Online Resource Use and Query Behaviour of English-to-Chinese Trainee Translators:

A Statistical Analysis

Yuxing Cai

University of Auckland, New Zealand

Supervisor: Dr. Vanessa Enríquez Raído

Faculty of Arts, School of Cultures, Languages and Linguistics

Co-supervisor: Dr. Thomas Yee

Faculty of Science, Department of Statistics

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Translation Studies, The University of Auckland, 2023.

Abstract

Online information is an essential resource for translators, hence there is a strong need to respond to this in translator training and research. In the existing literature students' web search processes have been investigated mostly from a non-developmental perspective, and there are only a few studies that have looked into whether and how translation trainees' web search processes change after pedagogical intervention. In this study the changes in trainee translator's resource use and search queries were statistically analysed. Data were collected at four different points in time within a translation practice course with English and Chinese.

In the main study a mixed-methods approach was adopted with four repeated measures and a single cohort of first-year Master's students in translation (N=19) tested remotely. Three pilot studies were conducted prior to the main study. Data collection tools in the main study included: one questionnaire, four source texts, a screen recorder, a key logger, and retrospective verbalization. A semi-automatic data processing and analysis method was developed which significantly expedited the research process. Two variables, resource use and query behaviour, were examined in the main study.

The main findings show that, over the period of the course, the participants moved away from electronic dictionaries to knowledge-based resources (e.g., encyclopaedias, corpora); in contrast, the time spent on search engines remained relatively stable over time. The diversity of dictionaries accessed thus decreased, while use of knowledge-based resources increased. The participants also spent less time using Baidu, but their use of other search engines (e.g., Google, Bing) remained steady. It was found that the participants paid less attention to lexical equivalents and more to contextual information. Additionally, search query time remained stable, yet changes in query complexity were unconfirmed. In terms of query language direction, participants constructed fewer source language (English) queries while the number of mixed language queries (combining source and target language) increased, and the number of target language (Chinese) queries remained stable. The results indicate that the participants still needed more training on advanced search techniques, more knowledge about how search engines work, and more deliberate query practice.

Key words: web search, resource use, query behaviour, translator training

Acknowledgements

First and foremost, I would like to dedicate my deepest gratitude to my supervisor, Dr. Vanessa Enríquez Raído, for her continuous support during the last four and a half years. She was the reason that I crossed half the earth, from the northern hemisphere to the southern hemisphere, to pursue my PhD. It has been my blessing having her during my PhD journey.

My greatest appreciation also goes to my co-supervisor Dr. Thomas Yee, who helped me tremendously with his expertise in statistical analysis. It was also through his encouragement that I have acquired new and rewarding skills, such as Rmarkdown and Latex, which I could never have imagined before.

Another co-supervisor I wish to address here with special thanks is Professor Changshuan Li, for his selfless support of my research design and data collection. Without his help, it would be impossible to collect the data I needed. In addition, special thanks go to Professor Sanjun Sun, who ignited my passion for empirical research, gave me unwavering support, and always made time for me when I needed him, regardless of how tight his own schedule was. Thanks also to Dr Su Stewart, the thesis proofreader, a great friend, who has always been there and ready to help.

I would like to thank Professor Luuk Van Waes for his time and patience. He was always willing to reply to my problems encountered using Inputlog (the main data collection tool in this thesis). Without him, the methodological component of this thesis could not have been achieved. Another person I am sincerely grateful to is Professor Michael Carl. Working as his research assistant before my PhD journey was the starting point in my realizing the true nature of "research". I also thank him for the opportunity he provided for me to go to Kent State University, US, to participate in the one-month Bootcamp.

Many thanks go to my family and friends, for their constant support and encouragement when I was going through the ups and downs of this journey.

Last but not least, I would like to show my gratitude to all the participants who took part in my research, for without them, this thesis would not have been possible.

Table of Contents

List of Tables

List of Figures

List of Models

List of Abbreviations

CAT	computer-assisted translation
EMT	European Master's in translation
GLMM	generalized linear mixed-effects models
IL	information literacy
LIS	library and information science
LMM	linear mixed-effects models
MTI	Master of translation and interpretating
PE	post-editing
SL	source language
TC	translation competence
TCA	translation competence acquisition
TfS	translation from scratch
TL	target language
TM	translation memory
TPR	translation process research

1. Introduction

In the practice and discipline of translation, the prevalence of the Internet, coupled with recent advancements in both information and communication technologies (ICT), and translation technologies, has meant that "the impact and penetration of the Internet have dramatically transformed the way translators carry out their documentary research and address their information needs" (Enríquez Raído, 2013, p.2, cf. Gough 2016, p.14). In today's digital world, translators increasingly operate in a highly-technologized environment which is driven by 'big data'. One of the core digital skills that translators must possess to make sense of this data, and hence succeed in their daily employment, is the ability to efficiently seek, retrieve, and evaluate information in order to acquire specialized knowledge, phraseology and terminology in a wide range of translation domains, quickly and with confidence.

This set of abilities, mainly known as information literacy in higher education, has driven life-long learning for several decades (Aharony et al., 2017; Lau, 2006), and has been integrated into national education policies in various countries, such as the United States (American Library Association, 1998, 2005; ACRL, 2000/2016), Australia and New Zealand (Australian and New Zealand Institute for Information Literacy, 2004), France (Ministry of National Education, 1997), Great Britain (SCONUL, 1999, 2004), and Finland (University of Helsinki, 2004) (as cited in Sales & Pinto, 2018, p. 187). It is also widely recognized as a vital component, in multi-componential models of translation competence. In the European Master's in Translation (EMT) model, for example, it is referred to as information mining competence in the 2009 EMT framework (EMT Expert Group, 2009), and was later updated and integrated in the area of technological competence in the 2017 EMT framework (EMT Board, 2017). It was identified as instrumental competence by the Process in the Acquisition of Translation Competence and Evaluation (PACTE) group in Spain (PACTE Group, 2011), and also as tools and research competence in the TransComp research project in Austria (Göpferich, 2009) (as cited in Enríquez Raído, 2011, p.1). In a recent post-editing competence model, it is called research competence (Nitzke & Hansen-Schirra, 2021).

The pivotal importance of this competence can be explained by the percentage of time spent on research when translating. For terminology alone, Champagne (2004, p. 30) discovered that

experienced translators spent about 20% to 35% of their working time on terminology activities, compared with 40% to 60% spent by novice translators. Schmitz and Gornostay (2014) reported that technical writers and translators spent up to one-third of their time on terminology work, and could sometimes take much longer. In more recent studies, Gough (2016) observed that the average research time spent in familiar domains was 30% while in unfamiliar domains, 36%; Hvelplund (2017a) reported around 20% of the total time was spent on searching digital resources (around 25% for specialised texts and 11% for literary texts); Paradowska (2020) found that the participants spent 39% time on web searching.

1.1 Rationale

Despite the importance of this competence, translators spend a large proportion of their time on searching, and there is no corresponding proportional research that has investigated translators' web searching behaviour. From the perspective of translator training, there is even less research that solely focusses on the changes of web search behaviours, or how this competence can be developed through training.

To fill this gap, the aim of this investigation is to determine whether, and how, trainee translators' use online information (i.e., online resources and search query behaviour), and to evaluate how these patterns of online search behaviour changed throughout the period of a translation practice course, and to use this information to provide insights for translator training.

The field of library and information science is a fundamental and leading discipline aimed at "understanding human information-related behaviours" (Spink and Cole 2006, p.1, as cited in Enríquez Raído 2011, p. 147). Translation comprises a significant and indispensable part of information-related behaviours, for the translator's role has been variously described as an information user, processor and producer (Pinto, 2008; as cited in Enríquez Raído, 2011, p.153). Like the study by Enríquez Raído (2011), the present study is positioned across the disciplines of both translation studies, and library and information science.

1.2 Key concepts and definitions

In this section, I first identify the scope of the present study by defining and distinguishing a

number of key concepts from an in-depth perspective, including the translation process, the difference between tools and resources, and competence and behaviour. I then distinguish between text processing and information searching, where information searching forms the focus of this thesis. I conclude by introducing the different terminologies relating to translation-oriented web searching in previous studies and justify the terminology used in this thesis.

1.2.1 Translation process

The translation process can be evaluated on a number of different levels, and according to one of the most cited definitions there are three levels of the translation process: first, the fundamental level consisting of mental states and operations, where translation is regarded as "self-contained sets of recursive, goal-oriented, conscious tasks" (Muñoz 2010, p.179). The second, closely related, level encompasses sub-tasks and observable operations, such as reading, typing, and revising. The third, the broadest, level starts from when the client first contacts the translator and ends when the translator is paid. In the present thesis methodology, unlike in a real workplace, there was no actual client, and the participants did not need to have any communication with clients. The translation process was therefore defined from the moment the translator started working on the source text until the completion of the target text (Hansen, 2003, p. 26). Operationally, in the present study, the translation process refers to the time from when the participants opened the Source Word file until the moment they clicked on the "save" button in MS Word on completion.

1.2.2 Tools and resources

As mentioned at the beginning of this introduction, translation technology has transformed the translation profession. In the *The Routledge Encyclopedia of Translation Technology*, the scope of translation technology includes both computer aided translation and machine translation (2015, p. xxviii). In one of the most recent definitions, translation technology is understood to include "word processors; spelling, style, and grammar checkers; the World Wide Web; corpus compilation and analysis tools; terminology management tools; translation memory tools (TM); translation management systems (TMS); and machine translation (MT)" (O'Brien & Rodríguez Vázquez, 2020, p. 264). On comparing the two definitions, it would

appear that the scope and definition of translation technology is inconsistent and therefore needs to be examined carefully and redefined to take into account these evolving technologies. The scope of the research in the present thesis is situated within the "technologized" translation environment, but the focus of this thesis is, primarily but not exclusively, the use of resources. Due to the sometimes subtle differences between tools and resources, it is particularly important to consider and define the differences between the two.

The following distinction is drawn from Gough (2016). As shown in Figure 1, tools and resources can be considered as two distinct branches within translation technology.

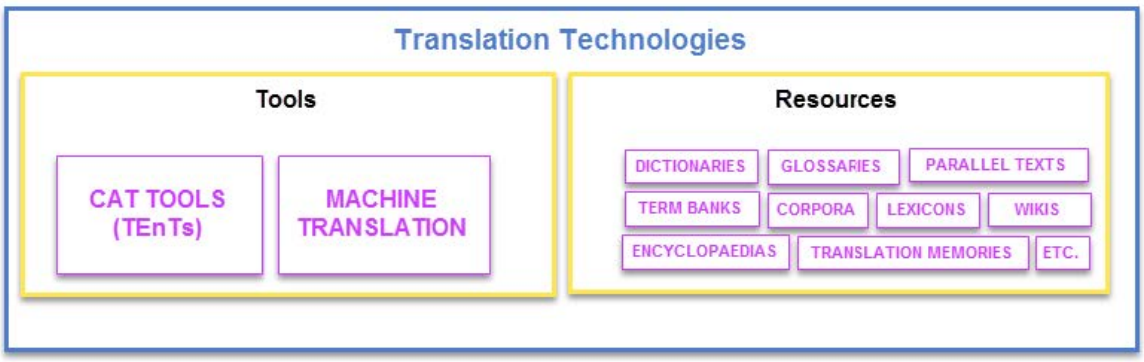

Figure 1 Translation technologies - tools and resources (from Gough 2016, p. 87)

This distinction is primarily based on the fact that, in the context of translation technology, a "tool" is about mechanisms and a "resource" is about content. Under Gough's definition, tools are understood to be "technologies serving many different particular purposes in the translation process, such as retrieving previously translated segments, organising and retrieving terminology, counting words, converting formats etc.", and resources are "technologies that assist in research associated with translation activities" (Gough 2016, p.89). Further, Gough argues that tools are applied to the "whole translation task", while resources are generally used "ad hoc" (p.89). Inside the tools branch of Gough's model, as shown in Figure 1, the CAT tools (or translation environment tools) and machine translation technologies are differentiated. This is aligned with the classical structure suggested by Alcina (2008), which divided tools into machine translation (MT) and computer-assisted translation (CAT).

In addition to Gough (2016) and Alcina (2008), Daems (2016) and Witczak (2021) also distinguished between online resources required for translation from scratch (without a

machine-translated product) and post-editing (post-edited the machine-translated product). This observation is based on studies where participants were given the tasks of both translation from scratch (TfS) and post-editing (PE), and participants in both studies were found to use different types of resources during the two translation processes (TfS or PE).

To clarify, although the scope of the present thesis falls within a highly-technologized environment, the focus of this thesis is on "resources".

1.2.3 Competence and behaviour

In order to identify the scope and main focus of this thesis, the concepts of competence and behaviour need to be clearly distinguished at the conceptual level. From the "behaviour iceberg" conceptual representation adopted by Kuznik and Hurtado Albir (2015), there is a "sea line", or horizontal line across the iceberg, which separates what can and what cannot be directly observed, see Figure 2. It is clear that "competences" (underlying entities) cannot be observed, while "behaviour" can be observed. Observable behaviour is the representation of unobservable factors, such as competences, aptitudes and abilities, personal traits, personal values, knowledge, culture and motivation.

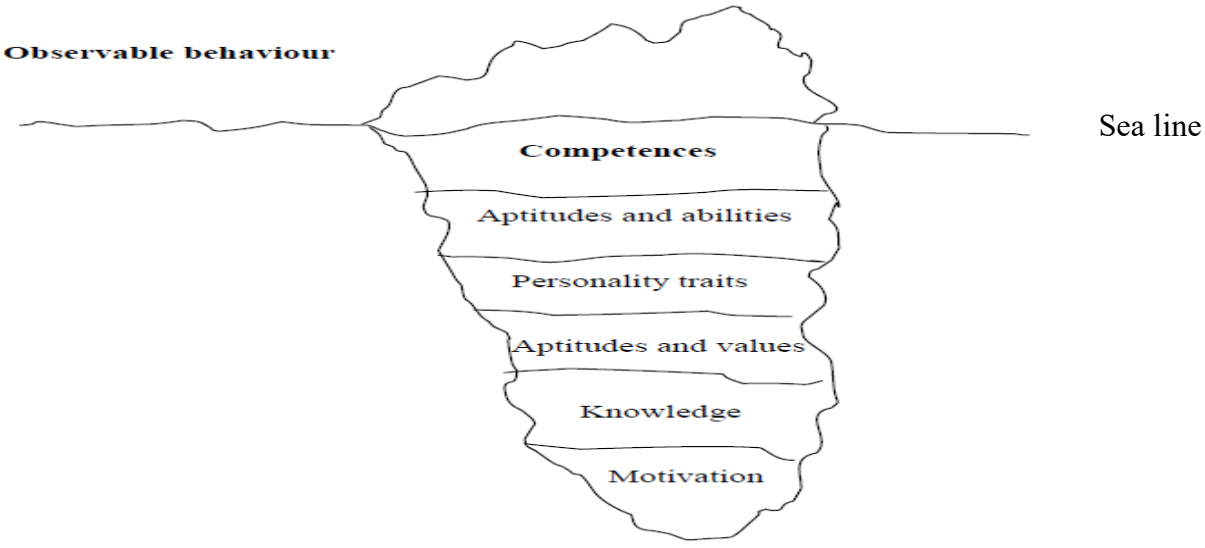

Figure 2 "Behaviour iceberg" (from Kuznik 2015, p.4)

To this end it was necessary to assess the changes in both the behaviour and competence of

translators. For clarification, in the present study "competence" is defined as an "underlying entity, and composed of attitudes, traits, knowledge, and motivation", while "behaviour" refers to "how (this competence) is expressed" (Kuznik & Hurtado Albir, 2015, p. 4). In other words, "translation is manifested through behaviour, (while) competences are a compendium of knowledge, skills, habits and motivation" (Dirube 2004, p.98 cf. Kuznik & Hurtado Albir, 2015, p. 5). Pinto and Sales (2008b) also agree that "competencies are repertoires of behaviour that are observable and can be transferred to the day-to-day working environment" (p.7). It is this "observable behaviour" that is investigated in this thesis.

1.2.4 Text processing and information searching

Following the definition and scope of the translation process outlined in Section 1.1.1, in terms of depth, the scope of this study is focussed mainly on observational behavioural data (i.e., actual translation process). In Gough (2016), the translation process was divided into two types of activities, "interactions with the texts (source and target) and interactions with resources" (p. 15). In this thesis, the two types of interactions are regarded as text processing (interactions with source and target texts) and information searching (interactions with resources).

1.2.5 Translation-oriented web search

In this section, I introduce the different terms related to translation-oriented web searches. In the field of information searching, "resources" are what users (here, trainee translators) interact with during translation. When translating, there is a clear distinction between internal resources, i.e., "human memory" (Gough 2016, p.18) and external resources, e.g., paper-based resources, online resources, or asking for help from other humans. As few print resources are now used for translation purposes (Kuznik, 2017), the focus of this thesis is on web-based resources, which are also referred to as digital resources (Hvelplund 2017), online resources (Gough 2016, Whyatt 2021), Internet-based resources (Gough, 2019), or electronic information resources (Olalla-Soler 2018). For the interactive use of web-based resources, it has been variously referred to as web search behaviours (Enríquez Raído 2011, Shih 2017, Paradowska 2020), online consultation behaviour (Cui & Zheng, 2020), search behaviour (Onishi & Yamada, 2020), information behaviour (Whyatt et al., 2021), and translation-related research activities

(Gough 2019). To maintain consistency and follow the mainstream terminology, web search is used throughout this thesis.

1.3 Structure of the thesis

The present thesis consists of eight chapters. Chapter 1 is the Introduction chapter which introduces the rationale for conducting empirical translation-oriented web search studies. The scope of this thesis is then defined by introducing the key concepts and definitions used throughout this thesis, and is concluded by introducing the overall structure of the thesis.

In Chapter 2 I review previous research relating translation-oriented web search studies from two disciplines, translation studies (TS) and library and information science (LIS). In LIS, translation-oriented web search studies are reviewed. In translation studies, studies from both a developmental perspective and non-developmental perspectives are reviewed.

In Chapter 3 I outline the two-dimensional theoretical framework of resource use and query behaviours, and explores how the resources are categorized, and the three aspects related to the query behaviours (i.e., query time, query complexity, and query search direction) are investigated.

In Chapter 4 I elaborate on the methodology and research design of this study. It is comprised of two parts. The first part introduces the three pilot studies which were conducted prior to the main study, and inform the methodology of the main study. The second part details the main study, including the contextual setting, the schedule and procedure, the data collection tools, the data processing methods, and the statistical methodology.

In Chapter 5 I present the results from the questionnaire, including the participants' background information, their self-declared frequency of resource use, and self-perceived level of web search knowledge.

In Chapter 6 I discuss the findings related to the first research question concerning the changes in the participants' use of resources during the four tasks.

In Chapter 7 I discuss the findings related to the second research question concerning the changes in query behaviours, related to query time, query complexity, and query search

direction.

Chapter 8 concludes the thesis, with a summary of the findings and the main contributions. Limitations and potential areas for further research are also suggested.

2. Previous research into translation-oriented web search

In the 1990's increasing attention was given to process-oriented research. In an article entitled *Process-Oriented Research into Translation and Implications for Translation Teaching*, Lörscher (1992) identified the need for undertaking empirical investigations into the translation process. To be more specific, investigating "hypotheses on what goes on in the translator's head" could "yield general insights into language processing", and benefit "the teaching of translation", for at that time, this was a newly established field (p. 145). The study of web searching in translation studies has traditionally been informed by empirical investigations of translation processes. In early process-oriented research, translation process scholars frequently borrowed methods from other disciplines, such as the "think-aloud" method from cognitive psychology, which was used to study various problem-solving and decision-making processes (Tirkkonen-Condit & Jääskeläinen, 2000). Think-aloud is a data collection method that "asks people to perform a task while stating directly what is going on in their heads/minds" (p. 18). Before computerised methods emerged, it had become the most widely used method in process-oriented translation research (Dam-Jensen & Heine, 2009, p. 10). With the development of new research tools and technologies, however, such as key logging software, the translation process could also take into account the "temporal dimension" (Hansen, 1999) and enabled new research avenues to open up. The more recent, more advanced data collection techniques include eye tracking, and the latest EEG,[1] fNIRS,[2] and fMRI technological advances (Jakobsen, 2020).[3]

With the development of new technologies, translation process research (TPR) became an established new research field within translation studies, with "a special descriptive, empirical, experimental approach to research based on close, technology-supported observation of translational (micro) behaviour" (Jakobsen, 2014, p. 65). While translation studies researchers have developed their own research tools, the development of subject-specific theoretical

[1] electroencephalogram

[2] functional near-infrared spectroscopy

[3] functional magnetic resonance imaging

frameworks was still lagging. TPR is a relatively new field, and "research questions and hypotheses have arisen within the field of translation, while the methods of data elicitation and analysis, as well as the theoretical frameworks, come from a variety of fields" (Jääskeläinen, 2011, p. 123). This was also emphasized by Alves (2015), who observed that "TPR still shows a tendency to borrow extensively from other disciplines" (p.18). The present thesis is embedded within the field of TPR, with a particular focus on web search processes embedded within the translation processes. While this research topic shares much in common with library and information science (LIS), in the following sections, I review the key literature as related to LIS and translation process research.

2.1 Translation-oriented web search studies in library and information science

LIS has a "highly interdisciplinary" nature (Case 2016, p.181) and occupations (such as health care providers, journalists, or lawyers) are often examined in terms of having a type of social role in LIS (Kurbanoğlu et al., 2018). However, neither the disciplinary cooperation between LIS and translation studies, nor translation as an occupation, have received much attention in LIS. In this section, I review the limited available literature on translation-oriented web search studies carried out by LIS researchers.

White et al. (2008) is one of only a few studies conducted within LIS research concerning translation-oriented web search behaviour (cf. Enríquez Raído 2011). Inspired by two task-based information behaviour studies from LIS, specifically Byström (2002) and Vakkari (1999), the aim of the study was to examine the hypothesis that "translators judge the complexity of the task and anticipate difficulty in doing the task, based on intrinsic task factors and extrinsic factors" (White et al., 2008, p. 591). This was both an exploratory and qualitative study, with the objective of determining how translation tasks affect web search behaviour, and what types of resources translators use when translating. It comprised two focus group sessions, one with nine junior professional translators, and one with ten senior professional translators. The data collected from these focus group sessions was then examined in terms of the effects of the intrinsic task factors on the task itself, and the extrinsic factors, including the translators' prior experience and language knowledge. In terms of the resources used, they found that although

dictionaries were the primary resource, a wide range of other resources were also accessed for supplementary information. Five main reasons for choosing a specific resource were differentiated from a qualitative analysis of the data and comprised; "the quality of the resource, the language of the source material, their comfort level or ability in that language, the translation task they [were] performing, and the availability of a source" (White et al., 2008, p. 587). Similar to this study, the investigation of the employment of resources is also an important component in this thesis.

From an LIS perspective the importance of information literacy is emphasized as it "forms the basis for lifelong learning", and is "common to all disciplines, to all learning environments, and to all levels of education" (Pinto & Sales, 2008a, p. 1). Pinto and Sales extensively published in translation information-related searching and firmly believed that "translators are not only information users, but also information processors and producers" (cf. Enríquez Raído, 2011, p. 1) and proposed a conceptual model for information literacy which gathered "skills, competences, knowledge and values, [and it is] based on the cooperation between the authors' expert knowledge of Information Science and professional translation practice" (Pinto & Sales, 2008b, p. 1). To improve information literacy instructions to translation trainees, and as also reported in Enríquez Raído (2011), Pinto and Sales (2007) convened an expert panel from a group of eight teachers and distributed a questionnaire to two cohorts of students (the Bachelor 4-year course was divided into two two-year cycles, and the information literacy related course was taught in the second year of the first cycle). A SWOT analysis (strengths+weaknesses+opportunities+threats) was used to determine improvements in information literacy instructions. In a subsequent study, Pinto and Sales (2008c) then used a semi-structured questionnaire to examine the strengths and weaknesses of students' actual information competences, and the importance of each in terms of their overall training needs from the perspective of translation trainers. In addition, Sales and Pinto (2011) also used a semi-structured questionnaire to look into the perceptions and needs in terms of information literacy from the perspective of professional translators (cf. Enríquez Raído, 2011). Pinto et al. (2016) then conducted a case study with translation trainees, asking students, in the Documentation Applied to Translation and Interpreting course, to complete the documentary research process as if they were going to translate the text (without actually translating the text), and to fill out a template to record how they carried out the preparatory research process. Three main types of mistakes were identified, 1) mistakenly conducting contextual-type searches

(such as cultural references) in dictionaries; 2) wrongly regarding sources of a general encyclopaedic nature (such as Wikipedia) as dictionaries; 3) considering Google as a type of source, rather than as an online search engine that gives access to further specific electronic sources. However, although Pinto et al. (2016) recognized Google as a type of resource, their recognition of the potential multiple functions of Google was still limited, for example, Google can also be used as a spelling type of dictionary. Most recently, from a meta-literacy viewpoint, Sales (2022) compared the assessment results and students' feedback from an information literacy course for Translation and Interpreting undergraduate students. The results emphasised the importance of "conscious and deliberate reflection" (p.9) in enriching students' information literacy, which is transferable to information literacy courses in other disciplines.

