995). Sternberg (1987)'s experimental study proved that the use of context resulted in superior learning of new words compared with simple vocabulary-memorisation training. The implication of this is straightforward for interpreters' terminological acquisition for specific interpreting assignments. The preparation documents sent to interpreters before a conference contain rich contexts where terms are used in genuine communication.

Learning terms in context could help the interpreters understand the meaning of the terms and how the terms are used (e.g. their collocations, grammatical, stylistic and pragmatic information, etc.). Through observing examples of the terms that are hard to interpret or that are not included in standard bilingual dictionaries, interpreters could deduce their meaning or understand nuances in their use, and identify suitable target terms accordingly. Learning terms in context also provides an environment to establish varied and rich links to other terminological items and concepts. It is important to note that interpreters need to interpret both the terms and the contexts in their interpreting assignments; therefore learning terms in context provides similar retrieval circumstances that interpreters will encounter later.

2.5.2 Deep semantic processing

Psychologists Craik and Lockhart (1972) proposed that memory is a by-product of the depth of processing of information. Shallow semantic processing like rote repetition/mechanical vocabulary memorisation does not lead to long-term retention of the vocabulary, whereas deep processing strategies such as semantic elaboration do achieve better vocabulary acquisition results (Craik & Lockhart, 1972; Hashtroudi, 1983; Ellis, 1995). This highlights that deeper semantic processing of information can aid memory.

Research on lexical semantics suggested that one’s lexicon is an interconnected meaning system, organised by various kinds of semantic relations. The most frequent relations are superordination (hypernyms), coordination (co-hyponyms), synonymy/antonyms and collocation (Aitchison, 1994, 2012). It is also reported that the richer and more varied the information linkages to a particular vocabulary item, the greater the chances of fast and accurate retrieval of the item (Segalowitz and Gabonton, 1995).

These explanations of semantic processing are useful for interpreters’ terminology acquisition. Semantic processing of relevant terms should by all means be deepened. The interpreters’ terminology preparation is often carried out under considerable time pressure before the beginning of the conference. Therefore, interpreters would be better able to balance and incorporate semantic processing with other activation activities. It would be beneficial for interpreters to make sense of the relationships between relevant terms through learning in context and explicitly establishing links to one another by grouping them together in termlists, from which they could also mentally prepare short talks on sub-topics. This learning process is like “weaving the knowledge web”.

2.5.3 Passive vs. active activation

Gile (2009) uses *the Dynamics of the Gravitational Model* to illustrate that different words have different levels of availability for an individual, ranging from those words which can be retrieved instantaneously and effortlessly from long-term memory to those to be “known” but unavailable at a given moment. The availability of the words is not static, but rises through activation and drops through deactivation. The Model comprises five rules:

“***Rule 1: The Centrifugal Principle*** If not simulated, words and rules tend to drift away from the center of the system.
Rule 2: Centripetal Effect of stimulation When used, words and rules tend to move inward.
Rule 3: Stimulation frequency and the centripetal effect The more frequently words and rules are used, the stronger the centripetal effect.
Rule 4: The centripetal effect of active vs. passive stimulation Active stimulation of a word or rule has a stronger centripetal effect than passive stimulation.
Rule 5: The escort effect and interference effect The centripetal migration of a word or rule generates the centripetal migration of other words or rules associated with it. ”

Table 3: Dynamics of Gravitational Model (Gile, 2009:229-231)

One of the most relevant concepts here is the two types of stimulation in *Rule 4: “passive stimulation”* (through hearing or reading) and *“active stimulation”* (through speaking or writing). Active stimulation (through speaking and writing) is more effective in increasing the availability of words than passive stimulation (through reading and hearing). This effect is well-known in foreign language teaching. As far as interpreters are concerned, most of their terminological preparation before a conference is normally through passive stimulation (i.e., reading instead of speaking). Since the efficiency of reading (the preparation documents) is limited in increasing the availability of terms according to the above model, more active stimulation approaches (speech production) should be employed. For example, interpreters could practise constructing and saying aloud meaningful sentences by using relevant terms (in both source and target languages).

2.5.4 Repetitive stimulation

In addition, the stimulation should be repeated in order to consolidate the initial vocabulary acquisition. As stated in Gile’s *Dynamics of the Gravitational Model*

(Rule3), the more frequently words are used, the more activated the words become (Gile, 1995, 2009). Moreover, the repetition should be done with context and deep semantic processing. Rote-memorisation and automatic repetition does not appear to be efficient in increasing availability (also see **Section 2.51 & 2.52**).

A relevant notion is “spaced repetition” to improve vocabulary learning. The idea behind it is that memory loss slows down considerably when a memorised item is reviewed at appropriate intervals. This idea has been implemented in a number of computer-assisted language learning solutions, enabling automated scheduling, presenting and testing of vocabulary at regular intervals, for instance, *the trainer module* of Rütten’s information management model (2003) (in **Section 2.44**) and some online flashcard applications such as Anki[2].

2.5.5 Automaticity

If the previous lexical activation conditions (learning in context, deep semantic processing, and active stimulation in consistent repetition) can be met, automaticity of terminology access and retrieval could possibly be achieved. Some studies in psychology and cognitive science mentioned automaticity of lexical processing, which means as a result of extensive practice, vocabulary performance becomes faster, more accurate/reliable and relatively effortless.

Automaticity of lexical processing is central to language fluency in language production. It can be understood as economical/efficient processing, and is beneficial for one’s overall language production (Segalowitz and Gatbonton, 1995).

Automaticity essentially involves a reduction in the consumption of attentional resources; consequently, more performance is automatised, and greater processing resources are available to focus on other aspects of language production, for instance, integrating information, the planning of future utterances, etc. (Perfetti, 1985; Segalowitz, 1986; Segalowitz et al., 1991). For simultaneously interpreting technical texts, automaticity of terminology access and retrieval could save both valuable time and processing capacity, and would benefit the whole interpreting process.

This study will adopt the above approaches and conditions which promote

---

[2]Anki is a free online flashcard application, oriented toward language-learning and other disciplines requiring memorisation. [Accessed 22 June 2015]. Available from: http://ankisrs.net/

lexical development to acquiring specialised terms for interpreting purpose. I will further discuss how particular preparation activities following the above approaches are implemented in my empirical study in **Section 7.2.1**.

2.6 Interpreters' specific needs

Based on the discussion so far, this study identifies four specific needs of interpreters regarding their terminology preparation for simultaneous interpreting assignments.

2.6.1 Quick term extraction

As it is mentioned in the previous section, interpreters may have to study all the meeting documents to get their terminology lists done prior to conferences. The actual collection of terms is still mostly done manually nowadays. However, reading all the meeting documents (copying and pasting terms) takes time. With limited preparation time, it would be helpful if assignment-based termlists could be extracted automatically, and interpreters could prioritise their terminology study on the conference subjects. It may save time and increase the efficiency of preparation as a whole.

2.6.2 Increasing the collection of useful documents

In fact, only focusing on the terms appearing in the conference documents may not be enough sometimes. For instance, quite often many other relevant domain-specific terms (not included in the documents provided by the conference organisers) are actually used by the speakers in free discussion and Q&A sessions. Therefore, interpreters should ideally have access to as many relevant texts (containing potentially relevant terms) as possible, but they only have very limited preparation time.

A solution to the above problem is that interpreters could compile their own electronic specialised corpora (a machine-readable collection of representative texts in a certain domain). Corpus compilation can be done both manually and with the help of automatic corpus compilation tools (e.g. web crawlers) (see **Chapter 3, Section 3.2**). Once the corpus is compiled, it is ready for further processing by automatic term extraction tools (see **Chapter 3**, **Section 3.3**).

2.6.3 An environment for deeper cognitive processing and adequate activation of terms

When interpreters learn something new from reading through the conference documents within limited time, they tend to "gobble up" unfamiliar terms and background knowledge without the chance of further digesting them well. Due to the lack of deep cognitive processing of those terminologies, from time to time, interpreters find that the terms they have included in their term lists are still not quite ready yet for their comprehension and production during interpretation. Therefore, specific steps need to be taken to ensure interpreters increase the readiness of those relevant terminologies for both comprehension and production during interpreting.

The solution proposed earlier (i.e. automatic corpus compilation & automatic term extraction) may shorten the time of searching for relevant terminologies in the first place. Valuable preparation time saved, interpreters can use the automatically-generated termlists as index and check the background information of unfamiliar terms from the list and check their equivalences in the TL. Interpreters can also use concordance tools (see **Chapter 3**, **Section 3.4**) as navigational aids for close reading and consolidating their learning of keywords in contexts.

2.6.4 Better terminology management for future use

Quite often simultaneous interpreters may find that those terms they prepared are easily forgotten soon after the particular assignment. After the conferences, busy interpreters would rather prepare for their coming assignments, leaving nearly no time to update the existing termlists after the conference, losing a chance to further consolidate the terms they have prepared, not to mention cross-referencing them with other terms and updating new information about the terms they have learned from the conference. In fact, terms and their relevant information (e.g. collocations, translation, context, etc.) can be re-used for future assignments on the same or similar topics. If the terms can be stored in a database, be referred to, updated and retrieved easily, it will save a lot of time for interpreters' preparation in the future. The personalised database is a useful learning resource throughout an interpreter's career.

There are several terminology management tools (see **Section 2.7.2**) that interpreters are possibly not aware of, and therefore, are not able to resort to. These tools can help interpreters create and manage terminology entries in their own termbases. Interpreters can further update, edit and review terms, and add more customised fields to the term entries (e.g. terms themselves and their variants, typical collocations, fixed expressions, possible translations, examples of their use, contexts, notes, etc.).

2.7 Interpreters' preparation using IT tools

Having discussed the interpreters' needs in their terminology preparation for simultaneous interpreting assignments, now I would like to provide an overview of research on using IT tools for interpreters' preparation.

2.7.1 Using corpora in interpreters' preparation

> "A corpus is a collection of machine-readable, authentic texts, sampled to be representative of a particular language or language variety" (McEnery et al., 2006: 5). "There are two broad types of corpora in terms of the range of text categories represented in the corpus: general and specialised corpora. General corpora typically serve as a basis for an overall description of a language or language variety. In contrast, specialised corpora tend to be domain or genre specific" (ibid., 15).

Compared with general corpora, specialised corpora can be used to address specific needs of interpreters in a particular domain more directly than general corpora. Specialised corpora can provide interpreters with information about authentic language use in specific domains. Several previous studies mentioned the use of specialised corpora as potential electronic tools for the interpreters, for instance, Rütten (2003), Fantinuoli (2006), Gorjanc (2009), etc.

Gorjanc (2009) discussed the use of corpus compilation tools to establish disposable web corpora for interpreters. He placed emphasis on using the WebBootCat tool for compiling specialised corpora for medical interpreters.

Fantinuoli (2006) compared term extraction results from both manually-collected corpora and automatically-crawled corpora from the Internet. The result showed that term extraction from manually-compiled and automated web-derived corpora led to comparable results. "Given how time-consuming it is to build a corpus by hand, automated web-based corpus construction is very promising way to reach good result with limited efforts" (ibid., 188).

Rütten (2003) suggested automatic term extraction for interpreters; however, her study neither tested the functions of any term extraction tools, nor further discussed the interpreters' perception of the usefulness of the tools.

The above studies have shown that using corpora could assist the interpreters' preparation in the following ways: first, domain-specific corpora could be built fairly quickly, and termlists can be generated automatically for interpreters; second, through using corpora, interpreters can learn terminology in authentic contexts; third, it is easy for interpreters to search, select and sort terminological data in corpora.

However, there are still many unanswered questions regarding using corpora in the interpreters' preparation, for example, how accurate automatic term extraction can be, whether term extraction tools can perform consistently in different languages, whether using corpora and corpus tools can make interpreters' preparation easier and more efficient, and whether using the tools can help interpreters perform better in simultaneous interpreting.

This study attempts to answer some of the above questions. We will test several term extraction tools for English and Chinese to see which tool offers comparatively better performance. We will discuss the trainee interpreters' perceptions of the usefulness of the automatically-generated lists in their interpreting preparation (**see Chapter 5**). We will also test whether using corpora and corpus tools can influence the trainee interpreters' SI performance (**see Chapter 6 & 7)**.

2.7.2 Terminology management tools

Terminology management tools, with a different focus from term extraction and corpus tools (as mentioned in **Section 2.71**), are specialised in compiling, storing, managing, importing and exporting glossaries/termbases[3], and allow looking up terms and term-related information (Durán-Muñoz, 2012). Various terminology management tools (e.g. SDL MultiTerm) have been developed for translators and terminologists as standalone tools to manage and control terminologies. Translation-

---

[3] A termbase is "an electronic collection of structured term entries in the form of individual or client-server databases of a relatively smaller size and with a more limited audience than a term bank" (Allard, 2012: 16).

oriented terminology databases can also be integrated with Translation Memory (TM) in computer-aided translation (CAT) systems to make sure terms are translated consistently for translators.

Terminology management tools have been commercially available on the market since 1980s (De Camp and Zetzsche, 2014:380). Moreover, the use of the tools and termbases in translation practice and training has been discussed extensively in literature (e.g. Wright and Wright, 1997; Bowker and Pearson, 2002; Bowker, 2003; Jaekel, 2000; Jaatinen and Jääskeläinen, 2006, etc.).

However, the use of term management tools by interpreters is very low. According to survey studies reflecting the general practice of professional (simultaneous) interpreters in terminology preparation, for example, Moser-Mercer (1992) and Jiang (2013), many practitioner interpreters still rely on fairly traditional resources, preparing their termlists by using loose paper or Word software. Only very few interpreters use glossary software or terminology management tools (e.g. Interplex and SDL Multiterm) for terminology documentation.

The discussion of developing the tools for interpreters' needs has been only a recent phenomenon. At the start of computer-assisted termbank development, Moser-Mercer (1992:507) rejected the assumption that "interpreters' needs are identical to those of translators and terminologists". She surveyed how conference interpreters manage terminology documentation and offered some guidelines for developing tools specifically for interpreters.

Rodríguez and Schnell (2009) reviewed the findings from two surveys conducted at Bologna University and the Sprachen & Dolmetscher Institut (SDI) in Munich. The two surveys focused on the use of computers and terminology management software in the interpreters' booth. The survey results indicated that many interpreters still use traditional tools (such as hard-copy glossaries with personal notes and standard reference books). According to Rodríguez and Schnell (2009), interpreters were disinclined to introduce computerised tools into their professional practice because of three possible reasons: there was no need for the tools; the tools for interpreters on the market were inadequate; or the interpreters had little knowledge of the tools available on the market.

Mohammadi (2013) presented a synopsis of the previous studies regarding the needs of different users of terminology management tools and termbases (with specific focus on translators and interpreters). He shared the same findings that conference interpreters expect distinct qualities from their termbases before, during, and after a conference. Their needs are different from translators and terminologists, for example, on speed of consultation, possibility of updating the terminology record in the interpreting booth, considerable freedom to define the basic structure of term records, and multiple ways of filtering data, among others.

Costa et al. (2014) provided a most up-to-date overview of current standalone terminology management tools for interpreters and conducted a comparative evaluation of eight terminology management tools[4] available on the market. The evaluation was on the completeness of features offered by the tools. The results showed that SDL MultiTerm and Intragloss are the two tools achieving the highest scores in the evaluation. SDL is the most expensive tool, and has been developed for more than twenty years. Intragloss has been developed by interpreters for interpreters and thus corresponds better to their needs. There are other web-based applications (e.g. Interpreters' Help), which can also be used for the same purposes (Rütton, 2014).

2.7.3 Reference term banks

Reference term banks are yet another type of terminology resources available for both translators and interpreters. According to Allard (2012), a (reference) term bank is an enormous termbase addressing a wide range of heterogeneous audiences encompassing companies, language learners, or even the general public. It is usually administered by major companies and governmental and international agencies. Some examples of reference term banks are TERMIUM®, InterActive Terminology for Europe (IATE) and the United Nations' Multilingual Terminology Database (UNTerm).

UNTerm[5], for instance, provides terminology in subjects relevant to the work of the United Nations. Information is provided in the six UN official languages, and there are also entries in German and Portuguese. The database is mainly intended for use by language staff of the United Nations to ensure accurate and consistent usage

---

[4] Costa et al. (2014) evaluates eight terminology management tools, ie. Intragloss (2014), InterpretBank (2014), Intraplex (2012), SDL MultiTerm (2013), AnyLexic (2009), Lingo (2011), UniLex (2007) and The Interpreter's Wizard (2011).

[5] [Accessed 22 June 2015]. Available from: http://untermportal.un.org/portal/welcome

in documents published by the Organisation. When a term is typed into its search engine (with a number of filters available to widen/narrow the search), it then yields relevant records in the term collections. The obvious advantage of UNTerm is the wide coverage of topics and accuracy of the terms and their translations.

I tried UNTerm's query function on two specialised topics (fast reactors and deep seabed minerals). The two subjects have been discussed in a series of workshops and conferences organised by relevant UN agencies. When typing "fast reactor" and "seabed mineral" in the query, UNTerm yields 37 and 32 records respectively. The search results include useful technical terms (e.g. *liquid-metal-cooled fast reactor*, *polymetallic sulphides* and *hydrothermal fluid*) and UN-specific terms (e.g. *International Seabed Authority* and *International Project on Innovative Nuclear Reactors and Fuel Cycles*).

However, many country or project-specific terms (e.g. *China Experimental Faster Reactor*) are not included in the UNterm database. I also tried searches on technical topics not specifically relevant to the UN (e.g. "cloud computing"). Much fewer search results were found in the database (e.g. there were only five results on "cloud computing").

It is evident that the large-size term banks administrated by companies or governmental and international agencies provide reliable terminology references for interpreters, yet they may not be necessarily specific enough for the tasks interpreters are involved in, and therefore may not meet the exact needs of individual interpreters. In this sense, interpreters may find terms directly extracted from conference documents more relevant to their preparation for specific assignments. Therefore, this study will focus much attention on procedures and tools that can assist interpreters to form and manage their own tailor-made terminology resources in their work environments.

The above sections (**2.7.1-2.7.3**) give an overview of studies on using IT tools for interpreters' preparation. In the rest of this study, I will only focus on the use of corpora and corpus tools in interpreters' preparation before their interpreting assignments.

2.8 The research goals and the originality of this study

2.8.1 Originality of this study

In this section, I will present the originality of this study in the following four areas.

2.8.1.1 User's investigation

Much of the research to date on using corpora and corpus tools to assist with the interpreters' terminology preparation has focused primarily on conceptual ideas and functions of specialised tools, for instance, Rütten (2003), Fantinuoli (2006) and Gorjanc (2009). However, not much attention has been paid to the actual user experience from the perspective of interpreters. Furthermore, there has not been any empirical study to test whether using corpora and corpus tools can help interpreters increase their preparation efficiency, and whether using the tools in preparation may have any impact on simultaneous interpreting performance. This study will investigate the effect of the tools on the trainee interpreters' performances by looking at the quality of their simultaneous interpreting outputs.

Díaz-Galaz (2012) and Díaz-Galaz et al. (2015) focused on the role of advance preparation in the simultaneous interpreting of scientific speeches. In her study, the experiment condition (30-minute study session of related materials) mainly followed Moser-Mercer (1992)'s traditional preparation procedure. This study aims to test the impact of three different preparation procedures (i.e. traditional preparation procedure, preparation with only term extraction tool, and preparation with both term extraction and concordance tools) on the trainee interpreters' simultaneous interpreting performance.

2.8.1.2 User's evaluation of automatic term extractors

The previous studies, such as Fantinuoli (2006) and Pignataro (2012) mentioned automatic term extraction. Pignataro (2012) aimed to use Word Smith Tools to detect from specialised texts as many noun phrases as possible for interpreters; however, her study did not include any form of evaluation of the accuracy and reliability of the automatically-extracted lists. Fantinuoli (2006) evaluated the level of specialisation and well-formedness of automatically-generated termlists; however, the evaluators in his study were terminologists rather than interpreters. However, Fantinuoli's study did not include any investigation into integrating term extraction evaluation into the practice of interpreting preparation. This study will involve trainee interpreters as end-users to compare and evaluate several term extraction methods and tools for Chinese and English in real interpreting assignments. A single tool with comparatively better performance will be selected to be used in the test groups of the SI experiments.

2.8.1.3 A scoring system on terminology accuracy in SI

This study has developed a scoring system on terminology accuracy for evaluating terminological performance in SI based on some existing interpreting/translation quality assessment systems, for instance, Barik (1971)'s categories of departures of translation, SAE J2450 translation quality metric (2001), BlackJack (2002), MeLLANGE (Secară, 2005), etc.

This scoring system, highlighting lexical accuracy in real communication, incorporates six error types and two degrees of departures (minor/serious) from the terms in ST. In addition, there are instances when terms in a sentence are all interpreted correctly, however the meaning at sentence level does not make sense. This scoring system guarantees that only a full score is given when the term itself is interpreted correctly and in the right context.

2.8.1.4 Gaps in interpreting training

Gile (2009:149-151) gave suggestions on how to raise trainee interpreters' awareness of ad-hoc knowledge acquisition in interpreting training. For example, he called for both demonstration and exercise to show that relevance of preparation to the interpreting tasks. He did mention that the use of the Web is a useful ad-hoc knowledge acquisition strategy, yet there was no mention of using corpus collection or automatic term extraction.

Gorjanc (2009) outlined a sequence of learning about online terminology resources and tools for interpreters (including corpus compilation, corpus analysis and terminology management), but his study did not discuss how relevant learning activities could be implemented in the existing professional interpreting training programmes.

Several recent research papers discussed teaching terminology and relevant electronic tools within the context of translation degree programmes (e.g. Sánchez-Gijón, et al., 2009; Montero Martínez and Faber, 2009; Alcina, 2009, etc.). Even though a number of such courses on terminology and the electronic tools have been available within translation and interpreting programmes at some universities and the translation students may have already benefited from them, the interpreting students still could not fully and realistically appreciate the usefulness of relevant electronic tools.

At the University of Leeds, relevant topics are covered in an optional module (*Corpus Linguistics for Translators and Interpreters*), as part of the one-year postgraduate degree programme and is open to both translation and interpreting students. The course is popular with the translation students, but only very few interpreting students who have great interest in technologies choose the courses. It is possible that the intensive interpreting programme has already kept the students very busy, or it may be the case that they are not convinced that the course is directly related to interpreters and their efforts in the course would improve their interpreting performance.

It is evident that the current technology content is not well-integrated in interpreting training at my institution. Therefore we need to find a way to bring terminology and technology more fully into interpreting training. This study will demonstrate that training on terminology preparation for technical meetings is a useful supplement to the already existing professional interpreting training. This study will explain how a workshop on terminology preparation for technical meetings is carried out in the last term of the 1-year MA interpreting training. The workshop not only introduces different preparation procedures and relevant tools that can assist preparation, but also provides hands-on experience involving actually preparing for technical meetings using different preparation procedures and tools.

2.8.2 Research goals and research questions

In summary, considering all the discussions above, I identify four major goals of this thesis.

- To investigate a possible terminology preparation procedure by using corpora and corpus tools **(Chapter 3)**
- To identify methods to measure performance of interpreters with respect to their use of terminology **(Chapter 4)**
- To evaluate the usefulness of current terminology extraction tools for interpreters' preparation **(Chapter 5)**
- To observe the impact of using the proposed preparation procedure and

corpus tools on simultaneously interpreting technical speeches **(Chapter 6 &7)**

The following research questions are related to the research goals listed above, and some entail a number of sub-questions.

1. How to integrate the use of corpus tools into the interpreters' preparation?
2. How to assess performance of interpreters with respect to their use of terminology?
3. Which term extractor offers comparatively better performance regarding term extraction in Chinese and English for the trainee interpreters?
4. Do the proposed preparation procedure and the tools have an impact on the trainee interpreters' SI performance?
    4.1 Does using the automatic term extractor during preparation affect the students' SI performance?
    4.2 Does using both the automatic term extractor and the concordancer affect the students' SI performance?
    4.3 Does the use of the corpus tools make the trainee interpreters' preparation easier and more efficient?

Chapter 3 - A corpus-based terminology preparation procedure and tools used

This Chapter aims to answer the first research question: how to integrate the use of corpus tools into the interpreters' preparation. I will investigate a corpus-based terminology preparation procedure and the electronic tools that can assist the interpreters' preparation.

3.1 Corpus-based terminology preparation pipeline

Based on literatures on the interpreters' terminology preparation (in **Chapter 2**), this study develops a corpus-based terminology preparation pipeline for interpreters, covering all the key elements of terminology-driven preparation (in **Section 2.4.5**). The pipeline includes:

1) Establishing interpreters' own corpora (formed by conference documents they receive from the organiser and terminologically rich text source collected from the Internet);
2) Automatically generating term lists from the established corpora;
3) Using concordance tools as navigational aids for close reading and consolidating learning of keywords in contexts.
4) Updating and managing terminologies for future use.

The preparation procedure and the electronic tools that can be used to assist the preparation are summarised below in the following table.

Step 1	**Step 2**	**Step 3**	**Step 4**
Corpus building	Automatic term extraction	Term exploration	Term management
•Web Crawlers •Manual collection	•Corpus analysis tools •Term extractors	•Concordancers	•Excel

Table 4: Terminology preparation procedure and the tools used

I will now introduce relevant tools, their main functions and the possible contributions to the interpreters' preparation.

3.2 Corpus building tools (web-crawlers)

Keyword-based web crawlers are corpus compilation tools based on the idea that the Internet provides a wealth of easily accessible specialised language data (Kilgariff and Grefenstette, 2003:333-347). Crawlers can build specialised corpora from publicly accessible documents on the web. Users first define a set of single-word or multiword seed queries, from which an initial corpus is created; terms are then automatically extracted from it and fed to the search engine to collect a bigger corpus. This type of tools would be particularly useful if interpreters can only get very limited preparation materials from the conference organiser.

Two web crawlers, namely, **WebBootCaT** and **Babouk**, and their possible contributions to interpreters' preparation will be discussed below.

3.2.1 WebBootCaT

WebBootCaT is a freely-accessible web-based crawler, which supports search in both English and Chinese and many other languages from online texts (in HTML). It takes only a few minutes to build a specialised corpus ("instant corpus"). The next step is to clean the corpus of all the HTML codes and tokenise it. The corpus, once produced, can be either downloaded or loaded into the Sketch Engine, a corpus query tool, for further exploration (Baroni and Bernardini, 2004; Baroni et al., 2006). WebBootCaT is particularly useful for small, short-term projects such as preparing for topic-based materials for interpreting assignments.

3.2.2 TTC's Babouk

Babouk, a module on the TTC Web Platform[6], is a focused web crawler for building domain-specific corpora. It also supports search in both English and Chinese. Since it is a focused crawler, Babouk finds relevant pages and retrieves all the links to the new pages. This suits the interpreters' needs when they are looking for domain-specific texts from the Internet. Babouk can gather as many relevant webpages as possible on a specialised domain defined by the user by means of seeds.

---

[6] The TTC Web Platform is a demonstrator of the result of the European financed TTC project (Translation, Terminology and Comparable Corpora). The platform allows users to create a processing chain by using three modules, from compiling a corpus to extracting monolingual terminology and generating bilingual terminologies. [Accessed 22 June 2015]. Available from: http://ttc.syllabs.com/

Babouk computes the relevance of webpages and filters out non-relevant documents. If a webpage is found to be relevant, all of its links are extracted and added to the crawl queue. Users can set stopping criteria to specify a maximum crawl depth and an upper bound time limit (Alonso et al., 2012:393).

While this strategy is theoretically sound, the crawling process might be slow. Babouk can take hours to crawl deeply for a corpus. It is too slow for an urgent interpreting task. Yet it can still be used as a useful tool if an interpreter has several conferences to prepare for in a row. The interpreter could prepare for one conference, meanwhile set their queries on Babouk for another conference's topic, leaving it there for at least an hour or so, allowing it to automatically harvest as many relevant webpages as possible. Once the corpus is established by Babouk, the interpreters can use it any time for his/her preparation for another meeting.

In this study, WebBootCaT is used for building (larger) English and Chinese corpora (of more than 100,000 tokens) on two specialised topics (fast reactors and deep seabed minerals) (also see **Chapter 5, Section 5.1.2**).

3.3 Tools to generate termlists

Once the specialised corpora are established, two types of tools could be used to automatically generate monolingual termlists, namely corpus analysis tools (e.g. AntConc, WordSmith Tools**)** and term extraction tools (e.g. TTC Term Suite Extraction, Syllabs Tools and Teaboat).

3.3.1 Corpus analysis tools

Corpus analysis tools, such as WordSmith Tools[7] (Scott, 1996, 1997) and AntConc[8] (Anthony, 2007) are programs which allow for producing lists of keywords or word-clusters from one or more texts, set out in frequency order.

Take AntConc for example: its "Keyword List" function allows to see which words appear more frequently in a corpus (e.g. specialised corpus) compared with

---

[7] WordSmith Tools is a PC software published by Lexical Analysis Software Ltd. and Oxford University Press since 1996. [Accessed 22 June 2015]. Available from: http://www.lexically.net/wordsmith/index.html

[8] AntConc is a corpus analysis toolkit developed by Laurence Anthony of Waseda University, Japan. [Accessed 22 June 2015]. Available from: http://www.antlab.sci.waseda.ac.jp/antconc_index.html

the words in a reference corpus (e.g. general corpus) specified by users. AntConc calculates the "keyness" and "frequency" of words using either the chi-squared or log-likelihood statistical measures (Anthony, 2004:9-11). In addition, its "Word Clusters" function has a lexical bundle option ("N-grams"), which allows to generate lexical units longer than single words (between 1 and 100 tokens). This function is potentially useful for interpreters as multi-word units (e.g. fixed expressions and proper names) are common in specialised communications.

A pilot study was carried out using AntConc to generate keyword lists and multi-word lists (en & zh) from comparable corpora on 'climate change'. The pilot study was to test AntConc's wordlist functions and the possibility to apply the tool in interpreters' preparation. It is found that AntConc's "keyword lists" contain mostly general (single) words (e.g. *"information", "development", "international"*, etc.), which are already known by the users. In the 'N-gram lists', ill-formed word clusters (e.g. *"under the", "by the conference of the"*, etc.) take a large part. There are also many non-term expressions (e.g. *"serving as", "access to", "be used"*, etc.) in the lists.

A possible reason is that tools like AntConc and WordSmith do not apply any specific lexical patterns/frameworks based on POS (part-of-speech) tagging, and therefore can neither distinguish term and non-term, nor focus on extracting terms. In addition, there are too many random word clusters in the lists. Revising or removing them manually from the lists is time-consuming. Apparently, AncConc's lists are not good enough for interpreters' preparation. Therefore, I had to explore other types of tools which are more specialised in extracting terms.

3.3.2 Term Extractors

Terminology extraction represents automatically extracting relevant terms from a given corpus. Distinct from "keyword list" and "word clusters" functions of corpus analysis tools, approaches to automatic term extraction make use of linguistic processors (part of speech tagging, phrase chunking) and/or statistical approaches to extract terminological candidates (Dunaevsky, 2015).

TTC Term Suite, Syllabs Tools and Teaboat, three individual term extraction modules on the TTC Web Platform[9], can extract both single and multi-word terms. Their lists seem to have less redundant information than AntConc's "N-gram" lists. Take TTC Term Suite for example, its monolingual term extraction result provides rich output, including term lemma, part of speech, lexical pattern, domain specificity, occurrence count, relative frequency, as well as different forms and variants of the same term. The terms are sorted by "domain specificity", the statistically-important information relevant to interpreters, who are more likely to find more topic-related terms ranking on top of the list. Single-word and multi-word terms (SWT & MWT) are integrated into one list with "domain specificity" as a consistent parameter (see **Figure 1**).

Id	Term	Term pilot	Part of speech	Pattern	Complexity	Spec.	Occurr.	Freq.	Forms list
917089	wind turbine	wind turbine	Noun	Noun-noun	Multi-word	188	1850	0.0080	Form: Occurr. wind turbine: 1148 wind turbines: 702
121371	blade	blade	Noun	Noun	Single-word	145	1425	0.0062	Form: Occurr. blade: 1019 blades: 406

Figure 1: Screenshot of TTC Term Suite's extraction result

The three term extractors on TTC platform have their own specifications, for example, they use different extraction approaches and POS patterns, and they rank terms based on different statistics. The following **Table 5** is a summary of the main technical features of the three term extractors.

9 The TTC Web platform is an online demonstrator. Using this web-based service, the user can compile monolingual/bilingual terminologies out of comparable corpora with the tools developed in the project directly on the web site, without having to download or to install the tools (Blancafort et al, 2013).

	TTC TermSuite	Syllabs Tools	Teaboat
Extraction approach	Knowledge-rich	Knowledge-poor	Knowledge-rich
POS patterns	noun & verb phrases	noun phrases	noun phrases
Complexity of terms	SWT & MWT	SWT & MWT	SWT & MWT
Ranking of terms	relative frequency & their domain specificity.	relative frequency	log-likelihood statistics
Operating system(s)	Windows	Windows	Linux

Table 5: Summary of the main features of the three term extractors (TTC TermSuite, Syllabs Tools & Teaboat)

All the three term extractors offer several optional parameters. "The tools can be used with the default setting by users that are less familiar with terminology extraction tools, whereas advanced users can configure the tools according to their needs" (Blancafort et al, 2013).

I will describe each tool in detail in **Chapter 5**, and provide a statistical evaluation of their term extraction performance respectively. I will also discuss the corresponding feedback obtained from the users, the trainee interpreters.

3.4 Concordancers

A concordance is a list of occurrences of a particular word, part of a word or combination of words, in its contexts drawn form a text corpus (Botley et al., 1996). Corpus analysis tools (e.g. AntConc, WordSmith, Sketch Engine[10], IntelliText[11], etc.) offer "Concordance" and "collocation" functions. Corpora are useful in vocabulary acquisition, not only because collocations (i.e. habitual co-occurrences of lexical items) can be measured quantitatively, but also because the KWIC (key word in context) format of corpus data exposes learners to a great deal of authentic data in a structured way (McEnery and Xiao, 2010).

---

[10] The Sketch Engine is a text corpus management and analysis software developed by Lexical Computing Limited since 2003. [Accessed 22 June 2015]. Available from: https://the.sketchengine.co.uk/

[11] IntelliText is developed by the Centre for Translation Studies (CTS) at the University of Leeds. The IntelliText project aims to facilitate corpus use for academics working in various areas of the humanities. The project is funded by the Arts and Humanities Research Council. [Accessed 22 June 2015]. Available from: http://smlc09.leeds.ac.uk/itweb/htdocs/Query.html

The use of corpora and concordancing in language education has been discussed extensively in the literature (e.g. Tribble and Jones, 1990, 1997; Woolls, 1998; Aston et al., 2004; Braun, 2007). Concordance tools have proven to benefit language learners in vocabulary acquisition. Corpora can provide more realistic examples of language usage that reflect the complexities and nuances of natural language (McEnery and Xiao, 2010).

Using concordance tools would be helpful to consolidate the learning of specialised terminology for interpreters. For interpreters, searching a specialised corpus[12] (formed by conference documents and online domain specific data) can reveal all concordance lines of the designated terms and all instances of a particular collocation. It saves interpreters looking up each occurrence. And checking through the concordance lines may help interpreters understand and memorise the terms better.

Sketch Engine is an online corpus query interface providing an array of easy-to-use functions. Sketch Engine is chosen for this study because it offers a wide range of useful features, including “Concordance” (line by line detailed view of the corpus contents) (**Figure 2**), “Word Sketch” (short summary of collocational behaviour of the search term) (**Figure 3**), “Thesaurus” and “Sketch Differences”, etc. Sketch Engine also incorporates the WebBootCat tool and allows users to create specialised corpora from Web instantly in English, Chinese and many other languages (Smith et al., 2008:2).

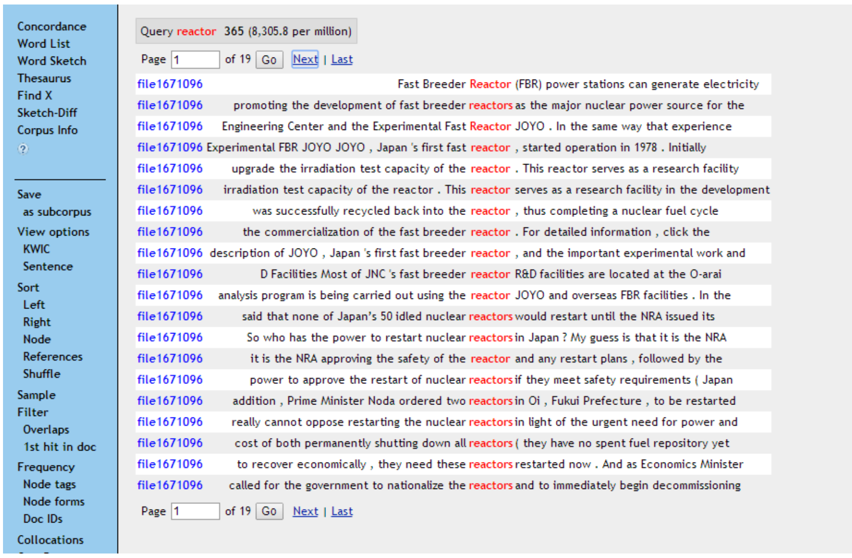


Figure 2: Sketch Engine's Concordance function

As shown in **Figure 2**, by using the concordance function, interpreters could navigate within a corpus, consolidating their learning of the keywords and getting more background information from the contexts where the keywords are used in. Through observing and comparing examples of words or phrases that are hard to interpret or that are not included in standard bilingual dictionaries, interpreters could deduce their meaning or understand nuances in their use, and identify suitable target terms accordingly.

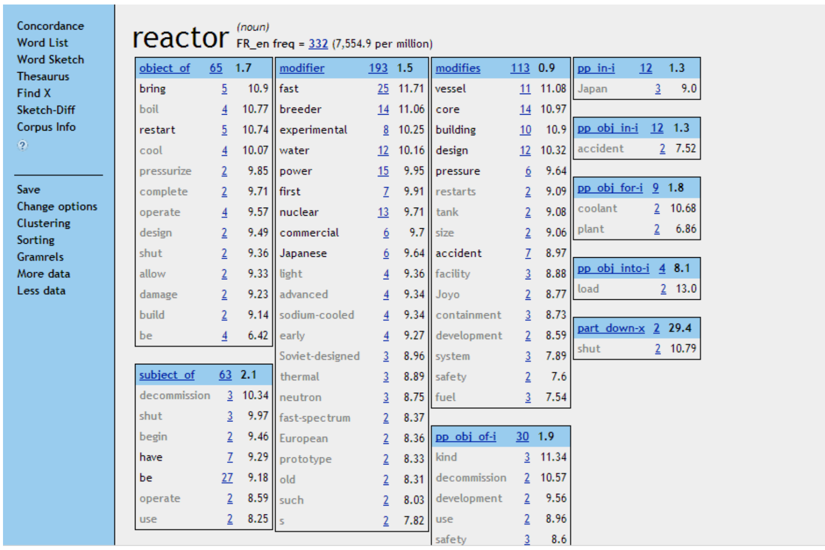


Figure 3: Sketch Engine's Word Sketch function

Figure 3 illustrates “Word Sketch”, a Sketch Engine’s unique feature. It is a one-page, automatic, corpus-derived summary of a word’s grammatical and collocational behaviour. For interpreters, “word sketch” could help to quickly detect frequent patterns/collocations of a term. With a further click of a patterns/collocation, all the concordance lines (authentic examples) in the corpus will be presented together. Then the interpreters could continue to learn the collocation in multiple contexts.