It is assumed that, since translation activity is a complex one (Jakobsen, 2014), even the purely lexical translation problems require the translator to be equipped with a thorough knowledge about how a particular word or phrase is syntactically and grammatically embedded within the text, and to have a thorough knowledge of the context (Nitzke, 2019, p. 151). In the following section, I review web search studies in translation studies.

2.2 Web search investigations in translation studies

Since the technological advances in research tools and resources from the second decade of the twenty-first century (2010's onwards), web searching has received a good deal of attention in translation studies and, as observed by Massey and Ehrensberger-Dow (2011a), "information behaviour" is "only now emerging as a significant area of research in Translation Studies" (p.208). As noted by Alcina (2008), although machine translation software opened/began the "the relationship between translation and the computer", the genuine boom in translation technologies was brought about by "the arrival of the Internet" (p. 79), which led to the development of electronic resources, such as online dictionaries and terminology databases. These are one of the foci in this thesis. During the past ten years, a number of web search studies have emerged in translation studies, such as Enríquez Raído (2011), Gough (2016), Shih (2017, 2019, 2021), and Paradowska (2020).

There are normally two main approaches to researching translation processes, including web searches embedded in them: one from a developmental perspective (e.g., longitudinal studies)

and one from a non-developmental perspective (e.g., comparison studies). Comparison studies tend to form the majority of studies, since studies from a developmental perspective tend to be extremely "cumbersome" (Göpferich, 2009, p. 26) and take a long time to complete and bear the risk of participant drop-out (Saldanha & O'Brien, 2014). In the following sections, I first review previous research into web search processes in translation studies undertaken from a non-developmental perspective, followed by a review of studies undertaken from a developmental perspective.

2.2.1 Studies from a non-developmental perspective

In this section, I examine non-developmental translation process studies which have included web search processes as a part of their investigations. Comparative studies that have compared two or more groups (e.g., professional vs. non-professional translators) are reviewed and then studies that specifically recruit trainee translators and professional translators as the only type of participants in the research groups.

2.2.1.1 Comparative studies

Non-developmental or comparative studies examined in this section include those comparing translators with different groups of participants (e.g., students vs. professionals and professional translators vs. foreign language teachers) and translation modes (i.e., translation from scratch vs. post-editing).

In terms of different participants, The PACTE research group (Process in the Acquisition of Translation and Competence and Evaluation) conducted empirical research to validate its proposed multicomponent model of translation competence, in which instrumental competence was considered a sub-competence. The variable "Use of Instrumental Resources (UIR)" was examined because "the data related to the instrumental sub-competence were particularly revealing" (2017, p.220). The Use of Instrumental Resources was defined as "the process in which documentation strategies are activated using resources in electronic format (web pages, dictionaries, encyclopaedias on CD-ROM, etc.)" (Kuznik 2017, p.221). Here, the variable was operationalized as the: number of resources, total time spent on searches, time taken on searches at each stage (orientation, development, and revision), number of searches, and the

variety of searches. The translation competence experiment compared the UIR between professional translators and foreign language teachers with no experience in translation. PACTE group recruited 35 professional translators and 24 foreign language teachers, and screen recorded the translation processes. They concluded that professional translators used a wider range of resources (such as websites, dictionaries and encyclopaedias on CD-ROM, etc.), spent more time on searching for information, and carried out a wider variety of searches. The translation competence acquisition experiment, which was conducted from a development perspective is reviewed in Section 2.2.2.

Perhaps one of the largest and most comprehensive projects in translation studies to examine the role of resource use in translation to date was the Translation Tools in the Workplace project which focussed on "the impact of Translation Memory (TM) and research tools and resources on professional translation processes" (Massey & Ehrensberger-Dow, 2011a, p. 195). This was a sub-project of the Capturing Translation Process project conducted from 2009 to 2012 by Zurich University of Applied Sciences.[4] This project used ethnographic observation, semi-structured interviews, a keystroke logger, screen recorder, eye tracker, and concurrent and cue-based retrospective verbalization to compare the behaviours of different groups of participants, i.e., beginners, advanced students, recent graduates, and professionals. The addition of the eye-tracking methodology opened a new dimension in the investigation of documentary research (Massey & Ehrensberger-Dow, 2011a, 2011b). With the help of an eye tracker, the information concerning what the participants were looking at, when, and for how long, enabled the researchers to identify possible problems the participants encountered while searching for information. For example, "looking but not seeing" was a classic issue identified with eye-tracking data (Massey & Ehrensberger-Dow, 2011b, p. 9). This means that the resources that could solve a participant's translation problems are presented on the screen, but the participants fail to identify or process them. Massey and Ehrensberger-Dow (2011b) emphasise that this phenomenon occurred repeatedly in their data at all levels of translation expertise. They concluded that, in line with their expectations, participants equipped with more translation experience, especially after taking a course on accessing resources, focussed more on employing their newly acquired research skills and their success rate improved. Contrary to

[4] See more information: https://www.zhaw.ch/storage/linguistik/forschung/uebersetzungswissenschaft/publikation-2013-capturing-translation-processes-final-eport.pdf

expectation, however, the advanced students and recent graduates were faster and more successful than the professionals, and employed more advanced search skills. The most logical reason being that the students and recent graduates had recently acquired fresh information seeking skills and up-to-date knowledge about the online resources and research techniques, with which professionals were not familiar. The findings suggested that the investigation of information behaviour could enrich the understanding of the development of translation competence and benefit the information literacy training.

Beyond the use of resources, Enríquez Raído (2011) was one of the first scholars to specifically address the web search behaviours of translation students in an exploratory, multiple-case-study with six participants: four postgraduate translation trainees who had enrolled in her introductory course on technical and scientific translation of English and Spanish, and who participated in the main study, and two professional translators who took part in a pilot study. The four trainee participants were asked to translate two scientific texts in weeks three and four of a 12-week long course. The data collection tools included background questionnaires, a screen recorder, an in-built key logger within the screen recorder, online search reports, and semi-structured interviews. The examination of the relationships between translation task attributes (text type and translation brief) and user attributes (translation expertise, online search expertise and domain knowledge) also suggested new insights and potential future directions for later studies. For example, the findings from Enríquez Raído (2011) suggested that translation expertise and task complexity impact on both the choice of online sources and the degree of iterative web search behaviour. Task-related attributes were also found to have a greater impact on the range of search behaviours than on their depth. Web search behaviours were conceptualised as "web search tasks", operationalized in turn as a search need, search goal, search process and search outcome. Two types of searching style "shallow" (checking and comparing) and "deep" (interactionistic) were identified. Enríquez Raído (2011) initiated the evaluation of web search behaviours and showed the feasibility of combining the various data collection tools for studying web search processes. Pedagogically, the findings suggested that online search reports are useful in raising students' awareness of translation and web searching, and have practical implications for translator training.

A number of subsequent studies have examined patterns of search behaviours, including Wang's (2014) unpublished PhD thesis, which is also based on a multiple-case study, albeit

comprising three groups of participants: one group with four first-year Master of Translation and Interpreting (MTI) students (novices), another group with four second-year MTI students (semi-professionals), and the third group with four professional translators (professionals). A screen recorder was also used to record the translation process and the participants were asked to think aloud after completing a translation task. It was observed that the more experienced the participants were, the more searches they conducted, the more resources they used, and the more target-language oriented queries (compared with source-language oriented) they constructed. In addition, it was concluded that with increasing professional translation experience, there was a strong correlation between the complexity of the participants' information search patterns (such as intensity and sophistication) and their translation performance.

Resource use patterns were addressed in a study by Wang and Lim (2017), who recruited two groups of participants comprising seven MA students (referred to as the "junior group"), and six professionals who taught on translation programmes and were experienced in professional translation (referred to as the "senior group"). The junior group, having been exposed to more sophisticated search techniques, were found to frequently construct queries with a combination of source and target languages, and used a wider range of resources. The senior group did not use any queries with mixed languages and used a narrower range of resources, although they reported using mixed language queries in their daily translation practice. The author assumed that these differences illustrated that the senior group tend to rely more on their memory and experience. Two of the six participants in the senior group however chose to handwrite on paper rather than use a computer, although they did know how to use the computer and how to access online search engines and dictionaries. This diversity in practice in the senior group demonstrates the lack of homogeneity in this group, and highlights the problems of small sample studies. In terms of translation expertise, all six were professional translators, four were professional interpreters but only two had National Accreditation Authority for Translators and Interpreters (NAATI), at levels three and four, regarded as advanced translator and certified translator.[5] In addition, unlike Enríquez Raído (2011), the number of years of translation experience or expertise for each participant were not reported, nor were the participant

[5] https://www.atl.com.au/naati/naati-accreditation/

attributes related to their resource use patterns.

Onishi and Yamada (2020) investigated the differences in searching behaviour between novice translators and professionals. Five university students with advanced English proficiency and with some experience in translation were recruited, hence considered novice, along with four professional translators with more than three years of professional translation experience. The searching behaviours of the two groups were found to be different in terms of the time spent on searching, the websites visited and time spent on them, the number of search engine query types (i.e. source language entry and target language entry), and the number of jumps from search results to individual websites. The overall differences were described as "shallow searching" for novice translators and "non-shallow searching" for professional translators. These search styles and their nomenclature were first identified and used by Enríquez Raído (2011), and subsequently reported in Shih (2017). According to the interview data from Onishi and Yamada (2020), the differences found in search styles may be due to different information seeking motivations, including: 1) to seek reliable information; 2) to understand the context; 3) to avoid mistranslation.

Issues specific to using information resources, but for solving cultural translation problems were investigated by Olalla-Soler (2018). Olalla-Soler (2018) specifically focussed on the solving of cultural translation problems by comparing the use of electronic information resources by translation students and professional translators. Compared with professional translators, translation students used a wider variety of resources, performed more queries and spent more time on queries, however their information-seeking processes were generally less efficient. It would therefore be helpful if translation trainers included learning activities during training, to help students improve the efficiency of their information-seeking strategies, especially in planning their searches, and their ability to adapt their searches when necessary.

Similar to Olalla-Soler (2018)'s domain-specific study, Sycz-Opoń (2019, 2021) targeted legal translation, and examined the information-seeking behaviour of translation students using an observation and think-aloud protocol. Participants were divided into pairs, where one participant assumed the role of the translator to translate a text and the other observed and recorded the translators' information-seeking behaviour. Through statistical analysis six features of searching behaviours were identified, namely: hunting for equivalents in bilingual

dictionaries, haste and impatience, conservatism regarding source preference, overreliance on external sources, uncertainty and doubts accompanying the information-seeking process, and following established paths (Sycz-Opoń, 2019). In a more recent study, Sycz-Opoń (2021) furthered the earlier discussion (Sycz-Opoń, 2019) by examining the strengths and weaknesses of each of the six features identified mentioned above and, more importantly, in suggesting a "mentor-apprentice model of teaching" compared with "standard group classes" to meet individual educational needs and ensure more efficient information seeking behaviour (p.149). Despite the potential methodological validity issues, problematic use of think-aloud protocols and students' direct observation, these findings enriched the current understanding of translation students' information-seeking behaviours on the web.

With regards to different translation modes or conditions, the increasing use of machine translation has focussed researchers' attention on the differences between translation-oriented web searches conducted while translating from scratch (TfS) and those carried out while post editing (PE) machine-translated output, such as Daems (2016), Nitzke (2019), or Witczak (2021). Although this present study did not include a direct comparison between TfS and PE, the research designs and the analysis of variables in those studies did provide insights.

One of the first studies to compare external resource use in the TfS and PE conditions involved a combination of CASMACAT (a key logging software), an eye tracker and Inputlog, used to record the entire translation process, and included the use of online resources (Daems, 2016). Ten Masters students translated four texts and post-edited four texts. Task order, and learning effects were minimised using a Latin square design. No differences were found between TfS and PE in terms of time spent on different types of resources, but a trend was observed for PE tasks to require less time on external resources than on translation tasks from scratch.

The differences in the types of external resources accessed and the cognitive load involved in Internet searches between two task types (translation from scratch and post-editing) was also investigated by Nitzke (2019), where two groups of participants (students and professionals) were compared. Despite the fact that the participants were explicitly asked not to "embark on time-consuming research" in the instructions they received (p.104), participants were observed to conduct more searches in post-editing tasks than when translating from scratch, and professionals searched less than students in both conditions. No differences were found

however for cognitive effort (average fixation duration), or for the range of consulted online resources (operationalized as the number of resource categories) for translation and post-editing.

Cognitive effort was further investigated in a PhD project conducted by Witczak (2021) in a mixed-method experimental study which investigated the information searching by translation trainees and English as a Foreign Language (EFL) students. Participants translated and post-edited two text types (operative-technical and informative-medical), and their PE processes were recorded using an eye tracker and key logger. The results showed that both groups spent more time in information searching when translating compared with when post-editing.

To sum up, the comparative studies reviewed in this section suggest that some differences in web search processes do exist between translators with different levels of translation expertise (e.g., students and professionals, translation trainees and EFL students) or different translation modes (TfS and PE).

2.2.1.2 Studies of trainee translators only

In this section, I review two studies, Shih (2017, 2019), which also following Enríquez Raído (2011) pioneering work into the study of translation-oriented web search behaviours, made a significant contribution in researching these behaviours. In Shih's (2017) exploratory study, six trainee translators were asked to translate a piece of scientific/technical text from English into Chinese using their own computers, their web search behaviours were recorded while verbalizing their thoughts concurrently. In this first study the focus was on describing and identifying commonly accessed web-resource types, such as search engines, online dictionaries, online machine translation, and online encyclopaedias. The findings are in line with Enríquez Raído (2011) in that, although dictionaries were still a primary and the most-frequently used resource, students should still be encouraged to use a combination of different types of web resources to maximize their chances of locating the most suitable and accurate resources, and to cross-check them. In addition to identifying the most commonly used queries related to the source text for translation, the types of search queries (except for the commonly composed of source language or target language alone) were categorized as: 1) a ST term and its corresponding (sometimes partially corresponding) TT; 2) a ST term and some question words

(in the target language); 3) a ST term and the name of the TL; 4) a provisionally translated sentence or title of the original ST. These four query categories enabled a wide variety of queries to be logged and analysed, for example "a tentatively translated ST title" as a query can be recorded as an effective strategy.

Compared with the investigation of the web search behaviour of six participants in the Shih (2017) study, the investigation of web search behaviour was extended to encompass web search optimisation with a larger cohort of eighteen postgraduate translation students in Shih (2019). Web search optimisation was defined as "a balanced combination of efficiency and effectiveness" (Shih 2019, p.2). 'Efficiency' was defined as the "minimum amount of time [spent] on web search episodes" and was considered to be a relative concept used to compare the same ST term for different participants (Shih 2019, p.2). 'Effectiveness' was defined as "the precision of [a] Target Language equivalent" which is largely judged based on the students' own verbalization (p. 2). All participants completed three texts, translated from English to Chinese, on three separate occasions in their own habitual working environment. The translation process was screen-recorded and their verbalizations while translating were also recorded. To describe the characteristics of optimal web searches, a theoretical framework was adopted based on information retrieval, which in turn was based on two types of actions: "primary actions", including query-related metrics (i.e., query formulation, query duration, and query abandonment), and "secondary actions" (i.e., dwell time and trail) comprising clicking and browsing-related metrics. More optimal web searches were characterized by a so-called "explorer's approach" and were deemed to be more related to "secondary" rather than "primary" actions.

2.2.1.3 Studies of professional translators only

Most of the studies described above compared participants with different levels of expertise in order to gain insight for translator training, however, these comparative studies could only provide indirect implications and applications, and it is only by recruiting professional translators, that can we know more about the professionals' search behaviours, and the results can directly inform translator training. Recruiting professional translators appears to be difficult due to the often-limited availability of research funding to pay for professionals' work and time. Despite this difficulty, some researchers have recruited professional translators as the

sole participants in their studies with different focuses, aims and objectives.

Gough (2016), for instance, explored the patterns of interaction between professional translators' and online resources using a two-stage, multi-method study. The first stage was a survey (N=540) that employed a questionnaire to gather general information concerning resource preferences accessed by professional translators. The second stage (N=16) employed a screen recorder, a Think Aloud Protocol, and a post-task questionnaire to investigate the nature and quantity of the resources actually used in a translation task completed by the participants. This is a quasi-naturalistic research design which took place at the participants' normal workplace and aided the ecological validity of this study. Two main taxonomies, the Resource Type User Taxonomy (RTUT) and the Taxonomy of Translator Research Styles (TTRS) were formulated. The former consisted of the Dictionary Enthusiast, the Parallel Text Fan, the Mixed User, and the MT Adopter, and the latter included the Explorer, the Prolific, the Methodical, the Economical, and the Understated. It was argued that the RTUT was more influenced by the available technology while the TTRS was more relevant to translators' inherent information behaviour. These taxonomies are useful for identifying resource style and research style in future studies.

Two years later, Hvelplund and Dragsted (2018) focussed on how, and to what extent, text genre affects search strategies. Nine literary translators and nine language-for-specific-purpose (LSP) translators translated four texts (two literary and two LSP texts, respectively). It was concluded that LSP translation required more searches than literary translation, however this may have been due to the greater number of low-frequency terms in the LSP texts than in the literary ones. The LSP translators were also found to rely more heavily on external resources.

More recently, Whyatt et al. (2021) compared the information behaviour as a function of translation direction of a cohort of 30 professional bidirectional translators. The participants translated two texts: one text into their native language (Polish) and the other into a major, non-native language (English). The data collection tools included a key logger, eye tracker, and screen recorder. The results suggested that translation direction and text type both had a significant effect on the resources used. To be more specific, the L1 to L2 translations demanded more effort than the L2 to L1, and the specialised texts demanded more effort than general texts. Information behaviour in a bidirectional translation (IBiBT) model was then built

to explain the dynamics involved.

Shih (2021) also investigated how professional translators navigate the web. Ten professional translators were recruited to translate a short piece of medical text from English into their respective mother tongues, with five translating into Chinese, and the rest translating into Arabic (N = 1), German (N = 1), Japanese (N = 1), Russian (N = 1), and Spanish (N = 1). The eye-tracked screen recording (eye movement superimposed onto the screen recordings) was used to provide cues for the participants to verbally report on their searches in retrospective reports. Despite the small cohort and different languages involved, the queries were categorized into five types based on the linguistic features observed in the screen recordings; 1) the use of natural language; 2) the use of a "casting-a-net" approach (i.e., use various queries to search online) ; 3) the use of a literal translation or potential "false-friend"; 4) the use of collocations in search engines and/or other specialised online resources; 5) the use of an ST term, a ST sentence, or even a whole ST paragraph. Based on retrospective verbal reports, for primary actions, the translator's query intent was categorized into: 1) locating a TL term; 2) locating further background information (about a specific term or about a field in general); 3) validating a TL term or expression; 4) seeking inspiration for alternative terms or expressions; 5) locating TL collocation; 6) post-editing intent; 7) language-specific intent. In short, these findings align with those already obtained in translation studies throughout first-generation, second-generation and third-generation cognitive research (e.g., Muñoz 2017). Shih developed, however, a further seven query intentions for primary actions as compared to Shih (2019) (see Section 2.2.1.1 above), and examined the relationship between query type and query intent. For secondary actions, the use of the eye tracker allowed an analysis of eye movements to promote a deeper understanding of the participants' browsing behaviours of search engine result pages, starting with their "scanning and skimming actions" subsequently followed by possible "reading actions".

2.2.2 Studies from a developmental perspective

Studies reported in this section were conducted from a developmental perspective (or self-claimed as studies from a developmental perspective), rather than strict developmental studies or longitudinal studies. This is because the scope of longitudinal studies is often vaguely or roughly defined as over "a relative long" period during which "several" measurements are

taken. This is understandable due to the differences between different disciplines and different research settings. To the best of my knowledge, there is currently no clear-cut scope in translation studies. A relevant and clear definition of longitudinal studies can however be found in education by White (2005), in which a longitudinal study is regarded to meet two criteria at the same time: 1) at least one year, 2) at least two measurements taken. In translation studies, some studies take a significantly shorter time than this. For example, Nunes Vieira et al. (2021) investigated the use of computer-assisted translation (CAT) tools and Alabau et al. (2016) targeted post-editing, which were only conducted for 16 weeks and 6 weeks respectively, although the studies were self-defined as longitudinal. In this section, studies conducted from a developmental perspective, or looking for changes in a period of time, are included in this section, regardless of the length of data collection time or the number of measurements taken.

TransComp, for example, was a longitudinal project aimed at exploring the development of translation competence over time (Göpferich, 2009). Translation competence was considered under six sub-competences: strategic competence, translation routine activation competence, tools and research competence, communicative competence in at least two languages, domain competence, and psychomotor competence, with the focus on the first three sub-competences, including the tools and research competence. These three sub-competences were chosen because compared to bilingual individuals without formal translation training, there might be some differences for professional translators (Göpferich, 2009). Tools and research competence were investigated in connection with the decision-making process since the use of online resources is embedded in decision-making processes (Göpferich, 2011).

The identification and development of competencies were also a focus of the Translation Competence Acquisition (TCA) experiment (PACTE 2011), reviewed in Section 2.2.1.1. The TCA experiment serves as a continuation of the Translation Competence (TC) experiment, which used the same data collection methods and analytical framework to investigate the Use of Instruments (UIR) from a development perspective (Kuznik & Olalla-Soler, 2018). The TCA experiment was a simulated longitudinal study that took simultaneous measurements from groups of first-, second-, third-, and fourth-year students and a group of recent graduates, with a total of 130 participants. The non-linear progression in the development of UIR was depicted as “a gentle, wave-like line”, with the first-year students at the lowest point, followed by a quantitative leap to second-year students, then a slight fall in the third- and fourth-year

students, followed by a rise again for graduates. In addition, the findings again suggested the need for translator training to include the efficient use of instrumental resources, to be more specific, the need to "establish an order and priority of searches, and evaluate their quality and the possibilities" (Kuznik & Olalla-Soler, 2018, p. 49).

The development of digital literacy in academic language learning contexts was further investigated by Mutta et al. (2014) in two studies. The first study concerned information seeking in the participants' first language and a foreign language, and the second investigated how to raise participants' awareness in information seeking. Participants (native speakers and L2 learners) were asked to complete two tasks: the first task was the translation of ten words and included writing down all the search paths employed, and the second was the translation of a short text and again included writing down the search paths employed. Between the two tasks, there was an awareness-raising intervention, where the main content was the discussion of the critical use of digital media and some awareness-raising tasks (e.g., finding words and comparing information). This study concluded, unsurprisingly, that the intervention had a positive effect on participants' digital literacy.

Following on from the multiple-case study conducted by Enríquez Raído (2011), Fernández (2015) conducted a similar case study, again with only four students over a period of one semester lasting 15 weeks. The four students were asked to translate two texts during the course, thus representing the only two points of measurement in this replication study, and the data collection methods included Think Aloud Protocols (TAP), screen recordings, pre- and post-class questionnaire. The study concluded that for these four students, previous web search training appeared to have a positive relationship with where and how information was sought. This raised the issue of how to determine how web search competence is acquired, and how this is related to the other translation competencies identified earlier.

This conceptual gap of how competence acquisition relates to multi-component competence models was further explored by Kiraly (2013), and Cheng (2017) attempted to address this issue by investigating the development of translation competence from a problem-solving perspective as opposed to the PACTE multi-componential competence model. In Cheng's PhD thesis (2017), the aim was to validate the proposed conceptualization through a one-year longitudinal study with four data collection sessions. The study was conducted with six

participants, using a translation task-based interview as the main research instrument. During four data-collection sessions in this one year, the six participants' information literacy was seen to improve in terms of the need for further information, the credibility of the accessed resources, and the way the resources were used (from reference to inspiration). To be more specific, the participants in the first data-collection session, did not show any awareness of, or tendency to consult, any further online resources and terminated their search activity as soon as they found a certain solution. In the second and third session participants showed a 'greater determination' in pursuing more in-depth information using mining behaviours/online search behaviour, and in the fourth session, their awareness of the need to search online information and to clear up the doubts in solution-proposing activities was evident. Over the four sessions, they also became more selective about different sources of information with the credibility parameter in mind, and consulted more institutional and authoritative websites and double checked the information they found. In addition, at first, they heavily relied on dictionaries and other online resources when proposing translation solutions without any modification, but gradually became more flexible and more independent when consulting online resources, and modified and proposed more creative translation solutions. In the first session, the majority of the participants were not motivated to search in order to clear their doubts; in the second session, some participants showed stronger determination to search and resolve their doubts; in the third session, participants demonstrated in-depth online search behaviours; in the fourth session, participants demonstrated their awareness for searching for further information to solve their problems in solution-proposing sub-activities. This development was manifested in their increasing awareness of the need for further searching and the need to evaluate and cross-check online resources. By the end of the one-year MA program, rather than relying excessively on online resources, they had become able to propose translation solutions more independently and creatively. Although only based on the observations of six students, these findings on the incremental acquisition in skills and knowledge are consistent with those obtained in previous studies dealing with competence.

The process of metacognitive skills development and web searching was also the focus of an unpublished PhD study by Chang (2018), who conducted a one-year longitudinal investigation of translation competence development. Ten postgraduate students translated three texts while verbalizing their thoughts concurrently. The translation processes were screen-recorded, and the recordings were used as cues in subsequent cue-based interviews. Over the space of the

year the students gradually moved away from dictionary-dominant searches to other online search features such as online images and online maps, and these students learned to cross check their search results. The students also began to spend more time looking for background knowledge rather than just finding simplistic target-text equivalents, and in addition, became more aware of the importance of web resources.

Another, more recent, unpublished PhD study conducted by Paradowska (2020) focussed on the development of information competence among translation students. It was a non-experimental case study, lasting two semesters and using direct observation of web search behaviours and analysis of the search histories of students on a general translation course, followed by a main study lasting one semester, specifically designed to improve the participants' information competence. Ten undergraduate students enrolled in a general English to Polish and Polish to English translation course participated in the main study. Two data collection points were implemented. One took place before the course and the other one after the course was completed. A multi-method approach was adopted, including questionnaires, screen recording, and translation journals and involved both within-case and between-case analyses. It was found that students initially did not have sufficient information competence, and the deliberate practice in the course helped develop students' information competence.

2.3 Summary

To sum up, in terms of web search investigations from a non-developmental perspective, the comparative studies examining differences between different levels of translation expertise (PACTE 2011, Enríquez Raído 2011, Massey & Ehrensberger-Dow, 2011a, Wang 2014, Wang and Lim 2017, Shih 2017, Shih 2019, Nitzke 2019; Onishi and Yamada 2020, Witczak 2021) suggest that differences in web search processes and behaviours do exist, for example other comparative studies have investigated; between different translation modes (i.e., translation from scratch and post-editing examined by Daems (2016) and Witczak (2021)) and in different domain specific translation (e.g., cultural translation investigated by Olalla-Soler (2018); differences in legal translation behaviours examined by Sycz-Opoń (2019, 2021)); and the results of comparative studies of professional translators (Gough (2016), Hvelplund and Dragsted (2018), Whyatt et al. (2021), and Shih (2021)). Identified similarities and differences

are cross analysed in order to explain the changes that occur as a result of a pedagogical intervention and in order to inform translator training understanding. In addition to these findings, previously reviewed research designs were also critically considered when designing the empirical component of this study, and are discussed in detail in Chapter 4 Methodology.

Studies conducted from a developmental perspective, as discussed in Section 2.2.2, Transcomp (Göpferich, 2009), PACTE (Kuznik & Olalla-Soler, 2018) and Cheng (2017), all focussed on translation competence as a whole and not solely on web searching. Fernández (2015) and Chang (2018) only recruited four and ten participants, respectively, and yet can be regarded as multiple case studies, and even though Paradowska (2020) was focussed on web searching and used inferential analysis, the sample size (ten participants) of the study is statistically unreliable. In addition, it has only two data collection points (before and after the pedagogical intervention) which makes the changes that took place during the process easily overlooked. It is therefore still crucial to conduct web search studies that recruit sufficient participants to reach data saturation where, regardless of how many participants are added to the sample, the results will not change. This is specifically relevant for hypothesis-testing studies that seek generalization.

After reviewing the web search investigations in the literature, it is clear that a study from a developmental perspective conducted with suitable research tools in an optimized ecological validity friendly environment is needed to enhance our current understanding in the changes of trainee translators' web search behaviours.

3. Theoretical framework

From the review of relevant literature in Chapter 2, it is evident that there is still a lack of empirical studies investigating the changes in web search process, an area which may directly and substantially inform our in-depth understanding of the translation process and benefit translator training. In this chapter, a two-dimensional theoretical framework is developed in order to systematically examine translation trainee's resource use and query behaviour, with conceptual and theoretical contributions from both translation studies and LIS.

3.1 Resource use

In LIS, resource use had attracted the attention of researchers, such as Byström (2002), even before the Internet age. Byström (2002) investigated the relationship between task complexity, information types and information sources. A more recent work, Saastamoinen and Järvelin (2017), based on Byström similarly examined the relationship between work task types, complexity, and dwell time on information resources.

Resource use in translation studies, has been investigated mainly in terms of resource categorisation, and there is no unanimous resource categorisation that fits all studies for various reasons. From a diachronic perspective, dictionaries have been a dominant resource since the beginning of translation practice but the "decentralisation" (Gough 2016, p.95) of dictionaries took place while other resources were emerging. Findings from Hirci's (2013) surveys conducted in 2005 and 2012, for instance, illustrate this phenomenon, suggesting that translation students were more likely to turn to a Google image search or various blogs than to search for information in traditional dictionaries (e.g., Oxford Pictorial Dictionary).

There are also major differences in available resources for different languages, and resources for translation between different language pairs are thus unequal in terms of content and availability, for some more popular, and more widely spoken, languages have richer resources than less commonly spoken language combinations. As pointed out by Whyatt et al. (2021), the status and prestige of languages and cultures vary, and the number of people speaking different languages are very different. It is therefore easy to imagine that the available resources

for English (a major language) are far greater than those available for Polish or Croatian (lesser-known languages). In the present thesis, the language pair under consideration is English-Chinese. With China now having one of the largest populations in the world, which makes up 18.47% of the world's population,[6] more and more foreign students are learning Chinese (or Mandarin) as a foreign language, and the translation resources for Chinese are now increasing exponentially.

The resource categorisation in empirical studies is generally based on their own collected data, for while the source texts are selected by the researcher which prompt translation-oriented web searches, the participants determine the actual resources used. Participants with different levels of expertise and responding to different types of stimuli, or translation genres, therefore require access to different resources. For example, Chang (2018) selected three tourism texts from an online travel website, which required searches concerning several place names. Some participants used Google maps, which the author categorised as an 'online map' for the main category and 'Google maps' for the sub-category. However, if the source text does not contain any location related information needs, the chances of using an online map, or Google maps would be low. The use of an online map would be rarely used in other studies.

In spite of all the issues which can complicate potential classifications of different types of resources available, the research community's interest in categorising these has not decreased. This has become particularly apparent with more types of resources (other than dictionaries) now becoming readily available due to the expanding use of the Internet and the constant development of new resources such as the use of search engines to conduct large volume searches, or online encyclopaedias to enable searches for more detailed information. To prepare students for the future employment market, it is now more important than ever that more knowledge and skills are gathered to determine what resources to use and how to effectively and critically use them.

In the present study, and similar to other empirical investigations in the field, resource categorisations were derived from the data collected. Resource use data was also collected from a questionnaire and was aimed at exploring the participants' self-rating of their frequency of

[6] Source: https://www.worldometers.info/world-population/china-population/#:~:text=China%20population%20is%20equivalent%20to,(and%20dependencies)%20by%20population.

resource use, and the other was aimed at examining the actual use of resources.

In the following sections, I first explain the resource categorisation used in the questionnaire before moving on to the categorisation for keylogging and screen recording data.

3.1.1 Resource categorisation used in the questionnaire

The questionnaire used in the main study and the rationale for its use are described in detail in the Section 4.7.1. In this questionnaire, there is one question about research use, which asked the participants to self-rate how often they use different types of resources when translating. Eight options were provided for them to choose from: search engines, web pages, dictionaries, machine translation, concordancers, encyclopaedias, online documents, and discussion forums.

3.1.2 Resource categorisation for keylogging and screen recording data

In this section I review how other empirical studies have categorised resources, and then present and justify the categorisation for keylogging and screen recording data in this thesis.

3.1.2.1 Review of categorisations in existing literature

In the following section, I review translation-oriented resource categorisations in LIS and TS, and explain how these classification systems will benefit the categorisation in the present thesis. I then introduce the categorisation of the actual resources used during the main study (i.e., key logging and screen recording).

The review of previous resource categorisations is made in a chronological order of data collection rather than publication date. This is because sometimes there is a significant time gap between the time the data was collected and the publication date, and this would confuse the chronological understanding of the development of resources use. For example, PACTE group (Hurtado Albir, 2017) conducted their translation competence (TC) experiment with translation teachers and professionals as their subjects in 2005, however the manuscript relating to the use of instrumental resources was not published until 12 years later in their book

Researching Translation Competence in 2017.

In order to justify the categorisation in this study, I divided the following categorisations into two types, one-level categorisations, and two-level categorisations. The former comprised a simple list of the resources that formed a number of categories, however the latter first categorises all the resources into different main categories, and then each main category was further categorised into subcategories.

White et al. (2008) was one of the earliest scholars in LIS to take an interest in translation-oriented resource use. They conducted exploratory and qualitative research using focus group sessions with professional translators. They divided the resources used into several categories (Figure 3). This study did not however mention when the focus group sessions took place, and the Internet use at that time (before the year of publication of the article) was not as widely used as it is now, which may be reflected in their observation that the scope of Internet dictionaries was "simply limited" (p.589).

Figure 3, from top to bottom, is arranged according to the frequency of resource use. It is clear that dictionaries were the primary sources accessed at that time, but other sources, such as Internet documents and cultural encyclopaedias, were becoming more accessible. This study can therefore be seen as reflecting a time when a wider range of resources were emerging and becoming available in addition to traditional dictionaries.

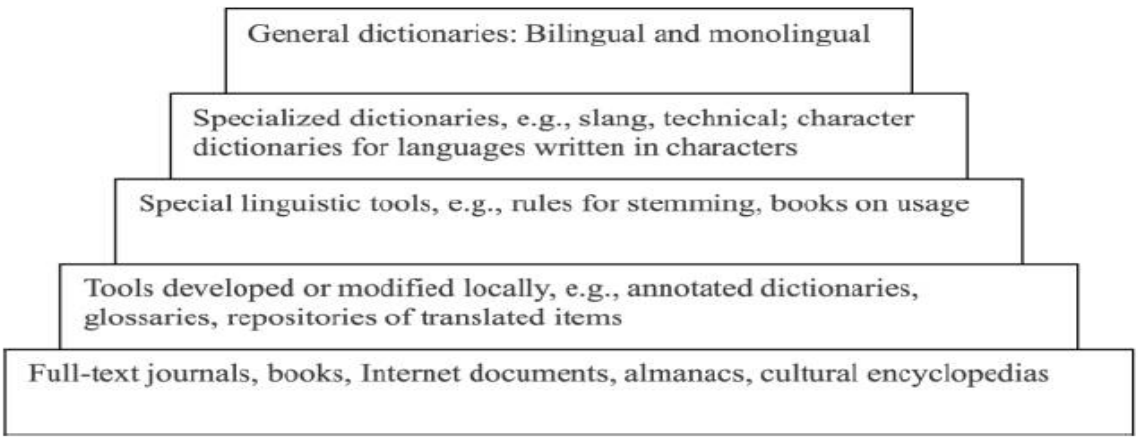

Figure 3 Resources consulted during a translation task (from White et al., 2008, p.587)

Daems et al. (2016) also broke down resources into categories in a study aimed at comparing

the differences in external resources used between Translation from Scratch (TfS) and Post-editing (PE) tasks, conducted from 2014 to 2015. The categorisation of resources was modified when presenting the data both visually and descriptively. Initially nine categories of external resources were defined: news, conversion, term bank, spelling, synonym, MT, encyclopaedia, dictionary, search, concordancer, and "others". The nine categories were only used however when visually presenting a general overview of the frequency and durations of those nine categories in bar charts. For further analysis, the relatively large number of categories were reduced to four because it was found to be difficult and unnecessary to compare between each category when a number of categories were only sporadically used. The four groups comprised: dictionary, concordancer, encyclopaedia, and "other", which were then used in further statistical analysis. The "other" category included the former categories: news, conversion, term bank, spelling, synonym, and MT.

Another one-level categorisation (a simple list of the resources combined into different categories) was employed by Wang and Lim (2017), who compared the range of resources used between a junior group and a senior group, see Table 1. The resources they listed were, however, more specific, such as the Youdao dictionary, Baidu dictionary, iciba, Haici dictionary, etc. This study, based on a one-level simple listing, concluded that the junior group used a wider range of resources than the senior group. On looking more closely at their results however, the wider range of resource use by the junior group (compared to the senior group) can be seen to be mostly related to dictionary use. For the junior group, the seven participants used an average of 1.7 different dictionaries while the six participants in the senior group used an average of 1.5 different dictionaries. These small sample groups of unknown homogeneity, and simple, highly specific, category listings however make any further analysis and conclusions difficult, including cross-study comparisons (Saldanha & O'Brien, 2014).

Next, I introduce the use of two-level resource categorisations which is a relatively recent progression. Here all the resources are first categorised into different first level (main) categories, which are then further categorised into second level categories. One of the earliest studies, by Pakkala-Weckström (2015), investigated the variety of sources used by second-year-undergraduate translation students to complete their homework in 2012. It was found that different genres and functions of texts required access to different types of resources. A two-level approach was employed, first to categorise the accessed resources into "Internet resources"

and “others”, and then to further categorise within the two first-level categories.

Table 1 Comparison of the use of web resources by the two groups of translators (from Wang and Lim, 2017, p.68) (dict = dictionary)

Junior group		*Senior group*	
	Web resources		*Web resources*
Sue	Baidu dict. (百度); iciba (愛詞霸); Haici dict. (海詞詞典); Google Translate; Google	Ruth	Baidu dict. (百度)
Andrew	Youdao dict. (有道詞典)	Jack	ichacha (查查詞典); Google Translate
Nancy	Haici dict. (海詞詞典); Google	David	Youdao dict. (有道詞典); Google
Nathan	Youdao dict. (有道詞典); Baidu dict. (百度); Wikipedia	Oscar	Youdao dict. (有道詞典); Google
Ellen	Cambridge English–Chinese (traditional) dict.	Andy	Baidu dict. (百度); Haici dict. (海詞詞典)
Carol	Youdao dict. (有道詞典); Google Translate; Google	Ivan	Baidu dict. (百度)
Ray	Youdao dict. (有道詞典); Oxford dicts; Linguee English–Chinese dict.; ichacha (查查詞典); Google		

Based on data collected in 2013, and using a fully Internet based categorisation, Wang (2014) compared the different translation resources used by novices, semi-professional and professional translators, while they translated from Chinese into English. The resources were first classified as dictionaries, web-based resources, corpora and machine translation. “Dictionaries” were then further classified as Free dictionary, Lingoes, Haici Dictionary, Thesaurus, Longman Dictionary and Oxford Dictionary; “web-based resources” were further classified as Baidu, Google, Wikipedia and Bing; “corpora and machine translation” were further classified as Google Translate, and Parallel Corpora.

The increasing numbers of primary and secondary classifications required increasingly complicated methods of analysis. One approach, taken by Gough (2016), was to analyse the resource categorisation “in a dual manner” (p. 123), where the primary categories (types) were compared and analysed first (e.g. dictionaries, glossaries, term banks), followed by an analysis of the second-level categories representing specific resources (actual resources such as Real Academia Española dictionary, ProZ Glossaries or Interactive Terminology for Europe). The development of the Resource Type User Typology (RTUT), mainly focussed on the classification according to the nature of the resources accessed.

Other, more recent examples of two-level categorisations include Chang (2018), whose data collection period lasted one year from 2014 to 2015. The categorisation in this study included dictionaries, machine translation engines, search engines, online encyclopaedias, travel websites, online images and online maps - the last three reflecting the needs of Chang's text selection of four tourism texts. Those seven categories were further divided into specific resources, for example, search engines were divided into Baidu, Google, and Yahoo. A study by Shih (2019) based on data collected from 2015 to 2016, first categorised resources as search engines, online dictionaries, machine translation engines and online encyclopaedias, and then further classified them into more specific resources. More recently, Onishi and Yamada (2020) categorised participants' resources into search engines, dictionaries, and non-dictionary, however their categories are somewhat unclear. It seems that "dictionary" and "non-dictionary" are the primary categories, but 'search engine' appears as a separate category. The evolution of two-level resource categories is summarised in Table 2 below.

The two-level categorisation, having greater flexibility and comprehensiveness, can be seen to outperform one-level categorisation in terms of the number, specificity and complexity of the categories. This flexibility can, to some extent, come at the price of the increased complexity of the analysis, however it also increases the depth and rigour of the studies. Despite a number of studies employing the two-level categorisation, none have been quantified on both levels, which would have led to greater insights into the width and depth of resources used.

Table 2 Summary of two-level resource categorisation

Studies	Data collection time	Main category	Sub-category
Pakkala-Weckström 2015	2012	Internet sources	1. information sites, which comprise library sites, online dictionaries, and for example Kotus, the Institute for the Languages of Finland; 2. Google (used for collocations); 3. Google Images; 4. Wikipedia 5. a diverse category I have called 'Other sites'. This last sub-category comprised parallel texts (e.g., for the recipes), official sites (Hampton Court, EU) and any other sites listed in the commentaries
		other sources	The category of 'other sources' covers everything mentioned in the commentaries that lies outside the Internet: visits to the library, browsing one's own or one's parents' bookshelf, and person sources, i.e. contacting experts or calling friends or relatives for information.
Wang 2014	2013	dictionaries	Free dictionary, Lingoes, Haici Dictionary, Thesaurus, Longman Dictionary, Oxford Dictionary
		web-based resources	Baidu, Google, Wikepedia, Bing
		corpora and machine translation	Google Translate, Parallel Corpora
Gough 2016	from July 2013 to March 2014	e.g., dictionaries, glossaries, term banks	e.g., Real Academia Espanola dictionary, ProZ Glossaries or IATE
Chang 2018	from November 2014 to May 2015	Dictionaries	Bing Dictionary,Youdao Dictionary,ICIBA Dictionary,Lingoes Dictionary,Oxford Dictionaries,Casio
		Machine Translators	Bing Translator,Google Translate
		Search Engine	Baidu,Google,Yahoo
		Online Encyclopedias	Wikipedia,Baidu Baike
		Travel Websites	Traveler's Digest, TripAdvisor, Booking, Qyer, Mafengwo, Ctrip
		Online Images	Google Images
		Online Maps	Google Maps
Shih 2019	academic years 2015/16 and 2016/17	search engines	Baidu, Google
		online dictionaries	ICIBA, Youdao, Lingoes
		online machine translation	Youdao Fanyi, Google Translate
		online encyclopaedias	Baidu Baike, Hudong Baike, Wikipedia
Onishi and Yamada 2020	Not mentioned	Search engine	(Not specifically named)
		Dictionary	Weblio,Eijiro on the web,Eigo-box,Google Translate
		Non-Dictionary	Original text,Wikipedia,National Geographic,University of Bristol,google image,Discovery,Nature Research,American Museum of Natural History,The page, Kyoryu-zukan,Blog.etsuterm,Logmi,Tow to Pronounce

3.1.2.2 Resource categorisation in the present study

To optimise the rigour and depth of the present study a two-level categorisation was selected. The first level primary categories were dictionaries, search engines, and knowledge-based resources. This was decided based on the previous literature and is further discussed in this chapter. The secondary level categories, were determined by employing a bottom-up approach, informed by the collected data. I explain the uniqueness of each primary category below.

Dictionary

Dictionaries were the earliest resources that translators have consulted while translating, and they are also one of the most direct ways to find an equivalent, especially in bilingual dictionaries. From a diachronic perspective, Kernerman (2013) summarised the various changes dictionaries have undergone:

- from print to digital
- from tangible to virtual
- from one dictionary for life to many simultaneously
- from one-size-fits-all to customised and personalised
- from a language product to multi-language services
- from paid to non-paid
- from privately-owned to publicly-funded/shared
- from passive reader to interactive user
- from human to machine
- from content to technology. (p.7, as cited in Gough 2016, p. 94-95)

Dictionaries have enjoyed a core position, but they also have a mixed reputation. Varantola (1998) observed several limitations of the dictionaries at that time, which can still be applied to the current situation.

• Dictionary makers aim at context-free descriptions of word use, whereas dictionary users resort to dictionaries to solve a context-dependent problem.

• Translators need equivalents, but they also need reassurance: therefore translators do not like to find equivalents which they do not recognise.

• Translators also need information relating to longer stretches of text than a single lexical item.
• Translators try to find non-dictionary type information in dictionaries because it is not readily and systematically available in other sources. (p.181)

In addition, some studies observed an overreliance on dictionary use (e.g., Sycz-Opoń 2019). This is probably due to students starting to use dictionaries when they first start to learn a new language, hence consulting dictionaries when they encounter problems is their first 'port of call', or maybe because it entails the least effort and is the most familiar resource. It is normally taken for granted that using dictionaries is extremely easy, however, this may not always be the case, as identified by Pastor and Alcina (2010; 2013), and Lew (2013), who discussed the search techniques for dictionaries and issues encountered when identifying the correct options.

More recently, the dominant role of dictionaries has been decentralised, as predicted by Varantola (1998), who suggested that "instead of expecting miracles from dictionaries we should aim at a more versatile conception of lexical and textual reference sources" (p.181). As Hirci (2013) suggested the two distinct resource categories: dictionary and non-dictionary, which also showed that dictionary is a distinct resource compared with others.

Search engines

Search engines are an important topic in LIS, for without them, it would be inconceivable for searchers to find information on the web (Varian, 2006). Despite the fact that there have been different evolutionary stages in line with different technology developments, i.e. pre-Internet, Internet (pre-Web), early Web, and Web 2.0, the essential service they provide is to "provide Internet users with access to important information by directing them to links to available online resources on a plethora of topics" (Tavani, 2020, p. 2). From the perspective of information retrieval, compared with other resources, search engines are unique for the following reasons: "the size of the document collection, the number of users, the need to satisfy queries on all topics, the need to satisfy different query types (which are not specified by the users), and the problems with unreliable or deceptive content" (Lewandowski, 2006, p. 261). Due to their uniqueness, search engines have become a primary category in data analysis.

Knowledge-based resources

The majority of knowledge-based resources provide text-based information, such as when consulted for contextual information, such as background information. Gough (2016) defined knowledge-based resources as "all resources that contain solely encyclopaedic information such as encyclopaedias, Wikis, compendia, or information databases." (p.103). In the present study, knowledge-based resources are resources that the participants used, but not including dictionaries and search engines. The two-level categorisation provides greater flexibility of resources categorisation, especially when further categorisation needed to be made within the collected data. With the rapid development of technologies, the resources have become dynamic and constantly changing, and the participants' use of them is therefor of current interest.