In **Chapter 4** and **7**, I will further discuss how the concordance functions are used to assist term activation during interpreters’ preparation.

3.5 Excel spreadsheets

I have reviewed a number of terminology management tools in **Section 2.7.2**. However, investigating term management tools was not within the scope of this study. Therefore, I asked the participants to manage their termlists by using Microsoft Excel in the SI experiments.

Excel is easy to access and operate. It provides a familiar environment and is readily available. Interpreters can put terms, their equivalents (in other languages) and comments in different columns. Interpreters can move the columns, add columns or use sort and filter function for organising their termlists. Terms can be grouped together for different interpreting assignments. Excel also has a helpful autocomplete feature: if a term has been typed in the same column, it will be offered to the users as a suggestion, so the term would not be entered twice. However, Excel cannot search for particular terms among different Excel files, which is often what interpreters need.

Chapter 4 - Methodology

We have discussed a terminology preparation pipeline for interpreters (including the preparation procedure and the tools used) in **Chapter 3**. This chapter reviews the main methodological approach of this study, and it is organised as follows.

Section 4.1 presents the research questions and objectives of the experimental research. The groups of participants will be described in **Section 4.2**. **Section 4.3** provides a detailed description of the process of selection and design of experimental speeches and preparation materials. **Section 4.4** explains how an independent variable (preparation time) is controlled in this study. **Section 4.5** presents the marking criteria applied to measure the dependent variables in this study (i.e. participants' terminological and simultaneous interpreting performance). Then the tasks and procedures are explained in **Section 4.6.**

4.1 Objectives, research questions and hypothesis

The objectives in the rest of the chapters are 1) to identify methods to assess the performance of interpreters with respect to their use of terminology 2) to evaluate the usefulness of terminology extraction tools for interpreters' preparation and 3) to investigate the impact of using the proposed preparation procedure and the corpus tools on simultaneously interpreting technical speeches.

The interpreters' terminology preparation by using corpus tools has been addressed in the literature. However, there has not been any empirical study to examine the effect of using the tools on interpreters' performance in simultaneous interpreting. This study compares the results of several term extractors by involving the trainee interpreters as evaluators. Then this study further examines whether using a single term extraction tool and a single concordance tool can help interpreters increase their preparation efficiency and improve their simultaneous interpreting performance.

Several studies have demonstrated that it is possible to observe the effect of preparation on simultaneous interpreting (e.g. Alonso Bacigalupe, 1999; Lamberger-

Felber 2001, 2003; Díaz-Galaz, 2012; Díaz-Galaz et al., 2015). These studies mainly compared the effects of preparation vs. non-preparation. For example, Díaz-Galaz's experimental study (2012) demonstrated ways to observe the effect of advance preparation in simultaneous interpreting, i.e. by measuring ear-voice span, accuracy of interpretation and reformulation strategies used.

Different from Díaz-Galaz (2012) and Díaz-Galaz et al. (2015), which focused on the role of 30-minute advance preparation in simultaneous interpreting (preparation vs. non-preparation), this study examines the role of different preparation procedures in SI, and we aim to find an optimal preparation procedure. We also want to look at participants' perception of using different preparation procedures and tools. The effects of using different preparation procedures and tools are measured by examining terminological accuracy and terminological errors during SI, holistic interpreting performance, as well as post-task recall of terms.

The main and specific research questions and hypotheses in the rest of the chapters are summarised below.

RQ1: How to assess performance of interpreters with respect to their use of terminology?

RQ2: How useful are term extractors for the trainee interpreters' preparation?

More specifically, which term extractor (TTC, Syllabs, Teaboat) has consistently higher precision rates in term extraction in English and Chinese on both topics (FR & SM) evaluated by the trainee interpreters?

RQ3: Do the proposed preparation procedure and the tools have an impact on the trainee interpreters' SI performance? (See **Table 6** for specific research questions and hypotheses of the third research question).

Specific research questions	Specific hypotheses
3.1 Does only **using an automatic term extractor** during preparation affect the students' SI performance?	➢ Only **using an automatic term extractor** within **limited preparation time (3 days)** results in better interpreting performance, as measured by term accuracy scores, holistic SI scores, numbers and types of term errors, post-task recall of terms.
3.2 Does **using both an automatic term extractor and a concordancer** affect the students' SI performance?	➢ **Using both an automatic term extractor and a concordancer** within **ample preparation time (9 days)** results in better interpreting performance, as measured by term accuracy scores, holistic SI scores, numbers and types of term errors, post-task recall of terms. ➢ Using both tools supports terminological performance in paragraphs that contains higher term density and in SI task into the B language, measured by numbers and types of term errors.
3.3 Does the use of the corpus tools make the trainee interpreters' preparation easier and more efficient?	➢ Only **using an automatic term extractor** helps save preparation time, as measured by the participants' real preparation time. ➢ **Using both an automatic term extractor and a concordancer** helps save preparation time, measured by the participants' real preparation time.
3.4 What is the role of term accuracy in SI performance on specialised topics?	➢ There is a strong correlation between term accuracy and SI performance. Term accuracy could be used as a predictor for SI performance.

Table 6: Summary of specific research questions and hypotheses of the third research question

This empirical study uses both quantitative and qualitative methods. It adopts a sequential design, starting from quantitative study (evaluation of automatically-generated termlists and the experiments) and qualitative study (focus groups).

Chapter 4 (**Section 4.5.3**) will answer the research question: "how to assess performance of interpreters with respect to their use of terminology". The evaluation of automatically-generated termlists answers the research question: “how useful term extractors are for the trainee interpreters’ preparation”, and it will be discussed in **Chapter 5**. The experiments are designed to answer “whether the proposed preparation procedure and the tools can influence the trainee interpreters’ SI performance”. The focus groups are expected to examine “whether the use of the tools makes the trainee interpreters’ preparation easier and more efficient”. All the participants’ interpretation and focus group discussions are transcribed and analysed. The interpreting performances and individual’s real preparation time in both control and test groups (of two SI tasks) are quantified and compared statistically to ascertain whether there are significant differences. If the test groups achieve significantly better results, the hypotheses are confirmed. The focus group discussions are summarised to reflect the students’ views on different preparation procedures and the tools used during their preparation. The data generated from the experiments and focus groups will be analysed and discussed in **Chapter 6 & 7**.

The data from both quantitative and qualitative methods are complementary and will be integrated to address different components of the research subject. As this study is mainly a product-oriented study, quantitative data from the experiments will be will be given “priority”, and qualitative data from focus groups will be supplemental in explaining the preparation process.

Table 7 summarises the experiment arrangements in both 2013 and 2014. The procedures will be further explained later in **Section 4.6**.

	2013	2014
April (Last week)	**Control group (FR)**	**Control group (SM)**
	Training on traditional preparation procedure	Training on traditional preparation procedure
	Preparation and SI task (FR) Focus Groups	Preparation and SI task (SM) Focus Groups
May	**Test group (SM)**	**Test group (FR)**
	Training on the use of term extraction tools Evaluation of auto-termlists (FR)	Training on the use of term extraction and concordance tools
June (First week)	Evaluation of auto-termlists (SM) Preparation and SI task (SM) Focus Groups	Preparation and SI task (FR) Focus Groups

Table 7: Arrangement of the experiments in both 2013 and 2014

4.2 Participants

This study recruited 22 trainee interpreters from the MA programme in Conference Interpreting and Translation Studies at University of Leeds in two consecutive academic years: 12 students from the cohort of 2012-2013 and 10 from the cohort of 2013-2014. All the participants were from mainland China, with the same language combination of Chinese A and English B, and little professional interpreting experience prior to the MA training programme. At Leeds, the students were recruited by the same recruiting standards and procedures over the two consecutive years. And the two groups received essentially the same training in conference interpreting from the same team of trainers, following the same curriculum. The students were invited to participate in this study in the second semester of their one-year MA interpreting programme.

In 2013, I recruited a group of **12 students**, they were instructed to prepare for the topic of **Faster Reactors** by using "traditional preparation procedure" without using any tools and then simultaneously interpreted two speeches on the topic (one in English and the other in Chinese). The same group of students received training on how to use term extraction tools in interpreting preparation, but one student decided to withdraw from the experiment. As a result, **11 students** stayed and were instructed to prepare for another topic on **Deep Seabed Minerals** by using automatically-extracted termlists during their preparation and then simultaneously interpreted for two speeches on the topic, one in English and the other in Chinese. After the preparations and the interpreting tasks, the students were engaged in focus group to discuss their preparation processes.

In 2014, I recruited a group of **10 students**, and they were instructed to prepare for the topic of **Deep Seabed Minerals** by using “traditional preparation procedure” (without using any tool) and then simultaneously interpreted on the topic. The same group (**10 students**) received training on how to use term extraction and concordance tools in interpreting preparation, before they were instructed to prepare and simultaneously interpret for speeches on **Faster Reactors** by using both automatically-extracted termlists and the concordancer. Focus group discussions were also conducted after each of two simultaneous interpreting tasks.

The decision of forming the control and test groups from students over two years was to ensure equal pedagogical treatments for the students of the same cohort. We have also endeavoured to ensure that no students receive preferential treatment when compared to their peers from the same cohort, and that all the students benefit from access to the training outcomes. However, this would necessarily mean that we need to guarantee that the participants being recruited in over the two year cohorts are of comparable (in terms of their levels of interpreting skills).

In addition, the inclusion of participants was absolutely voluntary, meaning that they could withdraw at any time after beginning the research. Eventually, 11 out of 17 interpreting students in 2013 and 10 out of 15 interpreting students in 2014 participated in all the relevant activities of this study. Participants were recruited on their consent. A *Consent Form* along with a *Participant Information Sheet* was distributed to all the willing participants before the start of the experiments. All the information that the researcher collected about the participants during the course of the research has been kept strictly confidential. Anonymisation has been applied in all the means of dissemination.

In the following sections (**4.2.1 & 4.2.2**), more detailed profiles of the participants will be presented in terms of participants’ interpreting competence, their prior knowledge about the topics, terminology and ideas of the experimental speeches.

4.2.1 Interpreting competence

In this study, the interpreting competence of the participants was measured by the trainee interpreters' consecutive and simultaneous interpreting exam results. The trainee interpreters were formally assessed at the end of both semesters in the MA interpreting programme. Our first experiment took place in late April, before which the students had been formally assessed in the first semester's exam (consecutive interpreting). The second experiment was arranged after the students took their second semester's exams (consecutive and simultaneous interpreting).

The results of the consecutive interpreting exam (first semester) and the consecutive and simultaneous interpreting exams (second semester) were used as pre-tests to see whether the participants in the two year cohorts have same level of interpreting competence.

Examiners assessed the students' performance according to MA marking criteria for interpreting skills, according to which a mark over 70% would suggest a solid performance to professional standard, and a mark below 50% would suggest that a student has not yet adequately mastered the skills required.

	2013	**2014**	**P-value**
en-zh Mean (STDV)	60.15 (7.60)	58.73 (5.80)	0.608 (P>0.05)
zh-en Mean (STDV)	55.85 (5.44)	53.91 (8.01)	0.505 (P>0.05)
Number of students: 12 students (2013), 10 students (2014)			

Table 8: Consecutive interpreting exam results (First semester)

	2013	**2014**	**P-value**
en-zh Mean (STDV)	61.92 (7.17)	57.00 (9.22)	0.166 (P>0.05)
zh-en Mean (STDV)	56.69 (7.38)	56.73 (7.96)	0.991 (P>0.05)
Number of students: 12 students (2013), 10 students (2014)			

Table 9: Consecutive interpreting exam results (Second semester)

	2013	2014	P-value
en-zh Mean (STDV)	55.85 (9.26)	55.45 (12.14)	0.931 (P>0.05)
Number of students: 12 students (2013), 10 students (2014)			

Table 10: Simultaneous interpreting exam results (Second semester)

The Tables (8, 9 and 10) present the mean and standard deviation values for the students' results in all the interpreting exams in 2013 and 2014 before they took part in the experiments. The T-test results show that there was no significant difference between the two year cohorts in terms of their interpreting competence reflected in all the formally-assessed interpreting exams.

4.2.2 Background knowledge of source speech's topics, ideas and terminology

After performing each of the experiment tasks (FR & SM), the participants were asked to fill out a questionnaire on their prior knowledge about the topics, terminology and the ideas of the source speeches before their preparation. This allows us to have a rough idea about how familiar participants were with the information contained in the source speech. Control measures would be taken if any participant was already knowledgeable enough of the topics, which might render preparation unnecessary.

We adopted the design of the background information questionnaire from Díaz-Galaz's PhD project (2012). Her questionnaire contains four items regarding which the participants indicated, on a scale of 1-5, their degree of knowledge about the source speech topic, terminologies used and the source speech idea, as well as participants' interest in the topic.

Our questionnaire contains a set of three items about prior knowledge on "the topic of the source speeches", "the terminology used in the source speeches" and "the ideas of the source speeches" before the participants' preparation for the relevant SI tasks. Participants indicate their "degree of knowledge" on a scale of 1-5 for each item. "1" is "very low knowledge" and "5" is "very high knowledge". The average ratings for each item are illustrated in **Figure 4** (for the topic FR) and **Figure 5** (for the topic SM).

	Topic	Terminology	Ideas
FR control Mean (STDV)	1.33 (0.49)	1.17 (0.39)	1.42 (0.51)
FR test Mean (STDV)	1.20 (0.42)	1.30 (0.48)	1.50 (0.53)
P value	0.502	0.491	0.713

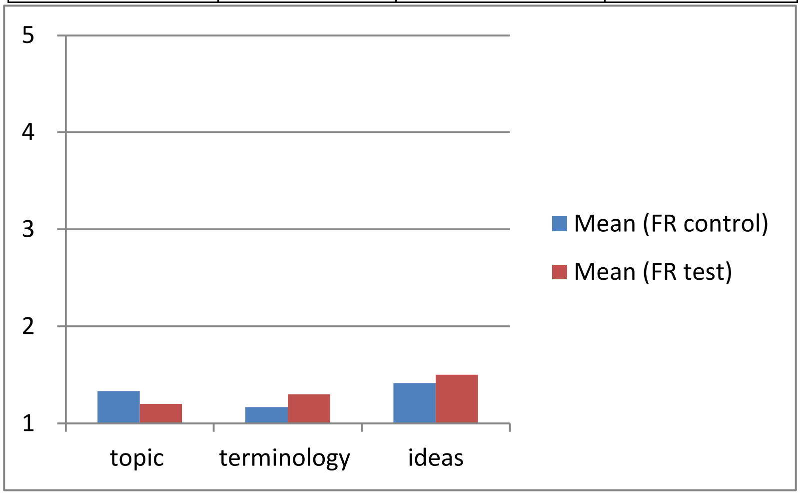


Figure 4: Mean ratings of prior knowledge about the source speech topic, terminology and ideas before preparation (Experiment II: FR) "1" is "very low knowledge" and "5" is "very high knowledge"

Figure 4 shows that for the topic of **Fast Reactors**, the average ratings on the prior knowledge (i.e. "topic of the source speeches", "terminology used in the source speeches" and "ideas of the source speeches") were close to "very low" in both groups. T-tests conducted revealed no significant differences between the two groups on the three items (P>0.05).

	Topic	Terminology	Ideas
SM control Mean (STDV)	1.20 (0.42)	1.10 (0.32)	1.40 (0.52)
SM test Mean (STDV)	1.27 (0.47)	1.09 (0.30)	1.45 (0.52)
P value	0.712	0.947	0.813

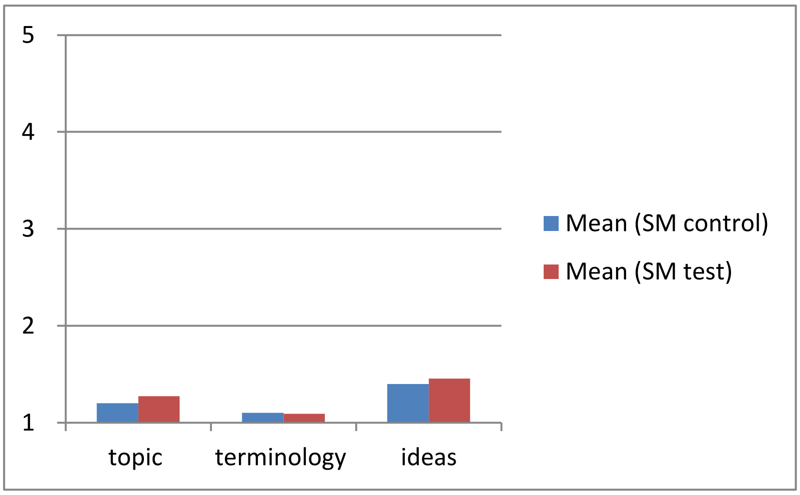


Figure 5: Mean ratings of prior knowledge about the source speech topic, terminology and ideas before preparation (Experiment II: SM) "1" is "very low knowledge" and "5" is "very high knowledge"

Figure 5 shows that for the topic of **Deep Seabed Minerals**, the average ratings on the prior knowledge were also close to "very low" in both groups. T-tests conducted showed no significant differences between the two groups on the three items (P>0.05).

In summary, the information presented in **Section 4.2** shows that the two year cohorts of students were homogenous in terms of native language, working language pair, interpreting and training experience. Furthermore, there was no significant difference between the two year cohorts in terms of their interpreting competence reflected in the formally-assessed interpreting exams of their MA interpreting programme. Participants in both years also reported having little or no knowledge about the topics, ideas and terminology of the experiment source speeches, and no significant differences were found between the two groups. Therefore we were confident to treat the two year cohorts as comparable groups in our experiments.

4.3 Materials

This section presents a detailed description of the selection and design of the materials used in this experimental study (including experimental speeches and preparation documents provided to the participants). During the selection and design process, special efforts were made in particular to maintain a balance between experimental control and ecological validity.

4.3.1 Selection of experiment speeches

We have considered several factors when selecting the experiment speeches.

The source speech topics had to be unfamiliar to the participants, that they had not been studied or discussed in participants' interpreting classes. As discussed in **Section 4.2.2**, the questionnaire results confirmed that participants had little or no knowledge about the topics, terminology and ideas of the source speeches.

It was important that speeches on one topic were interpreted only once (by each participant). A second interpretation could perhaps mask the effect of advance preparation. Therefore, two topics would have to be selected, one for each experimental condition. This study attempts to study the impact of different preparation procedures on interpreting into both the A and the B languages, therefore, source speeches would have to be in both English and Chinese. Four speeches in both Chinese and English on two specialised topics: Fast Reactors (FR) and Seabed Minerals (SM) were designed purposely for this experiment.

In order to ensure the authenticity of language and content of the experiment speeches, the speeches were first of all pre-selected from authentic speech transcripts and summaries of real-life international scientific meetings available on the Internet[13]. Then the selected materials were further edited to produce coherent and self-contained speeches for the experiments. The speeches contained an introduction, a main body and a conclusion. Special attention was paid to include the features of scientific oral discourse, as discussed in **Chapter 2 (Section 2.1.2)**, for instance, the speeches should contain specialised terminology and non-redundant

---

[13] Fast Reactors (FR) and Seabed Minerals (SM) are real topics discussed in relevant bodies of the UN. There have been a number of relevant conference documents stored in the online archives, which are publicly accessible. Fast Reactors (FR): http://www.iaea.org/NuclearPower/Meetings/2013/2013-03-04-03-07-CF-NPTD.html; Seabed minerals (SM): http://www.isa.org.jm/files/documents/EN/8Sess/Assembly/ISBA-8A-13.pdf. [Accessed 22 June 2015].

information (proper names, figures, acronyms, etc.). Nevertheless, the speeches should not be too technical, since the experiments were conducted on students rather than professional interpreters. Further editing of the pre-selected speeches (e.g. simplification of some content and syntactic structures, or lowering the density of terms used in certain paragraphs) was required whenever necessary (see the four experiment speeches in **Appendix A: 1-4**).

In addition, the English speech and the Chinese speech on the same broad topic (SM or FR) differed greatly concerning their focus and content (see the four speech outlines in **Table 11**). Moreover, the number of terms appearing in both speeches was kept to a minimum. For instance, there were only two shared terms in the two speeches on fast reactor (i.e. *"fast reactor" and "IAEA"*) and five shared terms in the two speeches on seabed minerals (i.e. *"seabed minerals"*, *"polymetallic sulphides"*, *"cobalt-rich crusts"*, *"mining"* and *"exploration"*). Such design was arranged to lower the potential bias created by the sequential presentation of the two speeches on the same topic.

Speech 1 (SM_en)
The exploration and mining technologies for new types of marine minerals

- Opening
- The distribution and basic features of polymetallic sulphides and cobalt-rich crusts
- The industrial use of the two types of mineral resources
- Research vessels and the exploration technologies
- The mining technologies
- Closing statement

Speech 2 (SM_zh)
China's exploration of new mineral resources in the international seabed area

- Opening
- The main responsibilities of The International Seabed Authority
- The regulations regarding exploring polymetallic nodules, polymetallic sulphides and cobalt-rich crusts in the international seabed area
- China's exploration of the three types of mineral resources in the international seabed area
- The methods to evaluate the environmental impact of exploring the marine mineral resources
- Closing statement

Speech 3 (FR_en)
The impact of Fukushima nuclear accident on fast reactor development in Japan

- Opening
- Fukushima nuclear accident
- Lessons learned from the nuclear accident

• Japanese authority's new safety rules • Japan's fast reactors • Japan's international cooperation in fast reactor R & D • Closing statement
Speech 4 (FR_zh) **The development of fast reactors in China** • Opening • Different generations of nuclear reactors in the world • The advantages of fast reactors • The development of fast reactors in China • The features of China Experimental Fast Reactor and its construction milestones • Closing statement

Table 11: The outlines of the four experimental speeches

Last but not least, we also made sure that all the specialised terminology used in the speeches could be found in the preparation documents provided to the students for their preparations.

The speeches were presented by native speakers, namely two female English speakers and a female Chinese speaker, to ensure proper language usage and clarity in delivery. The speeches were read from scripts with natural pauses. They were pre-video-recorded, so that the students could view the speakers from the screen while they were interpreting simultaneously.

The lengths of English speeches were measured by English words; Chinese speeches were measured by Chinese characters or syllables, as each Chinese character is of only one syllable (see **Table 12**).

As Li (2010:21) pointed out, "it is widely recognised that a rate between 100 and 120 words per minute (wpm) is optimal for English speeches. This translates into an optimal speed of 150-180 syllables per minute for Chinese speeches." According to this assumption, our speeches in English were delivered at slightly slower speeds.

	Size (words/characters)	Total time (min)	Delivery speed (w/m)
FR_EN	774	9.43	82
FR_ZH	1641	9.85	167
SM_EN	1001	12.5	80
SM_ZH	1813	10.07	180

Table 12: Basic information about the four experiment speeches

In this study, control and test groups using different preparation procedures and tools prepared for the same speeches, and then two groups' interpreting performances would be compared. The speeches on Seabed Minerals (SM) were used to test the effect of using only the automatic term extractor during preparation on simultaneous interpreting performance. The speeches on Fast Reactors (FR) were used to test the effect of using both the automatic term extraction and the concordance tools on simultaneous interpreting performance.

4.3.2 Terms in the speeches

As discussed in **Chapter 2**, this research focuses on terminology preparation and term performance. Therefore, it is necessary to differentiate between terms and general vocabulary (words). Terms or terminological units are used to label specialised concepts, while words are associated with general knowledge.

According to Gorodetsky (1990:117), in specialised languages, a term is a linguistic unit made of a single word or of a word combination, and is usually associated with the same conventional definition when used by speakers of a given specialised language. A terminological unit may also be a symbol, a chemical or mathematical formula, a scientific name in Latin, an acronym, an initialism, or the official title of an organization, an administrative entity or an individual's working title (Public Works and Government Services Canada, 2011).

We used the above description as a basis to recognise terms in the source speeches of this study.

4.3.2.1 Term categories in this study

In this study, terms in the source speeches were further categorised into **specialised terms (S)**, **general terms (G)** and **named entities (NE)**. **Category S** contained highly specialised terms relevant to the domain-specific topic (e.g.

“uranium-238”, *“decay heat removal system”*). **Category G** contained non-specialised terms commonly used in the field or general terms that are not specific to the subject field (e.g. *“performance tests”*, *“full power operation”*, *“nuclear accident”*). **Category NE** contained named entities, including organizations, locations names of persons, quantities, and other domain-specific proper names (e.g. *“International Project on Innovative Nuclear Reactors and Fuel Cycles”* or *“INPRO”*, *“China Experimental Fast Reactor”*).

	FR_EN	**FR_ZH**	**SM_EN**	**SM_ZH**
General Terms (G)	23	25	36	21
Specialised Terms (S)	20	53	71	39
Named Entities (NE)	12	4	2	14
Total	55	82	109	74

Table 13: Numbers of terms in different categories in the four experiment speeches

Table 13 summarises the numbers of terms in three different term categories in the source speeches. The same term used more than once in the source speeches was counted as one term here. Variations of a term were counted as same term (e.g. fast reactor, fast breeder reactor, and fast neutron breeder reactor were treated as the same term). However, when we observed the students’ term performance in simultaneous interpreting, our counting of terms was based on term occurrences. That is to say, the same term occurring in different contexts was treated as different terms (also see **Section 4.3.2.2** for term occurrences in the source speeches).

As shown in **Table 13**, the majority of terms used in the source speeches were specialised terms and named entities (S+NE). Our hypothesis is that specialised terms and named entities (S & NE) are challenging items to interpreters; better preparation may help reduce the challenges from the segments that contain specialised terms (S) and named entities (NE) in simultaneous interpreting (see **Chapter 6 & 7** for data analysis and discussion).

4.3.2.2 Density of terms

The delivery speeds of the source speeches, as discussed in **Section 4.3.1** (see **Table 12**), were controlled at comparatively slow speeds. Nevertheless, the density of terms was used to represent terminological difficulty of the source speeches.

“Lexical density” was measured in different ways in various studies. For example, Ure (1971) and Richards et al. (1992) defined “lexical density” as the ratio of lexical and grammatical words to the total number of words in a text; while in Díaz-Galaz et al. (2015), “lexical density” is defined as the proportion of content words in relation to the total number of words in a text.

In this study, we measured terminological difficulty by the density of terms rather than the density of content words. Therefore, the occurrence of terms was used as the numerator. In addition, we had the source speeches in both English and Chinese. In English, words are often separated from each other by blanks (white space); while Chinese is not a segmented language. In order to avoid tokenization ambiguity of Chinese texts, we decided to use total delivery time rather than total number of words as the denominator. Considering all these, we developed the following definition for the measuring density of terms for this study:

$$Density\ of\ terms = \frac{Occurrence\ of\ terms\ in\ the\ source\ speech\ (t)}{delivery\ time\ \ (m)}$$

After manual term selection, we counted the occurrences of terms for each speech. For instance, FR-en contains 145 terms per 9.43 minutes of delivery, therefore its density of terms is 9 terms/min. **Table 14** summarises the term density statistics of all the four experiment speeches. We tried to balance the terminological difficulty for the speeches in both languages, even if this was not always possible.

	Term occurrences (t)	**Total time (min)**	**Term density (t/m)**
FR_EN	86	9.43	9
FR_ZH	145	9.85	15
SM_EN	169	12.5	14
SM_ZH	138	10.07	14

Table 14: Term densities of the four experiment speeches

Furthermore, we also examined term densities in the source speech paragraphs. **Table 15** shows the mean and standard deviation values of term density in the source speech paragraphs, as well as the range of term density values for each source speech.

	FR_EN	FR_ZH	SM_EN	SM_ZH
Number of Paragraphs	14	12	16	13
Term density Mean (STDV)	9.06 (3.83)	14.16 (4.80)	13.67 (4.21)	13.20 (4.63)
Term density (Min, Max)	(2.38, 14.77)	(5, 20.37)	(6.8, 21.65)	(7.14, 20.37)

Table 15: Range of term densities in the experiment speech paragraphs

As **Table 15** shows, the paragraphs in each source speech have a wide range of term densities. This allows us to form a hypothesis that an increase of term density in source speech has a detrimental effect on interpreting processing and terminological performance, and furthermore, better preparation (perhaps using corpus tools) may help mitigate the effect (also see **Chapter 7**, **Section 7.3.2 and 7.3.3**).

We were aware that FR_en contained fewer terms and lower term density compared with the other three experiment speeches, as shown in **Table 13** and **Table 14**. We would take it into consideration when we analyse the experiment results.

4.3.3 Preparation documents

In order to control what participants would study during their preparation, we provided the participants with a set of preparation documents (comparable corpora in English and Chinese), which represented conference documents and relevant background documents provided by the conference organisers. **Table 16** shows the sizes of the preparation documents.

FR		SM	
EN	ZH	EN	ZH
9 texts	9 texts	9 texts	12 texts
42,006 words	30,174 words	20,533 words	40,545 words

Table 16: Size of preparation documents on FR & SM

Unlike Díaz-Galaz (2012) and Díaz-Galaz et al. (2015), whose preparation materials consisted of speech summary, information about the speaker, slide presentation based on the speech and a brief glossary that contained 30 specialised terms, the preparation documents in this study included: 1. itemed speech outlines (for both the English and Chinese speeches); 2. information documents on the outline items in both English and Chinese, including relevant research papers from experts and research institutes, reports from national and international authorities, as well as popular science articles, Wikipedia articles, specialised journal articles and interviews, etc. In the experiment groups of this study, participants were also provided with two monolingual termlists (en & zh) which were automatically generated from the preparation documents. The lists were expected to be further edited by the participants.

The speech outlines summarised the main content and reflected the structure of the source speeches. The information documents on the outlined items were original online documents selected by the researcher. All the above preparation documents were expected to provide relevant background information and necessary terminology information about the source speeches. Participants were asked to primarily focus on studying the preparation documents; meanwhile they were also allowed to check further information from the Internet when necessary.

Our purpose in preparing these items was to replicate the materials for reading done by professional interpreters before the actual conferences (also see Moser-Mercer, 1992; Gile, 1995; Abril & Ortiz, 1998; Donowan, 2001; Jiang, 2013).

4.4 Preparation time

According to Moser-Mercer's survey study (1992) among professional interpreters (AIIC members), 81% responded positively to receiving documents, generally 6-10 days (51%) and 3-5 days (24%) in advance.

Based on the research findings such as Moser-Mercer (1992) and the researcher's own experience as conference interpreter, this study decided to introduce preparation time as a control variable: limited preparation time (3 days) and ample preparation time (9 days).

In summary, two speeches on Seabed Minerals (SM) were used to test the effect of using only the automatic term extractor within **limited preparation time (3**

days) on simultaneous interpreting. Two speeches on Fast Reactors (FR) were used to test the effect of using both the automatic term extractor and the concordancer within **ample preparation time (9 days)** on SI performance.

4.5 Dependant variables

In order to obtain an objective view of the role of different terminology preparation procedures in simultaneous interpreting, participants' **SI performance scores**, **terminological accuracy scores**, **terminological error numbers** and **the degrees of departures** for each individual error category as well as **post-task recall of terms** were measured as dependent variables.

4.5.1 Judges

A panel of three judges including the researcher were involved in the assessment of the participants' performance. The three judges (with Chinese-A and English-B) were both practitioner interpreters and interpreting trainers at the University of Leeds.

The assessments were firstly conducted by the researcher on **holistic SI performance** and **terminology performance during SI** (according to the two individual marking criteria, which will be discussed later in **Section 4.5.2 & 4.5.3**). The other two judges then provided non-blind moderation by sampling. In other words, the second marker sampled work already first marked, and then checked the validity and consistency of the first marker's assessment. One judge moderated on the assessment of **holistic SI performance** by checking 50% of the total sample of participants' interpreting audio-recordings; while the other judge moderated on the assessment of **terminology performance during SI** by checking 30% of the total sample of participants' interpreting transcripts and the first marker's annotation of the students' term errors. The following **Figure 6** illustrates the procedure of assessment by the three judges.

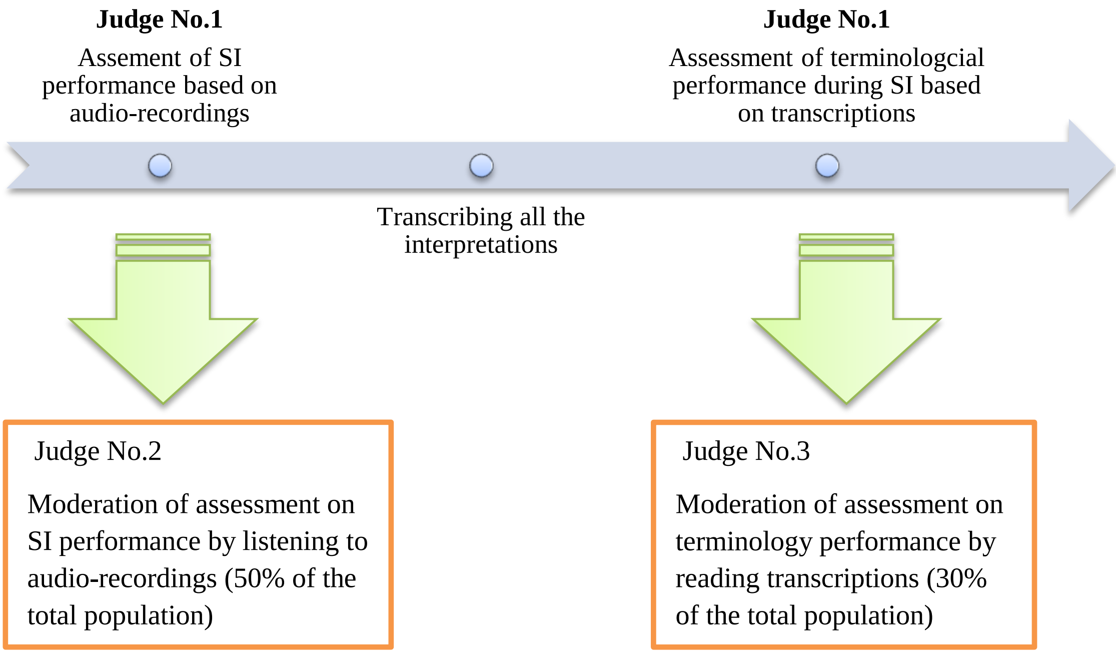


Figure 6: Procedure of assessments by the three judges

The two second markers (Judge No.2 and Judge No.3) were sent with all the necessary materials, including marking criteria, audio-recordings/ transcripts of participants' interpretations and the researcher's first marking/ annotations of term errors and degrees of departure. The two second markers (on two individual moderation tasks) were asked to check whether the relevant judging criteria and marking scheme were plausible and whether the first marker had reasonable and consistent judgement. The two second markers highlighted the disagreed items and provided their own judgments. Finally both the first and the second markers had discussions to resolve differences and produced agreed marks.

4.5.2 Holistic SI performance

The holistic SI performance score (%) was defined in this study as an all-encompassing score which reflected a participant's general interpreting performance on each source speech. Content, accuracy, target language use and delivery were all taken into consideration when a score was given to each individual participant. The assessments were carried out in line with the Marking Criteria in the MA Interpreting Final Exams at the University of Leeds, according to which a mark over 70% would suggest a solid performance to professional standard, and a mark below 50% would suggest that a student has not yet adequately mastered the skills required (see **Appendix A: 6**). The marking criteria used here was consistent with the criteria applied in getting the pre-test results in **Section 4.2.1** (the students' interpreting exam results).

As we have discussed in **Section 4.5.1**, the judges (both the first and second markers) were interpreting trainers at Leeds, who had also been involved in assessing the students' interpreting exams as examiners and therefore had been very familiar with the marking criteria applied in this study. Both the first and second markings were only based on listening to the audio-recordings of the interpreting outputs (with source speech on one track and interpretation on the other). After the first marking, the second marker moderated a sample of 44 interpretations, 50% of the total sample. The first marking results were all confirmed by the second marker. Therefore we were confident that our assessment results in this study were plausible.

4.5.3 Terminological accuracy and terminological error types

Another dependant variable in this experimental study is terminological performance during SI, which was observed by comparing terms in the source text and their correspondents in the target text. This study looked at interpreters' terminological performance in the following two area:

a. **Terminology accuracy**: The threshold for accuracy of terminology use in the TL was judged as acceptability by the TL audience (audience in specialised domain). **A terminological accuracy score (%)** was the ratio of the scores for interpretations of all the individual terms to the full score of the terms of that speech.
b. **Terminology errors**: terminology error categories, error numbers and degrees of departures for each error category.

4.5.3.1 Terminology error categories

As mentioned in **Section 2.2.3**, I reviewed the literature on frameworks used in interpreting and translation evaluation with special focus on error classification schemes including interpreting error taxonomies, such as Barik (1975, 1994), Altman (1994), Napier (2002, 2004), and translation quality metrics and evaluation tools, such as SAE J2450 (2001), BlackJack in Eckersley (2002), MeLLANGE in Secară (2005), etc. Unfortunately, none of the existing typologies could alone fully cater to the specific needs of this study. Adaptation of the existing typologies therefore seemed necessary.

In order to balance the total number of error categories at a manageable level, we decided to incorporate several categories from **BlackJack translation evaluation tool** and **MeLLANGE translation error annotation scheme** into a more general category. *"Inappropriate technical term in TT"*, *"inconsistent term in TT"* and *"wrong treatment of acronym/proper noun"* were incorporated into a single category – *"incorrect term"*.

Some of the error categories in translation quality matrix are written communication specific, which are not relevant to interpreting (oral communication). Therefore, this study decided to change *"misspelling"* and *"punctuation errors"* from **SAE J2450 Translation Quality Metric** into *"pronunciation errors"* for examining interpreter's performance.