3.2 Query behaviour

Query behaviour has only been sporadically researched in translation studies, such as Enríquez Raído 2011, Wang 2014, and Shih 2019, however, it is an important and more widely researched topic in LIS. As early as the 1960s, from the perspective of cognition and communication, Robert Taylor began to explore the potential disconnect between a searcher's information need and the formal query required to enable the information system to respond appropriately (Taylor, 2015). Taylor described the four levels of information need as:

> 1. First of all, there is the conscious or even unconscious need for information not existing in the remembered experience of the inquirer. It may be only a vague sort of dissatisfaction.
> 2. At the second level there is a conscious mental description of an ill-defined area of indecision. It will probably be an ambiguous and rambling statement.
> 3. At this level an inquirer can form a qualified and rational statement of his question. Here he is describing his area of doubt in concrete terms and he may or may not be thinking within the context or constraints of the system from which he wants information.
> 4. At the fourth level the question is recast in anticipation of what the files can deliver. (Taylor, 2015, p. 254)

In a similar vein, but from a psychological point of view, Nicholas Belkin proposed the Anomalous State of Knowledge (ASK) model, as the "basic motivator" for information searching, and the representation of the anomalous as a "query" (1982, p. 64). In the latest glossary in Case (2016), ASK is

> A recognised gap or uncertainty (i.e., an anomaly) in an individual's state of knowledge regarding a situation or topic. Faced with an ASK, the individual may attempt to address their uncertainty by requesting or consulting information; in that case, the person will then judge whether the anomaly has been resolved; if it is not resolved, another ASK may be generated, or the motivation to address it may be exhausted. (Case, 2016, p. 367)

In addition, Spink and Jansen identified a query as the "expression of a searcher's information problem" (2004, p.77, as cited in Enríquez Raído 2011, p. 279). Based on Taylor's concept, Belkin emphasised that queries would evolve as the answers they generate are evaluated. This is particularly relevant in translation studies given the complex and iterative processes involved in the completion of a translation task.

In translation studies, query behaviour also gained attention as "one of the most overt and explicit features of web search" (Shih, 2017, p.4). The scope of translation studies was then extended to encompass the location of where the query was submitted, including online dictionaries or newspaper archives (Enríquez Raído, 2011). Similarly, Shih (2017) also broadened the scope of queries to include queries entered outside search engines, as the aim of the study was to look into web search behaviour, i.e., in addition to search engine access.

In the present study, "queries" also refer to items typed into search engines, dictionaries, or any other searchable resources. However, those that fit either of the two following exclusion criteria will not be included in the quantitative calculation. The first exclusion criterion refers the location, for example, if it is entered into a machine translation engine (e.g., Google Translate), and the second exclusion criterion concerns the length of the query: for example, if it exceeds a whole sentence, paragraph or an entire text.

A number of studies have been undertaken in an attempt to address these problems. Enríquez Raído (2011) analysed search queries in terms of syntax, complexity, length, and effectiveness.

Mutta et al. (2014) focussed predominantly on where the queries were submitted and which language they were written in (Kuznik, 2017; Kuznik and Olalla-Soler, 2018) determined the number of queries constructed by each individual based on their information needs. Hvelplund (2017) focussed on the purpose and depth of queries and categorised queries as navigational queries (designed to reach a particular website), deep queries (designed to find relevant information), and shallow queries (designed to find relevant information without clicking further).

In the present study, query behaviour will be analysed in terms of query time, query complexity, and query search direction, as defined below.

3.2.1 Query time

In LIS, query time is defined as "time spent formulating queries and examining the search result pages" (Vakkari 2019, p.885). In translation studies, query time is rarely discussed, except by Shih (2019) who relates query time to "efficiency" (p.5), however, in this study, query time was only used as a relative metric and was not quantified. In the present study, query time represents the mean time a searcher spends on a query. The query time was calculated as total resource time divided by the number of queries per task and per participant.

3.2.2 Query complexity

In LIS, query complexity is used to examine the effect on web searching results (Jansen, 2000) or the frustration of the searcher (Field 2010). In translation studies, the complexity of the queries is generally represented by the number of advanced queries constructed. According to Enríquez Raído (2011), an advanced query is defined as "the use of advanced search options, operators and/or term modifiers entered by the user" (p.374). Wang (2014) did not directly use "query complexity" or "advanced query", and reported only two search operators, which were, the use of '*' and 'site' (i.e., UK). Unlike Enríquez Raído (2011), this study did not analyse the use of queries separately as it is only regarded as the characteristic of using search engines. More recently, closely related to the author's pedagogical web search intervention, Paradowska (2020) operationalised complex queries as: 1) containing more than three terms; 2) using advanced search queries, such as search operators and Google's advanced search; 3) using

meta-language; 4) combining both the source language and the target language. In the present study, query complexity was still measured according to Enríquez Raído (2011), comprising advanced search options, operators and/or term modifiers..

3.2.3 Query search direction

Query search direction has also been a variable of interest, and Wang (2014) used queries as one search strategy to search target language, and compared the target-language-oriented queries between novices, semi-professionals, and professionals, and found the target-language queries to be more effective from this perspective. Shih (2017) categorised queries as: 1) an ST term and its TT equivalent; 2) an ST term and some question words; 3) and ST term and the name of the TL; 4) a provisionally translated sentence or title of the original ST term. Gough (2016) further employed the initiated languages of queries to determine whether the research direction which is one element in the Resource Type User Typology (RTUT), was initiated in the source language (ST-oriented queries) or the target language (TT-oriented queries). The language direction of a query can therefore, to some extent, indicate the participant's search direction. In this study, the queries were divided into three categories according to the languages used: source language (Chinese), target language (English), and also a mixture of the source and target languages.

In conclusion, in this chapter a two-dimensional theoretical framework was introduced, which was derived from both translation studies and LIS. The research questions (defined in Chapter 4) concern both resource use and query behaviour. In terms of resource use, a two-level categorisation emerged as a logical progression from the resource categorisations in the literature. The first level consists of dictionaries, search engines, and knowledge-based resources, and the secondary level provides further categories within the three first-level categorisations, which were derived from the data collected. This two-level categorisation gives the flexibility to investigate resource use, which is especially useful during periods of rapid development of the resources. With regards to query behaviour, after reviewing the literature in translation studies and LIS, the three variables of interest (query time, query complexity, and query language direction) were chosen due to their relevance to information searching.

In the next chapter, I introduce the methodology of the present study which applies the theoretical framework proposed in this chapter to an empirical study.

4. Methodology

Technological evolution not only impacts translation practice but can also revolutionize the research communities, particularly by the development of new research tools to investigate the translation process. This also applies to the web search investigations in translation studies. In the present research, I adopted a mixed-method approach to investigate the changes in resource use and query behaviour during the first-year of a Masters course in Chinese-English translation. In this chapter, I first present the research questions of this thesis, and outline how the variables were operationalized in relation to the theoretical framework proposed in Chapter 3. I then describe the methodology of the three pilot studies conducted prior to the main study. The main study is introduced, including its contextual setting, the recruitment of participants, the data collection tools, schedule and procedures, and the collected data process is then described and discussed.

4.1 Research questions and variables

The central aim of this research was to gather empirical evidence of the changes in trainee translators' resource use and query behaviour during a first-year course in translation practice and to draw pedagogical implications for, and applications of, future translator training.

The specific research questions addressed were:

1. What type of resources do trainee translators use to complete their assigned translation tasks? How much time do they spend on each type of resource? How does the use of resources change over the course of a semester?

2. What type of search queries do trainee translators perform to find translation-oriented information? How much time do they spend on each query? How do their search queries evolve over the course of a taught semester?

The first research question concerns the choice of resources the participants made when translating. For this study, the categorisation of resources comprised dictionaries, search engines, and knowledge-based resources. To further investigate the diversity of these three

categories, each category was further broken down and analysed as a range of different sub-categories: dictionaries (e.g., Oulu dictionaries, Lingoes dictionaries, Merriam-Webster dictionaries), search engines (e.g., Google Search, Baidu Search, Bing Search), and knowledge-based resources (e.g., Concordancers and corpora, encyclopaedias, discussion forums) derived from the data collected. The rationale for this categorisation of resource use can be found in Section 3.1. The time spent consulting each category was measured in milliseconds, and then presented as hh:mm:ss (hour:minute:second) or mm:ss (minute:second) as appropriate. The dynamic of the three main resources were then analysed in terms of the percentage of the total time spent consulting each of them. The second research question concerns the query behaviours of the participants. The search queries were analysed in terms of query time, query complexity, and query language direction.

The variables for each research question, how they were measured, and the codes assigned in subsequent analyses are summarised in Table 3.

Table 3 Variables and operationalization

Main Variables	**Operationalization**	**Codes used in Models**
Resource use	time spent consulting dictionaries	dictionary_time/dictionary_pct
	time spent consulting search engines	searchengine_time/searchengine_pct
	time spent consulting knowledge_based resources	knowledge_time/knowledge_pct
Query behaviour	time spent per query	query_time
	query complexity (number of advanced queries)	N.A.
	query direction (source/target/mixed language queries)	SL/TL/MIXED_count; SL/TL/MIXED_pct

*N.A.: not applicable
*pct: percentage
*SL: source language
*TL: target language
*MIXED: mixed language (combine both source and target language)

4.2 Development of the research design

This doctoral research, including three pilot studies, was approved by the University of Auckland Human Participants Ethics Committee (UAHPEC) on 11 October 2018 (approval reference number 022125). The central goal of this research was to gather empirical evidence

of the changes of trainee translators' resource use and query behaviour during a first-year course in translation practice and to draw pedagogical implications for, and applications of, future translator training.

The chosen research design follows a mixed-methods approach with four repeated measures and a single cohort of translation trainees. In order to collect valid and reliable data, appropriate data-collection tools had to be selected. Several factors contributed to the selection criteria, with the overarching consideration being the participants' full access to online resources and the complete recording of the translation processes. In addition, it was necessary to ensure the "naturalness" of the trainee translators' behaviour, thereby maintaining their translation environment and behaviours as close to normal as possible. The size of the final dataset also needed to be manageable, given that there would be four repeated measures or data collection points in the main study. Bearing these considerations in mind, three pilot studies were conducted to test several research settings and data collection tools, prior to designing the main research study and the selection of appropriate data-gathering tools.

In the following sections, I critically discuss the evolution of the three pilot studies, with a focus on practical lessons learned and adjustments made. The data collection tools employed in the main study, together with conceptual considerations are discussed fully in Section 4.7.

4.2.1 First pilot study

The first pilot study was the starting point of the design of the main study. In this section, I introduce the research aims and questions in the first pilot study in relation to the main study. I then briefly present the data collection tools and how they were used (the procedure). After this, I elaborate on the data processing and results. Lastly, I reflect on the data collection tools in the first pilot study, particularly the use of the eye tracker.

4.2.1.1 Research aims and questions

The primary aim of the first pilot study was to test the reliability and validity of the data collection tools chosen to record and reproduce the translation processes, particularly interactions between the participants and the online resources they used. The secondary aim was to address the related research questions elaborated in Section 4.1. Compared with the

main study, which was designed to examine the changes in resource use and query behaviour during a semester-long course, the first pilot study examined the differences between two groups of four participants, one group was recruited from trainee translators who had attended the course, and the other, control group, was recruited from trainee translators who had not. The course (the contextual setting) was the same course as in the main study, and is introduced in detail in Section 4.4. The secondary aim concerns the following two research questions:

1. Are the resources used by the two groups different? Resources are analysed in terms of types of resources (i.e., dictionaries, search engines, knowledge-based resources).

2. Are the query behaviours performed by the two groups different? Query behaviour includes the query time, query complexity, and query language direction.

4.2.1.2 Research design

The first pilot study involved a simple two group comparative design: one group (N=4) had attended a translation course, and the other, control group (N=4) who had not. The course deals with both translation directions. Due to the limited scope of this proposed research, only one translation direction was examined, from Chinese into English. This was chosen in this pilot study because the market has a higher demand for this translation direction and translation-oriented web searching from the translators' mother tongue to their first foreign language is richer than in the opposite translation direction (Massey and Ehrensberger-Dow, 2013).

The variables in the first pilot study were as follows, the independent variable was the participants' training experience, and the dependent variables were the time/percentage they spent consulting different types of resources, the average time spent on queries, the number of advanced queries they constructed, and the number of queries they constructed in each language direction (source language queries, target language queries, and mixed-language queries).

Figure 4 shows the research design for Pilot Study 1.

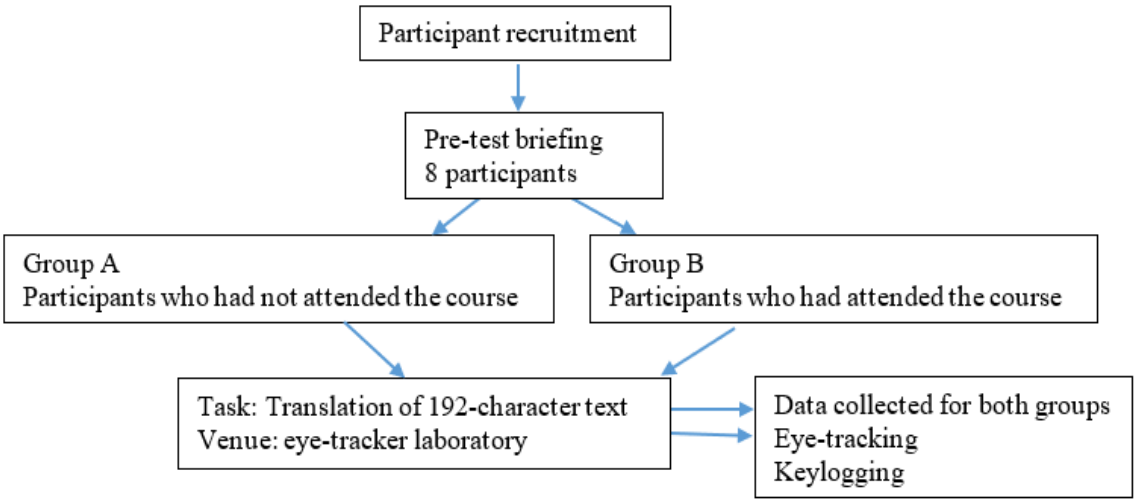

Figure 4 Research design for Pilot Study 1.

4.2.1.3 Data collection tools

There were three data collection tools used in the first pilot study, which were a single source text, a key-logger, and an eye tracker.

A legal document was chosen as the source text for the pilot study, entitled *Ad Exchange Services Framework Agreement Le.com, Leshi Internet Information & Technology Corp., Beijing* (source text is attached in Appendix A). It contained 192 Chinese characters (one title and two paragraphs) and a variety of potential translation problems which could trigger the participants to seek solutions by searching for information online. A translation brief was also provided as of "immediate use for publication".

In addition to the source text, the other two data collection tools used in the first pilot study were a key-logger (Inputlog, 8.0.0.6 version) and an eye tracker, which also had a built-in function for screen recording (Eyelink 1000 Desktop Mounted Head-free Remote Mode). The two tools were chosen for their applicability and suitability for recording and representing the resource use and query behaviour. The key logger (Inputlog) is a logging tool which records all keystrokes, mouse movements, and Microsoft Windows activities (Leijten & Van Waes, 2013). It was particularly useful in the context of this study due to its ability to log information concerning the participant's information search processes, as it can measure how much time

each participant spends on different resource types and records all the queries the participants construct.

Eye trackers have been used in the investigation of translation-oriented web search process since 2010. Along with many new breakthroughs, it has come with both challenges and inevitable drawbacks. Below I review the application of the eye tracker in web search process and justify why it was used in the first pilot study.

The advantages of an eye-tracker were demonstrated by Massey and Ehrensberger-Dow (2011a, 2011b), who used eye-tracking to find out why participants failed to solve specific types of translation problem. For example, eye-trackers can identify fixations, which are an indicator of visual attention. In Figure 5, data are shown representing a screenshot of the eye movement data produced by one participant undertaking a search for the English abbreviation, MoD (Ministry of Defence). The participant typed the query "MoD" and viewed the search engine result page. The participant's visual fixations are represented by red circles; the larger the red circle, the longer the fixation duration. The numbers within the red circles represent the order of the fixations. With Figure 5, it can be seen that most of the participant's attention was given to the first and second results, with barely any fixed attention on the third result. The third result does however provide the relevant background information that would have solved the translation problems the participants encountered, but the participant failed to identify it as a useful reference, hence leaving this problem unsolved.

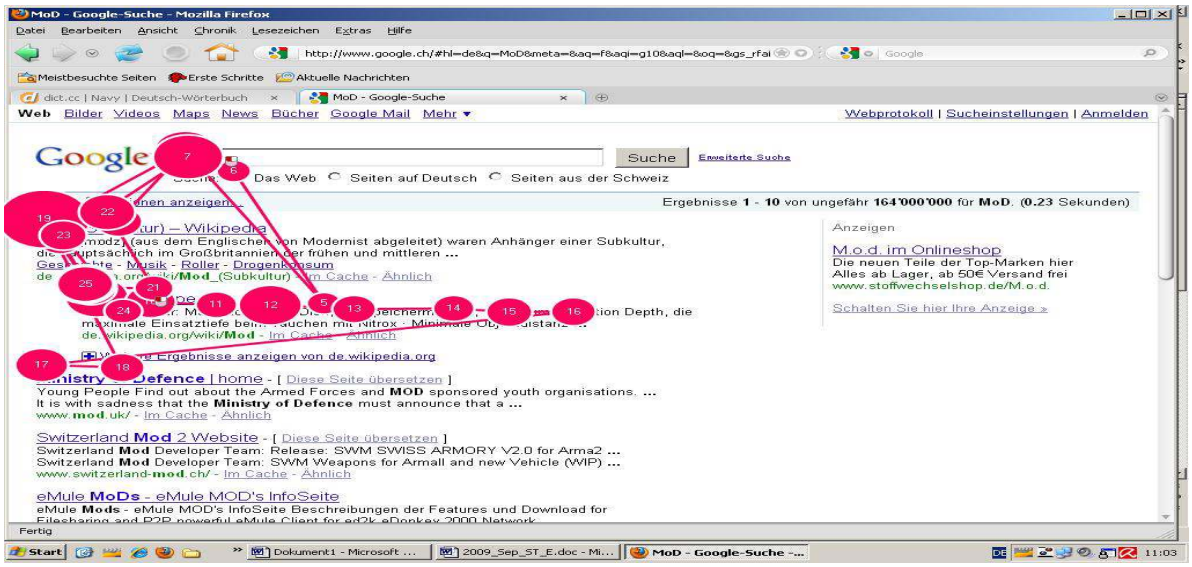

Figure 5 Fixations on a Google results page for the query "MoD", from (Massey & Ehrensberger-Dow, 2011b, p.7)

Another scholar who used an eye tracker to research translation-oriented web searching was Hvelplund (2017), where an eye-tracker was used to investigate the eye movement differences between resource consultation, translation drafting, and revision, and concluded that resource consultation constitutes a significant amount of the translation process. It was also found that longer fixations (which indicate more effortful processing) and larger pupils indicating a heavier cognitive load during consultation, compared with the translation phase and orientation phase (2017a). Hvelplund (2019) then used the same dataset to examine resource consultation during translation regarding the distribution of visual attention, variation in cognitive effort and processing flow, and empirically confirmed from visual attention data that, in terms of processing flow, consultation comprises a substantial part of the translation process. It was also verified that the resources could be introduced at several points in the workflow to help the comprehension of the source text and the production of the target text.

Instead of focusing on analysing statistical aggregated data, such as fixations, or pupil dilation, Shih (2021) used the recording of eye-tracked data as a video cue for retrospective verbal reports. The eye-tracked recording data represented the locus of "the eye movements (i.e., gaze) [that] were superimposed in real time onto the screen recording" which provided a "more holistic and contextualised overview" (p. 57).

The advantages of using an eye-tracker have also been acknowledged in a number of PhD thesis on the subject of translation-oriented web searching, where their limitations were also described. For example, Chang (2018) acknowledged that key loggers and eye-trackers can generate "hard data" which is more scientific and objective compared with "soft data" like TAP, especially the combined data from the simultaneous use of a key logger and eye-tracker. More specifically, in Gough (2016) situations are mentioned in which an eye-tracker would be of great assistance. Another example would be when the source text window and the resource window are displayed simultaneously on the screen, and it is hard to determine which window the participant is looking at. Second, when the participant is consulting a research query result, such as in Google, and it is difficult to access which option the participant chooses if there is no following click.

Based on the above-mentioned usages and its potential advantages, it was decided that the eye tracker should be used in the first pilot study. Before using an eye tracker to collect data however, it is important to decide how the screen should be split between text processing and web searching areas, since this directly impacts eye movement data. In previous empirical studies of web search process which employ data collected by an eye-tracker, the screen layout is usually fixed, with one half set up for the interaction with the texts to be translated (in the present study referred to text processing), and the other half set up for web searching. Once set up the screen layout cannot be adjusted, for when processing eye data, an Area of Interest (AOI) needs to be drawn to investigate the interaction with the text and the interaction with resources, hence areas of eye movements of the two windows cannot be overlapped. To my knowledge, all the previous empirical studies of web search process which employ data collected by an eye-tracker have fixed areas for text processing and information searching, such as Cui and Zheng (2020), Whyatt et al. (2021), and Witczak (2021).

In the first pilot study of the present research, the screen was also vertically divided into two equal windows, a so-called translation window on the right (Microsoft Word) and a search window (web browser of choice) on the left (see Figure 6).

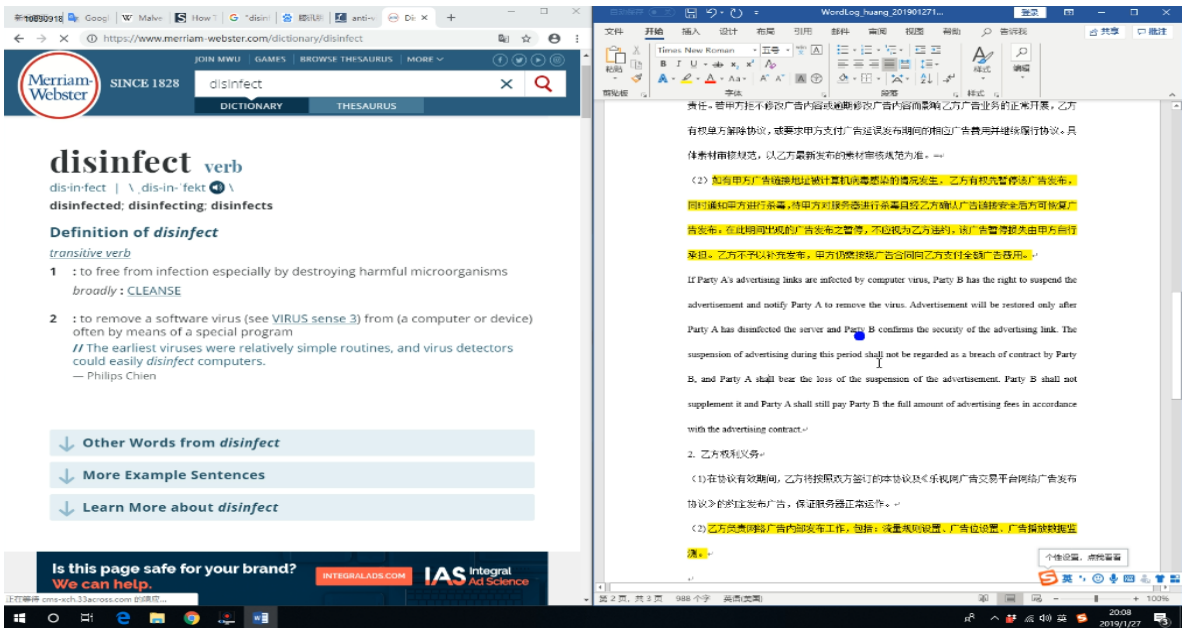

Figure 6 Example of the screen display for the Pilot Study 1

4.2.1.4 Procedure

In this section, I introduce the procedure for the first pilot study. After volunteering to take part, the participants were briefed on the aim and nature of the experiment and were then introduced to the eye-tracker. Given that the data for the dominant eye is normally considered sufficiently accurate and representative, with the data for the other eye deemed more distant from the validation points (Teixeira & O'Brien, 2017, p. 42), only the dominant eye was tracked. The Porta Test was used to determine the dominant eye,[7] or ocular dominance, for each participant (Chang, 2009).

The computer workspace used by each participant was then individually set up according to each participant's own working habits and preferences to maximize ecological validity, e.g., their preferred browser, font, and paragraph settings were offered to the participants. For all participants the monitor was a standard 20-inch flat screen Dell monitor (model S2421HS), and the screen was the same for all participants. The eye motion data collection rate was set at 500Hz (Eyelink, 2009). To calibrate the eye-tracker system, nine calibration points were chosen in accordance with the current literature on reading and translation research (Chang, 2009), as well as with the eye-tracker's operating manual (Eyelink, 2009). Once preparation was complete and the participants were familiar with the setup, they were asked to translate a 192-character Chinese text into English (i.e., translation into their second language or L2). The translation process was interrupted every 20 minutes to recalibrate the system in order to ensure the quality of the eye-tracking data (Hvelplund, 2014). The whole translation process was eye-tracker recorded and all key strokes were logged.

Finally, a small reflective disc was stuck with adhesive tape on to the centre of the participant's

[7] Based on the Porta Test, a participant is first instructed to use the index finger of his/her right hand to point at an object visible far afield with the entire right arm straightened and 'both' eyes open until s/he reported to the researcher that s/he is pointing at the object. Second, the individual is instructed NOT to move his/her arm. Third, the researcher instructs the participant to close his/her right eye (leaving the left eye open and arm still unmoved) and to report to the researcher whether or not the object has removed from the point of the index finger. Fourth, the researcher asks the participant to open both eyes, close his/her left eye (leaving the right eye open and arm still unmoved) and to report whether or not the object has removed from the tip of his/her index finger. If the object has not removed from the tip of the individual's index finger when a participant closes his/her right eye, then his/her ocular dominance is right-eye; the same rule is applicable to determining if someone is left-eyed (Chang, 2009, p. 84-85).

forehead to help track their eyes while the eye-tracker was operating in remote mode.

Before data collection commenced, all participants were allowed sufficient time to familiarize themselves with the screen recorder and eye-tracker, and to allow calibrations of the system to be undertaken.

4.2.1.5 Data processing

After collecting the data, the participants' keystroke data were exported from Inputlog, and eye-tracking data and screen recording data were exported from the eye-tracker (Eyelink). The originally plan was to synchronise/merge key data and eye data into one timeframe spreadsheet, but this failed since the interface was not yet available at the time of analysing the first pilot study data. This technical problem and its impact are discussed fully in Section 4.2.1.7. The Inputlog data and screen recording data are described in greater detail in Section 4.7, since the processing and analysis of these data will remain the same in the main study.

4.2.1.6 Results and analysis

In this section, I first report the findings pertinent to the research questions proposed in Section 4.1, where the main variables are resource use and query behaviour, as in the main study. I then discuss the validity, reliability and ecological validity of the methodology as the main aim of the first pilot study was to test the research design and its validity, as well as to solve any technical methodological issues that may arise. Due to the limited space of this thesis a full analysis of the data was not possible, and only a general analysis of the first pilot study findings is presented.

Resource use

Figure 7 illustrates the time spent on each type of resource per participant and the three colours represent the three types of resources the participants used: dictionaries (blue), search engines (orange), and knowledge-based resources (grey). The main findings of this pilot study show that the two groups of participants presented different patterns for resource use. Group A, P1 to P4 (P represents participant with the number of individual participant), who had not attended the course, used dictionaries for a mean time of 6min 26s, which was 3min 27s longer than

Group B (P5-P8, who had attended the course, mean 2min 59s). The differences were greater (31min 4s) for search engines and knowledge-based resources, none of the participants in Group A spent more time than any in Group B in accessing search engines (Group A mean 4min 9.5s, Group B 35min 13.5s.) and knowledge-based resources (Group A 4min 48.6s., Group B 28min 25.25s.).

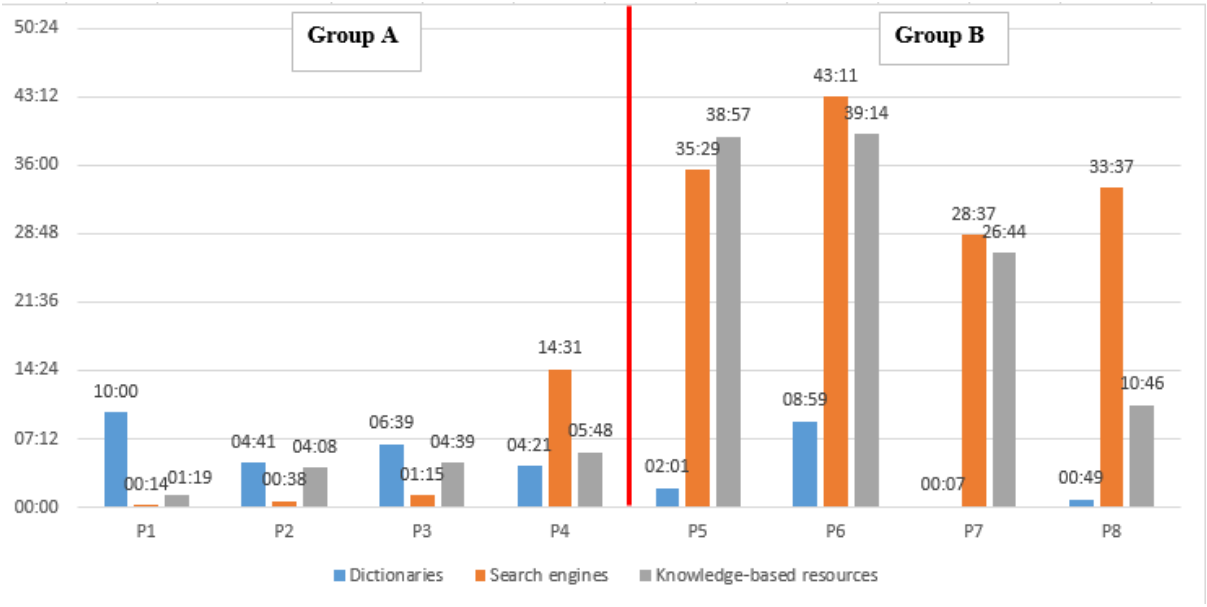

Figure 7 Time spent in each resource type per participant for Pilot 1 (time in minutes:seconds)

Figure 8 presents the percentage of time spent on each resource type. There is strong evidence that Group A spent a larger proportion time than Group B in consulting dictionaries (Group A mean 51.69%, Group B 3.62%). Although the group difference for the use of search engines and knowledge-based resources is not as large as for dictionaries, the difference is still quite clear (Group A 19.36% while Group B 54.90% search engines; Group A 28.95% while Group B 41.47% on knowledge-based resources).

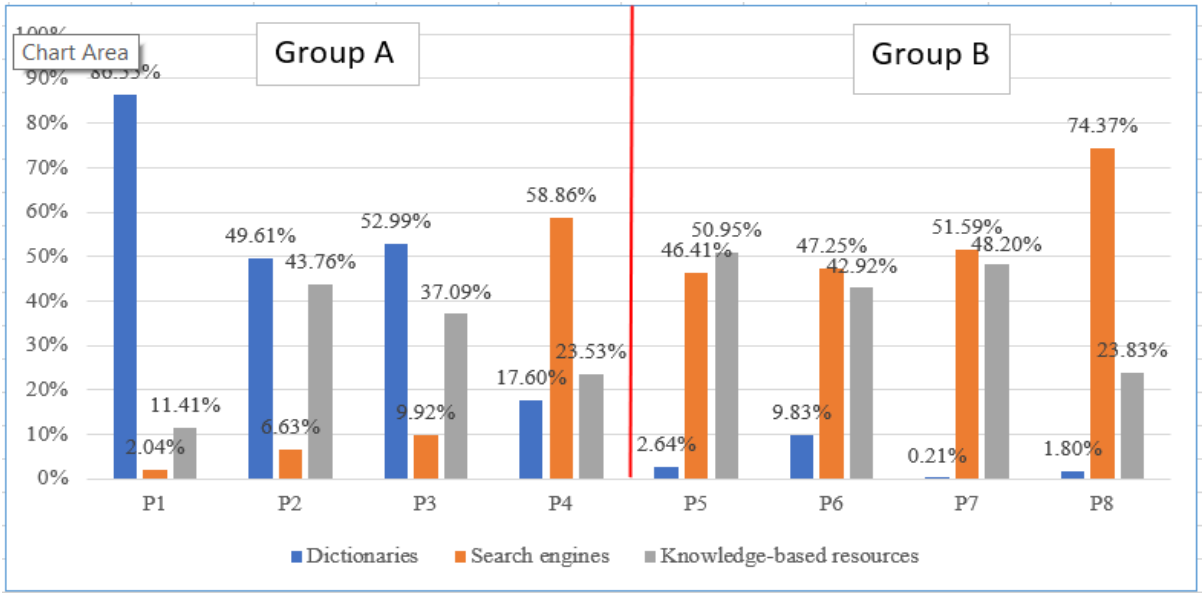

Figure 8 Percentage of time spent in each resource type per participant for Pilot 1

Query behaviour

The operationalization of query behaviours was the same as for the main study, which comprised query time, query complexity and query language direction. As Figure 9 shows for individual data, the mean query time spent by Group A (35.1 seconds) was less than by Group B (54.3seconds).

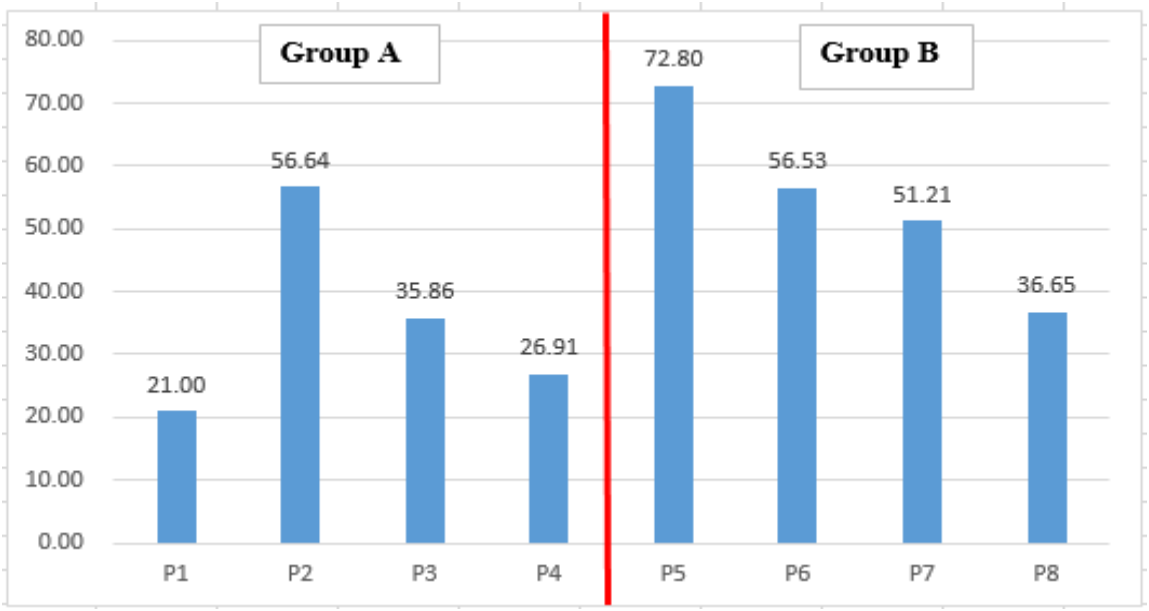

Figure 9 Average query time for Pilot 1 (time in seconds)

The findings for query complexity, shown in Table 4 below, in Group B, except for P5, the participants all used advanced search operators at least ten times, while none of the participants in Group A included any quotation marks or asterisks in their queries. On closer examination, two types of search operator were used, quotation marks (to narrow down search results) and an asterisk. The asterisk was only used by P8 from Group B, and only three from Group B used quotation marks, which may suggest that the instruction they had received on the translation course has informed three of the participants' query construction skills in terms of query complexity.

Table 4 The number of advanced search queries for Pilot 1

Group	**Group A**				**Group B**			
Participant	P1	P2	P3	P4	P5	P6	P7	P8
Number of quotations	0	0	0	0	0	13	10	9
Number of asterisks	0	0	0	0	0	0	0	1
Total number of search operators	0	0	0	0	0	13	10	10

The distribution of queries in terms of source language, target language and queries combining source and target languages is presented in Figure 10.

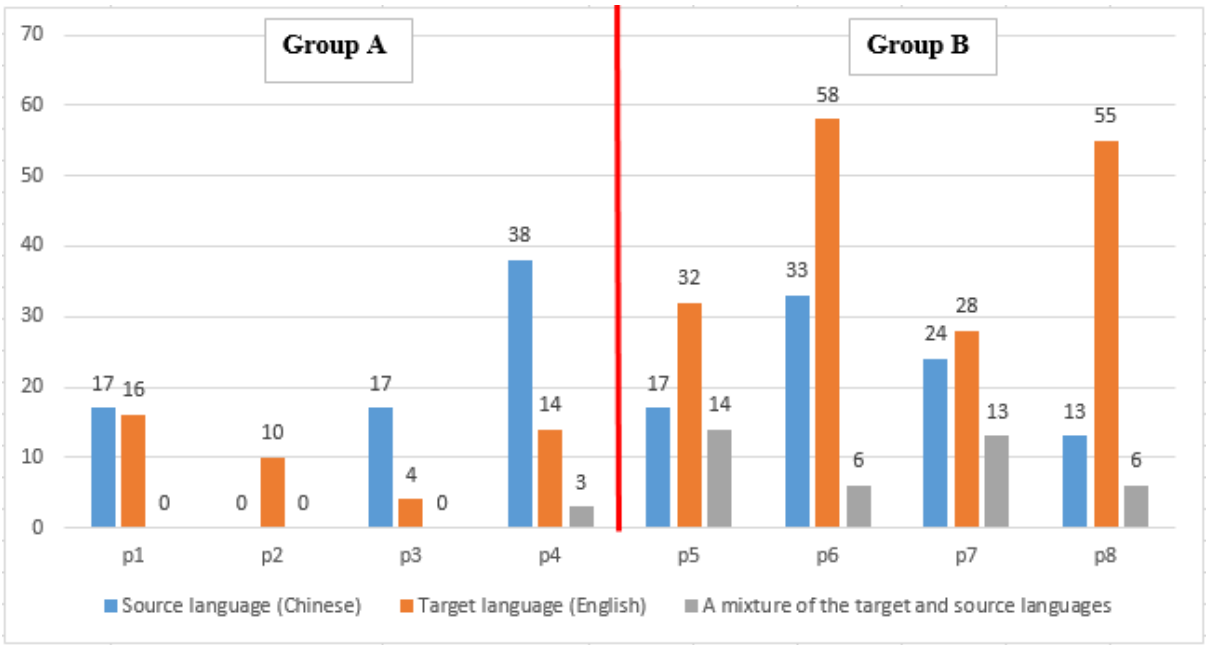

Figure 10 The number of source language, target language, and mixed queries per participant

Participants in Group B constructed more target language queries than source language queries (source language mean 43.25, target language mean 21.75), and participants in Group A performed more source text queries (source language mean 11, target language mean 18) with P6 as the only exception. It is therefore possible to conclude that the search instruction prompted Group B to construct fewer source language queries. In the queries composed of both source language and target language, Group B conducted more (mean 9.75), while in Group A P4, who was the only participant to conduct a combined query, conducted three.

In conclusion, from the first pilot study results, it can be seen that there was a clear difference in participants' resource use and query behaviour for Group A (who had not completed the course) compared with Group B (who had completed the course). The results indicate that the course (web search instruction) had an impact on the participants' resource use and query behaviour. Although there were insufficient participants to conduct inferential statistical tests to examine the significance of the differences, there is tentative evidence of a difference between the two groups, which justifies the need to undertake further research with a larger cohort to further investigate the magnitude and nature of these differences.

4.2.1.7 Discussion relating to eye tracker

The suitability of using an eye-tracker in the first pilot study, particularly the problems encountered and the reasons why it was finally removed from subsequent studies is discussed below.

First, as reported earlier in Section 4.2.1.5 a technical problem arose when processing collected data. The attempt to synchronise the eye-tracking data with the Inputlog data in real time failed. This was because the Inputlog interface connections with Eyelink were not compatible even after several rounds of trials, and numerous consultations with the developer at Inputlog and the engineer at Eyelink. This greatly limited the potential and possibility of using eye-tracking data. This was also observed by Shreve and Angelone (2010) as one of challenges in translation process research to analyse different data streams, such as eye-tracking data, key logging data, and even neural sources (e.g., fMIR, EEG). Saldanha and O'Brien (2014) further comment that independent analysis of each data stream may be one solution. This is particularly relevant to the current research, however if the eye-tracking data were analysed independently from the

Inputlog data, the data volume would then be hard to manage within the scope of this thesis, and also the degree of inference would be decreased.

Second, in addition to the synchronisation issue mentioned above, the intrusion of the eye-tracker posed a potentially detrimental threat to the ecological validity of the study. Since the eye tracker lab was not the participants' normal workplace, they were not familiar with the laboratory environment, which may have compromised their behaviours, particularly their search behaviours. Similar to Ehrensberger-Dow and Perrin (2009) study of professional translators who were less inclined to use the resources at their disposal in the laboratory than in their normal work place. In addition, the different screen layouts have an impact on the participants' eye tracking behaviour. The ecological validity in eye-tracking using different screen layouts was examined by Spinner et al. (2013) who compared the results of the same group of participants with two different screen layouts, and found that screen layout can have a significant effect on eye-tracking behaviours. This indicates that eye-trackers should be used with extreme care since "small changes in textual arrangement, font size, and other display features can lead to large changes in results" (p.409). The fixed screen layout may therefore also affect individuals differently and may therefore have an impact on web search processes and the interpretation of results.

Another issue to be aware of when using eye-trackers is that it is generally found that the reliability of eye-tracking data decreases with time due to the issue of drift. Drift occurs "when the recorded eye position and the true eye position become gradually asynchronous as a data-collection session progresses" (Hvelplund, 2014, p. 210). This phenomenon is particularly relevant for lengthy sessions. For instance, Göpferich (2009) rejected all the eye-tracking data in the Transcomp project because the experiment was too long (one hour), and Ehrensberger-Dow (2014) only analysed data for the first 15 minutes of their Capturing Translation Processes project. In the present research, in order to ensure all the collected data was accurate and reliable, the translation process was interrupted, and the eye-tracker recalibrated every 15-20 minutes. Since the participants spent an average of 57.39 minutes to complete the translation of 192 characters of text in the first pilot study, they were interrupted at least three times during each translation task in order to ensure the collection of reliable data. The influence of these interruptions is unpredictable and hard to measure, and is an inherent disadvantage of the eye tracker.

The last issue identified in the first pilot is purely practical in that using an eye-tracker can lead to a reduction in sample size, and study validity, because of the unreliable quality of the eye-tracking data. According to the quality assessment of eye-tracking data, gaze sample to fixation percentage (GSF), [8] if GSF is lower than 70%, then the eye data would be regarded as unacceptable (Feng 2017). In the first pilot study, three of the participants' data were lower than 70% and discarded.

In summary, after the strengths and weaknesses of an eye-tracker were carefully considered, I decided to remove this data collection tool from the subsequent studies. The Inputlog and screen recordings were however found to be effective in gathering data of resource use and quey behaviour, as further explained in section 4.7.

4.2.2 Second pilot study

In this section, I introduce the second pilot study in the sequence: research aim, the adjustments compared to the first pilot study, research design, and lessons learned in the second pilot study.

The aim of the second pilot study was to test the practicality and validity of the use of testing in a remote setting, and the additional technology and data collection tools. The second pilot study design was revised in line with the technical findings of the first pilot study. Apart from the removal of the eye tracker, other tools — namely, a questionnaire, screen recorder, and retrospective verbal reports — were added in order to obtain a more comprehensive picture of the participants' resource use and query behaviour. A convenience sample of four new participants was recruited for the study from the same cohort of Masters trainee translators from the Graduate School of Translation and Interpretation in Beijing Foreign Studies University. The study took place from 25 July to 28 July in 2020.

The adjustments to the second pilot study design compared to the first pilot study are listed below and the reasons for these changes and how they were applied in the second pilot study is also discussed below:

1) a change from a two-group comparative design to a single-group mixed-methods design;

[8] GSF is calculated as [number of gaze samples / number of fixation gaze samples]*100.

2) the removal of the eye tracker;
3) a change of test venue from a controlled laboratory setting to a remote naturalistic setting which could be any location where the individual participant was most familiar and accustomed to working;
4) the addition of a screen recorder (Screencast-O-Matic 2.5.6 version);
5) a step-by-step video tutorial was included to familiarized the participants with the procedure;
6) a questionnaire was added (to collect data pertaining to participants' background information and their self-declared web searching knowledge);
7) retrospective verbal reports were requested to elicit and record the participants' thoughts about the searches they had conducted after completing the translation of the source text;
8) the translation direction was reversed: from English to Chinese.

The first adjustment was applied due to the research aim of the second pilot study. As mentioned at the beginning of this section, the main aim of this pilot study was to test the practicality and validity of the use of a remote setting and the additional data collection tools, therefore a single-group study design would be more appropriate.

The second adjustment, the removal of the eye tracker, lead to the following three adjustments (from the third to the fifth). The decision to remove the eye tracker is fully discussed in Section 4.2.1.7.

It was possible to make the third adjustment (remote setting) due to the removal of eye tracker. The advantages and disadvantages of using a natural remote setting were considered, as follows. The remote setting allowed the participants to complete the translation task at their accustomed place of work, normally with their own computers or laptops. As mentioned in the ecological validity issue discussed in Section 4.2.1.7, and as Saldanha and O'Brien (2014) also affirm, an unfamiliar environment, such as a laboratory, or even classroom, which contains unfamiliar keyboards, monitors and software may have an impact on a participant's routine behaviour. This phenomenon is known as the observer's paradox (Xiao & Muñoz, 2020). To be more specific, this may elicit an improved performance (often called the Hawthorne effect), or on the contrary a compromised performance (the white coat effect). The remote test setting

allowed the participants to use their own computer set-up, with their preferred fonts, paragraph settings, browsers, and their preferred online resources. The remote location also brought flexibility and allowed the participants to exhibit their normal behaviour, and to complete the tasks at any time within a specific time interval. The added flexibility of remote testing was extremely important in allowing a larger participant cohort and decreased risk of drop-out. There are, however, also some disadvantages to data collection in a remote setting; for example, additional work to manually inspect the raw data. In this study, the start and completion points also had to be determined post-test and all unrelated data manually removed. In all however, the disadvantages of the additional manual data processing required are outweighed by the advantages of ensuring the normal behaviour and environmental validity.

The addition of screen recording, the fourth adjustment, was applied due to the screen recording functions of the discarded eye tracker (see Section 4.2.1.7 for more details).

The fifth adjustment was to ensure the consistency of the participants' preparation, by the addition of a video tutorial to make up for the absence of an onsite researcher.

The sixth adjustment was to add a questionnaire. The appropriateness of the questions asked and the clarity of the wording were first piloted. The background information questions in the questionnaire were successful, and can be found in Appendix C and are discussed further in Section 4.7.1. There was one question concerning the web search information:

Question:

> Are you familiar with any search engine operators (e.g., the operator "OR" to find one or more words in a search query) used to find an exact phrase (i.e., containing all the words in a search query)? Please list a maximum of 5 and describe their function, i.e., what they are used for. If you do not know any search engine operators, you can skip this question.

This question was discarded from the subsequent studies because all four participants in this second pilot study skipped this question. This may be because they were unfamiliar with the word "operator" and the unfamiliar term made the question difficult to answer. Further, participants who may have been familiar with the term may not have been able to think of any examples and hence also skipped the question. The design and rationale for the questions asked

are fully discussed in detail in Section 4.7.1.

The seventh adjustment was the addition of the retrospective verbal reports (required in English). As also mentioned earlier, the key logging and screen recording tools only collect behaviour data, and it is challenging to interpret these data in order to make inferences about the participants' actions and behavioural patterns. To compensate for this caveat, participants in the second pilot study were asked to engage in retrospective verbalization. To maintain consistency with the language used in this thesis and to avoid over-interpretation or under-interpretation, these reports were requested to be given in English.

In the second pilot study, the translation direction was reversed to being from English to Chinese. This was mainly due to the difficulty of measuring the source text, and this is elaborated in Section 4.7.2. The source text in the second pilot study was selected from the economic domain with 192 English words, specifically from the 2017 World Economic Situation Prospects (See Appendix B).

The research design for Pilot Study 2 involved a single-cohort, mixed-method study design, and is shown in Figure 11.