In addition, in this study, whether omission of a term is judged as an error would depend on whether the omission would affect/alter the original meaning in ST. If the omission of a term does not affect the original meaning, it would not be counted as an error; if a term is omitted and it results in slightly altered meaning or loss of meaning, it would be counted as a minor error; if the omission results in a significant loss or change in meaning, it would constitute a serious error.

Adapting and combining the taxonomies discussed above, this study defined the following six categories of errors primarily focusing on interpreters' terminological performance.

Error categories	Definition for each category
Incorrect Term (IT)	The interpreted term is either not acceptable in the TL or the interpreter mentions something other than the original term in the TT. (Wrong/ inappropriate term, abbreviation and acronym in the TL)
Omission (OM)	The term under observation is completely left out by the interpreter in the TT, which would cause degraded understanding of the message.
Inappropriate collocation (IC)	The collocation of a term is inappropriate or unacceptable. The use does not conform to the TL norms.
Grammatical Error (GE)	This error category covers ungrammatical use of a term (inc. error in tense, agreement).
Pronunciation Error (PE)	The interpreter doesn't articulate the term in the TL, which causes the loss the original message or causes confusion in the TT.
Semantic Error (SE)	Even though the term is interpreted correctly, the meaning of the sentence where the term is used is distorted from the original.

Table 17: Terminology error categories and definitions for each category

The categories **IT** and **OM** cover instances when a term is interpreted badly, or the term is left out by the interpreter. **IC** and **GE** are errors relevant to poor use of language, which covers ungrammatical use of a term and inappropriate use of its collocation. **PE** is pronunciation error. So far all the above five error categories stay at the lexical level.

There are instances when terms are all interpreted correctly in a sentence, but the meaning at sentence level is distorted or does not make sense. This would be categorised as semantic error **(SE)** in this study. The category **SE** is established to ensure that a term can only be awarded a full score when both the term and the context where the term is used in the ST are interpreted correctly.

The following table summarised all the possible situations for terminological errors in interpreting, which have been all covered by the error-based model developed in this study.

Term in the ST	**The context where the term is used in the ST**	**Possible error types**
Correctly interpreted (√)	Correctly interpreted (√)	---
Incorrectly interpreted (X)	Correctly interpreted (√)	IT OM IC GE PE
Incorrectly interpreted (X)	Incorrectly interpreted (X)	IT OM IC GE PE SE
Correctly interpreted (√)	Incorrectly interpreted (X)	SE

Table 18: Possible situations for terminology errors in interpreting

4.5.3.2 Degrees of departures and a terminology scoring system

Departures of interpretation from the original affect the meaning of what is said to different degrees. Some instances represent a very minor departure; in other cases, it is more serious. In this study, each category of errors are further divided into **two severity levels**, i.e. **minor error** and **serious error** which are given different weights. For instance, when a term is interpreted acceptably, a full score of **"2"** is awarded, **"1"** is given for a minor error and **"0"** for a serious error.

Abbreviations are used for annotating interpreters' errors: **"2"** for acceptable interpretation, **"IT-1"** for Incorrect Term-minor error, **"IT-0"** for Incorrect Term-serious error.

This study establishes different thresholds for serious and minor terminology errors. It also includes a positive category for acceptable interpretation. The following table **(Table 19)** illustrates the scoring system covering all the six error categories mentioned above.

Error category	Score	Scoring Criteria
"Incorrect Term" (IT)	**2**	The interpretation of the term is acceptable in the TL.
	IT-1	The term is interpreted inaccurately; it only slightly distorts the intended meaning.
	IT-0	The inaccuracy results in a substantial loss or change of the intended meaning.
"Omission" (OM)	**2**	No omission, or the omission of the term doesn't affect the original meaning.
	OM-1	The term is omitted, which results in slightly altered meaning or loss of meaning. The gist of what was said was maintained.
	OM-0	The omission results in a significant loss or change in meaning.
"Inappropriate Collocation" (IC)	**2**	The collocation use is appropriate in the TL.
	IC-1	The interpreter uses an inappropriate collocation of the term. But it does not affect the understanding of the message among the TT audience.
	IC-0	The collocation of the term is unacceptable in the TL. The misuse of the collocation causes serious confusion to the TT audience.
"Grammatical Error" (GE)	**2**	No grammatical error.
	GE-1	There is an ungrammatical use of a term. But it doesn't quite affect the understanding of the message among the TT audience.
	GE-0	The misuse of tense or agreement causes serious confusion to the TT audience.
"Pronunciation Error" (PE)	**2**	No pronunciation error is made.
	PE-1	The term is mispronounced, but the audience can still understand what is said without making too much effort.
	PE-0	The term is mispronounced. The mispronunciation causes serious confusion to the TT audience.
"Semantic Error" (SE)	**2**	No semantic error is made.
	SE-1	The term itself is interpreted correctly, but the overall message of a larger unit is not quite the same thing as the original. The gist of the message is retained though.
	SE-0	Even though the term is interpreted acceptably, the interpretation of the larger unit has a considerable difference in meaning from the original. The interpreter makes up something on the basis of some part of the text.

Table 19: Terminology accuracy scoring system used in this study

This study also followed the instructions from SAE International on ambiguous error types**. "**When an error is ambiguous, for example, when it can belong to more than one error categories, always choose the primary category" and "when an error is in doubt between serious and minor, always choose serious over minor" (SAE International 2001:11).

Finally, A terminological accuracy score (%) is calculated by adding up all the scores an interpreter received for all the individual terms, then divide it by the full score of terms in the speech:

$$\frac{number\ of\ terms\ scored\ 2\ x\ 2 + number\ of\ terms\ scored\ 1\ x\ 1}{total\ number\ of\ terms\ in\ source\ speech\ x\ 2}$$

The students' terminology performance in SI was assessed in line with the above scoring system. At the initial stage, a pilot assessment of six interpretations (three interpretations on FR-en and three interpretations on FR-zh) was carried out. The researcher consulted academic colleagues to validate the scoring system and also discussed with them on a few examples of ambiguous cases in the pilot assessment. It was thus reassured that the scoring system and the assessment were acceptable and the evaluation should carry on in the same manner.

As explained earlier in **Section 4.5.1**, two judges were involved in the assessment of the students' terminology performance. The first marking was done by the researcher. The second marker moderated a sample of 24 interpretations, 28% of the total population already first marked, in order to check whether the first marker had reasonable and consistent judgement. Then both judges had a discussion to resolve differences and produced agreed marks. All the data were then reviewed by the researcher once again to identify any inconsistency in the earlier assessment and correct them subsequently.

4.5.3.3 Annotation

After the assessment results were checked and updated, the researcher used **UAM CorpusTool**[14] to annotate terminological errors in the students' interpretation.

---

14 UAM CorpusTool was used and downloaded from http://www.wagsoft.com/CorpusTool/index.html [Accessed 22 June 2015].

The tool allows annotation of each text on multiple levels and multiple texts using the same annotation schemes of user's design (see **Figure 7**).

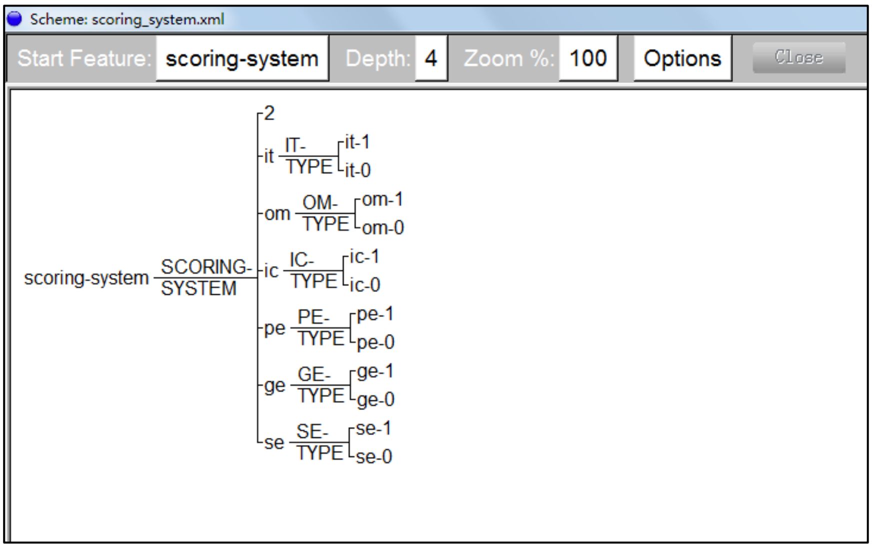


Figure 7: The UAM CorpusTool: the annotation system applied

After the annotation, we used the tool to search for instances of errors across different error categories. We also used a range of statistical analyses supported by the tool for revealing patterns of the annotation results. Below is a series of screenshots showing the major functions of the UAM CorpusTool (see **Figure 8-10**).

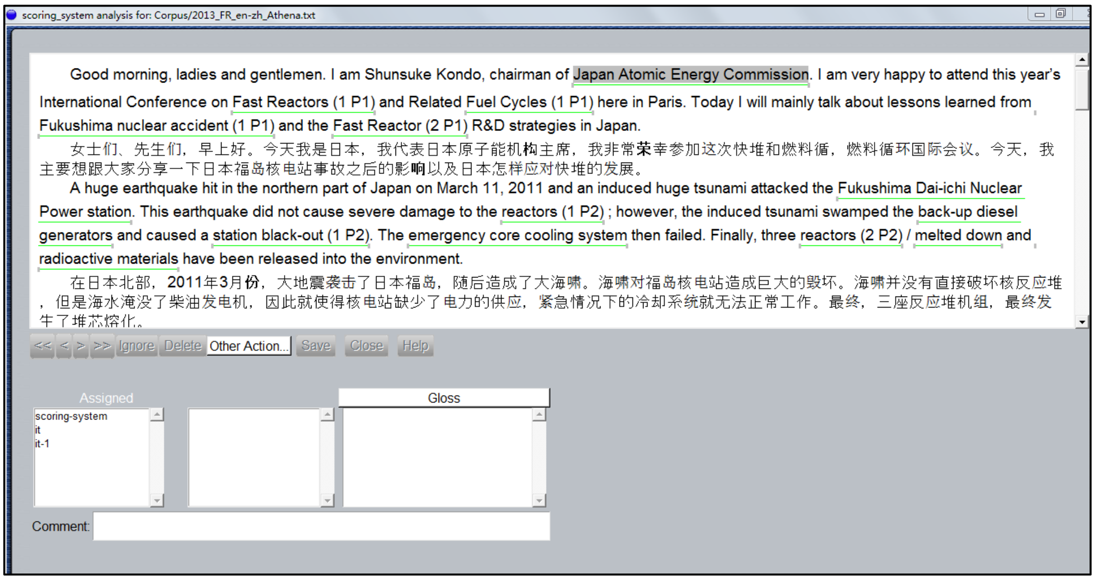


Figure 8: The UAM CorpusTool: annotating the errors in the English source text

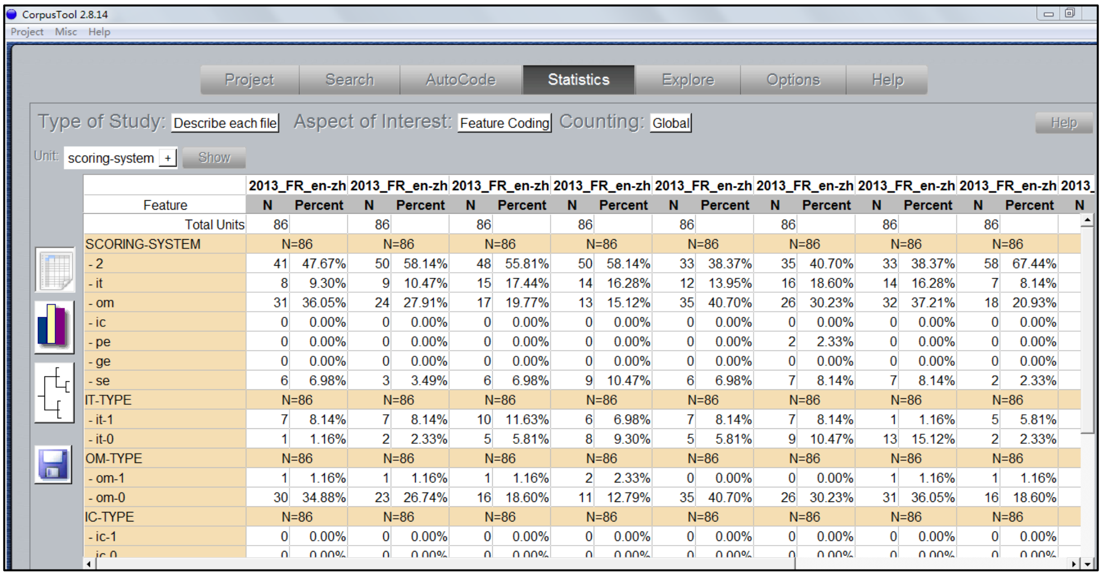


Figure 9: The UAM CorpusTool: statistics on errors in different categories

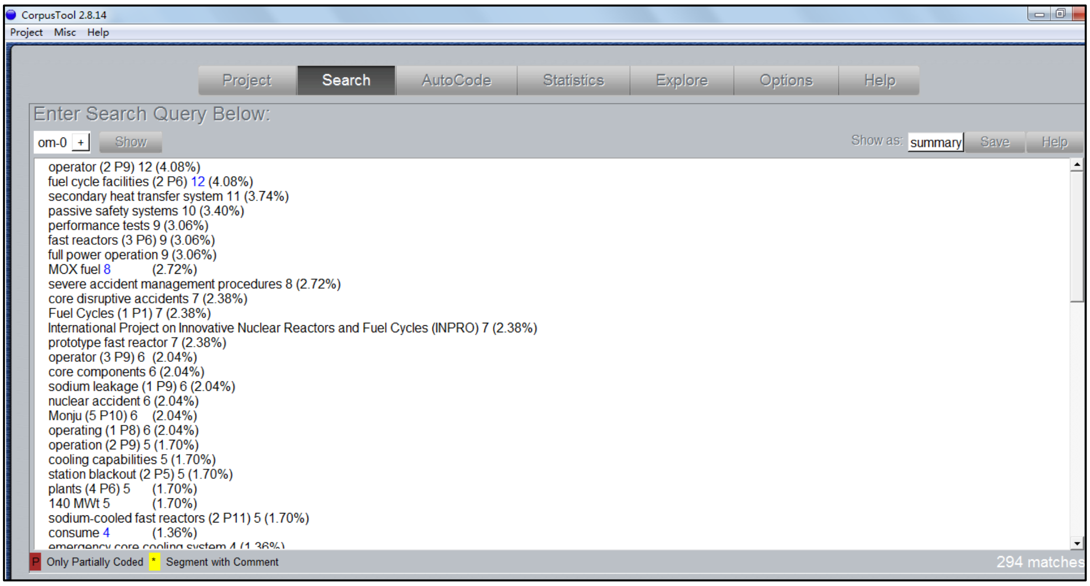


Figure 10: Search results and statistics of serious Omission Errors (OM-0)

4.5.4 Post-task recall of terms

As discussed in **Chapter 2** (**Section 2.5.2**), according to Craik and Lockhart's depth of processing hypothesis (1972), the deeper the processing of new words, the longer the retention of the words.

Based on this rationale, I also involved the students in term quizzes two months after their SI tasks. They were asked to fill out a quiz of 15 relevant terms on each of the two topics (FR & SM). The students were asked to write down in a printed form the Chinese translation of 15 English terms (for each topic) and to provide simple definitions of the terms in Chinese. Each term quiz has a total score of 15, the number of terms correctly translated would be calculated to see the effect of different terminology preparation procedure on the students' long-term memory of the terms. Moreover, their definitions of the terms would give us a rough idea on the depth of their understanding of the terms. Patterns of individual differences in cognitive capacities might also be observed.

All the terms in this term quiz were chosen from the students' termlists. In other words, all the terms were considered relevant by all the students during their preparation stage and had been activated to a certain extent for their SI tasks. Many of the terms also appeared in the source speeches that the students interpreted from in their SI tasks (see **Appendix A: 5** for the term quizzes on both topics).

4.6 Tasks and procedures

Having discussed the independent and dependent variables of the experimental study, this section will provide a detailed description of tasks and procedures of the two experiments.

4.6.1 Different preparation procedures

In this experimental study, we examined the impact of three different terminology preparation procedures on the student interpreters' simultaneous interpreting performance. The three preparation procedures were 1) traditional preparation procedure, 2) preparation with only the term extraction tool, 3) preparation with both term extraction and concordance tools.

4.6.1.1 Traditional preparation procedure

In this study, traditional preparation procedure is defined as interpreters' terminology workflow suggested by Moser-Mercer (1992), which requires reading the preparation documents, manually extracting the terms, followed by producing bilingual termlists and studying terminology for the interpreting tasks.

4.6.1.2 Preparation with only term extraction tool

Instead of reading the texts and manually extracting the terms, two monolingual termlists (each contained about 500 candidate terms) were extracted automatically and provided to the students. The students revised the auto-lists and produced bilingual lists of relevant terms specific to the interpreting task. They then studied the terms in the preparation documents for the interpreting tasks.

In order to avoid possible hidden variables in the experiments (e.g. the students' familiarity with the extractor, time spent on generating auto-lists), I provided the students with the auto-lists rather than asking them to use the automatic extractor themselves in the experiments.

4.6.1.3 Preparation with the use of both term extraction and concordance tools

Two monolingual termlists (each contained about 500 candidate terms) were extracted automatically and then were revised by the trainee interpreters to produce bilingual lists of relevant terms specific to the interpreting task. In addition, concordance tool was used to link the terms to their original contexts in the preparation documents in order to further activate the terms for the interpreting tasks.

4.6.2 Experimental settings

Human factors were taken into consideration in the experimental setup. We were aware that preparing for specialised topics with little prior knowledge of could be challenging for our participants (the trainee interpreters). In addition, the idea of corpora and corpus tools were completely new to them too. Using various corpus tools would require prior training (see **Section 4.6.3**) and a different way of thinking about terminology preparation for simultaneous interpreting. There were just too many new concepts to take in for the participants during the experiments.

In **Sections 4.6.1.2 & 4.6.1.3**, we could have asked the participants (in the test groups) to use an automatic term extractor during their preparation, but in the end, in order to avoid overburdening them with coping too many new things at the same time, we decided to streamline the preparation process by providing the auto-lists to the students directly.

As explained earlier in **Section 4.4**, this study introduced preparation time as a control variable: limited preparation time (3 days) and ample preparation time (9 days). And this study examined the three preparation procedures under these two time conditions (see **Table 20**).

	Groups	Preparation Method	Preparation Time
Experiment I (SM)	Control (2014)	Traditional	3 days
	Test (2013)	Using term extraction tool	3 days
Experiment II (FR)	Control (2013)	Traditional	9 days
	Test (2014)	Using both term extraction and concordance tools	9 days

Table 20: The two experiment settings

Each group in this experimental study followed the same sequence of activities (see **Figure 11**). First of all, the trainee interpreters received training on a particular terminology preparation procedure. Then in the pre-task briefing, the researcher announced the specialised topic (SM or FR) and provided basic information about the occasion and the speakers that the interpreters were going to interpret for. The researcher also provided to the interpreters the preparation documents in both English and Chinese, and explained the preparation procedure to be used for this task. Then the interpreters started their individual preparation subsequently. Their preparation included an initial preparation of termlists (individual), a group practice and further activation of terms and concepts (individual) (as defined in **Table 22**). After the preparation period (3 or 9 days), the interpreters were invited to simultaneously interpret two speeches (en & zh) on the specialised topic (FR or SM), and then had a focus group discussion with the researcher on their preparation process and their opinions on the terminology preparation procedure being used. Finally, two months after the interpreting task, the interpreters were invited again to take a term quiz of 15 terms on the same topic.

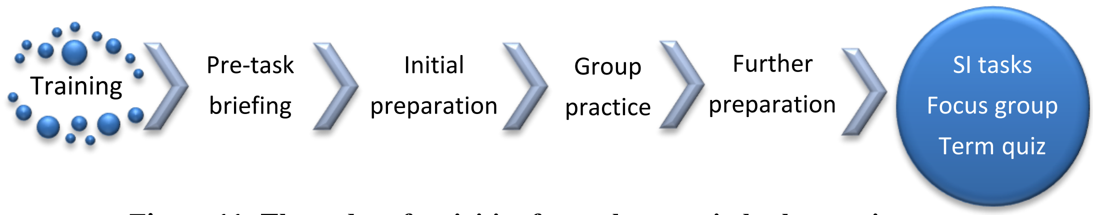


Figure 11: The order of activities for each group in both experiments

4.6.3 The trainings

The trainings were carried out outside the normal interpreting class as an extra-curriculum activity in a workshop on terminology preparation for technical meetings to the trainee interpreters. The workshop was arranged in the last term of MA interpreting training in 2013 and 2014.

The training focused on how to use term extraction and concordance tools. The students were expected to develop the skills to build their own terminology resources and activate relevant terms for simultaneous interpreting tasks in specialised fields. **Table 21** summarises the main contents of the trainings.

	The main contents of the trainings
First training	**The basics of interpreters' terminology preparation** ➢ What is terminology? ➢ Interpreters' terminology workflows ➢ Layouts of interpreters' termlists ➢ The use of Excel's sort & filter function (for organising termlists)
Second training	**How to use term extraction tools in interpreting preparation** ➢ Introduction to term extractors (TTC, Syllabs, Teaboat) ➢ Introduction to an annotation system (R/P/I/G/IL) for selecting relevant terms from auto-termlists ➢ Practicing annotating auto-termlists by applying the annotation system
Third training	**Training on the use of concordance tools in interpreting preparation** ➢ How to build one's own corpora? ➢ Introduction to various concordance tools ➢ Demonstration on how to use Sketch Engine's concordance function in interpreting preparation ➢ Practicing using SketchEngine to check terms and their uses in specialised corpora

Table 21: Trainings on terminology preparation for simultaneous interpreting

4.6.4 Pre-task preparations

As mentioned earlier, after relevant training was provided, the interpreters received a specialised topic, and then they started their pre-task preparation. **Table 22** demonstrates detailed activities that the interpreters were instructed to do during their preparation (i.e. initial preparation, group practice and further preparation) in all the groups.

	Initial preparation	Group practice	Further preparation
Control	➢ Reading through the speech agenda ➢ Understanding relevant background information from reading the preparation documents ➢ Highlighting terms while reading the documents and checking term equivalences ➢ Using Excel spread sheet to form an initial bilingual termlist	➢ Discussing logical systems of sorting terms in the termlists ➢ Brainstorming about relevant terms and information on the speech outlines ➢ Practising impromptu mini-speeches based on the speech outlines in the target langauge in small groups of 3-4 participants	➢ practising mini-speeches in the target langauge by using key terms learnt ➢ Practising sight translating some paragraphs in preparation documents ➢ Updating the long bilingual lists ➢ Learning the terms and equivalents by heart ➢ Producing a shorter list to be used in the booth
Test (SM)	➢ Reading through the speech agenda ➢ Annotating two monolingual auto-lists to select relevant terms ➢ Concentrating on learning the relevant terms while going through the preparation documents and checking term equivalences ➢ Using Excel spread sheet to form an initial bilingual termlists	➢ Same as control group	➢ Same as control group
Test (FR)	➢ Reading through the speech agenda ➢ Annotating two monolingual auto-termlists to select relevant terms ➢ Purposely reading about the relevant terms in the preparation documents by using Sketch Engine's concordance function; checking term equivalences ➢ Using Excel spread sheet to form an initial bilingual termlist	➢ Same as control group	➢ Using Sketch Engine to further check contexts and collocations of the terms as well as other relevant terms that might be closely linked to the terms already known. ➢ Other activites were the same as the control group

Table 22: The preparation activities

4.6.4 SI tasks

The simultaneous interpreting sessions were carried out in the interpreting laboratory at the University of Leeds. Each individual session (SI tasks on a specialised topic) lasted for one hour. All the participants signed an informed consent form that explained the purpose of the study and information about the experimental task.

There were two speeches (English and Chinese) on the specialised topic to be interpreted. They were of approximately similar length of 10 minutes. They were delivered and interpreted in one go with a break of 2 minutes between the speeches, and instructions were given beforehand. Participants simultaneously interpreted the speeches in front of an audience, and their interpretations were audio-recorded in a sound-proofed booth in the interpreting training facility, with the source speech on one sound track and the interpretation on the other. Both the original audio recordings and the transcripts of the interpretation would be used later to assess the participants' holistic interpreting performance and terminology performance during interpreting (also see **Section4.5.1-4.5.3**).

After the two interpreting tasks, participants were asked to fill in a questionnaire on their prior knowledge about the topics, terminology and the ideas of the source speeches before their preparation (also see **Section 4.2.2**).

4.6.5 Focus group

The study also had a group discussion session, which was conducted with the participants after the SI tasks. Each focus group had four to six participants apart from the main researcher, and each group interview lasted for about one hour. The interview was in Q&A fashion. Questions included both prompt questions and open-ended questions, which were asked by the researcher, and participants gave their own answers in a spontaneous open discussion. The focus group discussion was an open discussion on terminology preparation, during which every participant was heard.

This session was designed with the purpose of gathering information on the students' preparation process, the time they spent on their preparation and their opinions on currently used terminology preparation strategy, etc. The whole process was audio-recorded.

The process of data analysis for the focus groups followed Hale & Napier (2013), which involved:

1) Transcribing
2) Thematic analysis to identify overarching themes (see **Table 23**)
3) Content analysis to pull out representative quotes to elucidate various themes (see **Chapter 6, Section 6.3**)

1. Real preparation time used
2. Preparation strategy
3. The use of the automatically-generated lists
4. The use of the concordancer
5. Activation activities
6. Personalised short lists used in the booth
7. Challenges during prep
8. Opinions on the training

Table 23: The themes covered in the focus groups

4.6.6 Term quiz

Two months after the interpreting tasks, both the control and test groups were invited to take a term quiz, in which the participants were asked to write down Chinese translations of 15 English terms and provide simple definitions of the terms in Chinese. The 15 terms were specialised terms from the original preparation documents that had been provided to the participants for their preparation before the SI tasks. The term quiz was supposed to examine what the participants have remembered two months after their preparation and the SI tasks.

In summary, **Chapter 4** described the main methodological approaches of this study. This chapter also highlighted a method to evaluate performance of interpreters with respect to their use of terminology. The following four chapters will present and discuss the major findings of this study.

Chapter 5 - Evaluation of automatic term extraction

In **Chapter 3**, I described a corpus-based terminology preparation pipeline for interpreters covering corpus building, term extraction, term exploration and term management. I also reviewed several corpus tools that can potentially be used to achieve these goals. This chapter will mainly compare term extraction performance of three existing automatic extractors (TTC TermSuite, Syllabs Tools and Teaboat) from comparable texts of different sizes on two domains (fast reactors and seabed minerals). In **Section 5.1**, I will further describe how corpus building and term extraction are applied in this study. In **Section 5.2**, the results of numeric evaluation of three automatic term extractors will be presented. **Section 5.3** will discuss technical challenges in term extraction for interpreting purpose. Finally, a single term extractor will be selected to be used in the test groups of the SI experiments (also see **Chapter 6 & 7**).

5.1 Corpus collection and term extraction

5.1.1 Description of the procedure

5.1.1.1 Two specialised topics

As explained in **Chapter 4 - Methodology**, this study chose two specialised topics: fast reactors (FR) and Seabed minerals (SM). On each topic, two monolingual specialised corpora representing preparation documents were compiled in both English and Chinese by the researcher. **12 MA student interpreters** from the cohort of 2012-2013 were invited to prepare for SI tasks on these two topics. They were provided with the monolingual specialised corpora (En & Zh) for their preparation on each of the topics (FR & SM)

5.1.1.2 Three term extractors

The group of students started with the FR topic. They were asked to manually generate their own termlists from the provided corpora (FR1_En & Zh) in their preparation before simultaneously interpreting two speeches on the topic (En & Zh). After their SI tasks, they were then asked to evaluate the relevance of two monolingual lists (En & Zh) which were automatically generated by one of the three tools (TTC TermSuite, Syllabs Tools and TeaBoat). The purpose here is to see which tool could extract more relevant terms for the needs of the trainee interpreters.

We collected and compared the annotation results (on relevant terms) from the students and selected a single tool with comparatively better performance. We then invited the same group of students to prepare for the other topic (SM) with the use of automatically-generated lists in their interpreting preparation. Their annotation results (on relevant terms) were also collected to see whether the term extractor could have consistent extraction performance on the two topics.

5.1.2 Corpus compilation

There are two types of sources where the comparable corpora were from:

1) Conference documents (such as agenda) as well as relevant background documents provided by the conference organisers (in this case, provided by the researcher as the task organiser)

2) Specialised corpora collected from the internet using WebBootCat (Baroni and Bernardini, 2004; Baroni et al., 2006)

	FR0		**FR1**		**FR2**	
	En	Zh	En	Zh	En	Zh
Texts	1	1	9	9	81	86
size	774	1,641	42,006	30,174	206,197	129,350

	SM0		**SM1**		**SM2**	
	En	Zh	En	Zh	En	Zh
Texts	1	1	9	12	74	84
size	1,025	1,830	20,533	40,545	166,499	116,235

Table 24: The corpora used in this study (the size is in words for En, in characters for Zh)

Table 24 presents all the corpora we use in this study. They are of different sizes. The size of FR/SM0 and FR/SM1 reflects the typical amount of documents received by the professional interpreters in advance for their preparation.

FR0/SM0 was a single relevant document, representing the speech that the trainee interpreters were asked to interpret from in the SI experiments. Very often a text of this length is the only source of information given to the interpreters in advance.

FR1 (En & Zh) and **SM1 (En & Zh)** are comparable corpora, which represent conference documents and relevant background documents provided by the conference organisers, including speech outlines, research papers from experts and research institutes, reports from national and international authorities, as well as popular science articles, Wikipedia articles, specialised journal articles and interviews, etc. (also see **Chapter 4, Section 4.3.3**).

FR2 (En & Zh) and SM2 (En & Zh) are corpora collected by web crawling using Bootcat to send queries to search engines (Baroni and Bernardini, 2004; Baroni et al., 2006). WebBootCaT was used, since it is a user-friendly web-based tool. It takes only a few minutes to build a specialised corpus from the publicly accessible web pages and to extract its contents from the web pages. WebBootCaT is particularly useful for small, short-term projects such as preparing for topic-based materials for interpreting assignments.

For instance, to produce FR2 we started with a set of ten relevant keywords in English and Chinese as shown in **Table 25**, then used Bootcat to retrieve online resources and generate two corpora (FR2_En & Zh). All the keyword seeds were manually selected terms from the English speech-FR0 that the students were supposed to interpret from, and were therefore considered very relevant and important terms. The Chinese keywords are the translations of the English ones.

Seeds (En)	Seeds (Zh)
fast breeder reactor	快中子增殖反应堆
fission	裂变
decay heat	余热
uranium	铀
plutonium	钚
core damage	堆芯损坏
fukushima accident	福岛事故
nuclear waste	核废料
fuel cycle	燃料循环
coolant	冷却剂

Table 25: Parallel keyword seeds on Fast Reactors for FR2

Corpus pre-processing includes webpage cleaning (Baroni et al., 2008), as well as basic linguistic processing (i.e. lemmatisation, tokenisation and part-of-speech tagging, etc.). Lemmatisation is needed because the keywords in a glossary are expected to be in their dictionary form. Lemmatisation also helps in reducing data sparsity for the singular and plural forms, e.g., *sulphide deposit vs sulphide deposit(s)*. However, lemmatisation also leads to imperfect terms, e.g., *recognise type of marine resource,* which corresponds to *recognised type(s) of marine resources*. Tokenisation is a process of breaking a plain text up into meaningful constituent elements called tokens. In languages such as English, words are delimited by whitespace, this approach is relatively straightforward. However, tokenisation is more difficult for languages such as Chinese which has no word boundaries.

In this study, lemmatisation and tagging for English was done using TreeTagger (Schmid, 1994), while for Chinese we used “Segmenter”, an automatic tokenisation tool (Liang et al., 2010:46) followed by “TreeTagger” for POS (part-of-speech) tagging. However, there are errors in automatic tokenisation and POS tagging, and these errors would influence the next step of automatic term extraction. I will further discuss the types of errors in automatic tokenisation and POS tagging in **Section 5.3.1**.

5.1.3 Automatic monolingual term extraction

Monolingual term extraction is the process by which candidate terms are extracted from a monolingual corpus. Term extraction tools are particularly useful for new domains with few terminological resources publicly available. In this study, we compare the extraction results of three representative term extractors, namely, TTC TermSuite, Syllabs Tools and Teaboat on two domains, i.e., fast breeder reactors (FR) and deep seabed minerals (SM). All the three tools are modules developed by the TTC partners and are available on the TTC web platform. The service is available for some other languages as well: DE, ES, FR, LV and RU.

TTC TermSuite (Daille, 2012) is based on lexical patterns defined in terms of POS tags with frequency comparison against a reference corpus using specificity index (Ahmad et al., 1994). The tool extracts both single (SWT) and multi-word terms (MWT) outputs their lemmas, part of speech, lexical pattern, term variants (if any), etc. The candidate terms are ranked according to their relative frequency and their domain specificity. The most important feature of the TTC TermSuite is the fact that term candidates can be grouped with their corresponding term variants. **Figure 12** presents the output of TTC Term Suite on the TTC Web Platform.

Syllabs Tools (Blancafort et al., 2013) is a knowledge-poor tool, which is based on unsupervised detection of POS tags, following the procedure of Clark (2003:61-62), and on the Conditional Random Field framework for term extraction (Lafferty et al., 2001:284-285). In comparison to the knowledge-rich tool (e.g. TTC TermSuite) using POS tagger and hand-written rules to identify term candidates, the use of knowledge-poor methods just needs a big raw corpus, as well as a small corpus with manually annotated sentences (noun phrases) to train the term extractor.

Teaboat (Sharoff, 2012) does term extraction by detecting noun phrases using simple POS patterns in IMS Corpus Workbench (Christ, 1994:2-4) and by applying log-likelihood statistics (Rayson and Garside, 2000:2-3) to rank terms by their relevance to the corpus in question against the Internet reference corpora for English and Chinese (Sharoff, 2006:436) (also see **Chapter 3, Section 3.3** for more discussion on other possible tools to generate termlists).

52043	nodule	nodules	Noun	Noun	Single-word	130	84	0.0056	**Form** / **Occurr.** nodules 52 nodule 32
158834	international seabed	international seabed	Noun	Adjective-noun	Multi-word	118	76	0.0050	**Form** / **Occurr.** international seabed 76
11257	seamount	seamounts	Noun	Noun	Single-word	110	71	0.0047	**Form** / **Occurr.** seamounts 40 seamount 31
32209	sulphide	sulphide	Noun	Noun	Single-word	110	71	0.0047	**Form** / **Occurr.** sulphide 42 sulphides 29

Figure 12: Output of TTC Term Suite on the TTC Web Platform

5.2 Term extraction evaluation

5.2.1 Evaluation annotation system

Fantinuoli (2006) used five categories to investigate the level of specialisation and well-formedness of an automatically-generated candidate termlist:

1) Specialised terms that were manually extracted by the terminologist (and are contained in the reference term list);
2) Highly specialised terms that were not detected by the terminologist;
3) Non-specialised terms that are commonly used in the field of his study (medicine);
4) General terms that are not specific to the medical field;
5) Ill-formed, incomplete expressions and fragments.

Our annotation system extends Fantinuoli's study because the purpose of annotation in this project is to give the interpreters the possibility to extract relevant terms from all the candidate terms regardless of their levels of specialisation. Our premise is that interpreters may need relevant terms, both highly specialised and less specialised, in order to prepare themselves for a conference. The annotators are the end users of the list, i.e. the trainee interpreters who participated in this study. Since the interpreters are tasked with interpreting speeches in the domain, they need themselves to decide what is likely to be relevant instead of relying on the terminologists who describe the overall structure of the domain. The following is the five-category annotation system that we use in this research:

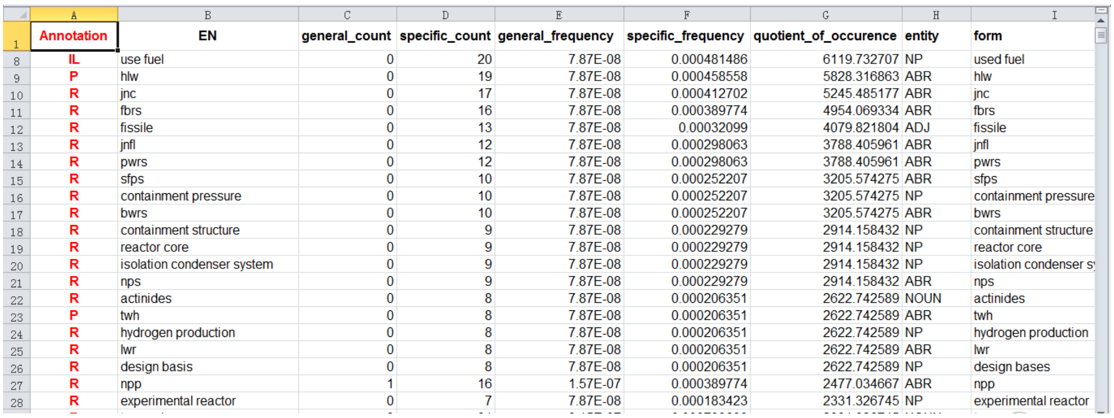
R	Relevant terms (terms closely relevant to the topic), *e.g. breed ratio, uranium-238, decay heat removal system*
P	Potentially relevant terms (a category between "I" and "R": they are terms; but annotators are not sure whether they are closely relevant to the topic of their assignment), *e.g. daughter nuclide, neutron poison, Western reactor*
I	Irrelevant terms (terms not relevant to the topic), *e.g. schematic diagram, milk crate*
G	General words (rather than terms), *e.g. technical option, monthly donation, Google tag, discussion forum;*
IL	Ill-formed constructions (parts of terms or chunks of words), *e.g. var, loss of cooling, separate sample container, first baseline data, control ranging*

Table 26: Five-category annotation system for automatic termlists

5.2.2 Evaluation results

It only took several minutes to generate a termlist after uploading the designated corpus onto TTC TermSuite, Syllabs Tools and TeaBoat. Each of them automatically generated corresponding monolingual termlists sorted by their term specificity scores. For all the tools we set the threshold of obtaining 500 terms (if possible), as a practical limit for all evaluation experiments.

The trainee interpreters were asked to annotate the list by using the annotation system above. The following **Figure 13** is a screenshot of an automatic termlist in an Excel spreadsheet as annotated by a student. The students reported that it took them about 60 minutes to annotate both lists (in En & Zh) on each of the topics (FR & SM). All the student annotators were briefed about what counts as a term and the annotation system before they started their evaluation of the term lists.



	A	B	C	D	E	F	G	H	I
1	Annotation	EN	general_count	specific_count	general_frequency	specific_frequency	quotient_of_occurence	entity	form
8	IL	use fuel	0	20	7.87E-08	0.000481486	6119.732707	NP	used fuel
9	P	hlw	0	19	7.87E-08	0.000458558	5828.316863	ABR	hlw
10	R	jnc	0	17	7.87E-08	0.000412702	5245.485177	ABR	jnc
11	R	fbrs	0	16	7.87E-08	0.000389774	4954.069334	ABR	fbrs
12	R	fissile	0	13	7.87E-08	0.00032099	4079.821804	ADJ	fissile
13	R	jnfl	0	12	7.87E-08	0.000298063	3788.405961	ABR	jnfl
14	R	pwrs	0	12	7.87E-08	0.000298063	3788.405961	ABR	pwrs
15	R	sfps	0	10	7.87E-08	0.000252207	3205.574275	ABR	sfps
16	R	containment pressure	0	10	7.87E-08	0.000252207	3205.574275	NP	containment pressure
17	R	bwrs	0	10	7.87E-08	0.000252207	3205.574275	ABR	bwrs
18	R	containment structure	0	9	7.87E-08	0.000229279	2914.158432	NP	containment structure
19	R	reactor core	0	9	7.87E-08	0.000229279	2914.158432	NP	reactor core
20	R	isolation condenser system	0	9	7.87E-08	0.000229279	2914.158432	NP	isolation condenser sy
21	R	nps	0	9	7.87E-08	0.000229279	2914.158432	ABR	nps
22	R	actinides	0	8	7.87E-08	0.000206351	2622.742589	NOUN	actinides
23	P	twh	0	8	7.87E-08	0.000206351	2622.742589	ABR	twh
24	R	hydrogen production	0	8	7.87E-08	0.000206351	2622.742589	NP	hydrogen production
25	R	lwr	0	8	7.87E-08	0.000206351	2622.742589	ABR	lwr
26	R	design basis	0	8	7.87E-08	0.000206351	2622.742589	NP	design bases
27	R	npp	1	16	1.57E-07	0.000389774	2477.034667	ABR	npp
28	R	experimental reactor	0	7	7.87E-08	0.000183423	2331.326745	NP	experimental reactor

Figure 13: Screenshot of a student's annotation of a monolingual auto-list

5.2.2.1 Inter-annotator disagreement

Table 27 shows the numbers of annotators (trainee interpreters) we involved for evaluating different automatically-generated termlists in this study.

	FR0_lists	FR1_lists	FR2_lists	SM1_lists
Syllabs	2	4	2	12
Teaboat	2	4	2	0
TTC	2	6	2	0

Table 27: The number of annotators for different auto-lists

We aim for consistency, yet inter-annotator disagreement does exist and there is a certain degree of subjectivity in annotation. To measure the level of agreement among annotators, we used Krippendoff's alpha (α) over the other available measures, because Krippendoff's α offers an extension of such measures as Fleiss' κ and Scott's pi (π) by introducing interval-scale ratings, thus making it possible to compute distances for the pairwise disagreements (Krippendorff, 2004:419). When measuring disagreement on categorical items, nominal-scale rating is normally used. However, the interval-scale rating is more suitable in our study, for it allows us to see the degrees of difference between the categorical items, e.g. the disagreement between **R** and **I** might be less severe than between **R** and **P**.

Auto-lists	FR1_TTC		FR1_Teaboat		FR1_Syllabs		SM1_Syllabs	
Language	En	Zh	En	Zh	En	Zh	En	Zh
Krippendorff's α	**0.541**	**0.500**	0.166	**0.435**	0.181	**0.662**	0.117	0.221
Annotators	6		4		4		12	

Table 28: Krippendorff's α for different auto-lists

The closer the value of Krippendorff's α to 1 means the higher the agreement is. The values in **Table 28** are relatively low, which means disagreement between the student annotators is quite high. We find that the most common cases of disagreement are between **R** and **P**. The boundary between **R** and **P** often depends on the amount of knowledge on the side of the annotator. However, the interpreters are learners of knowledge in a domain new to them, rather than domain specialists. In addition, there are terms for which it is not easy to make quick judgement without viewing the context (e.g. abbreviations, or general words with specific meaning in the domain). Quite surprisingly, the disagreement between **R** and **IL** is also quite high, since some annotators interpreted ill-formed sequences as a contribution to useful terms (*e.g. first baseline data* for *baseline data*). With more training on using the annotation system, we may reduce the discrepancies between **R** and **IL**. And for interpreters, it is on the safe side if they include in the first place the terms that they are not sure (whether relevant or not) into the category **P**. They could make further judgement on the possibly relevant terms (**P**) when they investigate more contexts to get familiar with the domain.

5.2.2.2 Evaluation results on FR

With the disagreement taken into consideration, our evaluation on the number of relevant terms was judged by the agreement between at least two annotators among two to six annotators for the topic of FR. This established the gold standard lists reported in **Table 29**.

	FR0_En	FR1_En	FR2_En	FR1_Zh
Syllabs	85/104 **(82%)**	309/500 **(62%)**	400/500 (80%)	156/500 **(31%)**
Teaboat	44/56 (79%)	232/376 **(62%)**	413/499 **(83%)**	141/450 **(31%)**
TTC	NA	136/500 (27%)	287/500 (57%)	119/500 (24%)

Table 29: The number of relevant terms (R) against candidate terms in the auto-lists (FR)

The annotation results from **Table 29** for the lists in English show that Syllabs generates more relevant terms than the other two tools from both FR0_En and FR1_En. Both Syllabs and Teaboat generate good numbers of relevant terms from FR2_En. In addition, Syllabs' and TeaBoat's English lists contain more specialised terms in the domain of FR, such as *defence-in-depth, once-through fuel cycle, and suppression chamber of the containment, etc.* However, these specialised terms with relatively low frequency are not included in the TTC's list. The terms included in TTC's list are more general terms, such as *steam, heat and leak, etc.*, which are likely to be already known by the trainee interpreters.

The English termlists from all the three tools contain a number of repetitions in the form of term variants, following Daille's definition as "an utterance which is semantically and conceptually related to an original term" (Daille, 2005:182). The automatically-generated termlists contain the following types of term variations, which are counted as individual term candidates scattered in the termlists:

Morphological variation: *bathymetry vs bathymetric (not different when translated into Chinese)*

Anaphoric variation: *pollymetallic sulphide deposit vs deposit*

Pattern switching: *meltdown of the core vs core meltdown; level of gamma radiation vs gamma radiation level*

Synonymy in variation: *deep sea mining vs deep seabed mining, seabed vs seafloor, ferromanganese crust vs iron-manganese crust*

One the one hand, these variations provide useful lexical information about terms, preparing the interpreters for what is possible in their assignment; on the other hand, the term variations need to be explicitly linked and grouped together, which is possible only in the TTC TermSuite tool. Identification of synonymic variations is probably more challenging than the other three types of term variants. Some synonymic variants were either not being able to be extracted successfully or not grouped together with their corresponding base terms. For example, *"MOX fuel"*, a term in the domain of fast breeder reactors, is a variation of *"mixed-oxide fuel"*. The TTC TermSuite extracted both *"MOX"* and *"MOX fuel"*, but *"mixed-oxide fuel"* is left out. Syllabs Tools extracted *"mixed-oxide fuel"*, but excluded *"MOX fuel"*.

The annotation results from **Table 29** for the lists in Chinese show that both Syllabs' and Teaboat's lists offer noticeably fewer relevant terms from FR1_Zh compared with the English lists. After further investigation, we find that Syllabs' Chinese list on FR1_Zh contains a large number of ill-formed constructions, including incomplete terms, e.g. *"水堆" (water reactor)*, *"里岛核电站" (Mile Island nuclear plant)* and longer chunks, e.g. *"最大程度上保证了钠"*, *"可用压水堆后处理得到的钚作为核燃料"*. Teaboat's list contains a number of general words, e.g. *"开发" (development)*, *"生产" (production)* or *"工程" (project)*. Both categories (G and IL) are frequent in the TTC's Chinese list. **Table 30** summarises the distribution of the five annotation categories for each automatically-generated termlist based on the students' annotation results.

	Syllabs_FR1		Teaboat_FR1		TTC_FR1	
	En	**Zh**	**En**	**Zh**	**En**	**Zh**
Total	500	500	376	450	500	500
R	309	156	232	141	136	119
P	90	73	33	61	48	32
I	15	5	19	7	3	4
G	56	46	73	191	310	209
IL	30	220	19	50	3	136

Table 30: The distribution of the annotation categories

5.2.2.3 Evaluation result on SM

On the basis of these results, we selected a single tool (Syllabs) with comparatively better performance in both languages to generate termlists on SM1 (En & Zh) and asked 12 annotators to select the relevant terms and learn the terms during their preparation to the interpreting task. Among the 500 candidate terms for English, 441 terms were agreed as relevant by at least two annotators, 266 terms were agreed by five annotators. Precision rates are 88.2% and 53.2% respectively. On the other hand, only 130 terms were agreed as relevant by two annotators from the 500 Chinese candidate terms. The precision rate for the Chinese list is 26%. The results basically replicate the previous findings on FR1.

The other pattern we observe from the current data is that the larger the corpus is, the more relevant terms the tools can generate (also see **Table 29**). When the corpus is of very limited size (e.g., FR0-En has only 774 words), the TTC TermSuite fails to generate any list for a "corpus" of only 774 words, while the Syllabs and Teaboat tools produce shorter lists of 104 or 56 terms respectively. The situation is similar to other studies, for example, Matsuo and Ishizuka (2004:166) which used small (single-document) corpora.

5.3 Discussion of the evaluation results

5.3.1 Reliability of the three term extractors

The results show the accuracy of the terminology extraction pipelines is not perfect, as its precision ranges from 27% on short texts to 88.2% on bigger corpora for English, 24% to 31% for Chinese. Among the three term extractors (TTC TermSuite, Syllabs Tools and Teaboat), Syllabs is more reliable in generating more relevant terms in English. All the three tools perform less satisfactory in generating relevant terms in Chinese. We hypothesise that at least three factors play an important role here:

5.3.1.1 Tokenisation errors

As mentioned earlier in **Section 5.1.2**, Chinese is written without explicit word boundaries, therefore the Chinese corpora need to be tokenised before automatic term extraction. Errors of the tokenisation process lead to difficulties in obtaining proper terms, e.g., *"一回路" (primary loop)* becomes *"一回" (once) "路" (road)*, also *"和非能动安全性" (and passive security)* becomes *"和非" (and not) "能动" (active) "安全性" (security)*, which reduces the chances of detecting *"非能动安全性" (passive security)* as a term.

5.3.1.2 Word-class ambiguity

Ambiguity in word class (verbs and nouns) in Chinese is high. This leads to POS tagging errors. For example, when nouns are treated as verbs, and this breaks the POS patterns for term extraction, e.g., *"示范堆" (demonstration reactor)* is treated as *"示范/vn 堆/v"*.

5.3.1.3 Flexible term patterns in Chinese

As discussed in **Section 5.1.3**, TTC and Teaboat are based on knowledge-rich term extraction approach. They use supervised POS tagging (LCMC tagset), extracting terms with such patterns as: "Adj+N", "Adj/adv/N+的/地/得+N", and "N+的/地/得+N". However, terms in Chinese are more flexible and exhibit more patterns than captured by the two term extraction tools we tested. For example, *"并网发电" (connect to the grid)* is potentially a useful term, which is correctly POS-tagged as *"并网/v 发电/vn"*, but not captured by the patterns in the tools.

Syllabs is based on probabilistic term extraction approach (knowledge-poor approach). We expected that for a language like Chinese, the knowledge-poor approach using unsupervised tagging may discover more term patterns than the knowledge-rich approach, however the precision rates on Chinese are similarly low no matter the knowledge-rich or knowledge-poor approaches are used (24-31%).

A possible reason is that the knowledge-poor approach depends entirely on frequency statistics. If a term only appears once or twice in a corpus, it would not be detected as a term. Due to its low frequency in the Chinese corpus, *"并网发电" (connect to the grid)* failed to be captured as a term by the knowledge-poor tool, Syllabs.

Furthermore, on the one hand, the knowledge-poor approach may discover more term patterns than the rule-based approach. On the other hand, human knowledge (rule-based approach) is normally more reliable than statistics obtained from a small sample. In other words, more term patterns overall means more noise in the list of extracted terms, thus negatively affecting precision, which is a trade-off effect between recall and precision (Sharoff & Hartley, 2012:336).

In summary, the first two factors on the unsatisfactory results in Chinese, i.e. tokenisation errors and word-class ambiguity, both concern text pre-processing. Further investigation might be helpful in finding out how the pre-processing steps affect the performance of the term extractors and which terms are affected by each source of errors. The third factor is that Chinese exhibits more term patterns than could be extracted so far. One possible way to improve extraction performance for Chinese is to combine both knowledge-rich and knowledge-poor methods in a semi-supervised setting.

5.3.2 Manual selection Vs automatic extraction of terms

For the interpreters, manually selecting terms from a single document of limited size (e.g. FR0_En=774 words) is possible. However, when conference documents amount to the size of FR1 (FR1_En=42,006 words), it took the trainee interpreters 8.2 hours on average to extract terms manually and to produce initial termlists, since they had to spend the majority of their time reading through fairly complex documents, copying the terms from the texts onto their own termlists and searching for unfamiliar terms.

With the use of automatically-generated termlists on the same preparation task, The students in the experiment group spent an average of 4.3 hours annotating and producing their initial bilingual termlists. Therefore nearly half of the time spent on reading and searching for terms could be saved for the interpreters to get familiar with the concepts relevant to the terms and further activate the terms for their simultaneous interpreting tasks.

Furthermore, if interpreters are given limited time for preparation, they would not be able to read through larger corpora of the size of FR2 (FR2_En=206,197 words) and produce termlists from them manually. That is probably when such tools we discussed in this chapter may have obvious advantages over the manual terms extraction by the interpreters. Moreover, in **Chapter 6 and 7** I will demonstrate that in addition to providing an automatically-extracted termlist, it is also beneficial to link the terms to their uses in the concordance lines of the corpus they have been extracted from (by using concordancer). This is expected to give the interpreters easy access to the context of the terms to see how they are used and get more background knowledge about the domain.

5.3.3 Feedback from the students

After doing annotation, the students offered their feedback on the termlists generated by the three term extractors in a group discussion.

The students commented on the usefulness and reliability of two monolingual lists generated by the extraction tool (Syllabs). They generally reported that the English termlist provided a good number of relevant terms on the topic. Some of them found the list "unexpectedly accurate and complete", and the presence of irrelevant words and the repetitions in the English lists "tolerable".

On the other hand, the students reported that the Chinese termlist offered much fewer relevant terms and contained quite a number of ill-formed constructions and repetitions compared with the English list. Therefore they felt the lists in Chinese were less reliable and less useful in their interpreting preparation.

Though the English list is much better than the Chinese list, the students pointed out some general problems. For example, the terms longer than four words/tokens (e.g. names of conventions, names of organisations and specific mining methods) were missing from the automatically-generated lists. Some of the important terms only appeared towards the very end of the automatic lists (containing about 500 candidate terms), which could be easily overlooked in annotation.

5.3.4 Extraction of proper names

Proper names (including names of organisations, names of places, names and titles of people) are equally important as terms for interpreters, yet many of them are not included in the automatically-generated lists by the three term extractors (TTC TermSuite, Syllabs Tools and Teaboat). Therefore, named entity extraction tools in addition to term extraction are needed to generate more complete lists for interpreters' use. This issue will be further explored in our future research. The POS patterns used by the named entity extractors are quite different from term extraction. They need to use keywords such as "organization", "association" and "river" to detect named entities and their variants. The frequencies of the name entities can be also quite low in a given corpus, therefore it is difficult for noun phrase extractors (e.g., Syllabs Tools and Teaboat) to detect them as terms, while the named entity extractors rely on extra information available in very large text collections, such as Wikipedia.

5.3.5 File formats, plain text, encodings

Finally, it is worth mentioning that all the tools we tested can only process plain text (including UTF-8). Nevertheless, all the meeting documents are normally in one of the word processing formats (.pdf, .doc, .xls or .ppt) other than .txt. Interpreters need to take some time to convert all the files they obtain from their customers into plain text before they can possibly use any tool mentioned above.

In summary, this chapter compared term extraction performance of three tools (TTC TermSuite, Syllabs and Teaboat) from comparable texts (in En & Zh) of different sizes on two domains (fast reactors and seabed minerals). A single term extractor (Syllabs) with comparatively better performance was selected to be used in the test groups of SI experiments. In the next two chapters, we will continue to investigate the impact of using the term extractor (Syllabs) on the trainee interpreters' SI performance.

Chapter 6 - Data Analysis

In **Chapter 2 and 3**, I reviewed several preparation procedures and possible electronic tools that could assist the terminology preparation for interpreters. This chapter presents the results of the experiments with the trainee interpreters using mainly two tools, i.e. an automatic term extractor (Syllabs) and a concordancer (Sketch Engine). The main objective is to investigate whether the use of the two tools can influence the trainee interpreters' SI performance. In other words, does the use of the tools help the trainee interpreters perform better in simultaneous interpreting?

As explained in **Chapter 4 - Methodology**, four SI tasks (En-Zh & Zh-En) on two specialised topics (deep seabed minerals & fast reactors) were designed for the experiments. Control and test groups were invited to do the same SI tasks but following different preparation procedures with the use of different tools.

In order to obtain an objective view on the influence of using different terminology preparation procedures and tools on SI performances, the following dependant variables were measured, i.e. a) holistic SI performance scores, b) terminological accuracy scores, c) term error numbers and the degrees of departures for each individual error category, d) post-task recall of terms. The scoring criteria and error categories have been discussed in **Chapter 4 (Section 4.5.2 & 4.5.3)**.

Also as explained in **Chapter 4 (Section 4.5.1)**, a panel of three judges were involved in this study. I was the first marker of all the performances. The other two judges were the second markers: one focused on holistic SI performance scores, and the other examined terminological accuracy scores which reflect term error numbers and the degrees of departures of individual error. The second markers sampled 30%-50% of work already first marked to check whether the first marker has reasonable and consistent judgement and then had discussions with the first marker to resolve differences to produce agreed marks. This chapter will present the validated results after the second marking.

The chapter is organised as follows. **Section 6.1** presents the overall effect of preparation using automatic term extractor within limited preparation time (3 days) on the students' SI performances. **Section 6.2** reports the effect of preparation using both automatic term extractor and concordancer within ample preparation time (9 days) on the students' SI performances. **Section 6.3** mainly reports the students' feedback on the use of the tools during preparation. Each section is followed by a summary of the results reported. Further discussion of the results will be provided in **Chapter 7**.

6.1 Experiment I: The effect of only using automatically-generated termlists during preparation on the students' SI performance

In this study, an experiment was designed to test whether using the automatically-generated termlists (by Syllabs) during limited preparation time (3 days) could affect the students' SI performance. Both the control and test group were invited to simultaneously interpret for a specialised topic: deep seabed minerals (SM:en-zh & zh-en). Both groups were briefed on the topic and were provided with relevant documents for their preparation 3 days before they participated in the SI tasks (also see **Chapter 4, Section 4.3.3** for more information on the preparation documents). The students in the control group used "traditional" preparation procedure and their preparation was done without the use of either term extractor tool or concordancer (also see **Chapter 4, Section 4.6.1** for the definition of "traditional" preparation procedure). In the experiment groups, participants were also provided with two monolingual termlists (en & zh) which were automatically generated from the preparation documents (by the term extractor, Syllabs).

6.1.1 Term accuracy

In this study, each individual participant's performance is given a term accuracy score based on the scoring system for terminological performance in SI discussed in **Chapter 4, Section 4.5.3.**

Table 31 presents data on the students' performance in the SI tasks on SM (Participants: n=21). The first row shows the means and standard deviation values of term accuracy scores in the two groups (control & test). The test group (mean=51.6%) performed better than the control group (mean=46.9%) in term accuracy. However, T-test result shows that mean term accuracy score of the test group is not significantly higher than the control group ($P>0.05$), which indicates that only using automatically-generated termlists during limited preparation time does not have a significant effect on the students' term accuracy in SI.

	Control Mean (SD)	**Test** Mean (SD)	**P-value**
Term accuracy score	46.9% (11.6%)	51.6% (15.2%)	0.129 (P>0.05)
Holistic SI performance score	53.3% (9.4%)	57.8% (9.0%)	0.058 (P>0.05)

Table 31: The students' performance (Experiment I: SM)

6.1.2 Holistic SI performance

In this study, we use the Marking Criteria in the MA Interpreting Final Exams at the University of Leeds to judge participants' holistic SI performance (also see **Chapter 4, Section 4.5.2**).

Table 31 also presents the mean and standard deviation values for holistic SI performance scores in the two groups (control & test). The test group has higher mean scores (57.8%) than the control group (53.3%), but the difference between the two groups is not significant (P>0.05). The result indicates that only using term extractor during limited preparation time has some positive effect on the students' general SI performance, but the effect is not significant.

6.1.3 Significant error categories

Table 32 presents the mean and standard deviation values for the error percentage of each error category (i.e. OM, IT, SE, GE, IC, PE). The result shows that among all the error categories, **Omission (OM)** and **Incorrect Term (IT)** were the top two categories with the most error counts in both control and test groups in both SI tasks (en-zh & zh-en) on SM. The third most common error type was **Semantic Error (SE)**. Compared with the first three error categories, there were far fewer occurrences of **Grammatical Error (GE)** and **Incorrect Collocation (IC)** and **Pronunciation Error (PE) in all the tasks**.

Error category	SM_En-Zh		SM_Zh-En	
	Control Mean (SD)	**Test** Mean (SD)	**Control** Mean (SD)	**Test** Mean (SD)
OM (%)	27.10 (12.1)	23.51 (7.8)	39.28 (8.4)	39.86 (8.0)
IT (%)	15.62 (4.7)	14.36 (4.7)	18.99 (5.1)	17.13 (4.0)
SE (%)	6.15 (3.1)	4.03 (3.2)	3.04 (1.5)	2.44 (1.3)
GE (%)	0 (0)	0 (0)	0.80 (1.0)	2.04 (1.6)
IC (%)	0.24 (0.4)	0.05 (0.2)	0.65 (0.8)	1.05 (0.9)
PE (%)	0.18 (0.4)	0 (0)	1.45 (1.4)	1.12 (1.7)
Total terms	169		138	

Table 32: Error percentages (Experiment I: SM)

Excel's t-tests (Two-Sample Assuming Unequal Variances) were conducted to test the significance of differences between control and test groups on the top three error categories. T-test results show that there is no significant difference in the number of **OM**, **IT** and **SE** between the two groups **(P>0.05)**. **Figure 14** illustrates that the margins of differences between the two groups on the three error categories were all quite narrow. The differences ranged from 0.58% to 3.59%. The figure also shows that both groups made noticeably more **OM** when interpreting into English than into Chinese. In **Chapter 7 (Section 7.3.3)**, I will further discuss the possible reasons for more omission errors when interpreting into the B language.

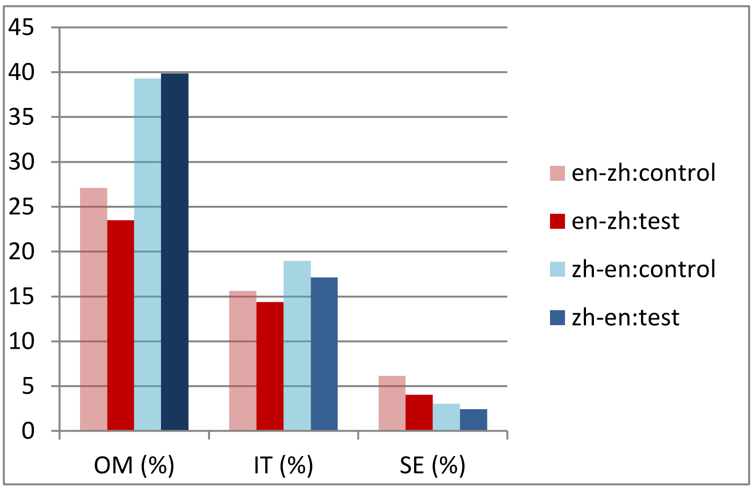


Figure 14: Mean error percentages of OM, IT & SE (Experiment I: SM)

6.1.4 Correlation between term accuracy and holistic SI performance

Excel's correlation coefficient tests were conducted to measure correlations between term accuracy and SI performance in the SI tasks into both language directions (en-zh & zh-en). **Table 33** presents the **correlation coefficient**, **coefficient of determination (R square)** and **standard error** in each group. Correlation coefficient and R square values were all over 0.80 (P value<0.05), which means there were significant positive correlations between term accuracy and SI performance in both the control and test groups and in both SI tasks (SM: en-zh & zh-en). In other words, term accuracy played an important role in holistic SI performance in this experiment.

	Correlation	R square	Standard error
Control: en-zh	0.900	0.809	4.537
Control: zh-en	0.923	0.853	3.683
Test: en-zh	0.971	0.942	2.633
Test: zh-en	0.909	0.826	2.973

Table 33: Correlations between term accuracy & SI performance (Experiment I: SM)

The scatter plots in **Figure 15** provide a visualisation of the correlations between term accuracy and holistic SI scores in the control group (n=20) and test group (n=22) respectively.

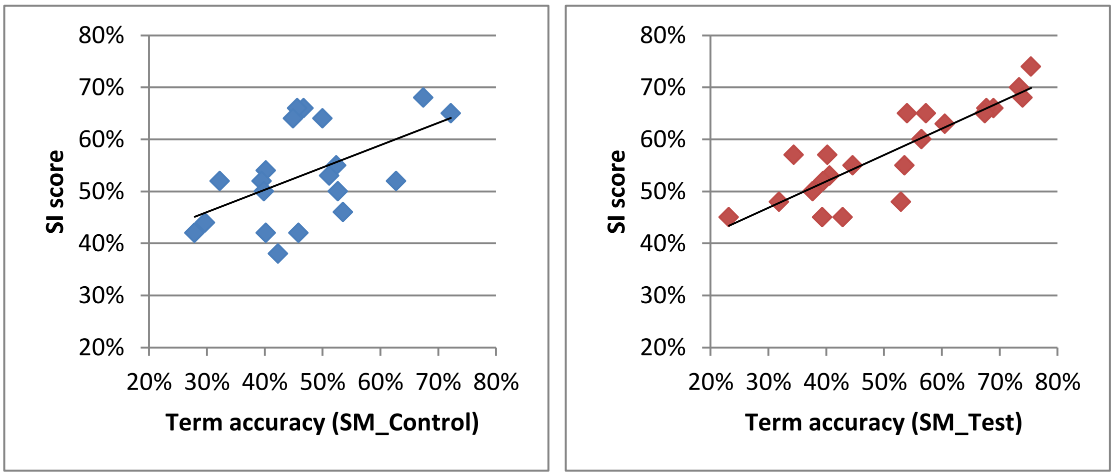


Figure 15: Correlations between term accuracy scores & holistic SI scores (Experiment I: SM)

When comparing the two groups, I find that on the one hand, the students who achieved over 60% in their holistic SI scores in the test group tended to have higher term accuracy scores than those in the control group. On the other hand, among the students who achieved lower than 60% in SI scores, there was not much difference in term accuracy scores between the two groups. To summarise, using the automatically-generated termlists during preparation with limited time (3 days) seemed to have some impact on term accuracy among some top students in the group (who achieved higher than 60% in SI scores). I will further discuss the students' feedback on using the automatic lists during preparation in **Section 6.3.2.2.**

6.1.5 Post-task recall of terms

Apart from examining the terminology performance in the SI tasks, I also investigate the impact of terminology preparation after the SI tasks, by examining the participants' recall of terms two months after the SI tasks. As explained in **Chapter 4 - Methodology (Section 4.5.4)**, two months after the SI tasks, the participants were asked to take a term quiz on the topic of SM. There are altogether 15 terms in English, all of which are specialised terms (category S terms) from the original preparation documents that had been given to the students before SI tasks. These terms were considered relevant by all the students (as all these terms were included in the students' termlists). In other words, these terms had been activated to a certain extent during their preparation stage. Many of the terms also appeared in the source speeches that the students interpreted from in their SI tasks.

In the term quiz two months after the SI tasks, the participants in both groups were asked to write down the Chinese translation of 15 English terms and provide simple definitions of the terms in Chinese. They were not allowed to refer to any other resources (preparation documents, termlists or online/paper dictionaries), as the term quiz is supposed to examine purely how many terms they have remembered two months after their preparation and the SI tasks.

	Control	**Test**
Mean (STDV)	**11.50 (2.27)**	9 (2.72)
P=0.02<0.05, Participants (n=21)		

Table 34: Term quiz results (Experiment I: SM)

Table 34 presents the mean and standard deviation values of the term quiz results of the two groups. Excel’s t-test (Two-Sample Assuming Unequal Variances) was conducted to measure the significance of difference between the two groups. T-test results show that the test group has significantly lower recall of the terms than the control group ($P<0.05$). In other words, the test group could remember fewer terms than the control group two months after the SI tasks. The control group has greater success of recollection of the terms two months after.

6.1.6 Summary

The correlation between term accuracy and SI performance was strong in both SI tasks (SM: en-zh & zh-en) in both groups (control and test). Term accuracy played an important role in holistic SI performance in this experiment.

In this experiment (Experiment I: SM), the test group (using automatically-generated termlists with limited preparation time) performed slightly better than the control group (without using any automatically-generated termlists) in both term accuracy and holistic SI scores, but there is no significant difference between the two groups. There is no significant difference on the number of OM, IT, and SE between the two groups either.

In the term quiz two months after the SI tasks, the test group could remember significantly fewer terms than the control group.

6.2 Experiment II: The effect of using both automatically-generated termlists and the concordancer during preparation on the students’ SI performance

In this study, another experiment was designed to test whether using both automatically-generated termlists and the concordancer with ample preparation time (9 days) could affect the students’ SI performances. Both the control and test group were invited to simultaneously interpret for two specialised speeches on fast reactors (en-zh & zh-en). Both groups were briefed on the topic and were provided with relevant documents for their preparation (see **Chapter 4**) 9 days before they participated in the SI tasks. The students in the control group used “traditional” preparation procedure, and their preparation was done without the use of either term extractor or concordancer. In the experiment groups, the students were also provided with two monolingual termlists automatically generated by Syllabs. Additionally, the students also used the concordancer, Sketch Engine to assist their preparation.

6.2.1 Term accuracy

Table 35 presents data on students' performance in the SI tasks on FR (Participants: n=22). The first row shows the means and standard deviation values of term accuracy scores in the two experiment conditions (control & test). Excel's t-tests (Two-Sample Assuming Unequal Variances) were conducted to measure the significance of difference between the two groups. T-test result shows that mean term accuracy score of the test group (58.8%) is statistically significantly higher than the control group (51.3%) (P<0.05).

The test group also had a smaller standard deviation (SD) in term accuracy scores. As standard deviation (SD) measures the amount of variation or dispersion from the average, a lower standard deviation (Test group SD=9.1%) indicates that term accuracy scores in test group tend to be clustered closely around the mean; a higher standard deviation (Control group SD= 14.1%) indicates term accuracy scores in the control group are more spread out over a large range of values.

In other words, the students in the test group generally had higher term accuracy scores, and they performed more similarly. It indicates that using both a term extractor and a concordancer during preparation helped the students achieve higher term accuracy during SI.

	Control Mean (SD)	**Test Mean (SD)**	**P-value**
Term accuracy score	51.3% (14.1%)	**58.8% (9.1%)**	**0.020(P<0.05)**
Holistic SI performance score	55.5% (11.2%)	58.3% (10.1%)	0.199 (P>0.05)

Table 35: The students' performance (Experiment II: FR)

6.2.2 Holistic SI performance

Table 35 also presents the mean and standard deviation values for holistic SI performance scores in the two groups (control & test). The test group has higher mean holistic SI scores (58.3%) than the control group (55.5%), but the difference between the two groups is not significant (P>0.05). The result indicates that the use of both term extractor and concordancer during preparation has some positive effect on the students' general SI performance, but the effect is not significant.

6.2.3 Significant error categories

Table 36 presents the mean and standard deviation values for the error percentage[15] of each error category (i.e. OM, IT, SE, GE, IC, PE). The result shows that among all the error categories, **Omission (OM)** and **Incorrect Term (IT)** were the top two categories with the most error counts in both control and test groups in both SI tasks (en-zh & zh-en) on FR. The third most common error type was **Semantic Error (SE)**. Compared with the first three error categories, there were far fewer occurrences of **Grammatical Error (GE)** and **Incorrect Collocation (IC)** and **Pronunciation Error (PE)** in all the tasks. And among the few cases of GE, IC and PE, more happened in simultaneous interpreting into the B language (in this case, from Chinese into English).

Error category	FR_En-Zh		FR_Zh-En	
	Control Mean (SD)	**Test** Mean (SD)	**Control** Mean (SD)	**Test** Mean (SD)
OM (%)	29.26 (9.7)	**19.77 (9.8)**	32.82 (14.1)	**23.79 (8.1)**
IT (%)	13.86 (4.9)	16.05 (3.0)	16.15 (4.7)	17.03 (4.0)
SE (%)	6.40 (2.7)	6.98 (2.5)	2.53 (2.1)	3.79 (1.7)
GE (%)	0 (0)	0 (0)	2.30 (1.6)	1.86 (1.1)
IC (%)	0.10 (0.3)	0.47 (0.8)	0.86 (1.0)	0.69 (0.7)
PE (%)	0.58 (1.2)	0 (0)	0.57 (0.9)	1.03 (1.2)
Total terms	86		145	

Table 36: Error percentages (Experiment II: FR)

Excel's t-tests (Two-Sample Assuming Unequal Variances) were conducted to test the significance of differences between control and test groups on the top three error categories. T-test results show that the test group made significantly fewer **OM** than the control group in both SI tasks (en-zh: P=0.018<0.05) (zh-en:

15 Error percentage of each error category: 'error counts in each error category' is divided by 'the total number of terms in the specific SI task'. Error percentage is used to compare across different SI tasks.

P=0.038<0.05). On the other hand, there is no significant difference in the number of **IT** and **SE** between the two groups (P>0.05).

When I further examine the data on **OM** in **Table 36**, I find that apart from the fact that the test group made significantly fewer omission errors in both SI tasks; the test group also had a particularly smaller standard deviation (SD) in the SI task into English (FR_Zh-En). A lower standard deviation (Test group SD=8.1) indicates that the numbers of **OM** made by the test group tend to be clustered closely around the mean; a higher standard deviation (Control group SD= 14.1) indicates that the numbers of **OM** made by the control group are spread out over a large range of values. In other words, the students in the test group generally made fewer **OM**, and they performed more similarly (in FR_Zh-En). This means that using both automatically-generated termlists and the concordancer during preparation were effective in reducing the number of OM, especially in simultaneous interpreting into the B language (English).

Figure 16 illustrates the differences between the two groups on the top three error categories. There are clearly fewer omission errors made by the test group who used both automatically-generated termlists and the concordancer to assist their preparation. On the other hand, there was no significant difference in the number of **IT** and **SE** made by the two groups.

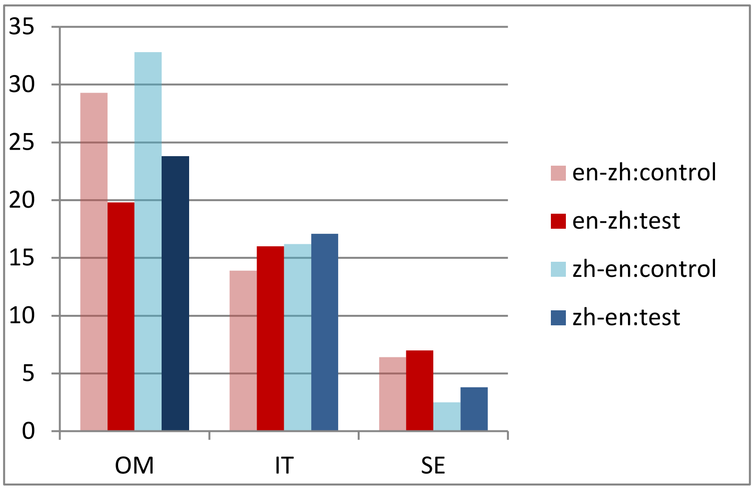


Figure 16: Mean error percentages of OM, IT & SE (Experiment II: FR)

6.2.4 Correlation between term accuracy and holistic SI performance

Excel's correlation coefficient tests were conducted to measure correlations between term accuracy scores and holistic SI scores in each group. **Table 37** shows that the correlations were strong in both groups (control & test) and in both SI tasks (FR:en-zh & zh-en) ($P<0.05$). In other words, term accuracy played an important role in holistic SI performance in this experiment, which is consistent with the experiment on SM.

	Correlation	R square	Standard error
Control: en-zh	0.933	0.871	3.642
Control: zh-en	0.969	0.939	3.340
Test: en-zh	0.829	0.688	5.414
Test: zh-en	0.929	0.864	4.495

Table 37: Correlations between Term accuracy & SI performance (Experiment II: FR)

The scatter plots in **Figure 17** provides a visualisation of the correlations between term accuracy scores and holistic SI scores in the control group (n=24) and test groups (n=20) respectively.

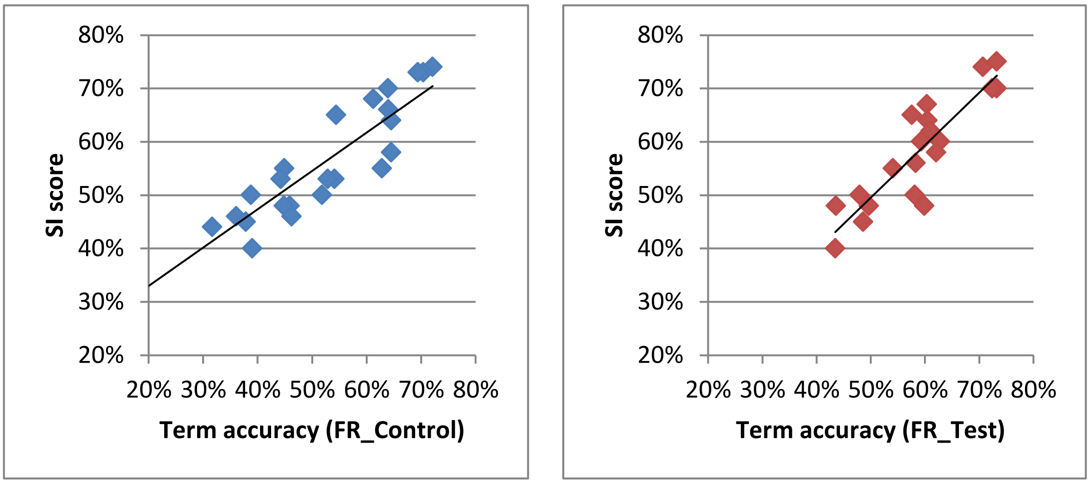


Figure 17: Correlations between term accuracy scores & holistic SI scores (Experiment II: FR)

When comparing the above two groups in **Figure 17**, it is evident that the test group had higher term accuracy scores than the control group. Moreover, the students who achieved below 60% in their holistic SI scores in the test group seemed to have higher term accuracy scores than the control group. This indicates that using both tools during preparation with ample time (9 days) generally helped the students improve their term accuracy, and the impact was more obvious on the average and poor students (who achieved lower than 60% in SI scores). I will further discuss the students' feedback on using both automatically-generated lists and the concordance tool during preparation in **Section 6.3.2.3**.