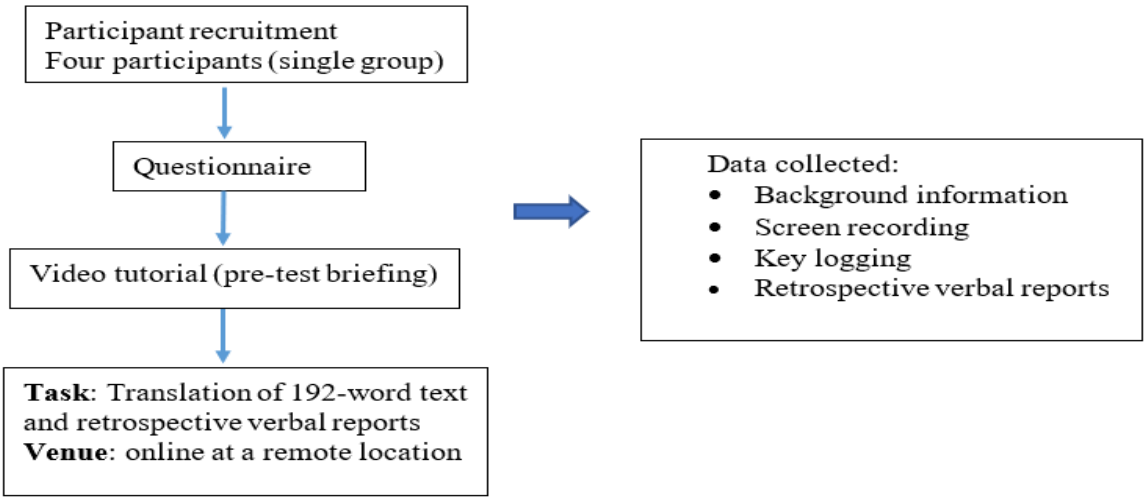

Figure 11 Research design for Pilot Study 2.

In summary, after the adjustments implemented in the second pilot study, the remote setting proved to be feasible, and the data collection tools proved to be efficient. As the second pilot study was near finalization with respect to setting and data collection tools, those are explained

and elaborated in Chapter 4 Methodology. In addition, due to thesis limitation, the results of the second pilot study will not be reported as they do not add to the final data interpretation.

4.2.3 Third pilot study

The aim of the third pilot study was to finalize the research design for the main study. After the first two pilot studies were conducted, the results and findings provided useful feedback. Some further minor adjustments were however necessary to ensure the proposed data collection tools were valid and reliable before commencing the main study.

The study design, data collection tools (including the source text), and procedure were all the same as for the second pilot study, however, the language required for the retrospective verbal reports and for all communications was changed from the participants' second language (English) to their first language (Chinese) in response to inaccurate and hesitant responses obtained in the second pilot study. After transcribing these retrospective verbalizations in the Pilot Study 2, some verbalizations were found to be superficial and inaccurate. For example, when asked to describe the type of information they had searched for, one participant only responded "I am not sure how to translate this term" rather than giving specific examples. This happened even when they had been provided with several examples of the type of information, they might search for in their introductory brief e.g., a source-term definition, a cultural aspect, an acronym, a specific collocation, a fluent idiomatic expression in Chinese, an equivalent term, etc. The participants claimed to have understood the examples provided in the initial instructions, however, it seemed to be difficult for them to describe their search items in some detail. This may have been due to their inability to describe or explain their searches in second language (English), or that they found it difficult to describe their searches in more technical and specific language. This phenomenon is also mentioned by Enríquez Raído (2011), who suggests that participants' may lack the adequate resources and/or terminology to accurately describe their translation problems and web search behaviour. Using their second language to verbalize and self-reflect is challenging for participants and they may lack the vocabulary, confidence and accuracy. Hence, the participants were required to verbalize in their first language (Chinese) in the following studies. The language used in the information sheet and instructional video was also changed to Chinese to ensure the accuracy of their understanding.

The video was also reduced from eight minutes to six minutes, a progress video bar was added to make sure the participants were aware of their whole progress and which stage they had reached, and Chinese subtitles were also added to make it easier to understand. Two new participants from the same cohort as the participants in the second pilot study were recruited for the third pilot study on 30 August 2020.

As a result of this change, the participants were able to complete the whole remote procedure, from downloading the two recording software, filling up the online questionnaire, complete the translation of the source text, verbalize their searches, and uploading the recorded files without any issues.

4.3 Overview of the main study

The aim of the main study was to investigate changes in trainee translators' resource use and query behaviour over the period of a semester. The study was a single cohort mixed methods repeated measures design, with four data collection sessions. The study design was the same as for Pilot Study 2 (Figure 11), except that there were four data collection sessions completed.

As can be seen in Figure 11, after the completion of the questionnaire, the participants undertook four tasks, evenly spaced throughout a single semester to examine their resource use and query behaviour over time. Each session followed the same format, consisting of a single translation task followed by a retrospective verbal report. Data was collected by a screen recorder and a key-logger, the retrospection was obtained verbally and recorded.

In the following sections, I first describe the contextual setting, participant recruitment, schedule and procedure. I then discuss the tools used to collect the data, and describe how the data were processed.

4.4 Contextual setting

In order to investigate the changes of resource use and query behaviour over time, a first-year compulsory course in translation practice offered at the Graduate School of Translation and Interpretation (GSTI), Beijing Foreign Studies University (BFSU), China was chosen for the main study. There are a two year program and a three year program, but the contextual setting

(the course) is the same. This course lasts for sixteen weeks over one semester from September to December and comprises 32 contact hours in total (2 hours a week). The ultimate aim of this course is to develop trainee translators' translation skills in non-literary translation, focusing on (re)search skills and critical thinking skills. The specific learning objectives included improving the understanding of the source text, the expression in the target text, and decision-making processes involved. The main teaching format includes lectures and translation assignments. The teaching contents were related to translation-oriented web searching and comprised an introduction to online search strategies, evaluation of online resources, comparison of trainee translators' web searching paths drawn from trainee translators' assignments, and a presentation of a desirable search process and search outcome by the lecturer. This course was established in 2005, with a recently published course text, *An English-Chinese Translation Course in Beiwai*, for translation educators, researchers, and trainee translators (Li & Wang, 2020).

At this point, I need to specify that, this course only served as a contextual setting for the observation of resource use and query behaviour, and was not the object of this study, nor was it the aim of this study to evaluate the effectiveness of the course. In addition, I was not in any way involved in the teaching, or teaching support for this course.

4.5 Participant recruitment

The participants were all first-year Masters trainee translators at GSTI, BFSU. They were all enrolled in the Chinese-English translation course mentioned above. They all had the same language background of Chinese as their mother tongue and English as their first foreign language. They had all completed a four-year Bachelor of Arts degree, mostly with an English major. More detailed background information, and their self-rating of searching related knowledge are provided in Chapter 5.

The participants were recruited on a voluntary basis and there were no exclusion criteria with all trainee translators in the course being potential participants. Potential participants received an invitation email (with the Participant Information Sheet and Consent Form attached in Appendix D and Appendix E), and trainee translators who were interested in taking part in this research replied via emails to request further participation information.

4.6 Schedule and procedure

Detailed instructions were emailed to all recruited participants (see Appendix F). Before the four data-collection sessions, the trainee translators were instructed to install and familiarise themselves with two pieces of software, i.e., Inputlog and Screencast-O-Matic. They were asked to translate four texts from English to Chinese, of around 200 words for each text, in their normal translation settings (either at the library or at home) on four separate sessions during the course from September 2020 to February 2021. The translation processes were screen recorded and Inputlog logged by the participants themselves, with both pieces of software running in the background without interfering with the participant's normal translation behaviour. There was no time limit, and they were allowed full access to the Internet, and the use of any resources or tools they were able to find on the Internet. After each translation, they were instructed to verbalize their searches according to the Instruction email they had received. The four texts were emailed to them at regular intervals and they were given deadlines by which to complete each translation to ensure the four data-collection sessions were evenly distributed throughout the course. After each data collection session, participants uploaded the screen recording file to the Screencast-O-Matic website, and emailed the Inputlog file (IDFX-file) and translated text (Word file) back to the contact email address.

This study was conducted from September 2020 to February 2021 at regular intervals throughout the course, see Figure 12.

The four sessions were carefully scheduled to avoid the participants' coursework deadlines and examination dates, and to distribute the translation tasks to be completed within ten days to control the exposure to the course contents.

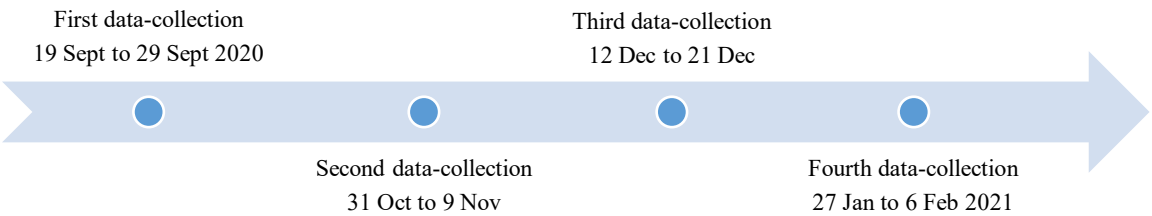

Figure 12 Data collection periods in the main study

4.7 Data collection tools

In this study a mixed-method approach was adopted, combining several data collection methods in order to elicit different types of data, which has become the norm in translation process research (Alves & Hurtado 2017; Halverson 2017, as cited in Xiao & Muñoz 2020). Data collection tools in the main study included: questionnaires, source texts, screen recorder, key logger, and retrospective verbal reports. In the following sections, each is described in detail.

4.7.1 Questionnaire

In translation-oriented web search studies, questionnaires have been widely used to collect pertinent background information on the participants (Enríquez Raído 2011, Gough 2016; Chang 2018). A questionnaire can also take into account the participants' needs, views or self-perceptions related to web searching. Pinto and Sales employed questionnaires to examine the view of trainers about the strengths and weaknesses of students' actual information competences (2008c), the self-perception of trainees (2007), and the needs of translators (2007).

In this present study, the aim of the questionnaire was to gather information concerning the participants' education and translation background and their self-declared web searching knowledge. The questionnaire was administered prior to the first translation task, and is described in detail below.

At the beginning of the questionnaire, I specifically emphasized that there were no right or wrong answers, and all answers would be kept confidential. This is because participants often do not choose the answer which reflects their reality, but rather select the answer that they think the researcher is expecting, or the one which reflects on them best (Pavlovic 2007, Saldanha and O'Brien 2014).

The questionnaire took approximately six minutes to complete. The questionnaire comprises two sections. In the first section, Questions 1 to 8 are about background information, and in the second section, and Questions 9 and 10 are about web searching information. In Questions 1 to 8, the participants were asked about their full name (Question 1), English language

proficiency (Questions 2 to 5), translation proficiency (Question 6), and their full-time translation experience (Question 7 to 8). Except for the question about their full name, the English language proficiency, translation proficiency, and full-time translation experience are all filter questions, which allow participants to only answer those questions that are applicable to them, and the follow-up questions could illicit more detailed information from the answers. In this way the questionnaire is more coherent and saves time for the participants.

Question 2 examines the participants' English language proficiency with four options: the Test for English Major 8 (TEM 8), the International English Language Testing System (IELTS), the Test for English as a Foreign Language (TOEFL), and 'others'. As a test at the level of a bachelor's degree with an English major, TEM 8 is the highest level of English language proficiency and the majority of the participants in the present study had achieved this qualification. Some of the participants also had IELTS and TOEFL qualifications to prove their English language proficiency and to enhance their suitability for employment or further study. Each option had follow-up questions. Questions 3, 4, and 5 requested the grades achieved in TEM 8/IELTS/TOEFL respectively, and if the participants chose the option 'others', then they were requested to specify which other English proficiency qualifications they had acquired.

Question 6 enquired about the participants' translation proficiency. The China Accreditation Test for Translators and Interpreters (CATTI) was chosen as an option as it is authoritative and widely accepted accreditation test in China. The follow up options included: CATTI-I, CATTI-I II, CATTI-III, and the participants could choose as many options as applicable.

Question 7 requested information about the length of the participant's translation experience, asking whether the participants had ever worked as a full-time translator. The term "full-time translator" was specifically chosen because "full-time" is sufficiently specific for participants to decide whether they had this type of experience or not. A response of "Yes" led to a more specific question (Question 8) about how long the participant had worked as a full-time translator, with the options of 'less than 1 year', '1-3 years', and 'more than 3 years. The participants were all trainee translators on the Master Program, and most did not have this type of experience. Some may have experienced a certain type of part-time experience or internship experience, but most of this work is usually underpaid, and they would not have been employed in a formal or professional way, and cannot be counted as having actual or formal translation

experience.

The second section included two matrix questions about web searching. One was about the self-declared types of resources they commonly used when translating, and the other was about their self-perceived level of web search knowledge. The matrix format was chosen because there are eight resources and four web search knowledge components to be rated, and this format is an effective way to cluster these items into two matrices that are easy for the participants to understand and answer.

For the first question in this second part of the questionnaire (Question 9), the resources categories were: search engines, web pages, dictionaries, machine translation, concordancers, encyclopaedias, online documents, and discussion forums, with self-rated scores of: 5 (very often), 4 (often), 3 (sometimes), 2 (rarely), and 1 (never). More details concerning the rationale for resources categorisation can be found in Section 3.1.

The second question in Section 2 (Question 10) concerns web searching and asks the participants to select their self-perceived level of knowledge. From a more general perspective about web searching knowledge, The Open University Digital literacy skills checklist (Open University 2012) [9] comprises four components; understanding digital practices, finding information, using information, and creating information. In translation studies, according to the requirements suggested by Gough (2019), translators should know: 1) what kind of information they are seeking; 2) where to seek; 3) how to seek; 4) when this information can be found. Hence, integrating the two sources, the elements of web searching in this question include: 1) what information can you find on the web; 2) where can you find the information you need, 3) when do you change search strategy, including stopping, 4) how do you evaluate the information you find, with self-rated scores of: 5 (extremely knowledgeable), 4 (very knowledgeable), 3 (moderately knowledgeable), 2 (slightly knowledgeable), and 1 (not at all knowledgeable).

Questionnaires were gathered online using Sojump, [10] a Chinese service provider engaged in

[9] See more information https://www.open.edu/openlearn/pluginfile.php/801340/mod_resource/content/3/Session%205%20Open%20University%20Digital%20and%20Information%20Literacy%20Framework.pdf

[10] See https://www.wjx.cn/ for details

online questionnaires. The Chinese service provider was chosen because, after a few attempts, the regular questionnaire providers used internationally, such as Survey Monkey, and Qualtric were not stable in China's mainland. The full questionnaire is attached in Appendix G.

4.7.2 Source texts

The aim of this research was to investigate the changes in trainee translators' resource use and query behaviour throughout the translation course using four data-collection sessions. In order to attribute the findings to the changes in resource use and query behaviour, and not due to differences in the texts, four comparable texts were selected. It is notoriously difficult to quantify the difficulty or complexity of a source text, and in the following sections, I elaborate on the criteria by which the source texts were chosen below.

Logically, the texts should be comparable in terms of text difficulty. Before beginning this discussion, several concepts need to be clarified. They are "translation difficulty", "translation complexity", and "text complexity". According to the distinctions made by Dahl (2004), difficulty is "a notion that primarily applies to tasks, and always relative to an agent: it is easy or difficult for someone; complexity as "an information-theoretic notion is more objective in the sense of being independent of use and can be computed mathematically" (p.39). To measure text complexity, the most influential and widely accepted framework was proposed by Jensen (2009), which has been used in many translation process research, such as Hvelplund (2011), Quinci (2013), Chang (2018), and Cui and Zheng (2020). This framework was particularly important for the preparation stage of selecting appropriate source texts for translation empirical research. The three text complexity indicators defined by Jensen (2009) were readability indices, word frequency, and non-literalness.

To start with the most important indicator, the readability index is comprised of six further indices: the Flesch-Kincaid, Coleman-Liau, Gunning Fog, SMOG, and Flesch Reading Ease. These are all mainly based on counts of syllables, words and sentences and used different formulae to calculate them (see Table 5).

Table 5 Readability index formulas

Indicators	Formulas
Flesch-Kincaid	11.8*syllables/words+0.39*words/sentences-15.59
Coleman-Liau	5.89*characters/words-0.3*sentences/(100*words)-15.8
Gunning Fog	0.4*(words/sentences+100*((words >= 3 syllables)/words))
Smog index	square root of (((words >= 3 syllables)/sentence)*30) + 3
Flesch reading ease score	206.835-84.6*syllables/words-1.015*words/sentences
ARI	4.71 x (characters/words) + 0.5 x (words/sentences) - 21.43

Readability indices for the four texts used in the main study are summarised in Table 6, where the indicators are given in the first column, followed by the four tasks. It can be seen that all the readability indices fall within a similar range, indicating that the four texts used in the present study can be considered comparable.

Table 6 Readability index for the four texts

Texts	**First task**	**Second task**	**Third task**	**Fourth task**
Title	The Next Great Migration: The Beauty and Terror of Life on the Move	Biodiversity, Sustainability and Human Communities: Protecting Beyond the Protected	The Evolution of Developmental Pathways	The Empire of Depression: A New History
Flesch-Kincaid	14.3	15.8	14.5	15.8
Coleman-Liau	14	15	13	12
Gunning fog	17.3	17.9	17.9	18.3
SMOG	13.2	13	13	14.1
Flesch reading ease	32.4	24.2	32.7	30.5
ARI	15.2	17	14.9	16

Word frequency, as the second indicator in text complexity, was also considered for the four chosen source texts. As has been determined, the more frequently a word is used in a language, the more likely that word is to be known to the participants (Read, 2000). The word frequency in this study was calculated based on two frequency bands in the BNC and COCA corpus.[11] The first was the K1 band with 1-1000 of the most frequently used words, indicating the words

[11] https://lextutor.ca/vp/comp/

are very common. The second band was the K2-K25 band with the 1001-25000 most common words, indicating the words are less commonly used. The word frequency information of the four texts is presented in Table 7.

The four texts can be seen to share a similar word frequency proportion in the two bands.

Table 7 Word frequency of the four texts

Frequency bands	Task 1	Task 2	Task 3	Task 4
K1	67.30%	69.70%	65.10%	68.50%
K2-K25	32.70%	30.30%	34.90%	31.50%

The third indicator of text complexity framework is non-literalness. This refers to the number of occurrences of non-literal expressions, i.e. “idioms, metaphors, and metonyms” (Hvelplund, 2011, p. 91). Since the four texts chosen were general texts as opposed to literal texts, the use of non-literal expressions was infrequent among them, and the total number of non-literal expressions are three, two, three, and two, respectively.

When choosing the texts, the degree of difficulty was also considered. For if the texts were too easy, every translation problem could be solved by using only dictionaries, and the aim of recording the usage of other resources could not be achieved. This was empirically proven in a study by Pokorn et al. (2020) where the participants admitted they did not feel the necessity to search for parallel texts when the text was simple. The texts therefore need to be sufficiently challenging to prompt a variety of resource use and query behaviour, however, if the texts are all too difficult, the changes in the behaviours might be variable and difficult to detect. As observed by Dragsted (2005), the differences between the novice and the professional is hard to capture when translating difficult texts, hence, the texts should not be too difficult nor too easy for the participants.

To ensure their comparability, the four texts were also from the same genre, shared the same function, and were used for the same translation purpose. Pakkala-Weckström (2015) used a tourist brochure, a popularized science column, two food recipes, an EU report, and a page from an art history textbook as a student assignment and examined the resources the trainee translators used, and empirically verified that genre has an impact on resources use. In this

study, the four texts were chosen from the same journal, *Nature*, and the same column, a book review, to ensure their comparability in terms of genre and function. In addition, task description is also related to how the translators approach the translation task, the participants were therefore instructed to translate the text "for publication purposes" across the four tasks.

In the previous three pilot studies, participants on average took around 70 minutes to translate the 200-word texts, which generated 200MB of memory. This is manageable as the length of the texts should allow participants to complete them in a reasonable time, without taking up too much PC memory. The lengths of the four texts of the main study were 211 words, 222 words, 213 words, and 221 words, respectively.

To estimate the text difficulty, two translation trainees with a similar background to the potential participants in the main study translated the four source texts. They rated the difficulty of the four source texts (from one to five) and highlighted potential translation problems or rich points (Nord, 1994) as further indicators of text suitability. The mean value of the estimated translation difficulty for the four texts was 4, which implies that the four texts were of a similar level of difficulty. The four source texts can be found in Appendix H.

4.7.3 Key loggings

In LIS, the core information needed when conducting search behaviour analysis typically involves: "the type of event (e.g., click, query); a timestamp denoting the time that the event occurred and the nature of the event; a unique identifier for the event and for the searcher; and the atomic information element that is being logged (e.g., URL or query text)" (White, 2016, p. 54). Following from best practice in LIS research, the same four types of information was recorded by the key logger chosen in this study, namely Inputlog.

In Translation Process Research (TPR) and Writing Process Research (WPR), the key logger is a tool widely used to record the processes of translation or writing. There are various types of key loggers used in translation studies to record the translation process, and they are often used together with other tools, such as eye-trackers or screen recorders: sometimes as the main tool, and sometimes as a complementary tool. Inputlog was chosen as the main data collection tool in this study. In the following sections, I first introduce the Inputlog, and then elaborate on its advantages and disadvantages.

Inputlog is a logging tool used to record all kinds of input, including from the keyboard, mouse, and program switches, within and outside its interface. The aim of Inputlog in this study was to record the resource use and the query behaviour of the research participants. Inputlog was chosen mainly because of its key function in recording keyboard and mouse activities outside of the Microsoft Word interface, and in providing time-based information for those activities and the window changes.

One of the difficulties encountered by translation scholars who conducted web searching studies without Inputlog involves the manual work of transcribing the online activities, which is prohibitively time-consuming and limits the number of participants that can be reasonably processed. As noted by Enríquez Raído (2011), the "most arduous, labour-intensive, and time-consuming" phase was the transcription of the screen recordings (p.220). Like any other screen recorder, the use of Inputlog helps mitigate this issue by providing time-based information. Daems et al (2016) is one of few studies that used Inputlog to examine differences in the external resources consulted by participants between translation from scratch and post-editing.

In Writing Process Research (WPR), Inputlog has been implemented to investigate web search processes, including the categorisation of external resources and the interactions among them. It is employed to investigate the resource used by a professional communications designer when creating a proposal for a telephone system to answer questions about government (Leijten et al., 2014). In a more recent study, Leijten et al. (2019) used Inputlog to examine how Masters students routinely search for information in source-based writing tasks. Inputlog in their study was used to provide the time the students spent on sources, which sources they most frequently consulted, and how often they switched between sources.

In this thesis, Inputlog was chosen as the main tool to record all online activities, due to its ability to identify and register the web browser used (e.g., Google Chrome), the active URL (e.g., www.google.be), the page title, the search queries used to activate the search, and the resulting web pages accessed subsequently. Another advantage of using Inputlog is that participants are already familiarized with its word processor (i.e., Word editing window) prior to their participation. This gives Inputlog a clear advantage over certain other log tools, such as Translog, which has its own interface, with half of the screen set up as the source language window and the other half as the target language window. When a system is new for

participants, the unfamiliar writing environment/window is very likely to have an impact on normal behaviours, and therefore "have (an) impact on research validity" (Saldanha & O'Brien, 2014, p. 133). Unlike Inputlog's logging environment which consists entirely of all MS Windows programs, Translog and Scriptlog are confined within their own editors (Scriptlog text editor and Translog editor). In addition, the size and position of the windows in Inputlog can be adjusted by the participant as personally required or preferred. Inputlog can therefore be viewed as a user-friendly tool, which can also be used remotely by the participants themselves in a normal workplace (Leijten et al., 2014).

Some screen recorders have basic key-logging functions, which can record keyboard activities, mouse clicks, and cursor movements with a timestamp, such as BB FlashBack Pro, however, Flashback's key-press data is restricted to keyboard and mouse data, compared with the Windows information, or query information available with Inputlog. In translation-oriented web search, Enríquez Raído (2011) used the BB Flashback's built-in keystroke-logging programs to create a time log of the keyboard and mouse activities as a basis for creating an online actions spreadsheet to complement the screen recording data. Onishi and Yamada (2020) also used BB FlashBack to record web search processes on this basis, and the spreadsheet was then enriched with data pertaining to searching behaviours such as typing, deleting, modifying, pausing, and searching. These basic key-logging functions can to some extent reduce manual work, however, Inputlog can record more detailed data.