6.2.5 Terms that the students commonly made serious errors during SI

I took a closer look at the terms that half of the students (5 students in each group) commonly made serious errors (OM-0, IT-0 & SE-0) in during the two SI tasks (FR: En-Zh & Zh-En). **Table 38** summarises the numbers of such terms by three different categories (S, G and NE).

As discussed in **Chapter 4 (Section 4.3.2.1)**, terms are categorised into specialised terms (S), general terms (G) and named entities (NE) in this study. **Category S** contains highly specialised terms relevant to the topic *(e.g. uranium-238, decay heat removal system)*. **Category G** contains non-specialised terms commonly used in the field *(e.g. performance tests, full power operation, nuclear accident)*. **Category NE** contains named entities, including organization and location and other domain-specific proper names *(e.g. International Project on Innovative Nuclear Reactors and Fuel Cycles (INPRO), China Institute of Atomic Energy, China experimental Fast Reactor)*.

	Total	S	G	NE
Control	89	51	29	9
Test	49	24	21	4

Table 38: Numbers and categories of terms that the students commonly made serious errors (Experiment II: FR)

Table 38 shows that among the terms that the students commonly made serious errors in during the SI tasks (FR), the test group made noticeably fewer serious errors in **specialised terms (S)** (control=51; test=24) and **named entities (NE)** (control=9; test=4), and moderately fewer errors in **general terms (G)** (control=29; test=21). It is evident that the control group after preparation still struggled a lot with specialised terms and named entities during the SI tasks. **Specialised terms**

and named entities (S & NE) posed a greater challenge to the control group during SI compared with the test group. In **Chapter 7 (Section 7.3.1)**, I will further discuss about these challenging terms, their distribution in the original speeches, the possible reasons for errors, as well as common and different coping strategies used by the two groups in interpreting these challenging terms. In **Chapter 7 (Section 7.3.2 & 7.3.3)**, I will further investigate how an increased level of term density in the source speech and working into the B language may affect the trainee interpreters' terminological performance in simultaneous interpreting.

6.2.6 Post-task recall of terms

Two months after the SI tasks, the participants were asked to take a term quiz on the topic of FR. Both groups were asked to write down the Chinese translation of 15 English terms and provide simple definitions of the terms in Chinese. The 15 terms in English are specialised terms from the original preparation documents that were provided to the students before the SI tasks. The term quiz is supposed to examine purely how many terms the participants have remembered two months after their preparation and the SI tasks.

	Control	Test
Mean (SD)	7.27 (3.52)	**10.00 (1.77)**
P= 0.02<0.05, Participants (n=20)		

Table 39: Term quiz results (Experiment II: FR)

Table 39 presents the mean and standard deviation values of the term quiz results of the two groups. Excel's t-test (Two-Sample Assuming Unequal Variances) was conducted to measure the significance of the difference between the two groups. The t-test result shows that the test group had significantly higher recall of the terms than the control group ($P<0.05$). In other words, the test group had greater success recalling terms two months after their preparation. The use of the tools helped them achieve superior learning result of relevant terms than the control group.

6.2.7 Summary

The correlation between term accuracy and holistic SI performance were strong in both SI tasks (FR: en-zh & zh-en) in both control and test groups. Term accuracy played an important role in holistic SI performance in this experiment, which is consistent with the experiment on SM.

It is also found when the students were given ample preparation time (9 days), using both automatically-generated termlists and the concordancer helped them better prepared for the SI tasks. The students, especially the average and poor students (who achieved lower than 60% in holistic SI scores) performed significantly better in term accuracy in the SI tasks than the control group using traditional preparation procedure (without using any tools). The test group also generally made significantly fewer Omission Errors (OM) in SI tasks. Through observing the terms that the students commonly made serious errors in, it is found that specialised terms (S) and named entities (NE) posed less terminological challenge to the test group during SI. It is also found that using both tools during the preparation helped the test group achieve significantly higher recall of terms in the post-task term quiz two months after the SI tasks.

It is also noted that the preparation with the use of the two tools only helped to slightly improve holistic SI scores in the test group (but not yielding any statistical significance). The use of the two tools did not help in reducing the numbers of Incorrect Terms (IT) and Semantic Errors (SE) in SI tasks either.

6.3 The students' feedback on the use of the tools during preparation

As explained in **Chapter 4 (Section 4.6.5)**, Focus groups ("group interviews") were used to complement other forms of data collection to investigate the trainee interpreters' views on the use of the tools and different preparation procedures.

Each focus group had 4 to 6 participants apart from the researcher, and each group interview lasted for about 1 hour. The interview was in Q&A fashion. Questions include both prompt questions and open-ended questions. The whole process was audio-recorded, then transcribed and summarised.

This section mainly reports findings from focus groups on the following issues: **6.3.1** real preparation time, **6.3.2** the participants' views on the tools and different preparation procedures, and **6.3.3** the features of the termlists used during simultaneous interpreting.

6.3.1 Real preparation time

Section 6.1 and **Section 6.2** discussed the effect of using different preparation procedures and tools on SI performances and post-task recall of terms. In this section, I will compare the real time that the participants spent on preparation in order to evaluate whether the use of the tools helped save preparation time.

As explained in **Chapter 4 (Section 4.6.1)**, this study examines three preparation procedures under two conditions:

- Traditional preparation procedure without using any tool; the preparation begins both 3 days and 9 days before SI.
- Preparation with the automatically-generated termlists, the preparation begins 3 days before SI.
- Preparation with both automatically-generated termlists and the concordancer, the preparation begins 9 days before SI.

6.3.1.1 Condition 1: the preparation begins 3 days before SI

In Experiment I: SM, the students in both groups (control & test) were assigned with the preparation tasks 3 days before the SI tasks, including preparing and updating bilingual termlists, understanding relevant background information about the topic and further activation of the terms and concepts. The students in both groups were asked to record the real time they spend on each preparation activity. In the focus groups, they reported their individual preparation time accordingly.

In the control group, the students followed the interpreters' workflow suggested by Moser-Mercer (1992), which we define as "traditional preparation procedure" in this study. The preparation procedure requires reading the texts and manually extracting the terms, followed by producing bilingual terminology lists and studying the list for the interpreting tasks.

In the test group, the students followed a different preparation procedure. Instead of reading the texts and manually extracting the terms, two monolingual termlists were automatically extracted and then revised by the student interpreters to produce their own bilingual lists of relevant terms. The students then studied the terms in the original preparation documents **but without using concordancer**.

Table 40 presents data on the students' real preparation time for the SI tasks on both topics (SM & FR). The first row summarises the mean and standard deviation values of the real preparation time by the control and test groups for the topic of SM.

The result shows no significant difference on the preparation time between the two groups (P>0.05). The test group and the control group spent more or less the same preparation time (almost 10 hours).

	Participants	Control	Test	P-value
SM (hours) Mean (SD)	21	9.60 (2.08)	9.83 (1.25)	0.768>0.05
FR (hours) Mean (SD)	22	18.77 (1.82)	**15.60 (3.03)**	**0.005<0.05**

Table 40: Real preparation time (Two experiments: SM & FR)

6.3.1.2 Condition 2: the preparation begins 9 days before SI

In the other experiment (Experiment II: FR), the students in both groups (control & test) were assigned with the preparation tasks 9 days before the SI tasks. The control group in this experiment followed "traditional preparation procedure". In the test group, the students followed a different preparation procedure. Two monolingual termlists were extracted automatically and then revised by the student interpreters to produce bilingual lists of relevant terms. In addition, a concordancer was used to link the terms to their original contexts in the preparation documents and assist the students to further activate the terms for interpreting tasks.

Table 40 also provides the mean and standard deviation values of the real preparation time by the control and test groups for the topic of **FR**. T-test result shows that the test group (Mean=18.77 hours) spent significantly less preparation time (about 3 hours) than the control group (Mean=15.60 hours) (P<0.05).

In summary, when limited preparation time (3 days) was given, the use of automatically-generated termlists alone did not help save preparation time; when ample preparation time was given (9 days), the use of both automatically-generated termlists and the concordancer helped save preparation time.

6.3.2 The participants' views on the tools and different preparation procedures

In this section I will further look into the participants' opinions on the following issues covered in the focus groups: **6.3.2.1** the traditional preparation procedure, **6.3.2.2** the preparation with the use of automatically-generated lists, and **6.3.2.3 t**he preparation with the use of both automatically-generated lists and the concordancer.

6.3.2.1 The traditional preparation procedure

The students in the two control groups in both experiments (FR & SM) used "traditional preparation procedure". One group was given a 9-day preparation time, the other group was given a 3-day preparation time. In the focus groups after the SI tasks, the students described how they implemented the preparation procedure in their own preparation and the particular challenges they found using the procedure.

They started the preparation by reading through the speech agenda and then the preparation documents provided to them. During this process, they learned about background information on the topic, manually selected relevant terms from the texts, checked term equivalences and produced their bilingual lists of relevant terms.

The students further familiarised themselves with the relevant terms through the following activities: 1) doing mini-speeches in the target language by using the key terms and the concepts learnt, 2) practising sight translating some paragraphs in the texts that they considered relevant to the speeches that they were going to interpret from (according to the speech agenda provided to them), 3) learning the terms and the equivalents in their own termlists by heart.

Most of the participants found reading preparation documents to get familiar with the background information and making mini-speeches in the target language quite useful. They all expressed they were more familiar with the terms and the background information on the topic than before any preparation.

On the other hand, they felt some terms had not been activated enough for the SI tasks that they participated in after the preparation. They said most of terms in the speeches that they interpreted from had actually been included in their own termlists. However, during simultaneous interpreting, they still struggled and hesitated at some specific and technical terms. They could not speak them out quickly enough in the target language.

They did find it challenging to prepare for the specialised topics that they had little prior knowledge of (even for the control group who was given relatively ample preparation time, 9 days). They found the preparation task overwhelming. "There were too many articles to read, too many new terms to remember and too many new concepts to digest, and the preparation time is so limited."

Many of them felt they spent too much time on reading the texts. They did manage to finish reading the majority of the preparation documents, but there was not enough time left to further review and activate the terms that they encountered from reading the documents. One student said: “I started reading the big files and it just took like forever. It was like going fishing in a sea without much clue, and it was quite stressful.”

6.3.2.2 Preparation with the use of automatically-generated termlists

The students in the test group (Experiment I), after interpreting for the topic of deep seabed minerals (SM), participated in the focus group, in which they described how they implemented the preparation procedure and shared their opinions on the two automatically-generated monolingual termlists they were using during their preparation.

The students were provided with two monolingual candidate termlists extracted automatically by **Syllabs**. They annotated the candidate termlists and then only concentrated on the terms which they considered relevant and important to remember. They then went through the preparation documents to learn the relevant concepts, check equivalences of the terms and formed their own bilingual termlists.

Same as the two control groups (mentioned above in **Section 6.3.2.1**), the students in the test group (Experiment I: SM) spent some time further activating the terms through doing mini-speeches in the target language by using the key terms and the concepts learnt. Some students also practised sight translating some paragraphs in the preparation documents. They also spent some time learning the terms and the equivalents in their own bilingual termlists by heart.

According to the students, the distinctive feature of this preparation procedure is that they started their preparation with annotating the automatic candidate termlists, from which they formed a basic idea of the subject matter and the terms needed for the interpreting task. They said that when using the traditional preparation procedure in the previous experiment, the primary goal of reading preparation documents was to collect relevant terms. Since two candidate termlists had been provided this time, the priority in their preparation became finding how these terms are used, their associations with each other and the background information behind.

When asked their opinions on the use of the automatically-generated lists in their preparation, the answers differed among the students. About half of them expressed positive attitude on the use of the automatic lists during preparation. They said they used the lists as an important indicator for the content of the preparation documents. The lists helped them prioritise their preparation on the most relevant terms and concepts.

The other half did not quite like the idea of using the automatically-generated termlists during their preparation. They expressed that "shifting the relevant terms from the candidate termlists" took time and was not quite effective. They said when annotating the candidate termlists at the very start of their preparation, they could not quite make sense of the terms without knowing much about the topic, and they were not sure whether the terms in the lists were relevant or not. They would rather use the time reading through the texts to make sense of the topic, and they preferred extracting their own termlists while reading through the texts. In fact, some of them just used the annotated termlists for "superficial understanding of the topic" at the start of the preparation. They then put the annotated lists aside and produced their own lists through reading afterwards.

Six participants (out of eleven) said using automatically-generated termlists saved their preparation time, the other five found using the lists didn't necessarily save their preparation time.

When asked whether they would consider using the automatically-generated termlists in their future preparation, most people said they prefer having the automatically-generated lists to start with the preparation task. And it would be better if the lists are shorter.

6.3.2.3 Preparation with both automatically-extracted termlists and the concordancer

The students in the test group (Experiment II), after interpreting for the topic of fast reactors (FR), participated in the focus group. They described how they implemented the preparation procedure and shared their opinions particularly on using the concordancer in preparation.

Same as in the test groups (Experiment I: SM) mentioned in **Section 6.3.2.2**, the students in the test group (Experiment II: FR) started their preparation with annotating candidate termlists. They then prioritised their reading around the key terms in the preparation documents with the aid of the concordancer (Sketch Engine). They learned the key terms in contexts, check equivalences of the terms and formed their own bilingual termlists.

Same as all the other groups (mentioned in **Section 6.3.2.1 and 6.3.2.2**), the students in this group also spent some time on the following activation activities: 1) doing mini-speeches in the target language by using the key terms and the concepts learnt, 2) practising sight translating some paragraphs in the preparation documents and 3) learning the terms and the equivalents in their own termlists by heart.

The use of the automatically-generated termlists together with the concordancer breaks the traditional linear reading process, and makes navigation possible. The students said they could find the terms in the preparation documents more quickly by using the concordancer. Previously when they used the traditional preparation procedure, they had to read a whole text to understand specific terms and they might easily be distracted during reading. This time checking and searching for terms became more focused and efficient.

The students told me they used the concordancer mainly to check contexts and collocations of the unfamiliar terms (e.g. abbreviations). They also used the concordancer to check other relevant terms that might be closely linked to the terms they had already known. For example, by searching for some general terms in the preparation documents on the Sketch Engine interface, they would be able to view more relevant technical terms appearing in the same contexts. And it helped them make sense of the relationship between the terms. They also checked the contexts of the same terms in the other language.

The students found the tool quite flexible to use. They could choose to see a certain number of or all the contexts that a specific term is used in, and they could easily compare different uses of one or two similar terms. They could also locate which text/file a specific term is frequently used in. They said the tool also provided a basis for doing other verbalised activation exercises, for example, mini-speeches and sight translation.

The students in this test group (Experiment II: FR) also told us they generally used the automatically-generated termlists as reference lists. Their own bilingual lists consisted of terms that were mainly constructed by themselves through reading and checking the texts. They revisited the annotated lists from time to time, and added more terms to their own lists.

They generally felt the terms that they prepared had been activated enough for the SI tasks. There were only a few terms that they didn't quite take notice of during preparation, therefore they were not able to interpret them correctly during the SI tasks. Most of the students felt the use of the automatically-generated termlists and the concordancer saved their preparation time, and the time was used much more efficiently.

6.3.3 The features of the termlists used during SI

In the focus groups, the students told us how they personalised their shortlists for the interpreting tasks and how they used the lists in the booth. The following are some distinctive features of their short termlists.

6.3.3.1 Length & fonts

Most of the students brought into the interpreting booth short lists containing about 10-30 terms on 1-2 pages. They included the important terms or challenging terms they were still less familiar with or might take too much time to recall during interpreting, for example, long terms (e.g. names of organisations) and terms that are difficult to pronounce (e.g. names of seabed minerals)

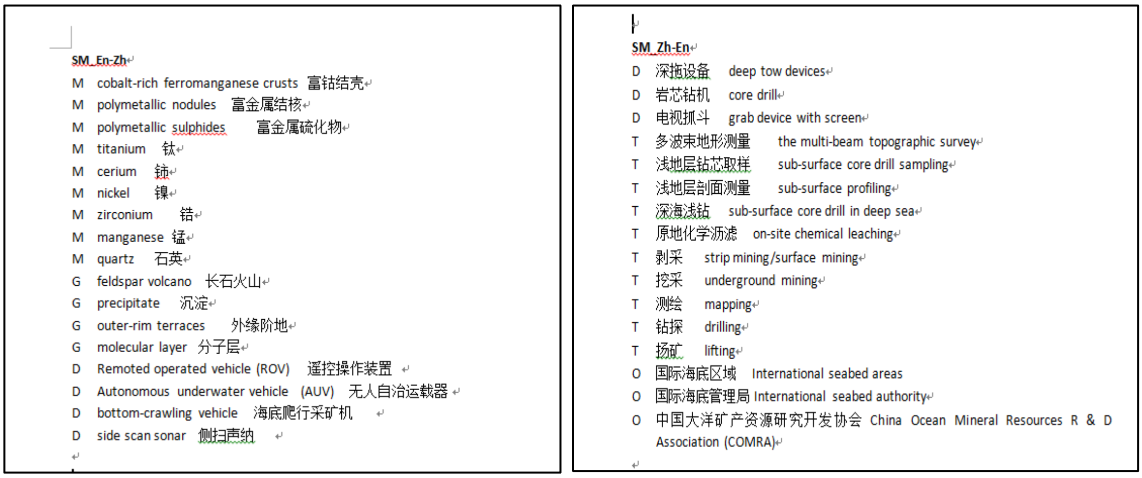
They use large characters, with certain items marked in bold texts or highlight colour.

6.3.3.2 Organisation of the short lists

Some of the students sequenced the terms according to the agenda. They had two parts. The first part was for SI task (English-Chinese) and second was for the other task (Chinese-English). They arranged both parts on a single page or listed terms on one page for each speech.

Some of them categorised the terms in the lists by using their own categorisation systems. Some of them only included organisation names and several less secure terms on one page. Some put most relevant and most-frequently used terms on the top of the list so they could find them quite easily.

Figure 18 is sample shortlist on seabed minerals, containing 33 terms organised in 2 pages following a categorisation system (M-metal, G-geographic term, D-exploration and mining device, T-exploration and mining technology, O-organization).



SM_En-Zh

M cobalt-rich ferromanganese crusts 富钴结壳
M polymetallic nodules 富金属结核
M polymetallic sulphides 富金属硫化物
M titanium 钛
M cerium 铈
M nickel 镍
M zirconium 锆
M manganese 锰
M quartz 石英
G feldspar volcano 长石火山
G precipitate 沉淀
G outer-rim terraces 外缘阶地
G molecular layer 分子层
D Remoted operated vehicle (ROV) 遥控操作装置
D Autonomous underwater vehicle (AUV) 无人自治运载器
D bottom-crawling vehicle 海底爬行采矿机
D side scan sonar 侧扫声纳

SM_Zh-En

D 深拖设备 deep tow devices
D 岩芯钻机 core drill
D 电视抓斗 grab device with screen
T 多波束地形测量 the multi-beam topographic survey
T 浅地层钻芯取样 sub-surface core drill sampling
T 浅地层剖面测量 sub-surface profiling
T 深海浅钻 sub-surface core drill in deep sea
T 原地化学沥滤 on-site chemical leaching
T 剥采 strip mining/surface mining
T 挖采 underground mining
T 测绘 mapping
T 钻探 drilling
T 扬矿 lifting
O 国际海底区域 International seabed areas
O 国际海底管理局 International seabed authority
O 中国大洋矿产资源研究开发协会 China Ocean Mineral Resources R & D Association (COMRA)

Figure 18: Screenshot of a student's shortlist (SM)

6.3.3.3 Display media

About 60% of the students said they took printed short lists into the booth. About 20% used iPads or iPhones to display their lists in the booth. Another 20% prepared shortlists during preparation, but decided not to use them in the booth. They said they were quite confident about all the terms they prepared, and they did not want to cause any distraction from using the lists while interpreting.

6.3.3.4 Using short lists in the booth

Nearly one third of the students said they referred to their lists quite often while interpreting. More than one third said they referred to the lists only occasionally. Less than one third told me they rarely referred to the lists (including those who intentionally excluded the lists outside the booth)

Most students found their shortlist very useful and they had no problem locating the terms in their lists during interpreting. A few students reported they had too many terms on the lists and they could not find the needed term efficiently.

I also find that most of the students who intentionally left the lists outside the booth had higher term accuracy and SI holistic scores. This means that on the one hand, they were more confident about their own preparation; on the other hand, those students who have better internalised the terms could actually perform better.

6.3.4. Summary

In summary, the data from the focus groups shows that when ample preparation time was given, the use of both automatically-generated termlists and the concordancer helped save about 17% (3 hours) of the students' average preparation time. When limited preparation time is given, the use of automatically-generated termlists alone does not make significant difference in preparation time.

All the groups (both control & test in both experiments) expressed that they were more familiar with the terms and the background information on the specialised topic than before any preparation.

In the two control groups, by following the "traditional preparation procedure", the students felt some terms had not been activated enough for the SI tasks after the preparation. During simultaneous interpreting, they still struggled and hesitated at some terms. They felt they spent too much time on reading the texts during the preparation, but there was not enough time left to further review and activate the terms that they encountered from reading the preparation documents.

In the test group (Experiment I: SM) who only used automatically-generated termlists, half of the students expressed the positive attitude towards the use of the lists during preparation as the lists helped prioritise the preparation of the most relevant terms and concepts. The other half of the students did not like the idea of using the automatically-generated lists during their preparation, as annotating the candidate termlists took time and did not work effectively for them. They preferred manually extracting terms while reading through the texts.

The students in the other test group (Experiment II: FR) reported the use of the automatically-generated termlists together with the concordancer worked well for them. Checking and searching terms was more focused and efficient. The students in this group generally expressed that the terms had been activated enough for the SI tasks. Most of them felt the use of the tools saved their preparation time, and their preparation time was used much more effectively.

In addition, the students in all the groups customised their own short termlists to assist simultaneous interpreting. They found their short lists useful and most of the students had no problem locating the terms in their lists during interpreting.

Chapter 7 - Discussion of results

This chapter aims to synthesise the key issues and the results obtained in the study by placing the findings in the context of theoretical and empirical frameworks reviewed in **Chapter 2 & 3**. The discussion will be presented in the following sections. **Section 7.1** discusses the students' termlists: what auto-lists are like, what has been done with auto-lists and how shortlists could be used during interpreting. **Section 7.2** focuses on activation of terms by different preparation procedures. **Section 7.3** elaborates on the impact of challenges in rendering the source speeches (e.g. specialised terms, high density of information and working into the B language) on the students' interpreting performance. **Section 7.4** presents the pedagogical implication of the findings.

7.1 Termlists

7.1.1 What are auto-lists like?

The automatic term extractors (Syllabs/ Teaboat/ TTC) generate monolingual termlists (**Figure 19**) of 500 terms in English and Chinese respectively. I measured the precision rates based on the number of relevant terms against candidate terms in an automatically-generated termlist. The evaluation results from **Chapter 5 (Section 5.2.2.2 &5.2.2.3)** show that Syllabs achieved the highest precision rates for English (88%), and is more reliable in generating more relevant terms. All three tools perform less satisfactory in generating relevant terms in Chinese. The precision rates ranged from 24 to 31%. The Chinese lists contained fewer relevant terms and more ill-formed constructions and repetitions compared with the English lists. The highest precision rates for automatic term extraction (for English) in this study are similar to other studies using small and medium size corpora, e.g. Fantinuoli (2006) and Haque et al. (2014). It is clear that the automatically-generated termlists are still not perfect so far. This study tries to investigate how the auto-lists can be used at their best to assist interpreters' preparation.

B	C	D	E	F	G	H	I
EN	general_count	specific_count	general_frequency	specific_frequency	quotient_of_occurence	entity	form
use fuel	0	20	7.87E-08	0.000481486	6119.732707	NP	used fuel
hlw	0	19	7.87E-08	0.000458558	5828.316863	ABR	hlw
jnc	0	17	7.87E-08	0.000412702	5245.485177	ABR	jnc
fbrs	0	16	7.87E-08	0.000389774	4954.069334	ABR	fbrs
fissile	0	13	7.87E-08	0.00032099	4079.821804	ADJ	fissile
jnfl	0	12	7.87E-08	0.000298063	3788.405961	ABR	jnfl
pwrs	0	12	7.87E-08	0.000298063	3788.405961	ABR	pwrs
sfps	0	10	7.87E-08	0.000252207	3205.574275	ABR	sfps
containment pressure	0	10	7.87E-08	0.000252207	3205.574275	NP	containment pressure
bwrs	0	10	7.87E-08	0.000252207	3205.574275	ABR	bwrs
containment structure	0	9	7.87E-08	0.000229279	2914.158432	NP	containment structure
reactor core	0	9	7.87E-08	0.000229279	2914.158432	NP	reactor core
isolation condenser system	0	9	7.87E-08	0.000229279	2914.158432	NP	isolation condenser s
nps	0	9	7.87E-08	0.000229279	2914.158432	ABR	nps
actinides	0	8	7.87E-08	0.000206351	2622.742589	NOUN	actinides
twh	0	8	7.87E-08	0.000206351	2622.742589	ABR	twh
hydrogen production	0	8	7.87E-08	0.000206351	2622.742589	NP	hydrogen production
lwr	0	8	7.87E-08	0.000206351	2622.742589	ABR	lwr
design basis	0	8	7.87E-08	0.000206351	2622.742589	NP	design bases
npp	1	16	1.57E-07	0.000389774	2477.034667	ABR	npp
experimental reactor	0	7	7.87E-08	0.000183423	2331.326745	NP	experimental reactor

Figure 19: Screenshot of Syllabs' auto-list in English (FR)

7.1.2 What has been done with auto-lists?

We agree with Rütten (2003) that termlists being extracted automatically should be revised by their users, the interpreters, who can then concentrate on those terms that are relevant and important to remember. In this study, the automatically-generated termlists were downloaded for further annotation and edited by the student interpreters during their preparation for SI tasks (see **Figure 20**).

A	B	C	D	E	F	G	H	
Annotation	EN	general_count	specific_count	general_frequency	specific_frequency	quotient_of_occurence	entity	form
IL	use fuel	0	20	7.87E-08	0.000481486	6119.732707	NP	used fuel
P	hlw	0	19	7.87E-08	0.000458558	5828.316863	ABR	hlw
R	jnc	0	17	7.87E-08	0.000412702	5245.485177	ABR	jnc
R	fbrs	0	16	7.87E-08	0.000389774	4954.069334	ABR	fbrs
R	fissile	0	13	7.87E-08	0.00032099	4079.821804	ADJ	fissile
R	jnfl	0	12	7.87E-08	0.000298063	3788.405961	ABR	jnfl
R	pwrs	0	12	7.87E-08	0.000298063	3788.405961	ABR	pwrs
R	sfps	0	10	7.87E-08	0.000252207	3205.574275	ABR	sfps
R	containment pressure	0	10	7.87E-08	0.000252207	3205.574275	NP	containment p
R	bwrs	0	10	7.87E-08	0.000252207	3205.574275	ABR	bwrs
R	containment structure	0	9	7.87E-08	0.000229279	2914.158432	NP	containment s
R	reactor core	0	9	7.87E-08	0.000229279	2914.158432	NP	reactor core
R	isolation condenser system	0	9	7.87E-08	0.000229279	2914.158432	NP	isolation cond
R	nps	0	9	7.87E-08	0.000229279	2914.158432	ABR	nps
R	actinides	0	8	7.87E-08	0.000206351	2622.742589	NOUN	actinides
P	twh	0	8	7.87E-08	0.000206351	2622.742589	ABR	twh
R	hydrogen production	0	8	7.87E-08	0.000206351	2622.742589	NP	hydrogen pro
R	lwr	0	8	7.87E-08	0.000206351	2622.742589	ABR	lwr
R	design basis	0	8	7.87E-08	0.000206351	2622.742589	NP	design bases
R	npp	1	16	1.57E-07	0.000389774	2477.034667	ABR	npp
R	experimental reactor	0	7	7.87E-08	0.000183423	2331.326745	NP	experimental

Figure 20: Screenshot of an annotation of Syllabs' auto-list in English (FR)

The students used the automatic termlists as reference lists and eventually formed their own bilingual termlists comprising roughly 150-200 terms that they considered relevant to the speech agendas provided at the start of their preparation. The students checked the terms in the preparation documents (en & zh), got familiar with both relevant concepts and the usages of the terms, and then recorded the information in their bilingual termlists. **Figure 21** is a sample bilingual list formed by a student based on monolingual auto-lists (en & zh).

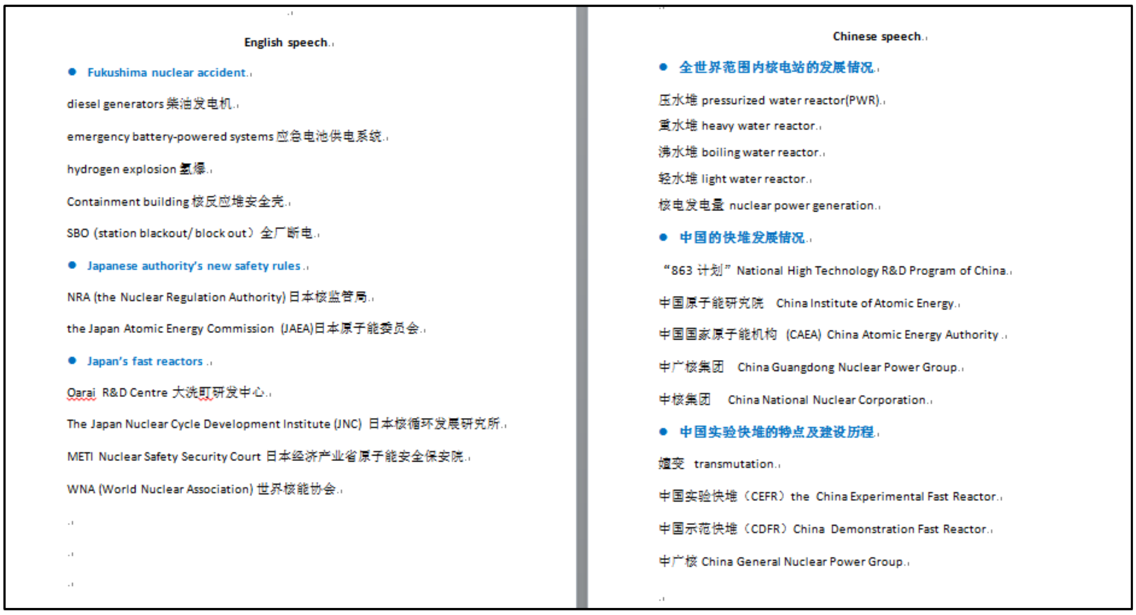
	B	C	D	E
1	**Term (English)**	**Term (Chinese)**	**Other relevant information**	**Other relevant information**
3	fissile	分裂性的		
4	primary containment vessel / PCV	第一层安全壳/反应堆外壳		
5	containment boundary	安全壳边界		
6	containment integrity	安全壳完整性		
7	breach of containment	安全壳破裂		
8	unit	机组		
9	reactor core	反应堆芯/堆芯		
10	lower head	底盖		
11	isolation condenser system	独立冷凝系统	和train（列）搭配	
12	scram	紧急停堆		
13	shutdown	冷停堆	bring to, reach, remained (shut down)	
14	suppression pool	抑压池	followed by temperature	
15	liquid metal fast breeder reactors / LMFBR	液态金属快堆		
16	Gas-Cooled Fast Reactor / GFR	气冷快堆		
17	helium	氦		
18	Sodium-Cooled Fast Reactor / SFR	钠冷快堆		
19	Lead-Cooled Fast Reactor / LFR	铅冷快堆		
20	breed ratio	增殖率		
21	fuel assembly	燃料组件		
22	hydrogen explosion	氢爆炸		
23	reactor pressure vessel / RPV	反应堆压力容器		
26	fuel cycle	燃料循环		
27	recombiner	复合器	autocatalytic/active hydrogen recombiners	自动催化/活性氢复合器
28	condenser	冷凝器		

English origin / Chinese origin / Units of measurement

Figure 21: Screenshot of a sample bilingual list (FR)

7.1.3 How could shortlists be used during interpreting?

The students then customised their own short termlists to assist simultaneous interpreting. The short lists (of about 10-30 terms) were concise versions of their long bilingual lists (150-200 terms). After preparation, the students could activate most of the terms in the long lists. In the short lists to be used in the booth, they only included the important terms or challenging terms that they were less familiar with or might take too much time to recall during interpreting. Customising shortlists is the last step of the students' terminology preparation and activation process. Most of them referred to their lists during interpreting from time to time, and they found their shortlists as useful reminders of relevant terms (also see the results from focus groups in **6.3.3**). **Figure 22** is a sample shortlist on fast reactors, containing 25 terms organised on 2 pages according to the speech agendas (FR: en & zh).



English speech

- **Fukushima nuclear accident**

diesel generators 柴油发电机

emergency battery-powered systems 应急电池供电系统

hydrogen explosion 氢爆

Containment building 核反应堆安全壳

SBO (station blackout/ block out）全厂断电

- **Japanese authority's new safety rules**

NRA (the Nuclear Regulation Authority) 日本核监管局

the Japan Atomic Energy Commission (JAEA)日本原子能委员会

- **Japan's fast reactors**

Oarai R&D Centre 大洗町研发中心

The Japan Nuclear Cycle Development Institute (JNC) 日本核循环发展研究所

METI Nuclear Safety Security Court 日本经济产业省原子能安全保安院

WNA (World Nuclear Association) 世界核能协会

Chinese speech

- **全世界范围内核电站的发展情况**

压水堆 pressurized water reactor(PWR)

重水堆 heavy water reactor

沸水堆 boiling water reactor

轻水堆 light water reactor

核电发电量 nuclear power generation

- **中国的快堆发展情况**

“863 计划” National High Technology R&D Program of China

中国原子能研究院 China Institute of Atomic Energy

中国国家原子能机构 (CAEA) China Atomic Energy Authority

中广核集团 China Guangdong Nuclear Power Group

中核集团 China National Nuclear Corporation

- **中国实验快堆的特点及建设历程**

嬗变 transmutation

中国实验快堆（CEFR）the China Experimental Fast Reactor

中国示范快堆（CDFR）China Demonstration Fast Reactor

中广核 China General Nuclear Power Group

Figure 22: Screenshot of a student's shortlist (FR)

7.2 Term activation by different preparation procedures

As reviewed in **Chapter 2**, knowing the terms does not necessarily mean the term is available for simultaneous interpreters' comprehension and production. Without sufficient preparation, the termlist itself could not guarantee good SI performance. Terminological preparation therefore needs to include not only collecting terms but also activating the terms to a certain level that ensures quick response and accuracy during simultaneous interpreting.

7.2.1 Implications of vocabulary acquisition for the interpreters' terminology preparation

In **Chapter 2 (Section 2.5)**, I have reviewed the nature of lexical development and how to promote lexical fluency in language learning. 1) Vocabulary acquisition develops when learning in context, in circumstances that make possible linking the new vocabulary to other terms and prior knowledge (Nagy & Herman, 1987; Sternberg, 1987). 2) Vocabulary acquisition could be consolidated by repetition. The more frequently words are used, the more activated the words become (Gile 1995 & 2009). 3) Shallow processing like rote learning or mechanical vocabulary memorisation does not lead to long-term retention of the vocabulary, whereas deep processing strategies such as semantic elaboration do achieve better vocabulary acquisition results (Bower & Winzenz, 1970; Craik & Lockhart, 1972; Hashtroudi, 1983). 4) Active stimulation (through speaking and writing) is more effective in increasing the availability of words than passive stimulation (through reading and hearing) (Gile 2009).

This study adopted the above approaches and conditions which promote lexical development to acquiring specialised terms for interpreting purpose. I will now present particular preparation activities following the above approaches and further discuss how I implemented the activities in the empirical study.

During the interpreters' terminology preparation, only checking definitions and equivalents of terms by using dictionaries is apparently not enough. Semantic processing should by all means be deepened. Relevant activities include 1) checking how the terms and their collocations are used in context, 2) making sense of the relationship between relevant terms and explicitly establishing ties to one another by grouping relevant terms together in the bilingual termlists (which I call "weaving the knowledge web").

The above activities are followed by all the three different preparation procedures being tested, i.e. 1) traditional preparation procedure, 2) preparation with automatically-generated termlists, and 3) preparation with using both automatically-generated lists and the concordancer. However, distinctive from the first two procedures, the third procedure implements the activation activities assisted by the concordancer (the Sketch Engine). It is used as "navigational tool" integrating reading the automatic termlists and the preparation documents (comparable corpora in both English and Chinese).

7.2.1.1 How could the concordancer be used?

The following figures **(Figure 23，24 and 25)** illustrate how Sketch Engine is used to assist learning terms on a specialised topic.

B30 | shutdown

Annotation	lemma	general_count	specific_count	general_frequency	specific_frequency	quotient_of_c	entity	form
R	containment pressure	0	10	7.87E-08	0.000252207	3205.574275	NP	containment pressu
R	bwrs	0	10	7.87E-08	0.000252207	3205.574275	ABR	bwrs
R	containment structure	0	9	7.87E-08	0.000229279	2914.158432	NP	containment structu
R	reactor core	0	9	7.87E-08	0.000229279	2914.158432	NP	reactor core
R	isolation condenser system	0	9	7.87E-08	0.000229279	2914.158432	NP	isolation condense:
R	nps	0	9	7.87E-08	0.000229279	2914.158432	ABR	nps
R	actinides	0	8	7.87E-08	0.000206351	2622.742589	NOUN	actinides
R	hydrogen production	0	8	7.87E-08	0.000206351	2622.742589	NP	hydrogen productio
R	lwr	0	8	7.87E-08	0.000206351	2622.742589	ABR	lwr
R	design basis	0	8	7.87E-08	0.000206351	2622.742589	NP	design bases
R	npp	1	16	1.57E-07	0.000389774	2477.034667	ABR	npp
R	experimental reactor	0	7	7.87E-08	0.000183423	2331.326745	NP	experimental react
R	shutdown	0	7	7.87E-08	0.000183423	2331.326745	NOUN	shutdown
R	suppression pool temperature	0	7	7.87E-08	0.000183423	2331.326745	NP	suppression pool t
R	breeder reactor	1	15	1.57E-07	0.000366846	2331.326745	NP	breeder reactor
R	fuel assembly	0	7	7.87E-08	0.000183423	2331.326745	NP	fuel assemblies
R	natural-phenomenon	0	6	7.87E-08	0.000160495	2039.910902	NOUN	natural-phenomenon
R	hydrogen explosion	0	6	7.87E-08	0.000160495	2039.910902	NP	hydrogen explosion
R	tsunamis	0	6	7.87E-08	0.000160495	2039.910902	NOUN	tsunamis
R	lwrs	0	6	7.87E-08	0.000160495	2039.910902	ABR	lwrs
R	reactor pressure vessel	0	6	7.87E-08	0.000160495	2039.910902	NP	reactor pressure v
R	fuel cycle	0	6	7.87E-08	0.000160495	2039.910902	NP	fuel cycle
R	condenser	1	12	1.57E-07	0.000298063	1894.202981	NOUN	condenser
R	fission product	0	5	7.87E-08	0.000137567	1748.495059	NP	fission products
R	direct disposal	0	5	7.87E-08	0.000137567	1748.495059	NP	direct disposal
R	beyond-design-basis	0	5	7.87E-08	0.000137567	1748.495059	NOUN	beyond-design-basi:

Annotation description | Syllabs list_en | Sheet3

Figure 23: Relevant terms in the English auto-list (FR)

The students could start with the terms from the English list (**Figure 23**), checking them on the Sketch Engine interface (see **Figure 24**). For example, "shutdown" is a relevant term. When typed into the Sketch Engine, all the instances of it in the preparation documents can be shown together. And when clicking on each concordance line, the term would be shown in a fuller context.

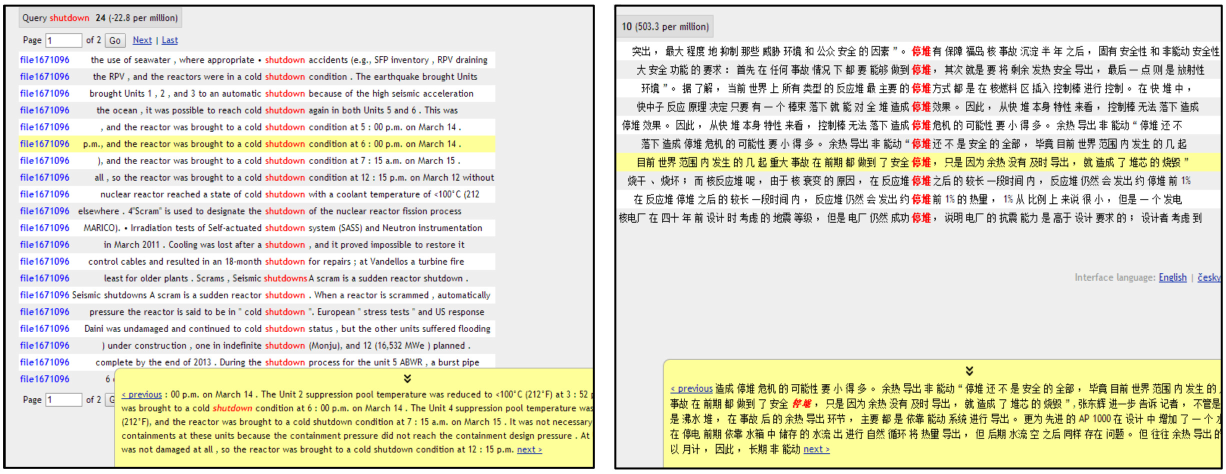


Figure 24: Concordance lines of 'shutdown' in both the English and Chinese corpora (FR)

Through reading the concordances, interpreters could easily find strong patterns and collocations of a term in both English and Chinese. For instance, collocation patterns of "shutdown" shown in the corpora include cold~, automatic~, seismic~, be brought to~, reach~, 冷停堆, 自动关停, 达到~, 安全~, etc.

Interpreters could also learn other relevant terms and concepts appearing in the same contexts of "shutdown" (e.g. decay heat removal system, fission reaction, Monju fast reactor, etc.). When necessary, they could check these terms in more contexts before they move to other terms in the monolingual lists. Finally, they could record what they have learned about the terms in their bilingual list (see **Figure 25**). This "recycling" learning process enables relevant terms being accessed from different routes and for several times, thus the terms could be learned thoroughly.

B13 shutdown

	B	C	D	E
1	**Term (English)**	**Term (Chinese)**	**Other relevant information**	**Other relevant information**
3	fissile	分裂性的		
4	primary containment vessel / PCV	第一层安全壳/反应堆外壳		
5	containment boundary	安全壳边界		
6	containment integrity	安全壳完整性		
7	breach of containment	安全壳破裂		
8	unit	机组		
9	reactor core	反应堆芯/堆芯		
10	lower head	底盖		
11	isolation condenser system	独立冷凝系统	和train（列）搭配	
12	scram	紧急停堆		
13	shutdown	停堆	cold~, bring to~, reach~, remain~, (shut down)	decay heat removal system, fission reaction, Monju fast reactor
14	suppression pool	抑压池	followed by temperature	
15	liquid metal fast breeder reactors / LMFBR	液态金属快堆		
16	Gas-Cooled Fast Reactor / GFR	气冷快堆		
17	helium	氦		
18	Sodium-Cooled Fast Reactor / SFR	钠冷快堆		
19	Lead-Cooled Fast Reactor / LFR	铅冷快堆		
20	breed ratio	增殖率		
21	fuel assembly	燃料组件		
22	hydrogen explosion	氢爆炸		
23	reactor pressure vessel / RPV	反应堆压力容器		
26	fuel cycle	燃料循环		
27	recombiner	复合器	autocatalytic/active hydrogen recombiners	自动催化/活性氢复合器
28	condenser	冷凝器		
29	non-fissile uranium-238	不可分裂铀-238		
	depleted uranium	贫铀		

English origin / Chinese origin / Units of measurement

Figure 25: A sample bilingual list (FR)

7.2.1.2 Active activation of terms

As mentioned earlier in this section, during lexical acquisition, active activation through speaking is supposed to be more effective than just passively reading through the preparation documents (Gile, 2009). Useful activities include 1) practising mini-speeches/ short talks in the target language by using the key terms and the new concepts, and 2) practising sight translation of some paragraphs in the texts relevant to the topics of the speeches. In this way, terms can be further activated in different occasions for several times until interpreters can retrieve the terms fast and accurately.

In the empirical study, the above active activation activities were introduced to the groups of using different preparation procedures. The activities and the rationales behind were demonstrated to the students before they started their preparation for the specialised topics (FR & SM). During their preparation, after the students have learned the terms by themselves, they practised delivering mini-speeches (based on the speech agendas) by using the key terms in the target language (in small groups of three people).

7.2.2 The results of activation by different procedures

A previous study by Díaz-Galaz (2012, 2015) has proven that preparation helps to increase term accuracy compared with non-preparation condition. My study aims to go one step further to explore the effects of three different preparation procedures, namely the traditional preparation procedure, preparation with automatically-generated termlists, and preparation with both automatically-generated termlists and the concordancer, and find which procedure works the best for student interpreters to activate the terms for simultaneous interpreting. We judge whether the terms are activated enough for SI tasks by examining the students' terminology accuracy and the number of terminological errors during simultaneous interpreting, as well as the recalls of terms two months after the interpreting tasks.

7.2.2.1 The main effect of different preparation procedures

Traditional preparation procedure vs. the procedure using auto-lists

The first experiment showed that **using the automatically-generated lists alone** during limited preparation time did not have significant effect on increasing term accuracy in SI tasks. There was no significant difference between the test group (using automatically-generated termlists during limited preparation time) and the control group (using traditional preparation procedure with limited preparation time) on the number of Omission Errors (OM), Incorrect Terms (IT), and Semantic Errors (SE). In the post-task term quiz two months after the SI tasks, the test group using automatically-generated termlists during preparation achieved lower recall of terms than the control group who used traditional preparation procedure.

Perhaps when only limited preparation time was allowed, the relevant terms were not processed deeply enough to ensure good term accuracy in both groups. However, compared with the new preparation procedure using automatically-generated termlists, the more familiar traditional preparation procedure contributed to better memory of the new terms among the students in the control group two months after the preparation task.

Traditional preparation procedure vs. the procedure using both auto-lists and the concordancer

In the second experiment, **using both automatically-generated termlists and the concordancer** with ample preparation time had significant effect on increasing term accuracy in the SI tasks. The test group (using both automatically-generated termlists and the concordancer with ample preparation time) made significantly fewer Omission Errors (OM) in SI tasks than the control group (using traditional preparation procedure with ample preparation time). In the term quiz, the test group also achieved higher recall of terms two months after the SI tasks.

In summary, when limited preparation time was given, using automatic termlists alone did not show better effect of term activation compared with the traditional preparation procedure. And the terms were less stable in the students' memory than those acquired by traditional preparation. When ample preparation time was given, compared with the traditional preparation procedure, using both automatically-generated termlists and the concordancer helped the students to better activate the terms for the SI tasks, and more terms could be remembered and translated correctly two months after preparation.

To further explore the effects of different preparation procedures, I integrate data from the four groups. The following **Table 41** present the average accuracy scores and the average ratio of the most common error types.

Preparation procedure	Preparation time	Term accuracy Score Mean (SD)	OM	IT	SE
Traditional preparation	limited time	46.9 (11.6)	33.19 (11.9)	17.31 (5.1)	4.60 (2.9)
Using auto-lists	limited time	51.6 (15.2)	31.69 (11.4)	15.75 (4.5)	3.24 (2.5)
Traditional preparation	ample time	51.3 (14.1)	31.04 (12.0)	15.00 (4.8)	4.46 (3.1)
Using auto-lists & the concordancer	ample time	**58.8 (9.1)**	**21.78 (9.0)**	16.54 (3.5)	5.38 (2.7)

Table 41: Term accuracy scores (%) and average ratios of common error types (%) by different preparation procedures

In terms of term accuracy, **using both automatically-generated termlists and the concordancer with ample preparation time** achieved the highest mean term accuracy scores among all the four procedures and conditions. Using both tools was also the most effective to reduce the number of omission errors **(OM)**. All the groups made similar number of **IT** and **SE**.

As discussed in **Chapter 4 - Methodology (Section 4.4)**, preparation time is an important factor influencing preparation result. In this study, 3 days and 9 days were set as preparation times reflecting real practice. When ample time was allowed (9 days), using both automatic term extractor and concordancer helped save about 17% of the students' average preparation time (Control=18.77h; Test=15.60h). When limited time (3 days) was given, the use of automatic term extractor alone did not help save preparation time (Control=9.60h; Test=9.83h) (also see **Table 40** in **Chapter 6**).

7.2.2.2 The students' views on different preparation procedures

As discussed in focus group in **Chapter 6**, **Sections 6.3.2.1-6.3.2.3**, the participants expressed their views on the preparation procedures in focus groups.

Traditional preparation procedure

The students in the two control groups followed the traditional preparation procedure. Their termlists were generated manually. They felt they spent too much time on reading the texts during preparation, and there was not enough time left to further review and activate the terms that they collected from reading. They felt the terms had not been activated enough for the SI tasks after the preparation. They reported during simultaneous interpreting, they still struggled and hesitated at some technical terms.

Preparation with automatically-generated termlists

In one test group, the students used automatically-generated termlists in their preparation. There were two types of opinions. Some students were positive about the automatically-generated lists, which helped them prioritise their preparation on the most relevant terms and concepts. Others didn't quite like the idea of using the automatically-generated lists during their preparation, as editing the lists (annotation) took time and was not effective, and suggested they would rather extract their own termlists while reading through the texts.

Preparation using automatically-generated termlists and the concordancer

The use of the automatically-generated termlists together with the concordancer breaks the traditional linear reading process, and makes navigation possible. The students said they could find the terms in the preparation documents more quickly by using the concordance tool. Compared with the traditional preparation procedure (reading a whole text to understand specific terms), the students found checking and searching for terms through using the concordancer became more focused and efficient. They also used the concordancer to check the contexts and collocations of the keyterms in both languages, and to make sense of the relationship between relevant terms (also see **Section 7.2.1.1** and **Figures 23, 24 and 25**).

In summary, using automatically-generated termlists and the concordancer within ample preparation time helped the students achieve better preparation result and received positive feedback from the students.

7.3 Challenges & coping strategies

Previous studies (Darò et al., 1996, Gile, 2009, Diaz 2012 & 2015) and many others mentioned segments that contain difficult features in the source speech (such as specialised terms, high density of information) have significant impact on simultaneous interpreting. **Sections 7.3.1** and **7.3.2** will focus on the impact of the first two features (specialised terms and high density of information). The hypothesis is that better preparation has beneficial effect on the process and performance of students' simultaneous interpreting, even on segments that contained the two difficult features.

In these two sections, I will mainly focus on how the group who achieved better preparation result in this study (the test group in Experiment II: FR, using both automatically-generated termlists and the concordancer) cope with various challenges from the source speeches in comparison with the control group (using traditional preparation procedure).

Apart from specialised terms and high density of information in the texts, another challenge might be from working into one's second language (the B language). This study focuses on a specific language pair (Chinese and English). **Section 7.3** will answer two research questions: 1) Is there any interaction between language directions and the types of errors in terminology use by the trainee interpreters? 2) Does better preparation help the students to improve their term performance in SI task into their B language?

7.3.1 Challenging terms

It is found that among the terms that the students commonly made serious errors (scored "0") in, most of those terms are specialised terms (S) and named entities (NE). Our data showed that these two types of terms posed greater challenge to the group using the traditional preparation procedure than to the group using both tools. It proved that better preparation (using both automatically-generated termlists and the concordancer) helped reduce the challenges from the segments that contained specialised terms (S) and named entities (NE) in simultaneous interpreting, even though NEs (including names of organisations, names of places, names and titles of people) were not always in complete/correct forms in the automatically-generated lists (by Syllabs) (also see Chapter 5, **Section 5.3.4**).

Having discussed the challenges from S and NE (in **Chapter 6, Section 6.2.5**), now I'd like to look at the possible reasons for errors, and the shared as well as and different coping strategies used by the two groups.

7.3.1.1 Unprepared, therefore non-activation

There are terms being left out during preparation therefore missed the chance of being activated before interpreting.

I checked the terms that the students commonly scored "0" in the students' manual termlists and the termlists automatically-generated by Syllabs.

Some of the terms, such as *"嬗变" (transmutation)* and *"余热" (decay heat)* were not included by the students in their manually-generated lists (in the control group). It was very likely that the students were overwhelmed by so many preparation materials to read and new concepts to digest within limited time, and they overlooked these relevant terms. During interpreting, the students opted to leave out those unfamiliar terms.

As for the test group (FR), some terms were not automatically extracted, and therefore not included in the auto-lists in the first place. For example, *"Generation IV International Forum (GIF)"* and *"secondary heat transfer system"* were not extracted as terms. Both of them only appear once respectively in the original English preparation documents, a corpus of 42,006 words. As I have discussed in **Chapter 5**, **Section 5.1.3**, due to the statistical nature of automatic term extraction, the accuracy of term extraction is subject to term frequency. If the terms only appear once or twice in the original preparation documents, the automatic term extractor would fail to detect them as terms. The terms were not recognised during reading afterwards either, therefore lost the chance of being activated. As a result, more than half of the students in the test group interpreted *"Generation IV International Forum (GIF)"* as *"国际原子能机构" (International Atomic Energy Agency)* or a "made-up" term *"国际核能组织" (international nuclear organization)*. These errors were annotated as IT-0. And *"secondary heat transfer system"* was omitted during interpreting by most of the students, and was annotated as OM-0. It seems that terms with low frequencies in the preparation documents are very likely being ignored by either manual collection or automatic extraction of terms. A bigger size of corpus may possibly increase the term frequencies and thus more terms are likely to be automatically extracted. This is probably when automatic term extraction has some advantage over manual term collection.

7.3.1.2 Insufficient activation

There are terms that had been selected and included by the control group in their termlists, but the students still made serious errors during interpreting.

For example, I notice that many students have included *"核临界"* as a term in their termlists. When simultaneously interpreting the segment *"…实现首次核临界" (…achieved initial criticality)* from Chinese into English, some students adopted literal translation and interpreted the segment as *"… realized its first critical"*. Many other students simply omitted the segment all together.

Another example is a term in Chinese *"并网发电" (be connected to the grid/ grid connection)*. The equivalents that the students had in the list were all in dictionary versions, for example, *"grid-connected power generation"*, or *"synchronization"*. During simultaneous interpreting, the segment *"…成功实现首次并网发电" (… was connected to the grid for the first time)* was interpreted as *"… managed to have connected power generation"* or *"… realized synchronized generating"*. The students applied the dictionary-version translation; however it was a bit too challenging to pronounce the complicated version under time pressure. Their interpretation was mechanical and did not make sense to the target audience.

It is evident that the students' "terminology work" in the control group stayed rather at a superficial level. Their preparation was very likely only focusing on checking definitions and equivalents, rather than getting familiar with how the term and the equivalents can be used in contexts. In other words, the activation is quite "shallow".

I also checked the same terms in the auto-list used by the test group. The two terms *"临界" (criticality)* and *"并网发电" (grid connection)* are both ill-formed in the Chinese auto-list. Yet it is interesting to observe that through annotation and further processing, many students in the test group recovered the ill-formed terms into the right form and included the English translations in their final bilingual lists. Then through using the concordancer, they also had the opportunity to check all the instances where the terms and the equivalents were used in both languages. In the final bilingual lists, they had more than one equivalents included for *"临界"*, i.e. *"critical (state); criticality"*, they also included a useful collocation in English *"achieve ~"* in a separate column headed by "other relevant information". Similarly, they had more than one equivalent for *"并网发电"*, i.e. *"synchronization; synchronize the generator with the grid; connect to the grid"*. It is evident that the two terms and many others have been checked, semantically processed and further activated before interpreting. In fact, most students in the test group interpreted the two terms accurately and effortlessly.

There are occasions in which the students from both groups recognised and included one variant of a term in the termlists, but failed to interpret the other variant of the term correctly during interpreting. In other words, activating one variant of a term does not mean all the other variants are equally activated. For example, *"mixed-oxide fuel"* was included in both of the manual and auto-lists, but it was the abbreviated version of the term *"MOX fuel"* (acronym) being used in the English speech (FR_en). Since the initial letters were pronounced as a single word, the students did not recognise it and failed to interpret the term correctly during interpreting.

7.3.2 High density of information

As mentioned earlier in **Section 7.3**, high density of information inherent in the source speech may have significant impact on simultaneous interpreting. Gile (2009) mentioned "high density of source speech increases processing capacity requirements, because more information must be processed per unit of time. This affects both Listening and Analysis Effort and the Production Effort".

In this study, the density of source speech is not reflected in speed of delivery. The delivery of the speeches is controlled at normal speed (FR-en=82wpm, SM-en=80wpm, FR-zh=167cpm, SM-zh=180cpm) (also see **Chapter 4**, **Section 4.3.1**). The high density of speech is reflected in high density of information content of the speech, especially in particular paragraphs. These paragraphs comprise a string of information elements with few functional words or low information density elements in between (e.g. ***Monju nuclear reactor*** *is a* ***prototype fast reactor****, which has* ***three coolant loops****, uses* ***MOX fuel*** *and is able to produce 714* ***MWt****……***).** Segments like this may easily cause saturation of processing capacity during interpreting and consequently lead to serious terminology errors.

This allows me to form a hypothesis that an increase of information density has a detrimental effect on interpreting processing and terminological performance, and furthermore, better preparation (using both auto-lists and the concordancer) may help to mitigate the effect.

7.3.2.1 Interactions between different preparation procedures & density of terms

Information density is counted as density of terms (of each paragraph in the speeches). I used the following definition for the measuring density of terms for this study:

$$Density\ of\ terms = \frac{the\ number\ of\ terms\ in\ a\ paragraph\ (t)}{delivery\ time\ \ of\ the\ paragraph\ (m)}$$

There are altogether 26 paragraphs in the original speeches (FR-en: 12 paragraphs; FR-zh: 11 paragraphs). The paragraphs have a wide range of density of terms: from 2.38 t/min to 20.37 t/min (also see **Chapter 4, Section 4.3.2.2**[16]).

In order to test the two hypotheses: 1) increasing density of terms in the original speech has a detrimental effect on terminological performance during SI, and 2) better preparation helps to mitigate the effect, **Excel's correlation coefficient tests** were conducted to measure the associations between **density of terms** and **the mean numbers of serious errors (score "0")**.

---

16 Average term density (speeches on FR): FR-en=9 t/m, FR-zh=15 t/m.

Figure 26 presents **Correlation Coefficient** and **Coefficient of Determination (R square**) for the control group (using traditional preparation procedure) and the test group (using both automatically-generated termlists and the concordancer) respectively. There were strong positive correlations between density of terms and the number of serious errors in both groups (**P value<0.05**). It confirms our hypothesis that increasing density of terms in the original speech had a detrimental effect on the students' interpreting processing and terminological performance.

Figure 26 also shows that the **Correlation** and **R square values** of the control group were higher than the test group, which means the positive correlation between density of terms and the number of serious errors was weaker in the test groups than in the control groups.

	Correlation	R square
Control (FR)	**0.842**	**0.709**
Test (FR)	0.792	0.628

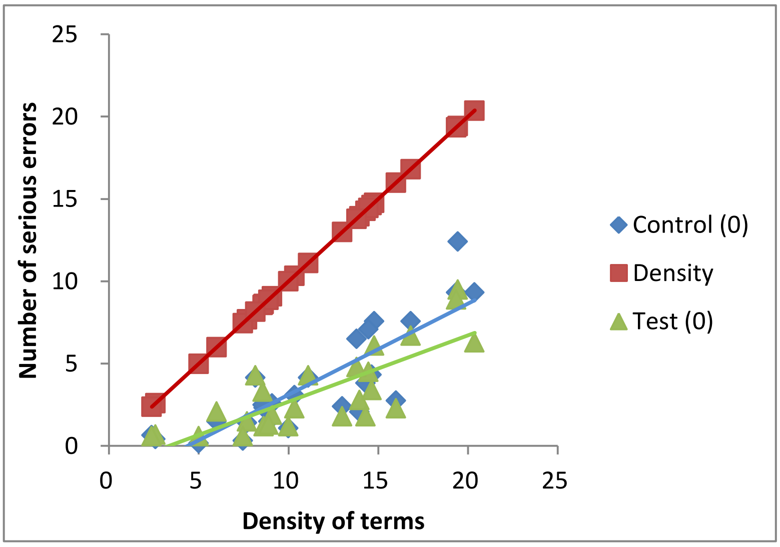


Figure 26: Correlations between density of terms & numbers of serious errors (Experiment II: FR)

The above scatterplot is a visualisation of the linear associations. The number of serious errors in the control group (represented by the blue line) was more closely associated with density of terms (red line) than the test group (green line). The scatterplot also helps to compare the performances of the two groups in specific paragraphs. The numbers of serious errors made by the control group (blue dots) were evidently higher than the test group (green dots) in the paragraphs of higher densities (13.00-20.37 t/m). This result proves that better preparation in the test group helped mitigate the detrimental effect of increasing density of terms in the source speech, in particular, it helped reduce the numbers of serious term errors in the paragraphs of high densities.

I further explore the correlations between density of terms and the numbers of three major error types, i.e. OM, IT and SE, respectively (see **Table 42**).

	OM		IT		SE	
	Correlation	R square	Correlation	R square	Correlation	R square
Control (FR)	**0.826**	**0.682**	**0.822**	**0.676**	0.234	0.055
Test (FR)	0.741	0.549	0.735	0.540	0.372	0.139
P-value (OM &IT)<0.05; P -value (SE)>0.5						

Table 42: Correlations between density of terms and the number of OM, IT and SE (Experiment II: FR)

The correlation and regression tests show that there were strong positive correlations between density of terms and the number of **Omission Errors (OM)** and **Incorrect Terms (IT)** respectively (P value<0.05). This applied to both groups. However, there was no linear association of term densities and the numbers of **Semantic Errors (SE)** (P value>0.05). That is to say **OM** and **IT** were two types of errors sensitive to condensed information inherent in the texts. When density increased, the students in both groups were more likely to make more Omission Errors (OM), and there were more terms being inaccurately interpreted (IT).

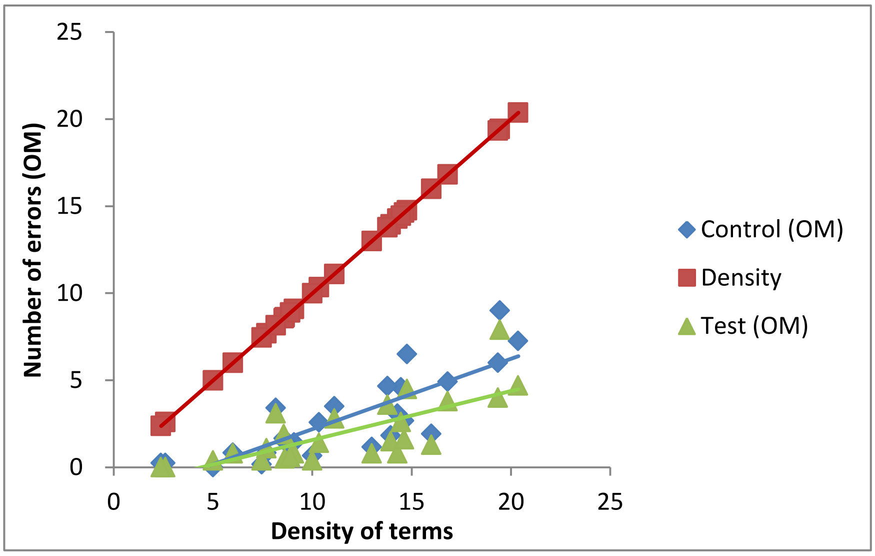


Figure 27: Correlation between density of terms & numbers of OM (Experiment II: FR)

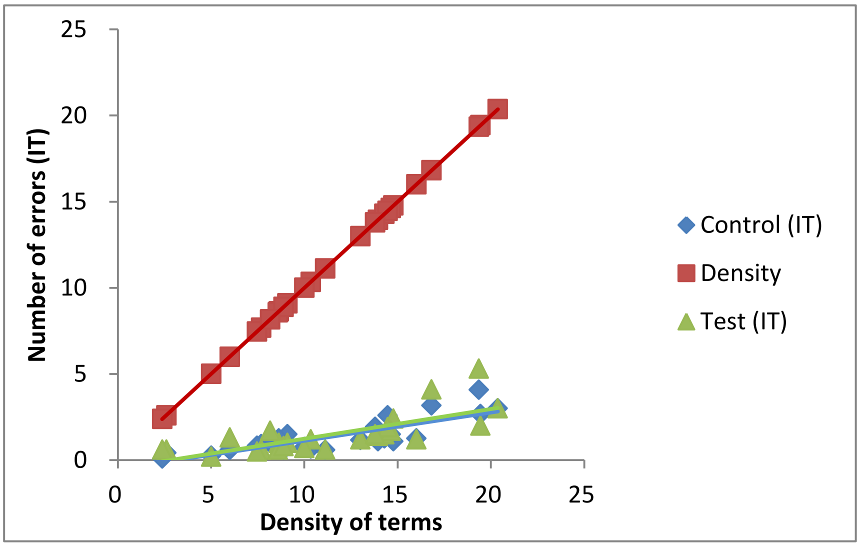


Figure 28: Correlation between density of terms & numbers of IT (Experiment II: FR)

The above two scatter plots **(Figure 27 & 28)** help us to compare the numbers of OM and IT of the two groups (control & tests) in specific paragraphs (with different term densities).

Figure 27 shows that the numbers of **OM** in the control group (represented by the blue line) was more closely associated with densities of terms (red line) than the test group (green line). In addition, the numbers of **OM** made by the control group (blue dots) were higher than the test group (green dots) in paragraphs with densities of 10.34-20.37t/m. This result proves that better preparation in the test group helped to mitigate the challenges from increasing density of terms to the number of Omission Errors (OM) during interpreting.

The pattern shown in **Figure 28** is however different from the previous two scatterplots **(Figure 26 & 27)**. The blue (control) and green (test) lines in **Figure 28** nearly overlapped, and the dots that represent the numbers of **IT** in each paragraph resembled quite a lot in both groups too. This means the strengths of linear associations between density of terms and the numbers of **IT** were quite similar in both groups. There was no obvious mitigation effect of better preparation on the number of **IT** errors.

In summary, the statistical analysis confirms the hypothesis that increasing density of terms in the original speech had a detrimental effect on the students' terminological performance during simultaneous interpreting. **OM** and **IT** were the two types of errors sensitive to condensed information inherent in the texts. When density of terms increased in the speech, the students were more likely to make more omission errors, and there were more terms being inaccurately interpreted. Better preparation helped to mitigate the detrimental effect of dense information in the source speech on the numbers of serious errors and omission errors made by the students; however, better preparation does not help decrease the challenge from high density of information to IT. It is possible that the more terms that the students interpreted, the more likely they made terminological mistakes.

7.3.2.2 Coping with challenging segments

In the following part, I will further examine the challenging paragraphs with higher density of terms and explore how the students coped with the challenges.

Dam (1993:302-306) proposed "condensing" as a necessary interpreting strategy in consecutive interpreting, which to a large extent is applicable to known elements, non-focalised elements (less important material of a sentence, which often can be easily inferred by the audience) and redundancies (such as expressions of repetitions, self-evident elements, etc.) and other features in the original speech. I need to make it clear that in this study condensing/omitting terms without changing or distorting the meaning of the text is not categorised as an omission error, and could be given a full score (depending on the situation). For example, when a term is mentioned a second time in the following sentence or an adjacent segment, and if the term is substituted by the pronoun *it* in the interpretation, they would not be counted as errors.

7.3.2.2.1 Coping with non-redundant segments

The "words" under observation in this study are terms. In the speeches that the students interpreted from, many specialised terms appear in non-redundant segments (i.e. the segments of no repetition, no redundancy, consisting of information elements put next to each other). Since both human memory and processing capacity are limited, when encountering such difficult segments, many students chose to simplify and generalise the information.

The following examples represent the challenging sentences/paragraphs where five (out of 10-12 students) made term errors. All the terms are highlighted with bold in the ST. Sample interpreting from the group (TT) as well as back translation (BT) are also provided.

ST: The new strategy will be formulated within a year. I think the major emphasis of **sodium-cooled fast reactors (T1)**' R&D should be on **full power operation (T2)** of **Monju (T3)**, **passive safety systems (T4)** and **severe accident management procedures (T5)**.
TT: 新的策略将在一年内制定完成，我们的重点应该放在**文殊快堆(T3)**上，**以及其他方面 (T4 & T5)**。 **BT:** New strategy will be formulated within one year. Our focus should be on **Monju fast reactor (T3)**, **and other aspects (T4 & T5)**.

The sentence (ST) above contains a list of specialised terms. TT is sample interpreting from the students in both groups. The interpretation was a condensed version. "**Full power operation** of **Monju**" was simplified as "**Monju"**, and two parallel items ("**passive safety systems"** and "**severe accident management procedures"**) were generalised by a non-term "other aspects".

The following is another example where information density is high in the ST and therefore resulted in omission errors in the students' interpretations. The table below presents the sample interpreting (and back translations) by the control and test groups respectively.

<table>
<tr><td colspan="2">ST: Monju nuclear reactor (T1) is a prototype fast reactor (T2), which has three coolant loops (T3), uses MOX fuel (T4) and is able to produce 714 MWt (T5). It started operation (T6) in 1994, but a sodium leakage (T7) occurred in its secondary heat transfer system (T8) during performance tests (T9) in 1995.</td></tr>
<tr><th>Control</th><th>Test</th></tr>
<tr><td>TT: 文殊核反应堆(T1) 有三个冷却回路系统(T3)，能够产生 714 兆瓦(T5)的电。
BT: Monju nuclear reactor (T1) has three coolant loops (T3), generates electricity 714 MWt (T5).</td><td>TT: 文殊(T1)是一个原型快堆(T2)，产电量是 714 兆瓦(T5)。但是在 1995 年试运行(T9)的时候，发生了钠泄漏事故(T7)。
BT: Monju (T1) is a prototype fast reactor (T2); Electricity capacity is 714 MWt (T5). But during performance tests (T9) in 1995, a sodium leakage accident (T7) occurred.</td></tr>
</table>

The control group relayed limited information in their interpretation. Only the first half of the paragraph was covered. Their processing capacity was clearly not sufficient to deal with the competing efforts (interpreting while listening to the incoming segments). By comparison, the sample interpretation by the test group had retained more specific information segments. The test group was lexically more resourceful due to increased term availability through sufficient preparation.

7.3.2.2.2 Coping with complex specialised concepts

It is found that many of the terms where the students made serious errors are in the sentences densely packed with complex specialised concepts. When much of interpreters' attention is on comprehending the message, it is likely that interpreters would either omit or mis-interpret the new terms that they have just learned through preparation. Hopefully when interpreters activate new terms sufficiently, they shall be able to work successfully under pressure.

The following example is from the Chinese speech on how a fast reactor works, where a number of terms are used consecutively. The table below illustrates the sample interpretations by the control and test groups respectively.

<table>
<tr><td colspan="2">ST: 核裂变 (T1) 所产生的热量 (T2) 可以通过液化钠 (T3) 带出堆芯 (T4)，然后再通过热交换器 (T5) 和蒸汽发生器 (T6)，产生高温 (T7)、高压 (T8) 的蒸汽 (T9)，最终推动汽轮机 (T10) 发电。
BT: Fission reactions (T1) produce heat (T2), which could be taken out of the reactor core (T4) by liquid sodium (T3). Then through heat exchanger (T5) and steam generator (T6), high-temperature (T7) and high-pressure (T8) steam (T9) is produced, and it finally pushes the turbine (T10) to generate electricity.</td></tr>
<tr><th>Control</th><th>Test</th></tr>
<tr><td>TT: The Heat (T2) producing can be taken away by liquid sodium (T3), and can produce steam (T9) to generate electricity.</td><td>TT: The heat (T2) generated from fission process (T1) can be removed by liquid sodium (T3). The heat will be transferred by heat exchanger (T5), and then pushes turbine (T10) to generate electricity.</td></tr>
</table>

The control group made a series of omission errors, leaving out terms (T4-T8). Only a gist of the original meaning is relayed. They used three terms, but they probably did not have the capacity or time to reformulate the information properly in the target language. Therefore their English interpretation was not quite easy to understand. By comparison, the better prepared test group made fewer omissions in interpretation. Their interpreting is easier to follow. The audience could get quite a lot of detailed and specific information from the interpretation. The test group also left out some detailed information, but they seemed to have well-planed choices about what to leave out and what to preserve, instead of leaving out elements at random forced by limitations of memory and time pressure inherent in interpreting.

In summary, dense information inherent in the texts to be interpreted, (i.e. non-redundant segments and complex specialised concepts) could cause saturation of processing capacity during interpreting, and consequently serious terminology errors were made. When dealing with the difficult paragraphs, both groups applied the strategy of compression and abstraction. Interpretation is characterised with being rough and less accurate.

I agree that condensing is a necessary strategy, in other words, the second best choice if total completeness of information cannot be achieved, especially when dealing with high density of information. It is evident that the lexical resources at the interpreter's disposal were quite limited when terms were less activated during preparation. In this situation, the interpreters did not have many choices but leaving out the specific terms and remaining abstract. Only a gist of the original message could be relayed. Compression and abstracting were used here as a 'rescue technique', while if terms were better activated during preparation, the message (of the source speech) could be conveyed with more details in place during SI. Lexical and syntactic compression was still a major strategy applied, yet information reduction was through selection. The interpreters were more in control because they had more resources at their disposal.

Therefore, increasing term availability would be useful to improve terminological performance during simultaneous interpreting. With better availability of terms, interpreters could be more resourceful. They could make some conscious and well-planned choices about what to leave out and what to preserve.

7.3.3 Working into the second language

Apart from the high density of information in the texts, another challenge is from working into one's second language (the B language). Although interpreters are often assumed to have achieved a good command of their working languages, numerous studies have demonstrated that simultaneous interpreting from the A to the B language and from the B to the A language may involve different processes and result in different products (Chang & Schallert, 2007:138).

For example, in terms of lexical processing, word-for-word translation from the A to the B language was found to be slower than from the B into the A language (de Bot, 2000). Producing L2 syntax was believed to be less automatic and require conscious monitoring (Bialystok, 1994, Ullman, 2001). It is also argued that simultaneously interpreting into one's L2 not only requires more effort but may also produce poorer results (Seleskovitch 1999). The trainee interpreters were found to make more serious errors leading to loss of information when interpreting difficult texts from their A to their B language (Darò et al., 1996).

7.3.3.1 Chinese and English specific problems

This study focuses on a specific language pair (Chinese and English). In this section, I will discuss some of the specific challenges when working from Chinese into English, and provide examples from the Chinese speech used in the experiment (FR_zh).

7.3.3.1.1 Different phonetic features of Chinese and English

Generally speaking, English words take up more space than Chinese. When words have one-to-one formal equivalents, the Chinese words are shorter, since the added dimension of tones allow for more semantic information per unit (character/syllable), and every single unit carries meaning (Setton, 1993:245). As regards to technical terms, they are normally expressed in fewer syllables in Chinese then in English. For example, *"原型快堆"* (4 syllables) become *"prototype fast reactor"* (7 syllables); the radioactive chemical elements *"铀"* (1 syllable) and *"钚"* (1 syllable) become *"uranium"* (3 syllables) and *"plutonium"* (3 syllables).

When interpreting from Chinese into English, the "swelling factor" is significant (Setton, 1993). It has an impact on interpreting (zh-en) under time pressure. Even though the Chinese speeches used in this study were delivered at reasonable speed (FR-zh=167 wpm, SM-zh=180wpm[17]), interpreters have to speak faster in English due to the "swelling factor". Alternatively, they have to summarise to produce a more succinct interpretation than the original when they cannot keep pace with the speaker.