The Inputlog does contain some drawbacks though; for example, it can only be used in a Windows operation system instead of an IOS operation system. In addition, it cannot log the use of translation memory or the actual queries entered into dictionary applications, but this can be compensated for by the simultaneous use of a screen recorder. This aspect will be further discussed in Section 4.7.4.

In all, Inputlog was particularly suited for the present study because it logs all interactions between the participants and the online resources consulted, it is unintrusive and behaves in a familiar way, maintaining the ecological validity of the data.

4.7.4 Screen recordings

Screen recording has become a customary tool in translation process research due to its

completeness and non-invasiveness (Angelone 2019). It is also generally considered to be the most commonly used method to record all on-screen activities in translation-oriented web searching research (Enríquez Raído 2011, Massey & Ehrensberger-Dow, 2011a; Massey & Ehrensberger-Dow, 2011b; Gough 2016; Hvelplund 2017, Kuznik 2017, Shih 2017, Olalla-soler 2018; Kuznik & Olalla-soler 2018; Chang 2018; Enríquez Raído et al. forthcoming). As Lauffer (2002) points out as early as 20 years ago, screen recording "was most useful in following the search paths and helping understand how and why the translators searched for information" (p.69). To be more specific, as explained by Göpferich (2009): "screen recordings seem to be particularly useful for analysing the research activities which form an integral part of translation processes, as they provide a detailed account of which electronic sources or web-sites the subjects are using during translation." (p. 173; as cited in Enríquez Raído 2011, p. 132). It is however still recognized as a vital tool to investigate the web search process (Enríquez Raído 2022). In the present study, a screen recorder was used as a supplementary tool rather than the main tool due to the advantages of the Inputlog elaborated in Section 4.7.3.

By recording all screen activities, the screen recorder registers and stores all the web search activities, including the dictionaries used, and the web browsers or searches undertaken in other applications. It can therefore compensate for the inability of Inputlog to register certain activities in the Translation memory tool or machine translation use as mentioned in Section 4.7.3.

A number of different screen recorders have been used in translation studies. Among them, the two most frequently used have been Camtasia (Gopferich 2009; Volanen 2015; Fernández 2015; PACTE 2017; Olalla-Soler 2018;) and BB Flashback (Enríquez Raído 2011; Shih 2017; Chang 2018; Onishi and Yamada 2020). Others, such as Screencast-O-Matic (Gough 2016; Zapata 2016; Angelone 2019); and Screen2Exe (Wang 2014); Morae (Witczak 2021) have also been used.

Two screen recorders were used in Shih (2019), QuickTime for Mac Operation system and BB flashback for Windows operation system, where consistency between the two systems was a pre-requisite. The Screencast-O-Matic (see https://screencast-o-matic.com/) has often been chosen because it can be accessed through a web browser directly and participants can upload recordings to the Internet instead of sending a large video file via email. It can also be used

both with the Mac and MS Windows operation systems, and is generally considered more user-friendly than the two mainstream systems Camtasia and BB flashback (especially when the participants need to use the screen recorder remotely and by themselves). Screencast-O-Matic was therefore chosen for this study, and since the free version only permits sessions of less than 15 minutes to be recorded, the premier team version was used, which allows 50 computers to access the account at the same time with no recording time limit.

4.7.5 Retrospective verbal reports

The direct observation of translation process research with key logger and screen recorder is now possible, however it is still difficult to infer what is in the participant's mind, which although not a focus of this study, is helpful in interpreting the results of the present study. Retrospective verbal reports were used as a supplementary tool to get a richer understanding of participant's resource use and query behaviour. Below I explain the rationale for choosing when the commentaries would be conducted (retrospective vs. concurrent), and in which form (verbal vs. written). A detailed description of the contents and instructions given to participants can be found in Appendix F.

For each task, the participants were asked to highlight the items in the source text that they searched for. Then, they were asked about their familiarity with the source text, and how difficult they found the source text to be. The participants were instructed to complete their retrospective verbal reports after the completion of each translation task. There were several reasons for asking the participants to report their thoughts after they had completed the translation tasks. The aim of the study was to compare trainee translators' resource use and query behaviour over the period of a taught course, and time-related factors (such as the time spent on each main type of resources) and post-test feedback are extremely important. Concurrent verbalization, however, would prolong task completion time due to the time used for verbalizing (Ericsson & Simon, 1984). Gough (2016) used a screen recorder to record the translation process, and participants were instructed to verbalize their thoughts at the same time as translating, and although the impact of concurrent verbalization on time-related factors was minimised by "subtracting the time where comments were made with no observable research activities", the slowing down of research activities when accompanied by audio-commentary was not possible to account for. The methodological shortcomings of concurrent "thinking

aloud" is that the additional verbalisation processes of talking can interfere with the translation process, raise awareness in problem solving or distract the participants. In a post-task questionnaire, designed to determine whether the Think Aloud Protocol (TAP) had an impact on the translation process and whether the TAP raised participants' awareness, the majority of the participants stated that the TAP compromises the translation problems, however some of them also thought that it raised their awareness, Gough (2016). Gough further commented that the TAP component also contributed to incomplete data.

In some qualitative and less time-related studies, however, concurrent verbalization worked well. For example, it helped to identify successful and unsuccessful searches, and pinpoint the importance of a knowledge of search engines. It can also identify the influence of affective factors such as frustration and mental fatigue (Shih, 2019; Chang 2018)

As emphasized in several places in this thesis, the importance of maximising ecological validity and maintaining participants' normal behaviour was paramount, however, concurrent verbalization would compromise the translation process itself (Jakobsen, 2014). Retrospective verbalization has been identified as an ecologically valid alternative to concurrent verbalization (Hansen, 2006) and, based on these considerations, retrospective commentaries were chosen for the methodology in the present thesis.

The time-consuming nature of the repeated testing elements of this study could easily increase the chances of participants dropping out, and the need to make the procedure as simple and as easy to complete as possible was imperative. The oral form of commentary was therefore chosen instead of a written format, for practical reasons, for when compared with the written form, the oral form is easier for the participants to complete. In addition, the oral format is a better approximation to an ongoing verbal commentary when compared with the written form to compensate for the "incompleteness" of TAP.

4.7.6 Video tutorial

As mentioned earlier, this study was conducted remotely (without the presence of the researcher) and the participants were therefore asked to install and use the two pieces of software (Inputlog and Screencast-O-matic) before completing the four translation tasks. In addition to reading the emailed instructions, the step-by-step video tutorial was also provided

to compensate for the researcher not being on site. Participants downloaded the video tutorial from a widely used Chinese video platform,[12] using the link provided in the instruction email. The video tutorial was 5-minutes-52-seconds in length.

The video tutorial had two parts, the first was an introduction to the whole procedure, and the second part consisted of a demonstration.

In the video tutorial, slides with supporting text were designed to introduce the procedures, including how to install the two pieces of software, instructions on how to translate the text, how to describe their thoughts according to the Instructions provided in the Instruction email, and how to upload the screen recordings and upload the Inputlog file. The demonstration provided step-by-step instructions for the participants, including examples of the types of web search related problems they might encounter such as abbreviations, people's names, and background information checks. It also introduced the available resources including search engines, web pages, encyclopaedias, and dictionaries. In the video tutorial a progress bar was added at the bottom of the screen to inform participants of their progress in terms of the four sections labelled: preparation, translation, description, and finishing.

In summary, the video tutorial was proven to be very successful in this study, especially for those with limited computer expertise.

4.8 Data processing

In this section, I first describe how I cleaned the collected data, and then elaborate on how I processed and coded the data.

4.8.1 Inclusion and exclusion criteria

In total, 42 participants were recruited in the main study, and the cohort for the main study comprised 19 participants. This is because, of the 42 participants, only 20 had access to a personal computer with the Windows operating system, required to install Inputlog (cf. Section 4.7.3). Hence, the data for the remaining 22 participants were not included in this thesis.

[12] https://www.bilibili.com/video/BV13541187Ef/

Another criterion was whether the individual participants had followed the instructions they had received by email and the video tutorial, which is an inherent risk in remote testing. Participants were required to complete the translation first, before describing their searches, however, one participant began the description as soon as they started translating and continued throughout the translation, thus having misunderstood 'retrospective verbal reports' as 'concurrent verbal reports'. This participant was consequently eliminated from the study after the first data-collection session.

Due to the remote setting in this study and the fact that there was no researcher at hand while the participants completed their tasks, all the collected data had to be thoroughly screened to check the validity and completeness of the data, as well as the integrity of the files. Following each of the four data collection sessions, the resulting screen recordings were viewed, and the key logging data were pre-processed to detect any problems with continuity or missing data. If the screen recording data, or key logging data were found to be problematic, the corresponding participant data was eliminated from the study. The immediate screening of the data helped to decrease the risk of collecting unusable data. This step only meant the preliminary screening of the data files; their further processing is explained in Section 4.7.3.

4.8.2 Questionnaire

The questionnaire results were downloaded and converted into an MS Excel file from the Sojump website. In the first section, concerning the participants' background information, the data were directly processed in Excel. The data from the second section concerning the participants' self-declared types of resources they commonly use, and their self-perceived level of web search knowledge were transferred into corresponding Likert scales for further quantitative analysis. The results from the questionnaire are presented in Chapter 5.

4.8.3 Process data analysis

As elaborated in detail in Sections 4.7.3 and 4.7.4, two tools were used to record the task completion process: Inputlog and Screen-o-Matics. In the next sections I describe how I processed the Inputlog files using its built-in analysis functions, and then how the Inputlog files and screen recording files were combined into a single file for each of the 76 observations,

which was later coded to address the research questions.

4.8.3.1 Keylogging data

Figure 13 presents a detailed flowchart of the Inputlog data processing adapted from the Inputlog manual (Leijten et al., 2019). Five modules, which are logging, pre-processing, analysis, post-processing, and replay, can be seen at the top, bottom, and on the right-hand-side of Figure 13. The elements used in this study are represented with solid lines, while the dotted line elements were not used. To be more specific, ID and time filter, Merge IDFX, Window filter, and general analysis function used are presented and explained below:

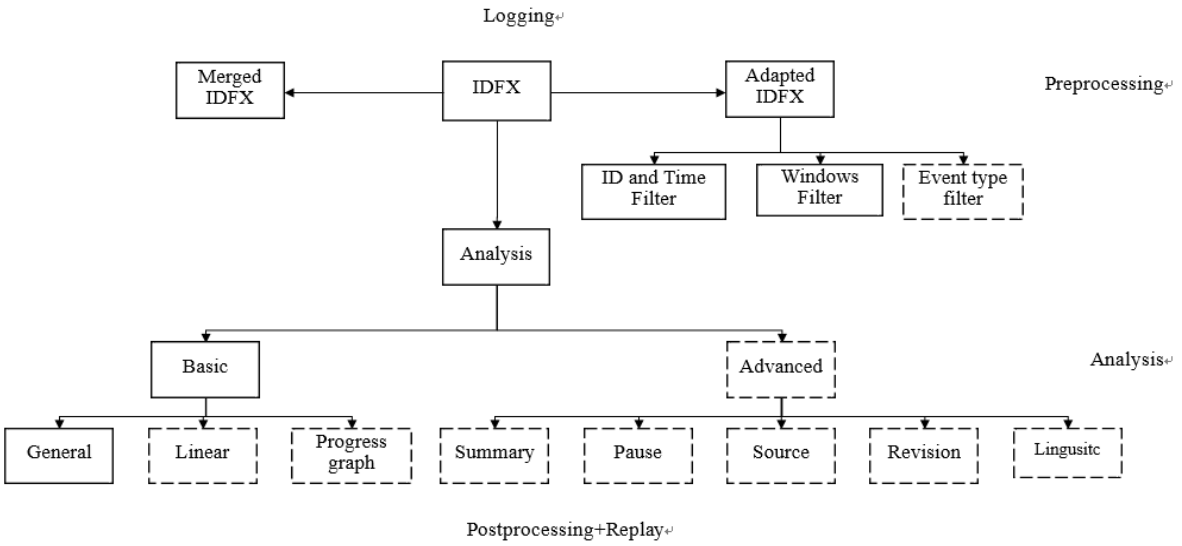

Figure 13 Detailed flow of Inputlog (adapted from Inputlog Manual)

- ID and time filter: crop the start/end time of logging files
- Merge IDFX: merge multiple logging files
- Window filter: select one or multiple Windows to include in further analyses
- General analysis: provides an XML file with a basic log file of the writing session in which every line represents an input action. For every input action (letter, function, mouse click or movement, voice input) the session information is stored together with an identification (ID) of the input, and the time stamps (in clock time and in seconds).

After the files had been completed by the participants and uploaded according to the instruction email and the video tutorial, these original Inputlog files were retrieved, cleaned and cropped to the start and end points of the translation process. The start and end points were defined as from the time the participants opened the source text Word file until they last clicked on the "save" button in MS Word. In addition to the recording of the translation process itself, the key logging files also included some periods of time outside the translation process, such as the process of highlighting the searches and the retrospective verbal reports which were conducted after translating the texts (see more in Section 4.6 about the procedure). In this way, those parts not included in translation process were cropped.

On two occasions, participants did not complete the same source text at one-time, once because the participant's laptop ran out of power and once because the participant had a course to attend before they had completed the whole translation task. For those two occassions, the separate Inputlog files for the same task and the same participant were merged.[13] The 76 valid Inputlog IDFX files were then filtered to extract any extraneous data that had come from outside the MS Word/Inputlog interface.

The data pertaining to the interaction between the participant and the source and target texts, and the data pertaining to the interaction between the participant and online resources were then separated. The Word external interface enables interactions with Internet browsers, such as Internet Explorer or Microsoft Chrome and other applications, such as the dictionary applications Lingoes and Oulu, and a new IDFX-file was generated for each participant recording all interactions with online resources. After pre-processing the 76 Inputlog files (19 participants x 4 tasks) the data were ready for further analysis.

The 76 Inputlog files (IDFX-file) were analysed and an XML file obtained for each one. This is a basic log file in a spreadsheet format where every row represents a separate input action of the keyboard or mouse, or a switch between windows (Leijten & Waes, 2019). The XML file was then converted to an MS Excel file to enables further analysis. The original XML file contained a lot more information, and only the four columns needed for the purposes of the present study were retained for later use (see Table 8). The column 'ID' contains the unique

[13] Completing one task at different occasions might impact the translation process to some extent, but the impact is minimized as the interruption time is limited (in one day).

number of each event (consecutively); the column 'startClock' represents the time in hours:minutes:seconds:milliseconds; the column 'type' contains the event type, such as keyboard or mouse click; and the column 'Window' which contains the type of window in use, and records scrolling and other cursor movement actions. A segment of spreadsheet is presented below in Table 8, which represents a period of time from 19m 35s 203ms to 20m 9s 328ms from a participant's test.

Table 8 Example of Inputlog file

ID	startClock	type	Window
2062	00:19:35.203	focus	Google Search - Google Chrome
2063	00:19:35.234	mouse	Movement
2064	00:19:35.953	mouse	Scroll
2065	00:19:35.953	mouse	Movement
2066	00:19:35.984	mouse	Scroll
2067	0:19:36	mouse	Movement
2068	00:19:38.453	mouse	LEFT Click
2069	00:19:38.727	keyboard	LCTRL
2071	00:19:39.120	keyboard	LCTRL + V
2072	00:19:39.352	keyboard	RETURN
2073	00:19:39.471	focus	regulatory genes - Google Search - Google Chrome
2074	00:19:40.047	mouse	Movement
2075	00:19:43.578	mouse	LEFT Click
2076	00:19:44.547	mouse	Movement
2077	00:19:44.859	focus	Regulator gene - Wikipedia - Google Chrome
2078	00:19:56.781	mouse	Scroll
2079	00:20:05.578	mouse	Movement
2080	00:20:07.531	mouse	Movement
2081	00:20:09.328	mouse	Movement

This spreadsheet can be used to reconstruct the web sesarch process. For example, in Table 8, at ID 2062, which happened at 19m 35s 203ms, the participant opened Google Search. After this, from ID 2063 to ID 2067 there were mouse movements. At ID 2071 and ID 2072, the participant used the keyboard shortcut "LCTRL + V", and pressed the RETURN key. At ID 2073 the participant entered "regulatory genes" into Google Search, and then at ID 2077, this participant viewed the contents in the Wikipedia. After this there were some more mouse movements to ID 2081).

Among the four columns, the column headed "Type" is particular important, it records the event type, including mouse activity, keyboard activity, and focus. According to Inputlog, a focus event in "Type" column is "a technical approach to the source interaction, in which each

computer window opened by the writer is registered and identified in the log file" (Leijten et al. 2019, p.565). Practically this indicates a window change. As shown in Table 8, there were three "focus" events (shown in red). ID 2062 indicates an activity took place in Google search, ID 2073 was in search engine result page (SERP) with the query "regulatory genes", and ID 2077 happened in a Wikipedia page. In this way, the column representing "focus" (or change into a new window) serves as the foundation for setting up the minimum unit in Inputlog, the "search step", which is similar to Gough's terminology of "research step" (Gough 2016), where a research step is "each instance of 'clicking' into another resource, or a sub-section of a resource, constitutes a 'step'" (p.64). In the current study, each instance representing a "focus" in column Type constitutes a search step. The "search step" and the "focus" are perfectly served for one can be considered as the definition, and the other as its operationalization. In the Excel file each row, or minimum unit, therefore represents a discreet search step, enabling further analysis. An example of search steps as represented in Inputlog can be seen in Table 9.

Table 9 An example of search step

A	B	C	D
ID	startClock	Type	output/Window
2062	00:19:35.203	focus	Google Search - Google Chrome
2073	00:19:39.471	focus	regulatory genes - Google Search - Google Chrome
2077	00:19:44.859	focus	Regulator gene - Wikipedia - Google Chrome

4.8.3.2 Screen recordings data

All the screen recording files were downloaded as MP4 files from the Screencast-O-Matic website/platform. Each screen recording contained two parts: the translation process, and the description of their searches according to the instructions they had received. This gave 76 screen recordings for the translation tasks and 76 screen recordings for retrospective verbal reports. The former were used in further coding of the Inputlog files, and the latter were transcribed through Speechmatics (see https://www.speechmatics.com/), and manually checked and prepared for further analysis.

4.8.3.3 Coding in the spreadsheet files

The prepared Excel files provided numerous possibilities to investigate the web search process.

For the purposes of this study, the focus was mainly on the participants' resource use and query behaviours.

Resources

As elaborated in Section 3.1.2, a two-level resources categorisation was applied in this analysis. Table 10 is an extension of Table 9, and built on search steps, where the first level category and the second level category were coded with the resource categorisation outlined in Section 3.1.2.2. This categorisation is based on the "Window" column, extracted from the Inputlog. In this example, the participant first conducted a search in Google Chrome, for which the first level category was search engines, and the second level category was Google Search. This participant also consulted Wikipedia, which, according to the rationale made in Section 3.1, belongs to the knowledge-based resources at first level category.

Table 10 Example of resource category coding

ID	startClock	Type	output/Window	First level category	Second level category
2062	00:19:35.203	focus	Google Search - Google Chrome	search engines	Google search
2073	00:19:39.471	focus	regulatory genes - Google Search - Google Chrome	search engines	Google search
2077	00:19:44.859	focus	Regulator gene - Wikipedia - Google Chrome	knowledge-based resources	Encyclopedias

According to the "First level category" and "startClock" columns, the time spent on one resource per participant and per task was automatically calculated using the MS Excel code (see Appendix I). The same code was also used for the second level category of resources. The processed data could now be analysed to answer the first research question concerning the changes in time spent on different types of resources over the four tasks.

Search queries

As defined in Section 3.2, the scope of the search queries includes all of the search queries conducted in the web browser as well as in dictionary applications. Table 11 is an extension of Table 9, where a "Query" column has been added to indicate which query the participants constructed. In row ID 2073 it can be seen, in the column headed "Window" that, the participants constructed a search for "regulatory genes" in Google search, and this query was added as the query in the column headed "Query" in row ID 2073.

Table 11 Example of query coding

ID	startClock	Type	output/Window	First level category	Second level category	Query
2062	00:19:35.203	focus	Google Search - Google Chrome	search engines	Google search	
2073	00:19:39.471	focus	regulatory genes - Google Search - Google Chrome	search engines	Google search	regulatory genes
2077	00:19:44.859	focus	Regulator gene - Wikipedia - Google Chrome	knowledge-based resources	Encyclopedias	

As mentioned in the discussion of the disadvantages of Inputlog, it cannot log the queries from outside the web browser. In the present research, this is specifically relevant when applied to dictionary applications. In the following, I give an example of a query of this type by presenting the screenshot taken from screen recording file (Figure 14) and the Inputlog file (Table 12). As shown in Figure 14, it is clear that this participant used the 欧路词典 (Oulu Dictionary) and input "Sadowsky" as a query. This query cannot, however, be logged by Inputlog in Table 12, which means the query "Sadowsky" cannot be found in the Inputlog file, however, if the precise time was recorded when the participant first opened the Oulu dictionary, then this can be used to identify the use of the shortcut "LCTRL+V" as a query in Oulu dictionary, where from Inputlog file alone this query could not have been determined. It is then necessary to go back to the screen recording, and visually observe what the query was from the keyboard activities which are recorded. For example, in the red rectangle in Figure 14 below, this query cannot be found directly in the spreadsheet generated by the Inputlog's general analysis function (see Section 4.8.3). However, Inputlog logged the participants mouse activity, "LCTRL+V". Then, in combination with the use of screen recording data, the query "Sadowsky" constructed in the dictionary application was identified.

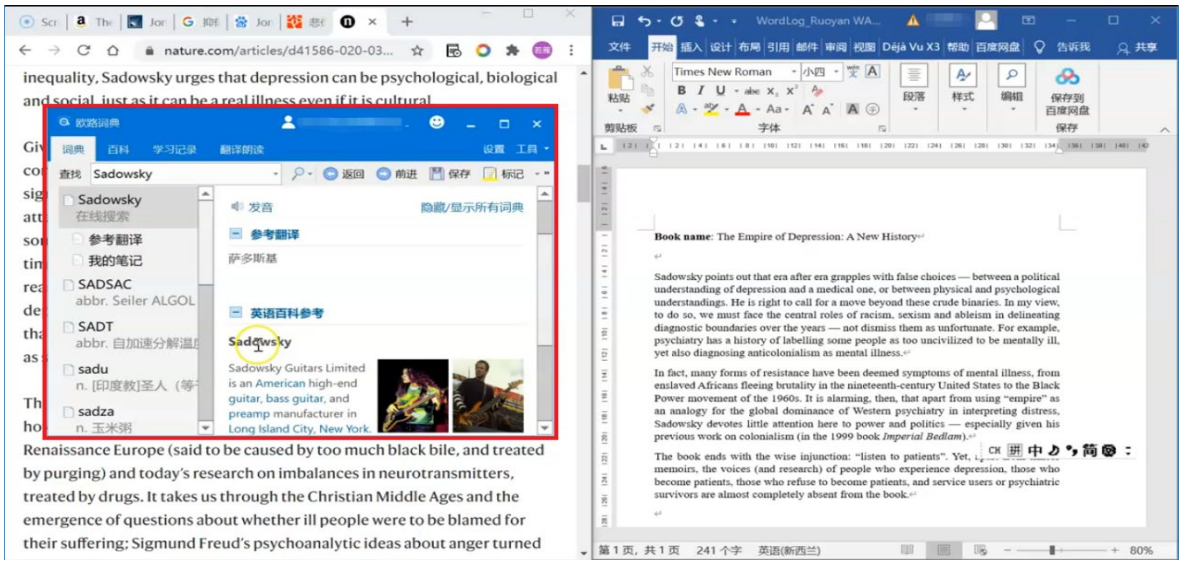

Figure 14 Example of screen shot of dictionary application

Table 12 Example of Inputlog where the log query cannot be seen in applications other than the web browser

	A	B	C	D
	id	startClock	type	Window
	1029	00:09:30.812	focus	欧路词典
	1030	00:09:30.922	mouse	LEFT Click
	1032	00:09:31.031	mouse	Movement
	1033	00:09:31.331	keyboard	LCTRL
	1035	00:09:31.440	keyboard	LCTRL + V
	1036	00:09:31.453	mouse	Movement
	1037	00:09:31.625	mouse	Movement
	1038	00:09:32.528	keyboard	RETURN
	1039	00:09:33.062	mouse	Movement
	1040	00:09:34.297	mouse	Scroll
	1041	00:09:36.062	mouse	Movement
	1042	00:09:36.390	mouse	Scroll
	1043	00:09:36.390	mouse	Movement
	1044	00:09:36.422	mouse	Scroll
	1045	00:09:36.547	mouse	Movement

With the supplementary role of screen recording, all the queries were identified in the column headed “Query”, enabling further analysis.