7.3.3.1.2 Syntactic difference between Chinese and English

Generally speaking the two languages share a basic sentence structure (S-V-O). When interpreting simultaneously from Chinese into English, interpreters could apply linearity/segmentation as the main coping strategy. In other words, they could follow the sentence structure and lexical choices in the ST with minimum change. However, in many other cases, Chinese sentence structures are more flexible than English. Interpreters have to reformulate sentence structures when interpreting from Chinese into English. They have to rely heavily on waiting and anticipation to overcome problems caused by word order difference between the two languages.

For instance, sentences without any subject are quite common in Chinese, since the subject of a sentence is often omitted when it is self-evident or uncertain. But in English, every sentence (but an imperative sentence) must contain a subject (Hu & Tao, 2013:634). Interpreters must choose between active and passive constructions

---

[17] Each Chinese word (character) is of only one syllable. This study measures by syllables in Chinese. "It is widely recognised that a rate between 100 and 120 words per minute (wpm) is optimal for English speeches. This translates into an optimal speed of 150-180 syllables per minute for Chinese speeches" (Li, 2010: 21).

when interpreting these zero-subject sentences.

ST:	**燃料(T1)**中的**铀 238(T2)**就转换成了**钚 239(T3)**，**钚 239(T4)**为**易裂变核素(T5)**。这样一来，就实现了**核燃料(T6)**的**增殖(T7)**。
BT:	(In) **nuclear fuel (T1), Uranium 238 (T2)** turns into **plutonium 239 (T3)**, **plutonium 239 (T4)** is **fissile material (T5)**. In this way, (it) achieves **nuclear fuel (T6)**'s **breeding (T7)**.
TT:	**Uranium 238 (T2)** in **nuclear fuel (T1)** turns into **plutonium 239 (T3)**, **which (T4)** is **fissile material (T5)**. In this way, **the fuel (T6)** is **bred (T7)**.

In the above example from Chinese speech (FR-zh), the underlined sentence (in the ST) does not have any subject, featuring a "verb + object" structure. The interpreter must instantly determine a subject to use when interpreting into English. If *"核燃料" (nuclear fuel)* is used as a subject, a passive construction needs to be used, and the original sentence order has to be changed too.

Some verb constructions are often used in Chinese, but when interpreting into English, the sentence structures have to be changed dramatically. For example, Chinese sentences using BA, JIANG, DUI, XIANG constructions often have long and complicated objects. In the following example (ST), a JIANG construction (meaning "to make", "to take"…) is used. If the original word order is followed, the English interpretation would be wordy and clumsy (see **BT**). To produce an acceptable interpretation in English, the interpreter needs to change the word order and eliminate redundancy (see **TT**).

ST:	**快堆(T1)**可以将**压水堆(T2)**所产生的**长寿命废料(T3)嬗变(T4)**为**短寿命核素(T5)**。
BT:	**Fast reactors (T1)** can make **pressurised water reactor (T2)** generated **long-lived nuclear waste (T3) (be) converted (T4)** into **short-lived nuclide (T5)**.
TT:	**Fast reactors (T1)** can **convert (T4) long-lived nuclear waste (T3)** from **pressurised water reactor (T2)** into **short-lived nuclear waste (T5)**.

In addition, Chinese does not share many grammatical features that English has. Unlike English, Chinese verbs have no inflectional endings, such as those for tense, person and number agreement. Nouns have no markers of number, gender and case. Therefore, simultaneously interpreting from Chinese into English probably requires the interpreters, in many cases, conscious monitoring their language production (in English) in order to avoid grammatical errors. As identified in Giles' Effort Models (2009), when an interpreter's cognitive effort is engaged in one aspect of the task, other aspects can be at risk of being adversely affected.

In summary, due to differences between the two languages in terms of phonetic, syntactic and grammatical features, when interpreting from Chinese into English simultaneously, interpreters probably need to employ much of their analytical skills and engage in extensive transformation in order to convey the original message. Interpreters also need to speak faster and to avoid committing serious language errors.

7.3.3.2 Results from the experiments

As discussed in **Chapter 4 - Methodology, Section 4.2**, the participants of this study were MA trainee interpreters at University of Leeds. They all had Chinese as their mother tongue, and worked bi-directional (English-Chinese and Chinese-English) throughout their training. Their B language (English), while fluent, is still non-dominant with respect to their mother tongue (Chinese).

The data from the experiments with the trainee interpreters allows us to see the effects of language direction (English-Chinese and Chinese-English) on terminology performance during SI. My research question is whether there is any interaction between language directions and types of errors in terminology use committed by the trainee interpreters. For this purpose, I compare error percentages of different types of errors committed in interpreting from Chinese (the A language) to English (the B language) and from English to Chinese.

7.3.3.2.1 The impact of language directions on error categories

As mentioned in **Chapter - 4 Methodology**, **Section 4.5.3**, this study classifies errors (in terminology use) into six categories, i.e. **OM, IT, SE, GE, IC, PE**. I calculated the means and standard deviation values for error percentage of each error category for all the SI tasks (also see **Chapter 6, Section 6.1.3** and **Section 6.2.3**). In **Table 43 & 44**, I further compare the difference of the error percentages between working B-A and A-B on two specialised topics in control and test groups respectively. The significant differences are highlighted with **bold.**

Error category	Control		Test	
	en-zh Mean (SD)	**zh-en** Mean (SD)	**en-zh** Mean (SD)	**zh-en** Mean (SD)
OM (%)	27.10 (12.1)	**39.28 (8.4)**	23.51 (7.8)	**39.86 (8.0)**
IT (%)	15.62 (4.7)	18.99 (5.1)	14.36 (4.7)	17.13 (4.0)
SE (%)	**6.15 (3.1)**	3.04 (1.5)	4.03 (3.2)	2.44 (1.3)
GE (%)	0 (0)	**0.80 (1.0)**	0 (0)	**2.04 (1.6)**
IC (%)	0.24 (0.4)	0.65 (0.8)	0.05 (0.2)	**1.05 (0.9)**
PE (%)	0.18 (0.4)	**1.45 (1.4)**	0 (0)	**1.12 (1.7)**

Table 43: Error percentage (Experiment I: SM)

Error category	Control		Test	
	en-zh Mean (SD)	**zh-en** Mean (SD)	**en-zh** Mean (SD)	**zh-en** Mean (SD)
OM (%)	29.26 (9.7)	32.82 (14.1)	19.77 (9.8)	23.79 (8.1)
IT (%)	13.86 (4.9)	16.15 (4.7)	16.05 (3.0)	17.03 (4.0)
SE (%)	**6.40 (2.7)**	2.53 (2.1)	**6.98 (2.5)**	3.79 (1.7)
GE (%)	0 (0)	**2.30 (1.6)**	0 (0)	**1.86 (1.1)**
IC (%)	0.10 (0.3)	**0.86 (1.0)**	0.47 (0.8)	0.69 (0.7)
PE (%)	0.58 (1.2)	0.57 (0.9)	0 (0)	**1.03 (1.2)**

Table 44: Error percentage (Experiment II: FR)

The results from **Table 43 & 44** demonstrate that the students made more **Omission Errors (OM)** when interpreting into their B language (English) than into their A language (Chinese). On the topic SM, this difference is particularly significant (in both control and test groups) **(P<0.05)**. As explained in **Section 7.3.3.1**, it is probably due to the “swelling factor” when interpreting from Chinese into English, and the interpreters had to comprise and generalise to keep pace with the speaker. Therefore they tended to leave out the terms and expressions that they were less familiar with or not able to retrieve immediately from memory.

The two tables also shows that the students made similar numbers of or only slightly more **IT** when interpreting into English than into Chinese on both topics

(P>0.05). There might be a trade-off effect between **OM** and **IT** when interpreting into Chinese (the A language). On the one hand, the students probably made fewer term omissions, on the other, the more terms they attempted to interpret; the more possible they interpreted them incorrectly.

As regards the other four categories of errors, I find that the students generally made significantly lower semantic errors (SE) when interpreting into their B language than into their A language (P<0.05). While they made significantly more Grammatical Errors (GE), Inappropriate Collocations (IC) and Pronunciation Errors (PE) when interpreting into their B language than into their A language (P<0.05). This result is in agreement with the Lee (2003) that interpreters committed more language-use errors but fewer meaning errors when interpreting from A to B.

7.3.3.2.2 The impact of preparation procedures on error categories (Chinese-English)

As discussed in **Section 7.2.2**, using automatically-generated lists and the concordancer within ample preparation time generally helped the students achieve better preparation result (with higher term accuracy scores achieved and fewer serious term errors committed). The next research question is whether preparation through using auto-lists and the concordancer could help the students to improve their term performance when working into their B language. For this purpose, I further examine which category (or categories) of errors could be significantly reduced in the SI task into the B language through preparation using both the auto-lists & the concordancer (see **Table 45**).

Preparation procedure	**Control group** **Traditional preparation** Mean (SD)	**Test group** **Auto-lists & the concordancer** Mean (SD)	**P-value**
OM (%)	32.82 (14.1)	**23.79 (8.1)**	**0.038**
IT (%)	16.15 (4.7)	17.03 (4.0)	0.319
SE (%)	2.53 (2.1)	3.79 (1.7)	0.068
GE (%)	2.30 (1.6)	1.86 (1.1)	0.231
IC (%)	0.86 (1.0)	0.69 (0.7)	0.314
PE (%)	0.57 (0.9)	1.03 (1.2)	0.171

Table 45: Error percentages of six error categories (Experiment II: FR_ zh-en)

Table 45 shows that the group using both the auto-lists and the concordancer made significantly fewer **OM** than the group using traditional preparation procedure (P<0.05), while the two groups (with or without using both the auto-lists and the concordancer) did not have significant difference in the number of errors in the other categories (i.e. IT, SE, GE, IC, PE) (P>0.05).

As discussed in **Section 6.2.3**, apart from the fact that the group using both the auto-lists and the concordancer (test group) made a significantly lower percentage of Omission Errors (OM) in the SI task into English (the B language), I also find that the group had a particularly smaller standard deviation (SD) than the control group (without using any tools). A lower standard deviation (SD=8.1%) indicates that the percentages of OM made by the test group tend to be clustered closely around the mean (23.79%). In other words, the students in the test group generally made fewer OM, and they performed more similarly in FR_Zh-En. This means that using both the auto-lists and the concordancer during preparation were quite effective in helping the students reduce the number of OM in SI task into their B language (English).

In summary, the results provide evidence to support that language direction has an impact on types of errors made during simultaneous interpreting. The participants generally made more Omission Errors (OM) when working from Chinese into English (A-B) than working from English into Chinese (B-A). They also committed significantly more language-use errors (GE, IC and PE) and fewer meaning errors (SE) when working into their B language. Better preparation through using automatically-generated lists and the concordancer helped to significantly reduce the number of Omission Errors (OM).

7.4 Pedagogical implications of the findings

7.4.1 Training workshop on terminology preparation for technical meetings

As discussed in **Chapter 4, Section 4.6.3**, the empirical data was collected throughout a workshop (about 2 months) on terminology preparation for technical meetings to the trainee interpreters in the last term of their MA interpreting training in 2013 and 2014. With training on how to use term extraction tools and concordance tools, the student interpreters developed the skills to build their own terminology resources and activate relevant terms for simultaneous interpreting tasks in specialised fields.

Through this workshop, the students were more aware of effective preparation procedures that work for themselves. The hands-on experience of using different preparation procedures to prepare for two specialised topics, as well as retrospective group discussions afterwards helped the students to clarify their understandings of terminology preparation for simultaneous interpreting. Their preparation practice at the workshop also motivated them to search for the preparation strategies that best suit themselves. It is evident that the students became more confident and reflective in using the tools and the preparation procedures. Generally speaking, they became less stressed and more professional in their preparation for specialised fields.

Our statistical analysis suggests that the pipeline/preparation procedure involving using both the term extractor and the concordancer yielded better preparation result than the traditional preparation procedure, and generally helped to improve the trainee interpreters' term accuracy during simultaneous interpreting. It is important for both students and trainers to be aware that electronic tools, when used properly, can assist the interpreters' terminology preparation to achieve an enhanced performance.

On the other hand, it is clear to us that more efficient terminology preparation (through using both the term extractor and the concordancer) is not a "magical cure" for all errors. Our data shows that the preparation procedure only helped to improve holistic SI score to a certain extent (but not yielding any statistical significance). The use of the two tools did not help in reducing the numbers of Incorrect Terms **(IT)** in SI tasks either.

7.4.2 Term accuracy as judging criteria in training

Our results in **Chapter 6 (Sections 6.1.4 & 6.2.4**) show there were strong correlations between the students' term accuracy and SI performance in the SI tasks on both topics (SM & FR) in both control and test groups. These results proved that term accuracy plays an important role in SI performance on specialised topics.

However, as we all know, term accuracy is not the only criterion to judge an interpreter's simultaneous interpreting performance. Other important criteria include sense and logical consistency, appropriate use of target language, delivery and presentation, etc. Therefore I also explored among what kind of students, term accuracy may play more prominent role than other factors influencing their interpreting performance. I integrated all the data relevant to the students' term and SI performances in this study to further observe the correlation between term accuracy and SI performance.

The scatterplot in **Figure 29** provides a visualisation of the correlation between term accuracy and SI performance among all the participants in all the four SI tasks (n=86). As expected, there is a strong positive correlation between term accuracy scores and holistic SI scores (P<0.05).

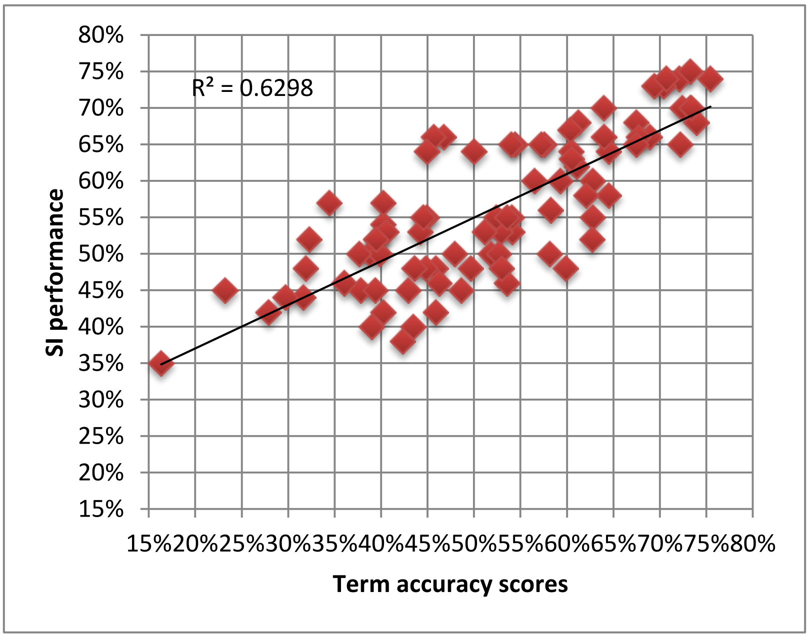


Figure 29: Correlation between term accuracy & holistic SI performance (All SI tasks)

Figure 29 shows that among the top students who achieved over 66% in holistic SI score (66-75%), the correlation between term accuracy and holistic SI performance were stronger. The students' SI performance became more dependent on term performance (the scattered dots are less spread). While among the majority of the students who achieved lower than 65% in SI score (65-35%), their SI performance was comparatively less dependent on term performance (the scattered dots are more spread). This probably means that among the top students who had solid interpreting skills and language capacity in place, term accuracy was a prominent factor influencing on their SI performance. For the top students, increased term accuracy directly helped to enhance their SI performance. However, among the majority of the students, other determining factors, such as logical coherence, appropriate use of target language, delivery and presentation, might also have critical influence on their SI performance. In other words, for the majority of students, the contribution from increased term accuracy on their SI performance might not be as obvious as for the top students. However, if not having given any chance to prepare, average and poor students could have performed a lot worse. Training on terminology preparation helps to raise their awareness on the importance of preparation and terminology use in interpreting. Terminology accuracy is something they could focus more on in their future professional career as interpreter.

Suppose the top students are approaching to the level of professional interpreters. Similar results may be observed among the professionals. Term accuracy could possibly be used as a benchmark to judge professional interpreter's SI performance on specialised topics. Yet it has to be examined in future studies.

It is also found that during simultaneous interpreting of the specialised texts, the students attempt to interpret everything they process effortlessly, especially in those more generic parts in a speech (e.g. opening and closing or the parts with less technical contents). They also tend to "parrot" the redundancies, repetitions and non-meaningful segments from the source speech. However, the students do not always have the capacity to deal with those specialised contents, where technical terms and concepts are concentrated in. As a matter of fact, those technical contents provide more useful information to the specialist audience in real technical conferences. Terminological accuracy is considered more important in technical meetings than in general meetings (Moser 1995, 1996). Therefore, if the students' interpretation is wordy and with less real substance, it may not being able to meet the real audience's needs.

Finally, the trainee interpreters should be provided with guidance on terminology preparation for technical meetings. Training in this regard could be a useful supplement to the already existing professional interpreting training. And if possible it would be more appropriate to provide such training after the basic interpreting skills have been mastered by most of the students.

Chapter 8 - Conclusions

8.1 Synopsis

This study focuses on interpreters' advance preparation for technical meetings and the application of corpus tools in the interpreters' preparation. In order to address the issues, this study formulated four research objectives. This section summarises the findings in relation to each of them.

The first objective was to investigate how to integrate the use of corpus tools into the interpreters' preparation. Based on various studies on the interpreters' terminology preparation, this study experimented with a corpus-based terminology preparation pipeline for interpreters covering **a)** corpus building, **b)** automatic term extraction, **c)** term exploration and **d)** term management (see **Chapter 3**). Based on this pipeline, interpreters could establish their own corpora with documents from conference organisers and terminologically-rich text source from the internet (by using Web Crawlers). Interpreters then use an automatic term extractor to generate termlists from the established corpora, and use a concordance tool as navigational aid for close reading and consolidating learning of key terms in contexts. Finally they could use term management tools to update and manage their own terminology resources for future use.

The second objective was to identify methods to assess performance of interpreters with respect to their use of terminology. In **Chapter 2**, I reviewed literature on frameworks used in interpreting and translation evaluation with special focus on error classification schemes. However, none of the existing typologies could alone fully cater to the specific needs of this study. In **Chapter 4**, to address the need of evaluating the trainee interpreters' terminology accuracy in simultaneous interpreting, I suggested a scoring system to assess terminological accuracy in real communication. The scoring system incorporated **six error types** (Incorrect Term, Omission, Inappropriate Collocation, Grammatical Error, Pronunciation Error and Semantic Error) and **two levels of error weight** (1-minor error, 0-serious error).

The third objective of this study was to evaluate the usefulness of automatic term extractors for interpreters' preparation. More specifically, this study aimed to evaluate which term extractor (TTC, Syllabs or Teaboat) has consistently higher precision rates in term extraction from comparable texts (of different sizes) in two languages (English & Chinese) and on two topics (FR & SM). The evaluators were the end users of the automatically-generated termlists, i.e. the trainee interpreters who participated in this research. **Chapter 5** reported the evaluation results of the term extractors. The accuracy of the three term extractors was not perfect, as their precision rates (measured by the number of relevant terms out of the total number of automatically extracted terms) ranged from 27% on smaller corpora to 88.2% on bigger corpora for English, 24%-31% on Chinese. Among the three existing tools, Syllabs was more reliable in generating more relevant terms in English. All the three tools perform less satisfactory in generating relevant terms in Chinese. Based on the evaluation results, a single term extractor (Syllabs) with comparatively better performance was selected to be used in the empirical study of this research.

The fourth objective was to investigate the impact of using the proposed preparation procedure and the corpus tools on simultaneously interpreting technical speeches. This broad objective was narrowed down to four specific objectives: a) to observe the effect of **only using automatically-generated termlists** during preparation on the students' SI performance, b) to observe the effect of u**sing both automatically-generated termlists and the concordancer** on the students' SI performance, c) to investigate whether the use of the corpus tools made the trainee interpreters' preparation easier and more efficient, and d) to observe the role of term accuracy in SI performance on specialised speeches.

First, the results (in **Chapter 6**) showed a consistent strong correlation between term accuracy and SI performance on both topics (FR & SM) in both language directions (en-zh & zh-en) and among all the groups (control & test). It proved the widespread consensus that term accuracy does play an important role in SI performance on scientific/technical speeches.

Second, the results (in **Chapter 6**) showed that **only using automatically-generated termlists** within **limited preparation time (3 days)** did not have significant effect on increasing term accuracy in SI tasks (the average term accuracy score was increased by 4.7%). Two months after the SI tasks, the group who used the auto-lists even had significantly lower post-task recall of terms than the group without using any tools (by an average of 17%). In terms of participants' real preparation time, when limited preparation time (3 days) was allowed, the group **using the automatic term extractor** spent more or less the same preparation time as the group without using any tool (almost 10 hours).

On the other hand, **using both automatically-generated termlists and the concordancer** with **ample preparation time (9 days)** led to significantly better term accuracy in the SI tasks (the average term accuracy score was increased by 7.5%). Two months after the SI tasks, the group who used both tools had significantly better post-task recall of terms than the group without using any tools (by an average of 18%). When ample preparation time (9 days) was given, the group **using both automatically-generated termlists and the concordancer** with **ample preparation time** spent significantly less preparation time than the group without using any tool. Using both tools helped to save about 17% (3 hours) of the students' average preparation time.

Furthermore, we were able to observe (in **Chapter 7**) how an increased level of term density in the source speech and working into the B language affected the trainee interpreters' performances, such as deteriorating term accuracy in interpretations. However, we also observed that using both auto-lists and the concordancer helped to mitigate the detrimental effects of increased term density in the source speech and working into the B language on the numbers of serious errors and omission errors in interpretations.

In summary, **using both automatically-generated termlists and the concordancer** within **ample preparation time** helped the trainee interpreters achieve comparatively better results. It generally helped improve the trainee interpreters' terminological performance during simultaneous interpreting by increasing term accuracy scores by 7.5% and reducing the number of omission errors by 9.3%. We should also note that the terminology preparation (through using both the term extractor and the concordancer) is not a "magical cure" for all errors. Our data shows that the preparation procedure only helped to improve the students' holistic SI scores by 2.8% (but not yielding any statistical significance).

8.2 Contributions

8.2.1 User's investigation

Much of the research to date on using corpora and corpus tools to assist the interpreters' terminology preparation has focused primarily on conceptual ideas and functions of specialised tools, for instance, Rütten (2003), Fantinuoli (2006) and

Gorjanc (2009), however little attention has been given to the actual user experience from the perspective of the trainee interpreters.

This study has demonstrated not only a number of tools and functionalities available, but the way in which these functionalities are combined to be used for a special user group: the trainee interpreters.

This study has been one of the very few empirical studies to examine whether using corpora and corpus tools in preparation may have any impact on simultaneous interpreting performances. The results suggested that the preparation procedure using both the automatically-generated termlists (by Syllabs Tools) and the concordancer (Sketch Engine) helped improve the trainee interpreters terminological accuracy during simultaneous interpreting.

All the tools and functionalities used in the experiments have clear and easy-to-operate user interface, and they do not require any programming skills. However, new users of corpora or corpus tools (such as the trainee interpreters in this study) need prior training and may experience a steep learning curve before the tools could be used efficiently. Future studies on user experience could focus on ergonomic issues of using several tools/functionalities together (i.e. web crawler, term extractor, concordance and term management tool) in the interpreters' preparation, especially among the new users.

8.2.2 Evaluating term performance

This study demonstrated a method to evaluate performance of interpreters with respect to their use of terminology. The scoring system with hierarchical organisation provided a structured approach and explicit guidelines to conduct evaluation of term performance in interpreting, and it has proved to be usable and useful for interpreting trainers. The same scoring system could be applied in similar researches in the future.

8.2.3 Implications

For interpreters and interpreter trainers, this study offered practical guidance on the interpreters' terminology preparation for technical meetings. This study demonstrated that training on terminology preparation by using comparable corpora could be a useful supplement to the already existing professional interpreting training. It is important for both trainees and trainers to be aware that corpora and corpus tools, when used properly, can assist the interpreters' terminology preparation and achieve an enhanced performance.

On the computational side, this paper revisited concepts and technical challenges in terminology extraction, such as tokenization errors of Chinese texts and flexible term patterns in Chinese. The paper also provides room for further investigation into integration of term extraction and the practice of interpreter training.

8.3 Limitations

Despite these encouraging results, I am aware of the limitations of this study.

8.3.1 Small sample sizes

The scope of this PhD study was limited. There were inevitable constrains on the availability of resources, including participants, time and institutional arrangements. For example, this study only managed to recruit 22 participants in two consecutive academic years (12 students were recruited in 2012-2013 and 10 students were recruited in 2013-2014). The research findings were therefore preliminary, and replication studies with sufficiently larger size of samples are needed to confirm the research findings.

8.3.2 Limited range of experiment conditions

An interesting control variable, namely preparation time (3 or 9 days) was introduced, but it made the research design complicated at the same time. Due to the limited scope of the study, we did not have the capacity to involve the participants in two other possible experiment conditions. In order to see the interaction between preparation time and using different tools on the preparation results, the two other conditions will be looked at in the next stage of this research, i.e. a) using only automatically-generated lists in ample preparation time and b) using both automatically-generated lists and the concordancer within limited preparation time.

8.3.3 Control variables

In the experiments, there was still a compounding effect of several control variables which might affect the results of participants' performance, i.e. variation in term density and delivery speed of the four experimental speeches. For example, the English speech on Fast Reactors had lower term density (9t/m) than the other three speeches (14-15t/m). In addition, the two English speeches were delivered at slightly slower speeds (80-82 w/m) than the two Chinese speeches (167-180 c/m). Being aware of that, I did not treat the four experiment speeches as comparable speeches in data analysis. In future studies, I should make further efforts to control variation in term density and delivery speed in order to minimize the compounding effect of control variables.

8.3.4 Subjectivity of assessment

In order to avoid biased judgement of holistic SI performance, I assessed the participants' holistic interpretation before annotating their terminological errors.

Furthermore, I was aware that in order to avoid subjectivity, the data should ideally be annotated blind. However, out of necessity, I carried out every step of the process, from data collection, transcription to data analysis. Two second markers were asked to check whether I had reasonable and consistent judgement, but the second marking was not done in blind. Using blind annotation in the future can help to reduce subjectivity.

8.4 Future work

I have identified several topics which can be taken further in future works. First, I should continue to search for possible tools to assist term extraction and term acquisition for interpreters, and consider the integration of the tools in the already used pipeline. I could also investigate to what extent the amount of preparation time may affect the preparation results.

The current research results only apply to the trainee interpreters. I plan to apply the same experimental design to examine the preparation results of professional interpreters, and also look at professional interpreters' feedback on using the corpus tools.

This research primarily focuses on the effect of different preparation procedures. In other words, it is a product-oriented study. I can apply other theoretical frameworks (for example, from cognitive psychology and psycholinguistics) to learn more about interpreters' terminology acquisition process and preparation results (both process and product oriented researches).

In short, this thesis offers interesting findings and implications for interpreting training, as well as offering directions for researching both conference interpreting and computational terminology.

Appendix A

Appendix A1: The English speech script (FR)

Good morning, ladies and gentlemen. I am Shunsuke Kondo, chairman of **the Japan Atomic Energy Commission**. I am very happy to attend this year's International Conference on **Fast Reactors** and Related **Fuel Cycles** here in Paris. Today I will mainly talk about lessons learned from **Fukushima nuclear accident** and the **Fast Reactor** R&D strategies in Japan.

A huge earthquake hit in the northern part of Japan on March 11, 2011 and an induced huge tsunami attacked the **Fukushima Dai-ichi Nuclear Power station**. This earthquake did not cause severe damage to the **reactors**; however, the induced tsunami swamped the **back-up diesel generators** and caused a **station black-out**. **The emergency core cooling system** then failed. Finally, three **reactors/ melted down** and **radioactive materials** have been released into the environment.

According to the **IAEA**'s investigation committee, the Japanese government and **Tokyo Electric Power Company (TEPCO)** failed to prevent such disaster from happening because both parties had been reluctant to invest time, effort and money in protecting against a natural disaster considered unlikely: they had been overly confident that events beyond the scope of their assumptions would not occur.

There are mainly two important lessons we have learned from the **Fukushima accident** among many others. Firstly, when designing **Nuclear power plants**, the impact of external events such as earthquakes, volcanoes, storms, tides and tsunamis should be evaluated and reflected in the design of the **plants**. **Nuclear power plants** should be able to withstand greater than **design-basis** natural disasters.

Secondly, **emergency preparedness** should by all means be improved. Proper **countermeasures** should be conducted to keep **off-site power supply** and **cooling capabilities** during emergencies in order to prevent the **station blackout** and **core disruptive accidents**.

Based on the lessons learned from **Fukushima accident**, **the Japanese Nuclear Regulatory Authority** is currently reviewing new safety rules. All the **operators** will have to comply with the new rules and upgrade their **plants**. The new safety rules for **fast reactors** and **fuel cycle facilities** based on the same principle will be published later this year.

In terms of **Fast Breeder Reactors,** they can generate electricity while producing more **fuel** than they **consume**. In Japan, where natural resources are scarce, we are therefore promoting **fast breeder reactors** as the major **nuclear power** source for the future.

In Japan, there are two **fast reactors**, **Joyo** and **Monju**. **Joyo nuclear reactor**, Japan's first **fast reactor** is an **experimental sodium-cooled fast reactor**, which had been **operating** at 50 **MWt** since 1977 and boosted to 140 **MWt** in 2003. It has been **shut down** since 2007 due to damage to some **core components**.

Monju nuclear reactor is a **prototype fast reactor**, which has **three coolant loops**, uses **MOX fuel** and is able to produce 714 **MWt**. It started **operation** in 1994, but a **sodium leakage** occurred in its **secondary heat transfer system** during **performance tests** in 1995. The **operator**'s inappropriate information management in the face of this event led to a loss of public trust in the **operator.** It also caused **Monju** to **shut down** for almost 15 years.

With lessons learned from the **sodium leakage accident**, **Monju** carried out a lot of **modification** work in 2007. It restarted its **commissioning test operation** in May 2010 with a view to completing it in 2013. Unfortunately, Japan suspended the project after the **nuclear accident** in 2011.

We are currently reviewing our strategy for **sodium-cooled fast reactors**' R&D to make it compatible with the new **safety regulation**. This strategy will be formulated within a year. I think the major emphasis of **sodium-cooled fast reactors**' R&D should be on **full power operation** of **Monju**, **passive safety systems** and **severe accident management procedures**.

In terms of international cooperation, Japan has been working together with members of the **Generation IV International Forum (GIF)** and **the International Project on Innovative Nuclear Reactors and Fuel Cycles (INPRO)** to lay the groundwork for the **fourth generation fast reactor system**.

Japan will also make utmost effort to share its experience and lessons learned from the severe accident at Fukushima with the world, as it is the responsibility of Japanese nuclear community to contribute to strengthening **nuclear safety** worldwide.

To conclude, there have been so far three major **nuclear accidents** in the history of **civil nuclear power generation**: **Three Mile Island accident** in USA,

Chernobyl disaster in Ukraine and the recent **Fukushima accident** in Japan. As the previous two **accidents**, **Fukushima** offers both challenges and opportunities. **Nuclear power**'s contribution to electricity in Japan will probably not return to the level before 2011 anytime soon. But in the long run, we hope to turn the lessons we learned into valuable momentum to improve **nuclear safety** for the whole country. Thank you!

Appendix A2: The Chinese speech script (FR)

女士们、先生们：大家早晨好！非常高兴参加此次第十三届国际**快堆**及**燃料循环**研讨会。我是**中国原子能科学院**的徐銤，非常期待和各国专家就**快堆**的研发交换意见。

多年来，各国的实践证明：**核能**是安全、清洁、高效的能源。**第二代核电站**和**第三代核电站**已经在全世界范围内广泛地应用。截止至 2010 年底，全世界 31 个国家共运行 **439 座核电站，总装机容量**达到 **372 千兆瓦**，约占世界发电总量的 17%。

不少国家也正在积极研发**第四代核电站**。**第四代核电站**的安全性更高、经济性更好。**第四代核电站**主要是以**快中子堆**为主，**一至三代的反应堆**呢，主要属于**热中子反应堆**。**热中子反应堆**主要依靠**铀 235** 作为**燃料**。但很可惜，**铀 235** 的储量是十分有限的，仅占**天然铀**的**储量**的 0.7%。而**铀 238** 呢，占到占到 99.3%，却很难被有效的利用。

第四代反应堆，也就是我们刚才解释了，这种**快中子堆**，又称作**快中子增殖反应堆**，主要采用的是**铀**和**钚**作为**燃料**。那么**快堆**呢，有以下两点主要的优势：首先第一，**快堆**可以充分利用**铀 238**，可以大幅地提高**铀资源**的利用率。和以前的**压水堆**相比较，**铀资源**的利用率可以由 1%左右提高到 60%，甚至 70%。

那么，**快堆**的工作原理是这样的：首先，由**快中子**引起**核裂变链式反应。燃料**当中的**铀 238** 就被转变成了**钚 239**，**钚 239** 为，嗯，**易裂变核素**，这样一来呢，就实现了**核燃料**的**增殖**。**核裂变**所产生的**热量**可以通过**液化钠**带出**堆芯**，然后再通过**热交换器**和**蒸汽发生器，**产生**高温/高压的/蒸汽**，最终推动**汽轮机**发电。这是快堆的第一点优势，就是它可以大幅度地提高**铀资源**的利用率。

那么，第二点优势就是：**快堆**可以通过**嬗变**大幅度地减少**长寿命/**、**高放射性/核废料**的储量。所谓**嬗变，**就是通过**核反应，**将一种**核素/转换**为另一种**核素**的过程。**快堆**可以将**压水堆**所产生的这种**长寿命的核废料/嬗变**为**短寿命的核素**，这样呢，就可以使**放射性废物**对环境的影响从百万年降低到几百年，另外呢，需要最终**处置**的**核废料**的量也可以大大地减少。

基于以上这两点显著的优势，中国呢，是不断地推进快堆的研，嗯，**快堆**技术研发和快堆的工程建设，因为这对中国的**核能**可持续发展呢，是有着非常重要的意义的。

中国自上个世纪 60 年代开始研究**快堆**技术。那么，中国的**快堆**发展战略呢是一个三步走的战略：也就是第一步发展**实验快堆，**第二步发展**示范快堆，**第三步建设**大型先进快堆**。**中国实验快堆**是**国家“863”计划**的重大项目，是中国**自主设计/**、**建造/**、**调试运营**的第一座**快中子反应堆**，是我国**快堆**发展的第一步。

中国实验快堆的**热功率**为 **65 兆瓦**，**发电功率**为 **20 兆瓦。实验快堆**所采用的是**液态金属钠**为**冷却剂**，采用的是**钠-钠-水三回路**设计。这座**快堆**呢，是目前世界上为数不多的具备有发电功能的**实验快堆**之一。那么这座**快堆**的**设置**也已经接近了**商用快堆**。这个项目呢，是由**国家科技部**主管**，**由**中国原子能科学研究院**具体实施。这座**实验快堆**就建在我们**中国原子能科学院**在北京房山的基地内。

下面让我们来看一下**中国实验快堆**的几个主要突出的**安全特性**：首先第一就是这座**实验快堆**无论是在**反应堆/运行的状态**下，还是**事故状态**下，都可以保证最大限度地排除**堆芯熔化**的可能性。第二，这座**快堆**的**一、二回路/冷却**使用的是**液态钠**。**钠**具有良好的**热传输性**，因此可以将**堆芯**的**热量**顺畅地带出**堆芯**。第三点，这座**反应堆**采取的是**非能动系统**，比如说这座**反应堆**采用了**非能动余热处理系统**，用于排除**反应堆**内的**余热**。

刚才这三点呢，是实验，**中国实验快堆**的几个很突出的，这个，**安全特性。**下面再让我们来看一下**中国实验快堆**的主要的建设历程。2000 年 5 月**中国实验快堆**的**核岛/浇灌/混凝土**；2005 年**堆本体**开始**安装**；2009 年首次**装料**；2010 年实现首次**核临界**；2011 年 7 月成功实现首次**并网发电**。

今后呢，我们将在**中国实验快堆**的基础上，积极地推进**大型商用示范快堆**的工程建设。并且呢，我们还将努力地发展**快堆**及其**闭式燃料循环**技术，从而最终达到**核能**在中国的可持续发展。为了达到这些目标，只靠中国本国的**快堆**建设经验是远远不够的。因此，我们希望同**国际原子能机构**及各成员更加紧密地合作，学习各国在**快堆**研发和建设方面的经验，从而提高核电技术在中国的发展。

Appendix A3: The English speech script (SM)

Good morning, ladies and gentlemen. My name is Sarah Dodd, Professor of Marine Geology from the Department of Earth Sciences of the University of Oxford. I am very happy to attend this Annual Workshop on the **mining/** of s**eabed mineral resources**, organised by **the International Seabed Authority** in Jamaica. In my talk, I will provide some updated background materials on two **marine mineral resources**, **cobalt-rich crusts** and **polymetallic sulphide deposits**. I will also speak about **exploration** and **mining** technologies for the two kinds of **deposits**. Then I'll make a few comments at the end on decision making in **mining** from a geologist's perspective.

A scientific revolution in the 1970 and 1980s transformed our views of the **earth's movements**. It significantly expanded our knowledge of **marine minerals**. Previously we only knew that the **ocean basins** are important sources of **manganese nodules**, **heavy metals,** for example, **tin** and **gold** and **gemstones**, especially **diamond**. Thanks to the scientific revolution, there is a newly recognised type of **marine mineral resource** called **polymetallic sulphides** and containing **copper, zinc, silver** and **gold** in varying amounts.

Polymetallic sulphide deposits are formed over thousands of years at sites along a **global active volcanic mountain range** on the **seafloor**. This **volcanic mountain range** extends through all the **ocean basins** of the world. Since 1979, **polymetallic sulphide deposits** have been found at **water depth**s of up to 3,700 **metres** and in a variety of **tectonic settings** on the **seafloor,** including **mid-ocean ridges** and also **seamounts.**

Another newly recognised type of **marine mineral resource** is the **cobalt-rich iron-manganese crusts** that are **deposited** over millions of years on the surface of **inactive underwater volcanoes**. Now these **crusts** occur throughout the global ocean on **seamounts**, on **ridges** and **plateaux**. **Crusts** are important as a potential resource primarily for **cobalt**, but also for **titanium, nickel, platinum, manganese** and other metals. The thickest and most **cobalt-rich crusts** occur at **depths** of about 800-2,500 **metres** under the sea.

Due to the high **concentration** of **base and precious metals**, **polymetallic sulphide deposits** and **cobalt-rich iron-manganese crusts** in the **seabed area** have recently attracted the interest of the international mining community.

The primary uses of **manganese**, **cobalt** and **nickel** are in the manufacture of **steel**. They are used to add **specific properties** to **steel**, such as **hardness, strength** and **resistance to corrosion**. **Cobalt** is also used in the electrical, communications, aerospace and tool manufacturing industries. **Nickel** is used additionally in chemical plants, petroleum refineries, electrical appliances and motor vehicles.

Marine mining of the two **mineral deposits** appears to be feasible and enjoys some distinctive advantages. For example, it can be economically attractive, considering that the entire **mining system** is **portable** and can be moved from **site** to **site**. An investment in **mining systems** and **ships** would thus not be tied to a single location as is the case on land.

Scientific research on **sulphide deposits** and **crusts** is being carried out by academic and governmental institutions worldwide. **Exploration** requires sophisticated, multipurpose **research vessels**, using advanced technologies such as **deep-sea mapping equipment**, **remotely-operated vehicles**, **photographic and video systems** and **sampling/ and drilling devices**. Leading countries in this field are Australia, Canada, Germany, Japan, the Russia Federation, the United States and the United Kingdom.

Although **exploration** technologies are fairly well-developed, at least in terms of **scientific exploration**, **mining** technologies have been around for a very short time. Many of them are just conceptual ideas. Now I will speak about **mining** technologies for the two kinds of **seabed deposits** individually.

Research and mining, technology for **mining/ crusts** are only in their infancy. To locate specific **sites** for **exploitable crusts**, prospective **miners** will first have to develop detailed **maps** of **crust deposits** and **small-scale seamount topography**. Once **sampling sites** have been chosen, **miners** can **deploy** / **dredge hauls** and **core samplers** to **identify/ crusts,** their **sediment types** and **distribution**. So far, only a very few **crust deposits** in the **Pacific Ocean** have been **mapped** and **sampled**.

Crust mining is technically difficult because **crusts** are **attached** to **substrate rocks**. For successful **crust mining**, it is essential to **recover** the **crusts** without **collecting/ substrate rocks**, which would significantly **dilute** the **ore grade**. In terms of **crust recovery,** I personally would prefer the following method**;** we can use a **bottom-crawling vehicle** equipped with **articulated cutters** to **fragment** the **crusts**.

In this way, we can minimize the amount of **substrate rock/ collected**. Then the

crusts can be **lifted** to the **surface vessel** by a **hydraulic-pipe lift system**. Some other institutions have also suggested some new and innovative systems for **crust recovery**. All the suggestions offer promise but need to be further developed.

Now let's take a brief look at **mining** technologies for **polymetallic sulphide deposits**. Although the technology for **recovering/ seafloor polymetallic sulphide** does not exist, theoretically we could adapt some well-developed schemes for **recover**ing/ **diamonds** in **shallow offshore waters** and some schemes for **recover**ing/ **manganese nodules** in **deep ocean basins** for future **sulphide mining**.

Some researchers have already envisaged a **continuous recovery system** using **rotating cutting heads** to **extract** and **grind** the desired **minerals** from the **seabed**, before **lift**ing all the **slurry** to **mining vessels** and then **transport**ing it to a **processing plant.** Again, the suggestions however have not yet all been fully tested.

As a geologist, I have some experience and insider knowledge of **ocean mining**. I should say that decision making on whether to **mine** these two types of **mineral deposits** is not just made by collecting a few **samples** and analysing them. A great deal of data, as well as **risk assessment** will be involved before a decision is made.

In conclusion, **deep sea mineral/ exploration** and **mining** is innovative, exciting and forward looking. However, we are still at an early stage in the **exploration/** of **the seabed minerals**. There are still many unanswered questions. It is necessary for us all; from academic and governmental institutions worldwide as well as from private sector mining companies, to collaborate in our efforts to fill the information gap concerning various aspects of the **seabed minerals**. The work should be coordinated by **the International Seabed Authority**, in order to assure compliance with regulations and to avoid unnecessary duplication of effort. Thank you!

Appendix A4: The Chinese speech script (SM)

各位女士们、先生们，大家早上好。我是来自**中国大洋矿产资源研究开发协会**的总工程师周宁。非常高兴参加**国际海底管理局**组织的这次国际**海底资源/勘探**专题研讨会。我想利用这个机会为各位代表简单介绍的一下中国在**国际海底区域**内针对三种主要**矿产资源**这所进行的**勘探**活动。这三种**矿产资源**呢，包括：**多金属结核**、**多金属硫化物**和**富钴结壳**。

大家都知道，**深海**是地球上最后一片尚未开发的区域，海洋深处蕴藏着大量的**矿产资源**，很多**矿产资源**都集中在国家，恩，各国管辖的区域之外，那么这一部分区域呢，是**国际海底区域**，它的面积约为 **2．5 亿平方公里**，约占地球表面积的 49％。根据**《联合国海洋法公约》**的具体规定，这一区域以及它所蕴藏的丰富的资源是全人类共同的财产，由**国际海底管理局**代表全人类进行管理，对**"区域"**内的**矿产资源**进行**勘探**和开发。

近年来，**国际海底管理局**的核心工作就是为**国际海底**的三种新**资源**制定**勘探规章**。**管理局**于 2012 年审议通过了**《"区域"内富钴结壳探矿和勘探规章》**，在此之前，**管理局**分别于 2001 年和 2010 年审议通过了关于**多金属结核**和**多金属硫化物**的**勘探规章**。这三个**规章**的设立为开发**"区域"**内的这三种重要**矿产资源**提供了法律依据。

下面呢，让我们再来看看中国在**"区域"**内对以上三种**资源**的**勘探**情况。早在 20 世纪 80 年代，我国就开始在**太平洋**的**国际海底区域**，对**多金属结核**进行了系统科学调查。1991 年,**中国大洋协会**在**联合国**正式登记为"**国际海底先驱投资者**"，获得了 **15 万平方公里**的**多金属结核/开辟区**。目前获得**先驱投资者**地位的国家还有日本、韩国、印度、英国、法国、德国、俄罗斯和美国。中国呢，是第五大投资国。

90 年代，**中国大洋协会**在 **15 万平方公里**的**开辟区**进行了 10 个航次的调查研究工作，最终，优选出了 **7.5 万平方公里**的**多金属结核区**。

2001 年，**海管局**与**中国大洋协会**正式签订了**《专属勘探合同》**。这个合同的签订，就意味着中国在**东太平洋**上拥有面积达 **7.5 万平方公里**的**多金属结核/专属勘探矿区**。并且在今后进入**商业开采**的时候，中国也将具有**优先开采权**。

2011 年，**管理局**又核准了**中国大洋协会**申请的**多金属硫化物/专属勘探矿区，**这个**专属勘探区**是位于**西南印度洋**面积为 **1 万平方公里**的一片**海域**内。这是中国获得的第二块享有**专属勘探权**和**优先开采权**的**矿区**。

另外，2012 年 7 月，**中国大洋协会**根据刚刚通过的**《"区域"内富钴结壳探矿和勘探规章》，**率先向**管理局**提交了一份申请书，要求进行**富钴结壳 /勘探**工作。申请区域位于**西太平洋**。**西太平洋/ 海山区域**是目前所知的**富钴结壳资源**分布最为**密集**的地区。这是**中国大洋协会**在过去 15 年的调查基础上选定的。这些调查工作主要由**"海洋四号"**、**"大洋一号"**和**"海洋六号"/科考船**完成的。调查所使用的技术手段包括：**多波束测深系统**，**海底摄像**、**钻心取样**，等等。

今年，也就是 2013 年，**海管局**的第 19 届大会将审议并决定是否核准上述申请。如获核准，中国将获得面积为 **3000 平方公里**的**富钴结壳资源/专属勘探矿区**。这将是我国在国际海底领域取得的又一项重大成就。

与**大洋资源**/**勘探**同等重要的另一项工作，就是**评估/勘探**和**采矿**对**大洋环境**所带来影响。应该说呢，**海底区域**开发要想不造成任何的污染，是相当困难的。**勘探**和**开采**所采用的各种技术手段，比如**机械**、**化学**、**电解**、**激光**等技术是会造成对**海洋环境**的污染的。另外，许多**矿石**本身是含有**放射性元素**或**重金属的**（如**铬**、**汞**、**铅**、**镍** 等），在**矿石/ 开采**或**碎裂**之后，有毒有害物质呢，会大量扩散，从而造成对周围**海域**的**环境**破坏。

因此，国际社会对**深海采矿**引起的环境问题一直都很关注。我们**大洋协会**的工作的基本原则就是在保护**海洋环境**的前提下，开发**国际海底区域**内的**资源**。**大洋环境**调查也是我们在**专属勘探区**的一项重要的工作。

中国大洋协会目前正使用多种设备和手段，对**勘探**区域的**环境基线**进行综合性调查，所谓**环境基线**研究就是指：在**采矿**活动尚未开始前，**收集**尽可能多的环境资料，**测量**出**海洋环境**中**生物**、**化学**、**地质**和**物理要素**的一些基本数据，以便确定**基线**，用于同**采矿**开始之后的状况进行比较。对**环境基线**的调查可以为帮助我们准确地评价人类活动对**海底生态系统**的影响。同时呢，我们还将调查相关**海域**中的**海底栖息生物**、**浮游生物**、以及**微生物**的**种类**和**分布**情况，初步**评估/海底采矿**对海洋**生物多样性**有哪些潜在的威胁。

以上呢，就是中国在**国际海底区域**内所进行的主要**勘探**活动。我们**中国大洋协会**愿意继续参加**海管局**今后所开展的各类科研和**勘探**活动，并作出积极

贡献。最后，我代表中国代表团向东道国牙买加所给予各国代表团的盛情接待表示衷心的感谢。谢谢各位。

Appendix A5: Quiz of terms (FR & SM)

The following is a list of terms about "fast breeder reactors". Please write down the Chinese translation beside each of them and provide a simple definition of the term in Chinese.

	Term in English	Chinese translation	Definition in Chinese
e.g .	coolant	冷却剂	由化学物质组成，用于冷却反应堆。
1	sodium-cooled fast neutron breeder reactor		
2	plutonium		
3	depleted uranium		
4	loop		
5	fission chain reaction		
6	core meltdown		
7	initial criticality		
8	decay heat removal		
9	reactor containment vessel		
10	steam generator		
11	heat exchanger		
12	minor actinides		

13	spent fuel		
14	closed fuel cycle		
15	cold shutdown		

The following is a list of terms about “deep seabed minerals”. Please write down the Chinese translation beside each of them and provide a simple definition of the term in Chinese.

	Term in English	Chinese translation	Definition in Chinese
e.g .	nickel	镍	一种化学元素。
1	polymetallic sulphide(s)		
2	cobalt-rich ferromanganese crust		
3	manganese nodule		
4	International Seabed Authority		
5	core sampler		
6	bottom-crawling vehicle		
7	sediment		
8	seamount		
9	research cruise		
10	ocean basin		
11	topography		
12	baseline study		
13	substrate rock		

14	cerium		
15	hydraulic-pipe lift system		

Appendix A6: The marking criteria applied to evaluate the participants' holistic SI performance

Marking Criteria in the MA Interpreting Final Exams
(University of Leeds)

70- A solid performance to professional standard
62- A promising performance showing good techniques
50- An adequate performance showing basic skills
Bilateral, Consecutive and Simultaneous Interpreting
72% just a distinction
To achieve 70% or higher, a student's interpretation should:

- show a very high degree of reliability in relaying meaning
- be entirely coherent as discourse
- show command of appropriate TL expression
- achieve a standard of presentation which demonstrates mastery of the skills involved in keeping pace and addressing an audience

68% almost distinction
65% clearly a merit
62% just a merit
To achieve 60%, a student's interpretation should:

- relay meaning with few distortions and few unwarranted omissions or additions
- be coherent as discourse
- conform generally to the TL norms in terms of lexis, syntax and idiom (marginal underperformance on this criterion may be condoned where performance on the three other criteria is suitably high)
- achieve a standard of presentation which is generally successful in keeping pace and addressing an audience

58% almost a merit
55% clearly pass
52% just a pass
To achieve 50%, a student's interpretation should:

- relay meaning without systematic distortion and without major unwarranted omissions or additions

- be mostly coherent as discourse
- achieve a standard of TL expression which does not impede communication to a significant extent
- achieve a standard of presentation which shows some evidence of ability to keep pace and address an audience

BELOW THE PASS MARK
48% almost a pass
To achieve 40%, a student's interpretation should:

- relay the basic meaning
- achieve a basic level of coherence
- demonstrate a standard of TL expression which does not fully impede communication
- achieve a minimum standard of presentation

To achieve 30%, a student's interpretation should:

- relay a gist of the original meaning
- be at least minimally coherent
- demonstrate a standard of TL expression which makes some sense

25%, a student's interpretation should:

- relay elements of the original meaning
- include some coherent discourse
- convey some ideas

Appendix A7: The scoring system adopted to evaluate the participants' terminological performance

Error category	Score	Scoring Criteria
"Incorrect Term" (IT)	**2**	The interpretation of the term is acceptable in the TL.
	IT-1	The term is interpreted inaccurately; it only slightly distorts the intended meaning.
	IT-0	The inaccuracy results in a substantial loss or change of the intended meaning.
"Omission" (OM)	**2**	No omission, or the omission of the term doesn't affect the original meaning.
	OM-1	The term is omitted, which results in slightly altered meaning or loss of meaning. The gist of what was said was maintained.
	OM-0	The omission results in a significant loss or change in meaning.
"Inappropriate Collocation" (IC)	**2**	The collocation use is appropriate in the TL.
	IC-1	The interpreter uses an inappropriate collocation of the term. But it does not affect the understanding of the message among the TT audience.
	IC-0	The collocation of the term is unacceptable in TL. The misuse of the collocation causes serious confusion to the TT audience.
"Grammatical Error" (GE)	**2**	No grammatical error.
	GE-1	There is an ungrammatical use of a term. But it doesn't quite affect the understanding of the message among the TT audience.
	GE-0	The misuse of tense or agreement causes serious confusion to the TT audience.
"Pronunciation Error" (PE)	**2**	No pronunciation error is made.
	PE-1	The term is mispronounced, but the audience can still understand what is said without making too much effort.
	PE-0	The term is mispronounced. The mispronunciation causes serious confusion to the TT audience.
"Semantic Error" (SE)	**2**	No semantic error is made.
	SE-1	The term itself is interpreted correctly, but the overall message of a larger unit is not quite the same thing as the original. The gist of the message is retained though.
	SE-0	Even though the term is interpreted acceptably, the interpretation of the larger unit has a considerable difference in meaning from the original. The interpreter makes up something on the basis of some part of the text.

Bibliography

Ahmad, K., Davies, A., Fulford, H., and Rogers, M. 1994. What is a Term? The semi-automatic extraction of terms from text. In: Snell-Hornby, M., Pöchhacker, F., and Kaindl, K. (eds). *Translation studies: An interdiscipline*. Amsterdam: John Benjamins Publishing Company, pp. 267–278.

Aitchison, J. 1994. *Words in the mind: An introduction to the mental lexicon (2nd Edition)*. Oxford: Blackwell

Aitchison, J. 2012. *Words in the mind: An introduction to the mental lexicon (4th Edition)*. Chichester: Wiley-Blackwell.

Alcina, A. 2009. Teaching and learning terminology: New strategies and methods. *Terminology* **15** (1), pp. 1-9.

Allard, M. G. P. 2012. *Managing terminology for translation using translation environment tools: Towards a definition of best practices.* PhD dissertation, University of Ottawa.

Alonso Bacigalupe, L. 1999. Sobre la doble recepción del mensaje y la preparación del texto: Resultados de un estudio experimental en IS. In: Álvarez Lugrís, A. and Fernández Ocampo, A. eds. *Anovar/Anosar: Estudios de traducción e interpretación, Vol. II.* Vigo: Servicio de Publicacións, Universidad de Vigo, pp. 11–25.

Alonso, A., Blancafort, H., De Groc, C., Million, C. and Williams, G. 2012. METRICC: Harnessing comparable corpora for multilingual lexicon development. In: *Proceedings of 15th EURALEX International Congress, Oslo.* [Online]. [Accessed 22 June 2015]. Available from: https://halshs.archives-ouvertes.fr/halshs-00725224/document

Altman, J. 1994. Error analysis in the teaching of simultaneous interpretation: a pilot study. In: Lambert, S. and Moser-Mercer. B. eds. *Bridging the gap: Empirical research in simultaneous interpretation.* Amsterdam: John Benjamins, pp. 25-38.

Anthony, L. 2004. AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. In: *Proceedings of IWLeL 2004: An Interactive Workshop on Language eLearning.* [Online]. [Accessed 22 June 2015]. Available from: http://www.laurenceanthony.net/research/iwlel_2004_anthony_antconc.pdf

Anthony, L. 2007. *Antconc 3.2.1: A free text analysis software.* [Online]. [Accessed 22 June 2015]. Available from: http://www.antlab.sci.waseda.ac.jp/

Aston , G., Bernardini, S. and Stewart, D. 2004. eds. *Corpora and language learners*. Amsterdam: John Benjamins.

Barik, H. A. 1975. Simultaneous interpretation: Qualitative and linguistic data. *Language and Speech.* 18, pp. 272-298.

Barik, H. A. 1994. A description of various types of omissions, additions and errors of translation encountered in simultaneous interpretation. In: Lambert, S. and Moser-Mercer, B. eds. *Bridging the gap: Empirical research in simultaneous interpretation.* Amsterdam: John Benjamins, pp. 121-138.

Baroni, M. and Bernardini, S., 2004. Bootcat: bootstrapping corpora and terms from the web. In: *Proc. of LREC2004*. Lisbon.

Baroni, M., Chantree, F., Kilgarriff, A. and Sharoff, S. 2008. Cleaneval: a competition for cleaning web pages." In: *Proc. of the Sixth Language Resources and Evaluation Conference, LREC 2008*, Marrakech.

Baroni, M., Kilgarriff, A., Pomikálek, J., Rychlý, P. 2006. WebBootCaT: Instant domain-specific corpora to support human translators. In: *Proceedings of EAMT 2006*, pp. 247-252.

Berruto, G. 1974. *La sociolinguistica*. Bologna: Zanichelli.

Bialystok, E. 1994. Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition.* **16**, pp. 157–168.

Blancafort, H., Bouvier, F., Daille, B., Heid, U. and Ramm, A. 2013. TTC web platform: from corpus compilation to bilingual terminologies for MT and CAT tools. In: *Proc. of Conference Futures in Technologies for Translation (TRALOGY II),* Paris.

Botley, S., Glass, J., McEnery, T. and Wilson, A. eds. 1996. *Proceedings of Teaching and Language Corpora 1996. Technical Paper 9.* University Centre for Computer Corpus Research on Language, Lancaster.

Bower, G. H. and Winzenz, D. 1970. Comparison of associative learning strategies. *Psychonomic Science.* **20**, pp. 119-120.

Bowker, L. 2003. Teaching translation technology: Towards an integrated approach. *Tradução e Comunicação.* **12**, pp. 65-79.

Bowker, L. and Pearson, J. 2002. *Working with specialised language: A practical guide to using corpora.* London/New York: Routledge.

Braun, S. 2007. Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL.* **19** (3): 307-328.

Bühler, H. 1986. Linguistic (semantic) and extralinguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua* **5** (4), pp. 231-235.

Cabré, M. T. 1996. Terminology today. In: SOMERS, H. ed. *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager.* Amserdam/Philadelphia: John Benjamins

Cabré, M. T. 1998. *Terminology: Theory, methods and applications.* Amsterdam/Philadelphia: John Benjamins

Chang, C. and Schallert, D. L. 2007. The impact of directionality on Chinese/English simultaneous interpreting. *Interpreting.* **9** (2), pp. 137-176.

Christ, O. 1994. A modular and flexible architecture for an integrated corpus query System. In: *COMPLEX'94*, Budapest, pp. 1-10.

Cokely, D. 1992. *Interpretation: A sociolinguistic model*. Burtonsville, MD: Linstok

Press.

Costa, H., Paster, G. C. and Durán-Muñoz, I. 2014. A comparative user evaluation of terminology management tools for interpreters. In: *Proceedings of the 4th International workshop on computational terminology*. Dublin, pp. 68-76. [Online]. [Accessed 22 June 2015]. Available from: http://www.aclweb.org/anthology/W14-4809

Craik, F. and Lockhart, R. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour.* **11**, pp. 671-684.

Craik, F. I. M. and Lockhart, R. S. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*. **11**, pp. 671-684.

Daille, B. 2005. Variations and application-oriented terminology engineering. *Terminology* **11**(1), pp. 181–197.

Daille, B. 2012. Building Bilingual Terminologies from Comparable Corpora: The TTC TermSuite. In: *Proc. of 5th Workshop on Building and Using Comparable Corpora at LREC* 2012.

Dam, H. V. 1993 Text condensing in consecutive interpreting. In: Gambier Y. and Tommola, J. eds. *Translation and Knowledge*. Turku, Finland: University of Turku, pp. 297-313.

Darò, V., Lambert, S. and Fabbro, F. 1996. Conscious monitoring of attention during simultaneous interpretation. *Interpreting.* **1** (1), pp. 101-124.

de Bot, K. 2000. Simultaneous interpreting as language production. In: Englund Dimitrova, B. and Hyltenstam, K. eds. *Language processing and simultaneous interpreting*. Amsterdam/Philadelphia: John Benjamins, pp.65–88.

De Camp, J. and Zetzsche, J. 2014. A history of translation technology in the United States. In: Chan, S. ed. *Routledge Encyclopedia of translation technology*. New York: Routledge, pp. 375-391.

Déjean Le Féal, K. 1990. Some thoughts on the evaluation of simultaneous interpretation. In: BOWEN, D. A. M. ed. *Interpreting-Yesterday, Today and Tomorrow*. Binghamton (NY): SUNY, pp. 154-160.

Díaz-Galaz, S. 2012. La influencia del conocimiento previo en la interpretación simultánea de discursos especializados: Un estudio empírico. PhD dissertation, Universidad de Granada.

Díaz-Galaz, S., Padilla, P. and Teresa Bajo, M. 2015. The role of advance preparation in simultaneous interpreting. *Interpreting.* **17** (1), pp. 1-25.

Dunaevsky, C. 2015. *Term extraction and tools.* [Online]. [Accessed 22 June 2015]. Available from: http://termcoord.eu/useful-links/free-term-extractors/

Durán-Muñoz, I. 2012. Meeting translators' needs: translation-oriented terminological management and applications. *The Journal of Specialised Translation* **18**, pp. 77-92.

Eckersley, H. 2002. Systems for evaluating translation quality. In: *Multilingual Computing and Technology*. **3**(3), pp. 39-42.

Ellis, N. C. 1995. The psychology of foreign language vocabulary acquisition: Implications for CALL. *Computer Assisted Language Learning.* **8**(2-3), pp. 103-128.

Fantinuoli, C. 2006. Specialized corpora from the web and term extraction for simultaneous interpreters. In: Baroni, M. and Bernardini, S. eds. *WaCky! Working Papers on the Web as Corpus.* Bologna: Gedit, pp. 173–190.

Gile, D. 1995. *Basic concepts and models for interpreter and translator training*. Amsterdam/Philadelphia: John Benjamins.

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

Gorjanc, V. 2009. Terminology resources and terminological data management for medical interpreters. In: Dörte, A. and Pöllabauer, S. eds. *Spürst Du, wie der Bauch rauf-runter? Fachdolmetschen im Gesundheitsbereich*. München: Meidenbauer, pp. 85-95.

Gorodetsky, B. Y. 1990. Cognitive aspects of terminological phenomena. *Terminology and Knowledge Engineering.* (**Vol. 1**), pp. 117-123.

Haque, R., Penkale, S. and Way, A. 2014. Bilingual termbank creation via Log-likelihood comparison and phrase-based statistical machine translation. In: *Proceedings of the 4th International workshop on computational terminology.* Dublin, pp. 42-51. [Online]. [Accessed 22 June 2015]. Available from: http://www.aclweb.org/anthology/W14-4806

Hartley, A., Peng, G., Mason, I. and Perez, I. 2004. *Peer and self-assessment in conference interpreter training.* Subject Centre for Languages, Linguistics and Area Studies. The Higher Education Academy, UK.

Hashtroudi, S. 1983. Type of semantic elaboration and recall. *Memory and Cognition.* **11**, pp. 476-484.

Hu, K. and Tao, Q. 2013. The Chinese-English conference interpreting corpus: Uses and limitations. *Meta: Translators' Journal.* **58** (3), pp. 626-642.

ISO 1087-1:2000, *Terminology work — Vocabulary — Part 1: Theory and application*.

Jaatinen, H. and Jääskeläinen, R. 2006. Introducing IT in translator training: Experiences from the COLC project. In: Pym, A., Perekrestenko, A. and Starink, B. eds. *Translation technology and its teaching (with much mention of localization)*. Tarragona: Intercultural Studies Group, Universitat Rovira i Virgili, pp. 83-88.

Jaekel, G. 2000. Terminology management at Ericsson. In: Sprung, R. ed. *Translating into success: Cutting-edge strategies for going multilingual in a global age.* Amsterdam/Philadelphia: John Benjamins, pp. 159-171.

Jiang, H. 2013. The interpreter's glossary in simultaneous interpreting: A survey. *Interpreting*. **15**(1), pp. 74-93.

Kalina, S. 2005. Zur Dokumentation von Maßnahmen der Qualitätssicherung beim Konferenzdolmetschen. In: Heine, C., Schubert, K., Gerzymisch-Arbogast, H. eds. *Translation theory and methodology.* Tübingen: Jahrbuch Übersetzen und Dolmetschen, pp. 253-268.

Kalina, S. 2015. Preparation. In: Pöchhacker, F. (ed.). *Routledge Encyclopedia of Interpreting Studies*. London/New York: Routledge, pp.318-320.

Kilgariff, A. and Grefenstette, G. 2003. Introduction to the special issue on the web as corpus. *Computational Linguistics*. **29**, pp. 333-347.

Kopczynski, A. 1980. *Conference interpreting: Some linguistic and communicative problems*. Poznan: Adam Mickiewicz University Press.

Kopczynski, A. 1994. Quality in conference interpreting: some pragmatic problems. In: Snell-Hornby, M., Pöchhacker, F. & Kaindl, K. eds. *Translation studies-interdiscipline: Selected papers from the translation studies congress 1992.* Amsterdam/Philadelphia: John Benjamins, pp. 189-198.

Krippendorff, K. 2004. Reliability in content analysis: some common misconceptions and recommendations. *Human Communication Research* **30**(3), pp. 411-433.

Kurz, I. 1989. Conference interpreting-user expectations. In: *Coming of Age: Proceedings of the 30th annual conference of the American Translators Association.* Medford NJ: Learned Information, pp. 143-148.

Kurz, I. 1993. Conference interpretation: Expectations of different user groups. *The Interpreters' Newsletter*. **5**, pp. 13-21.

Kurz, I. 1994. What do different user groups expect from a conference interpreter? *The Jerome Quarterly*. **9**(2). pp. 3-7.

Kurz, I. 2001. Conference interpreting: Quality in the ears of the users. *Meta: Translators' Journal*. **46**, pp. 394-409.

Lafferty, J., McCallum, A. and Pereira, F. 2001. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: *Proceedings of ICML-01*. pp. 282-289.

Lamberger-Felber, H. 2001. Text-oriented research into interpreting. *Hermes, Journal of Linguistics* **26**, pp. 39-64.

Lamberger-Felber, H. 2003. Performance variability among conference interpreters: examples from a case study. In: Collados Aís, A., Fernández Sánchez, M. and Gile, D. eds. *La evaluación de la calidad en interpretación: investigación*. Granada: Comares, pp. 147-168.

Lee, Y. H. 2003. *Comparison of error frequency in simultaneous interpretation A to B and B to A (Korean-English)*. Unpublished DEA (pre-doctoral) thesis, University of Geneva, Ecole de Traduction et d'Interprétation.

Li, C. 2010. Coping strategies for fast delivery in simultaneous interpretation. *The Journal of Specialized Translaiton*. **13**, pp. 19-25.

Liang, M., Li, W. and Xu, J. 2010. *Using corpora: a practical coursebook.* Beijing: Foreign Langue Teaching and Research Press.

Livingston, S., Singer, B. and Abramson, T. 1994. Effectiveness compared: ASL interpretation versus transliteration. *Sign Language Studies*. **82**, pp. 1-53.

Mark, G. and Cattaruzza, L. 1995. User Surveys in simultaneous interpretation: a means of learning about quality and/or raising some reasonable doubts. In:

TOMMOLA, J. ed. *Topics in interpreting research.* Turku: University of Turku, pp. 37-49.

McEnery, A. & Xiao, Z. 2010. What corpora can offer in language teaching and learning. In: Hinkel, E. ed. *Handbook of research in second language teaching and learning (Vol. 2)*. London / New York: Routledge.

McEnery, T., Xiao, R. and Tono, Y. 2006. *Corpus-based language studies: An advanced resource book*. London: Routledge.

Mohammadi, M. 2013. *Meeting translators' terminological needs.* [Online]. [Accessed 22 June 2015]. Available from: http://linguistech.ca/display4353

Montero Martínez, S. and Faber, P. 2009. Terminological competence in translation. *Terminology* **15** (1), pp. 88-104.

Moser, P. 1996. Expectations of users of conference interpretation. *Interpreting **1**(2)*, pp. 145-178.

Moser-Mercer, B. 1992. Banking on terminology conference interpreters in the electronic age. *Meta: Translators' Journal*. 37(3), pp. 507–522.

Nagy, W. and Herman, P. 1987. Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In: McKeown, M. and Curtis, M. eds. *The nature of vocabulary acquisition.* Hillsdale NJ: Lawrence Erlbaum Associates, pp. 19-35.

Napier, J. 2002. *Sign language interpreting. Linguistic coping strategies.* England, UK: Douglas McLean.

Napier, J. 2004. Interpreting omissions: A new perspective. *Interpreting.* **6**(2), pp. 117-142.

Nation, P. 1993. Vocabulary size, growth and use. In: Schreuder, R. and Weltens, B. eds. *The bilingual lexicon*. Amsterdam: John Benjamins, pp. 115-134.

Ogden, C. K. and Richards, I. A. 1923. *The meaning of meaning*. Routledge, London.

Perfetti, C. 1985. *Reading ability.* New York: Oxford University Press.

Picht, H. & Draskau, J. 1985. *Terminology: An introduction.* Guildford: The University of Surrey.

Pignataro, C. 2012. Terminology and interpreting in LSP conference: A computer-aided vs. empirical-based approach. In: Kellett Bidoli, C. J. ed. *Interpreting across genres: Multiple research perspectives*. Trieste, EUT Edizioni Università de Trieste, pp. 125-140.

Pöchhacker, F. and Zwishenberger, C. 2010. *Survey on quality and role: conference interpreters' expectations and self-perceptions.* [Online]. [Accessed 22 September 2015]. Available from: http://aiic.net/page/3405/survey-on-quality-and-role-conference-interpreters-expectations-and-self-perceptions/lang/1

Public Works and Government Services Canada. 2011. *ARCHIVED 2.5.2. Recognition of Terminological Units.* [Online]. [Accessed 22 June 2015]. Available from:
http://www.bt-tb.tpsgc-pwgsc.gc.ca/btb-pavel.php?page=chap2-5-

2&lang=eng&contlang=eng

Rayson, P. and Garside, R. 2000. Comparing corpora using frequency profiling. In: *Proc. of the Comparing Corpora Workshop at ACL 2000*, Hong Kong. pp. 1-6.

Richards, J. C., Platt, J., Platt, H. 1992. *Longman dictionary of language teaching and applied linguistics.* London: Longman.

Rodríguez, N. and Schnell, B. 2009. A look at terminology adapted to the requirements of interpretation. *Language Update*, **6** (1), pp. 21. [Online]. [Accessed 22 June 2015]. Available from: http://www.btb.termiumplus.gc.ca/tpv2guides/guides/favart/index-fra.html?lang=fra&lettr=indx_autr8KcA5x7sBRyw&page=9oHAHvmFzkgE.html

Rütten, A. 2003. Computer-based information management for conference interpreters-or how will I make my computer act like an infallible information butler? In: *Proc. of Translating and the Computer*, *Aslib 2003*, London. [Online]. [Accessed 22 June 2015]. Available from: http://www.mt-archive.info/Aslib-2003-Rutten.pdf

Rütton, A. 2014. *Booth-friendly terminology management programs for interpreters – a market snapshot*. [Online]. [Accessed 22 June 2015]. Available from: http://blog.sprachmanagement.net/?p=265

Rütton, A. 2015. Terminology. In: Pöchhacker, F. (ed.). *Routledge Encyclopedia of Interpreting Studies*. London/New York: Routledge, pp.416-417.

SAE International, SAE J2450. 2001. Translation quality metric, Society of automotive engineers. Warrendale, USA.

Sager, J. C., Dungworth, D. and McDonald, P. F. 1980. *English special languages: Principles and practice in science and terminology.* Wiesbaden: Brandstetter.

Sánchez-Gijón, P., Aguilar-Amat, A., Mesa-Lao, B. and Pahisa Solé, M. 2009. Applying terminology knowledge to translation: Problem-based learning for a degree in translation and interpreting. *Terminology.* **15** (1), pp. 105-118.

Schmid, H. 1994. Probabilistic part-of-speech tagging using decision trees. In: *Proc. of International Conference on New Methods in Language Processing*, Manchester.

Scott, M. 1996. Wordsmith Tools, Oxford: Oxford University Press.

Scott, M. 1997. Wordsmith Tools version 2, Oxford: Oxford University Press.

Secară, A. 2005. Translation Evaluation - a state of art survey. In: *Conference Proceedings of eCoLoRe/MeLLANGE Workshop*. Leeds, UK, pp. 39-44.

Segalowitz, N. 1986. Second language reading. In: Vaid, J. ed. *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives.* Hillsdale, N.J.: Erlbaum, pp. 3-19.

Segalowitz, N. and Gabonton, E. 1995. Automaticity and lexical skills in second language fluency: Implications for computer assisted language learning. *Computer Assisted Language Learning.* **8**(2-3), pp. 129-149.

Segalowitz, N. Poulsen, C. and Komoda, M. 1991. Lower level components of reading skill in higher level bilinguals: Implications for reading instruction. *AILA Review*. **8**, pp. 15-30.

Séleskovitch, D. 1986. Comment: Who should access an interpreter's performance? *Multilingua*, **5**(4), p. 236.

Séleskovitch, D. 1998. *Interpreting for international conferences (3rd revised edition)*. Arlington, VA: Pen and Booth.

Seleskovitch, D. 1999. The teaching of conference interpretation in the course of the last 50 years. *Interpreting.* **4** (1), pp. 55–66.

Séleskovitch, D. and Lederer, M. 1989. *Pédagogie raisonnée de L'interprétation*. Paris: Didier édudition.

Setton, R. 1993. Is non-intra-IE interpretation different? European models and Chinese-English realities. *Meta: Translators' Journal.* **38** (2), pp. 238-256.

Sharoff, S. 2006. Open-source corpora: using the net to fish for linguistic data. *International Journal of Corpus Linguistics* **11**(4), pp. 435–462.

Sharoff, S. 2012. Beyond translation memories: finding similar documents in comparable corpora. In: *Proc.Translating and the Computer Conference*. London.

Sharoff, S. and Hartley, A. 2012. Lexicography, terminology and ontologies. In: Mehler, A. and Romary, L. eds. *Handbook of Technical Communication (HAL8)*, Berlin/Boston: Walter de Gruter GmbH, pp. 317-346.

Smith, S., Sommers, S. and Kilgarriff, A. 2008. Learning words right with the Sketch Engine and WebBootCat: Meaningful lexical acquisition from corpora and the web. In: *CamTESOL Conference on English Language Teaching: Selected Papers*, pp. 28-42.

Sternberg, R. J. 1987. Most vocabulary is learned from context. In: McKeown, M., and Curtis, M. eds. *The nature of vocabulary learning.* Hillsdale NJ: Lawrence Erlbaum Associates, pp. 89-106.

Tribble, C. & Jones, G. 1990. *Concordances in the classroom: A resource book for teachers.* London: Longman.

Tribble, C. & Jones, G. 1997. *Concordances in the classroom: Using corpora in language education (new edition)*. Athelstan, Houston TX.

Ullman, M. T. 2001. The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition.* **4** (1), pp. 105–122.

Ure, J. 1971. Lexical density and register differentiation. In: Perren, G.E. and Trimm, J.L.M. eds. *Applications of linguistics: selected papers of the 2nd International Congress of Applied Linguists.* London: Cambridge University Press, pp. 443-452.

Varantola, K. 1986. Special language and general language: Linguistic and didactic aspects. *Alsed-LSP Newsletter 9.* **2**(23), pp.10-19.

Vuorikoski, A. R. 1993. Simultaneous interpretation—user experience and expectation In: PICKEN, C. (ed.) *Translation—the Vital Link: Proceedings of the XIIIth World Congress of FIT (vol. 1).* London: Institute of Translation and Interpreting, pp.317-327.

Vuorikoski, A. R. 1998. User responses to simultaneous interpreting. In: Bowker, L., Cronin, M., Kenny, D. and Pearson, J. eds. *Unity in diversity? Current trends in translation studies.* Manchester: St. Jerome Publishing. pp. 184-197.

Wadensjö, C. 1998. *Interpreting as interaction.* London: Longman.

Will, M. 2007. Terminology work for simultaneous interpreters in LSP Conferences: model and method. *MuTra 2007 – LSP Translation Scenarios.* [Online]. [Accessed 22 June 2015]. Available from: http://www.euroconferences.info/proceedings/2007_Proceedings/2007_Will_Martin.pdf

Winston, E. 1989. Transliteration: What's the message? In: Lucas, C. ed. *The sociolinguistics of the deaf community.* Washington, DC: Gallaudet University Press, pp. 147-164.

Woolls, D. 1998. Multilingual parallel concordancing for pedagogical use. In: *Teaching and Language Corpora.* Keble College, Oxford. pp. 222-227.

Wright, S. E. and Wright, L. D. 1997. Terminology management for technical translation In: Wright, S. E. and G. Budlin. Eds. *Handbook of terminology management: Basic aspect of terminology management.* Amsterdam/Philadelphia: John Benjamins, pp. 147-159.

Yates, J. and Orlikowski, W. 2002. Genre systems: Structuring interaction through communicative norms. *Journal of Business Communication.* **39**, pp. 13-35.

Zauberga, I. 2005. Handling terminology in translation. In: Károly, K. and Fóris, Á. eds. *New trends in translation studies: In honour of Kinga Klaudy.* Budapest: Akadémiai kiadó, pp. 107-116.