4.9 Statistical methods

Before elaborating any of the statistical methods used, it is necessary to point out that the thesis subtitle described this as “a statistical analysis”. The aim of this section is to ensure the transparency of the statistical analysis in a scientific way, however readers without specific knowledge of these statistical procedures can skip this section without affecting their understanding of this thesis, as the statistical results are explained in plain language below each table and figure.

As detailed in Chapter 4, the raw data were cleaned, processed, and coded. In total, the final data set comprised the data from 76 observations (19 participants × 4 tasks). The 19 participants completed the four tasks independently, which guaranteed that the data points in each group were independent of each other. Descriptive and inferential statistics were used to address the research questions and to analyse the variables which are summarized in terms of the types of data and the codes used in the models, as summarized in Table 13. The statistical analysis procedures were conducted in RStudio Version 1.4.1106. (RStudio Team, 2020).

Table 13 Types of response variables

Response variables	Types of data	Variable name
Time spent on dictionaries	continuous	dictionary_time
Time spent on search engines	continuous	searchengine_time
Time spent on knowledge_based resources	continuous	knowledge_time
Percentage of time spent on dictionaries	continuous	dictionary_pct
Percentage of time spent on search engines	continuous	searchengine_pct
Percentage of time spent on knowledge_based resources	continuous	knowledge_pct
Time spent per query	continuous	query_time
The number of advanced queries	discrete	advanced_count
Number of source language queries	discrete	SL_count
Number of target language queries	discrete	TL_count
Number of mixed queries	discrete	MIXED_count
Percentage of source language queries	continuous	SL_pct
Percentage of target language queries	continuous	TL_pct
Percentage of mixed queries	continuous	MIXED_pct

4.9.1 Outliers

Before any statistical analysis could be undertaken, the outliers needed to be handled cautiously. Boxplots were used to detect and visually present potential outliers, using the '`boxplot.stats()`' function in grDevices package (Tukey, 1977).

Example:

```
outlier_searchengine_time <- boxplot.stats(websearch$searchengine_time)$out
```

According to boxplot outlier analysis any data point more than 1.5 inter-quartiles above the third quartile or below the first quartile, was regarded as a potential outlier. If a potential outlier is identified, the original files (keylogging files and screen recording files) should be scrutinised for possible causes. If an outlier is caused by technical or measurement errors (e.g. the data from the screen recorder or the key logger was corrupted during the process), and does not measure what it should measure, it should be removed (Larson-Hall, 2016). However, in the present study, none of the potential outliers were caused by technical or measurement error. In contrast to technical errors, the potential outliers can also be "unique" data points, which "contribute to a complete understanding of the phenomenon of the study" (Newton & Rudestam, 1999, p. 197). In the present study, these unique outliers represented a longer time

spent on certain resources or more queries constructed in a certain language direction; they were a part of the data produced by the participants, and therefore retained in the dataset.

4.9.2 Descriptive statistics

To present the descriptive analysis results, figures were used to portray overall trends and to give a visual representation of the data sets, specifically bar charts, pie charts, line charts, boxplots, and effect plots. Bar charts were used to demonstrate the underlying characteristics of the data sets; pie charts were used to present the percentage of the time spent on the three main resources and the percentage of the number of three types of language direction queries employed during the total four tasks; line charts were used to observe the trend of percentage change (percentage of time spent on each main resource and the number of each query direction) during the four tasks; boxplots (also known as box-and-whisker plots) were used for task comparisons in terms of median and interquartile range. Tables were used to summarise central tendency (mean and median) and dispersion (minimal, maximal, standard deviation, and interquartile range). The '`Summarise()`' function from the dplyr package (Wickham, 2021) was used for descriptive statistics, however, descriptive statistics can only provide "surface features" of a dataset, and more in-depth inferential statistics were needed.

4.9.3 Inferential statistics

Inferential analyses were used to investigate whether the differences (between the four data collection sessions) observed in descriptive analyses were significant. In this thesis, mixed-effect regression models were chosen. This includes both fixed and random effects. The fixed effect variable is systematic and holds across all observations, while the random effect variable has an idiosyncratic effect (Mellinger & Hanson, 2017). In this present study, a random-effect variable was also included due to the variation between participants, which helps to explain more of the residual. The individual differences were visualized with the '`xyplot()`' function from the lattice package (Sarkar, 2008). Mixed-effect regression models were chosen instead of traditional ANOVA due to the aim and the setting of this study. The significance of the differences found between the four data collection sessions was then compared, representing a quasi-naturalistic research design. This assumes a high degree of experimental

control (Balling, 2008; Hvelplund, 2011), and reduces the sample size when dealing with missing observations via listwise deletion (Brown, 2021).

To be more specific, Linear Mixed-effects Models (LMMs) and Generalized Linear Mixed-effects Models (GLMMs) were adopted (McCulloch et al., 2008; Balling & Hvelplund 2015). The two are extensions of a simple linear model. A LMM was applied for variables with continuous data (time-related variables) while GLMMs (count-related variables) was used for discrete data, as shown in Table 13. The mathematical equations for LMM and GLMM are as follows:

LMM mathematical equation

$$y_{ij}^{1/2} = \beta_0 + a_i + bt_j + \varepsilon_{ij}$$

In this equation, $y_{ij}^{1/2}$ is the time and percentage spent on dictionaries/search engines/knowledge-based resources; query time; and the percentage of source/target/MIXED language queries. $i = 1, \ldots, n = 19$ for participant, j = 1, 2, 3, 4 for four tasks, β_0 is the intercept, $a_i \sim N(0, \sigma_a^2)$ independently. Here, t_j can be modelled as equally-spaced time points (as a continuous variable), else as a factor with levels called t_1, t_2, t_3, t_4.GLMM mathematical equation

$$\log \mu_{ij} = \eta_1 = \beta_0 + \alpha_j + \delta_i$$

$\mu_{ij} = E(Y_{ij})$ is the number of the source/target/MIXED language queries, β_0 is the intercept, a_i are the task effects (j = 1, 2, 3, 4) treated as a factor, δ_i are the participants' (random) effects for $i = 1, \ldots, n = 19$, $\delta_i \sim N(0, \sigma_a^2)$ independently. Here, $Y_{ij} \sim \$Pois(\mu_{ij})$.

GLMM (for both Poisson and negative binomial) is the above GLMM equation coupled with

$$\log k = \eta_2 = \beta_{00}$$

Where β_{00} is the intercept, and k is the index parameter for handling overdispersion with respect to the Poisson. Then $Y_{ij} \sim \$NBinom(\mu_{ij}, k)$. If k is infinity, then the negative binomial becomes the Poisson. This model assumes a common index parameter over the entire

data set. It is justified because the data is hard to permit an overly complicated analysis involving random effects or covariates affecting k.

As the aim of the study was to investigate changes in dependent variables between the four tasks, the explanatory variable (also known as the independent variable) was the Task ID (T_ID). The variables measured in this study were response variables (also known as the dependent variables). The random-effect variable in this instance was Participant ID (P_ID). The Task ID (T_ID) and Participant ID (P_ID) were categorical variables, and all other response variables were numeric. The linear mixed-effects models were constructed using the '`lmer()`' function from lmerTest package (Kuznetsova et al., 2017). An example of this model was defined as below:

```
searchengine_time_lmem <- lmerTest::lmer(searchengine_time) ~ T_ID + (1|P_ID) , data = websearch
```

The information on the left side of "<-" is the model name for the particular variable of interest, in this case the time spent on search engines. On the right side, "lmerTest" is the package used, and "lmer" is the name of the function used to fit the linear mixed-effects model. The function in brackets, is the variable of interest in the model: here T_ID is the explanatory variable, "(1|P_ID)" specifies the random effects and "data = websearch" indicates the data frame in this thesis.

Assumptions were checked for each LMM model and transformed with a Box-Cox transformation (Box & Cox, 1964). This consists of a family of power transformations that incorporates and extends the traditional options (e.g., log or square root mentioned above) to help researchers easily find the optimal transformation method and remediate deviations from the assumptions of the linear regression models. The '`boxcox()`' function from the MASS package was used (Venables & Ripley, 2004). Taking time spent on search engines as an example, in Rstudio, Figure 15 is the output of a Box-Cox transformation of time spent on search engines.

```
boxcox(searchengine_time ~ T_ID, data = websearch, lambda = seq(-0.25, 1.25, length = 10))
```

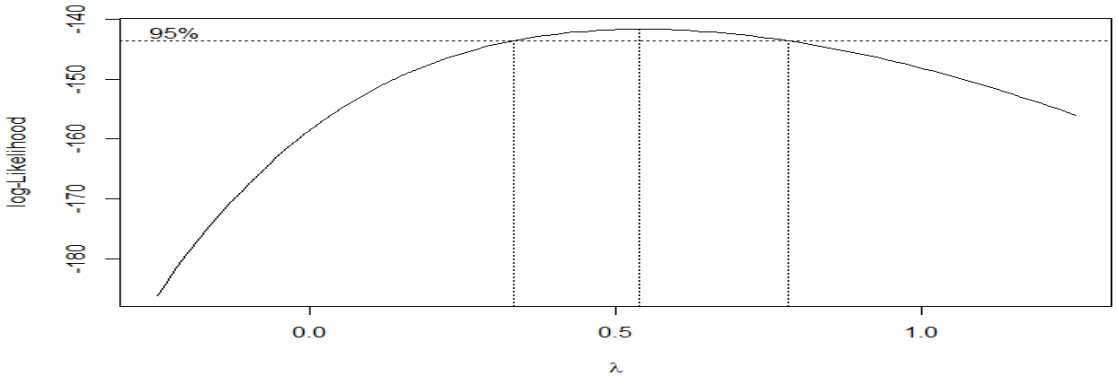

Figure 15 Box-cox transformation of time spent on search engines

The horizontal axis represents the transformation power λ and the vertical axis represents the log-likelihood of the corresponding power transformation. The optimal transformation is the one with the highest log-likelihood. It is easy to see that the optimal/ideal value of λ is about 0.5. The two vertical dashed lines on either side of the optimum are confidence limits; any value of λ in between are 'permissible'. It was shown that 0.5 (square-root) was good for most responses. Then, according to

> λ = 1.00: no transformation needed; produces results identical to original data
> λ = 0.50: square root transformation
> λ = 0.33: cube root transformation
> λ = 0.00: natural log transformation
> λ = -1.00: reciprocal (inverse) transformation and so forth. (Osborne, 2010, p. 4)

The square root transformation would be a good choice for this variable. The rest of the variables were transformed by the same technique.

The overdispersion of each Poisson GLMM model was checked using the '`check_overdispersion`' function from performance package (Lüdecke et al., 2021). If the overdispersion was detected, then the negative binominal model was also built to compare the "goodness-of-if" (Levshina 2015, p. 194) between the poisson model and the negative binominal model. To be more specific, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were compared for the two models (ibid). For example, the following Table 14 shows the comparison between the AIC and BIC for the number of the

query consisting of the source language. It can be seen that the AIC and BIC value of the model SL_count_glmernb is smaller than the model of SL_count_glmer. Thus, the SL_count_glmernb is a better model.

Table 14 Example of the comparison of GLMM models (poisson and negative binominal)

	AIC	BIC
SL_count_glmer	578.9	590.5
SL_count_glmernb	550.1	564.1

The GLMM models built in the thesis were all checked, the better models are chosen and presented in the results below.

5. Questionnaire results

In this section, the results of the questionnaire are reported. The rationale for the questionnaire design can be found in Section 4.7.1, and its processing in Section 4.8.2. The structure of this section logically follows the structure of the questionnaire. The participants' background information is described first, including their English language proficiency, translation proficiency, and translation experience. Then, the participant's self-rating of their frequency of online resource use for translation is presented, and their perceived level of searching for translation-oriented knowledge is described.

5.1 Background information

The participants' background information concerning their proficiency in the English language. To guarantee the participants' anonymity, I used "P (participant)+number" throughout this thesis.

Table 15 Participants' declared level of English proficiency and translation experience

Participant	TEM 8	IELTS	TOEFL	CATTI	Translation Experience
P1	N.A.	7.0	N.A.	CATTI-II	None
P2	N.A.	8.0	N.A.	CATTI-II	None
P3	N.A.	N.A.	N.A.	CATTI-II	None
P4	N.A.	7.5	N.A.	CATTI-III	None
P5	60	N.A.	N.A.	CATTI-III	None
P6	N.A.	N.A.	N.A.	CATTI-II	None
P7	N.A.	N.A.	N.A.	CATTI-III	None
P8	N.A.	8.0	N.A.	CATTI-II	None
P9	72	N.A.	N.A.	CATTI-II	None
P10	N.A.	N.A.	N.A.	CATTI-III	None
P11	78	N.A.	N.A.	CATTI-II	Less than one year
P12	N.A.	8.0	N.A.	CATTI-II	None
P13	N.A.	N.A.	N.A.	CATTI-III	None
P14	N.A.	N.A.	N.A.	CATTI-III	None
P15	N.A.	N.A.	N.A.	CATTI-III	None
P16	N.A.	N.A.	N.A.	CATTI-III	None
P17	N.A.	7.0	N.A.	None	None
P18	N.A.	N.A.	N.A.	CATTI-III	None
P19	N.A.	N.A.	N.A.	CATTI-II	None

*N.A.: not applicable

The most suitable criteria for the participants' English proficiency would normally have been considered to be the Test for English Majors-Band 8 (TEM 8)[14] because it is the most widely accepted test in China. The TEM 8 is usually taken in the fourth year of an English bachelor's degree. In addition, this would normally have been the most recent test taken before the data collection sessions in 2019 and 2020, however, the TEM 8 was cancelled in 2020 due to Covid 19, and the majority of the participants were unable to sit the exam. Only three participants who took the exam in 2019 in the fourth year of their bachelor's degree were able to provide their TEM 8 results.

IELTS is another test used to measure participants' English language proficiency. Six participants took this test, with their overall band scores ranging from 7 to 8. Specifically, two participants scored 7, one participant scored 7.5, and the remaining three scored 8. The IELTS website classifies candidates with scores of 7 and 8 as a "good user" and "very good user" of English. [15] A third English proficiency test is the Test of English as a Foreign Language (TOEFL),[16] but none of the participants in this dataset had taken it. The participants' level of English proficiency could therefore not be estimated on the basis of previous tests, however, their successful pass of the post-graduate entrance exam could, to some extent, indicate that their English proficiency was at the same level.

To rate translation proficiency, the China Accreditation Test for Translators and Interpreters (CATTI) is the most authoritative and widely accepted accreditation test in China.[17] Among the 19 participants, 18 were certificated either at CATTI-II or CATTI-III (around intermediate level), which indicates they had obtained a similar level of translation proficiency.

Hardly any of the participants had any full-time experience, as expected, since they were still full-time students, with only one participant having experience of full-time translation, which had lasted less than one year.

[14] http://tem.fltonline.cn/

[15] https://takeielts.britishcouncil.org/teach-ielts/test-information/ielts-scores-explained

[16] https://www.ets.org/toefl

[17] http://www.catticenter.com/cattiksjj/1848

Based on the information above, it is reasonable to assume that the participants had similar backgrounds, concerning their English proficiency, translation proficiency, and translation experience.

5.2 Self-declared frequency of resource use

Figure 16 presented the participants' self-declared frequency of online resource use vis-à-vis different types of resources for translation purposes is presented, here eight types of online resources are listed, with self-rated scores of: 5 (very often), 4 (often), 3 (sometimes), 2 (rarely), and 1 (never). The justification for this categorisation is elaborated in Section 3.1.1.

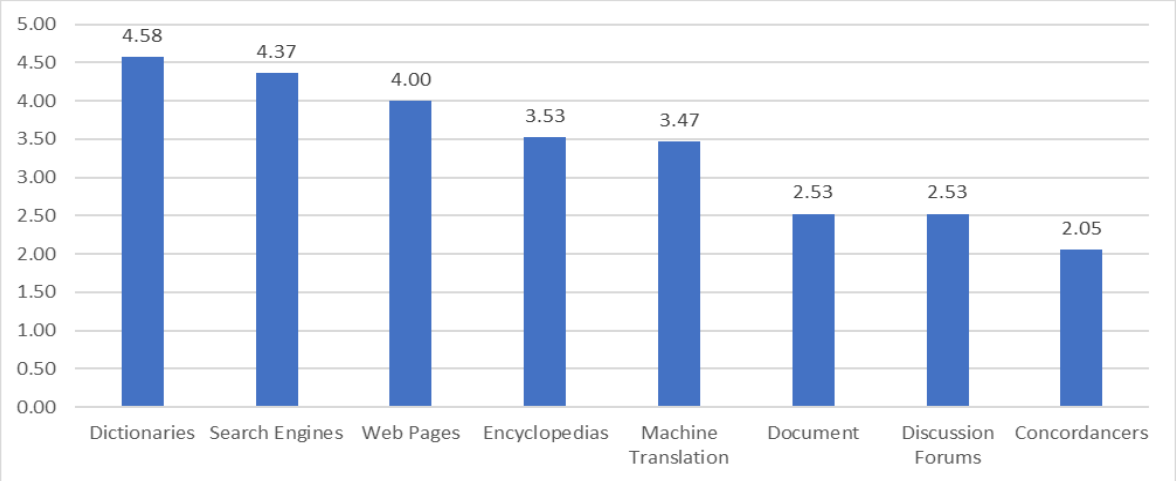

Figure 16 Self-declared frequency of online resources

Dictionaries were found to be the most used online resource from the eight types of resources, followed by search engines and web pages. The average of these three resources was rated between four and five (mean 4.32), which suggests they were used "often" and "very often". The average of Encyclopaedias and machine translation (mean 3.5) were rated between three and four, suggesting they were used "sometimes" and "often". The remaining four resource types, i.e., knowledge, document, discussion forums, and concordancers were, on average (mean 2.37), rated either two or three, suggesting they were only accessed "sometimes" or "rarely".

5.3 Self-perceived level of web search knowledge

Finally, the participants were also asked about their level of knowledge in terms of: a) what information they could find; b) where they could find information; c) when to change search strategy, including stopping searching; and d) how to evaluate the information, with self-rated scores of: 5 (extremely knowledgeable), 4 (very knowledgeable), 3 (moderately knowledgeable), 2 (slightly knowledgeable), and 1 (not at all knowledgeable). These results are shown in Figure 17.

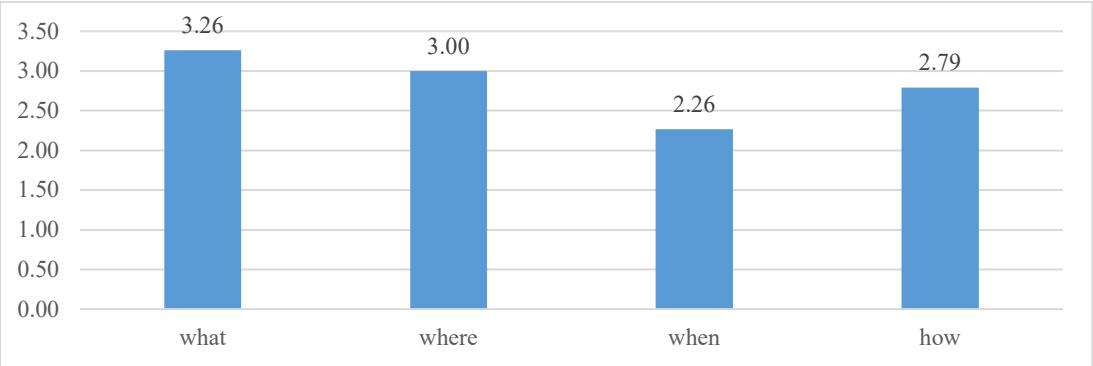

Figure 17 Self-perceived level of web search knowledge

In Figure 17, it can be seen that the highest self-rated element was the knowledge of what information they could find, which scored 3.26, while the lowest self-rated element was the knowledge of when to change the search strategy scored 2.26. How to search scored 2.79 and where to search scored 3.00. In the light of the data, the participants seemed to be ill-equipped with enough knowledge for translation-oriented web searching, and weak decision-making skills concerning when to change strategy (including stopping searching) were identified.

The self-declared data concerning participant's resource use and web search knowledge are only briefly reported here and are discussed in greater detail in Chapter 6, together with the observed data collected in the four data-collection gathering sessions.

6. Resources use

In this chapter, I address the first research question relating to the changes in the use of resources. The resources used in the four translation tasks by 19 participants (4*19 observations) were categorised into three main types and then further classified into more detailed resources in a 'bottom-up' fashion. The first level of analysis is based on their functions, and the three main resources were: dictionaries, search engines, and knowledge-based resources. At the second level, each main category was further classified into 22 different dictionaries, six different search engines, and five types of knowledge-based resources. The detailed justification of this categorisation can be found in Section 3.1.

In the following section, I give an overview of the distribution of the main types of resources accessed during the four tasks. Then, I present and interpret the changes in the time spent (and percentage of time spent) accessing each main resource during the four sequential tasks. After this, more detailed analysis within each main resource is investigated. To conclude the chapter, results from the questionnaire related to resource use will be triangulated with findings from screen recordings and key data logs.

Figure 18 gives an overall summary of the resource distribution by presenting the mean time per task participants spent accessing each resource type throughout the four tasks, in terms of absolute value and the percentage contribution to overall resource time.

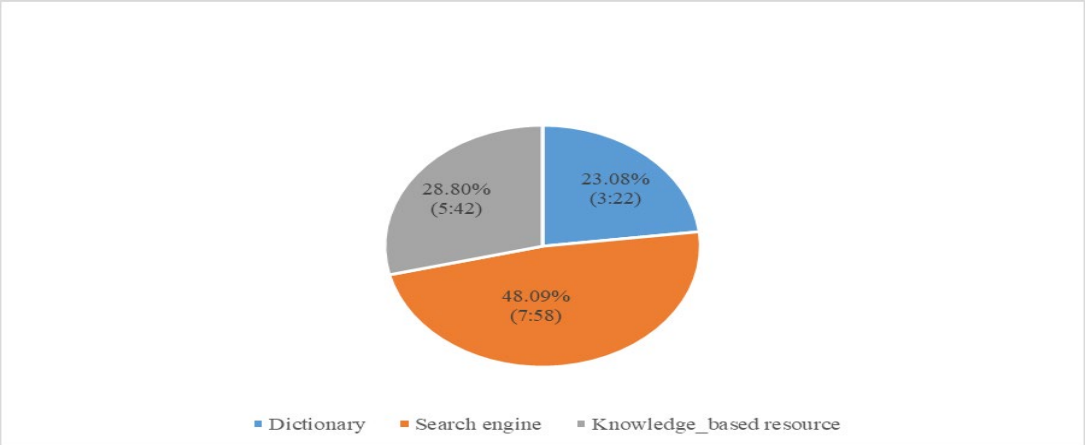

Figure 18 Mean time spent accessing the three main resources for the four tasks (time in minutes:seconds)

It is clear that search engines were the most used resource with participants spending 48.1% of their total resource time accessing them; the average was 7min 58s for one task per participant. The knowledge-based resources were the second most used resource, comprising 28.8% of the total resource time; with a mean of 5min 42s. Dictionaries were the least used resource, only taking up 23.1% of the combined three types of resource, with participants only spending 3min 22s on average per task accessing dictionaries of any kind.

To investigate whether there were any differences of resource use between the four tasks, the data were further broken down by each task and resource. In addition, to facilitate comprehension, the variety of each resource, for the ad hoc four tasks and four tasks separately, were also examined.

6.1 Dictionaries

Despite the fact that the results presented above showed that the dictionary was the least used resource in this study, this was one of the most hotly debated resource in translation-oriented web searching studies between 2000 and 2010 (Bowker & Pearson, 2002; Livbjerg & Mees, 2002; Pastor & Alcina, 2010; Varantola, 1998). In this section I report on time spent, and the percentage of time spent on dictionaries, and then closely examine the different dictionaries the participants used.

6.1.1 Time spent consulting dictionaries

Boxplots were created to visually compare the distribution of time spent on consulting dictionaries during the four tasks (see Figure 19).

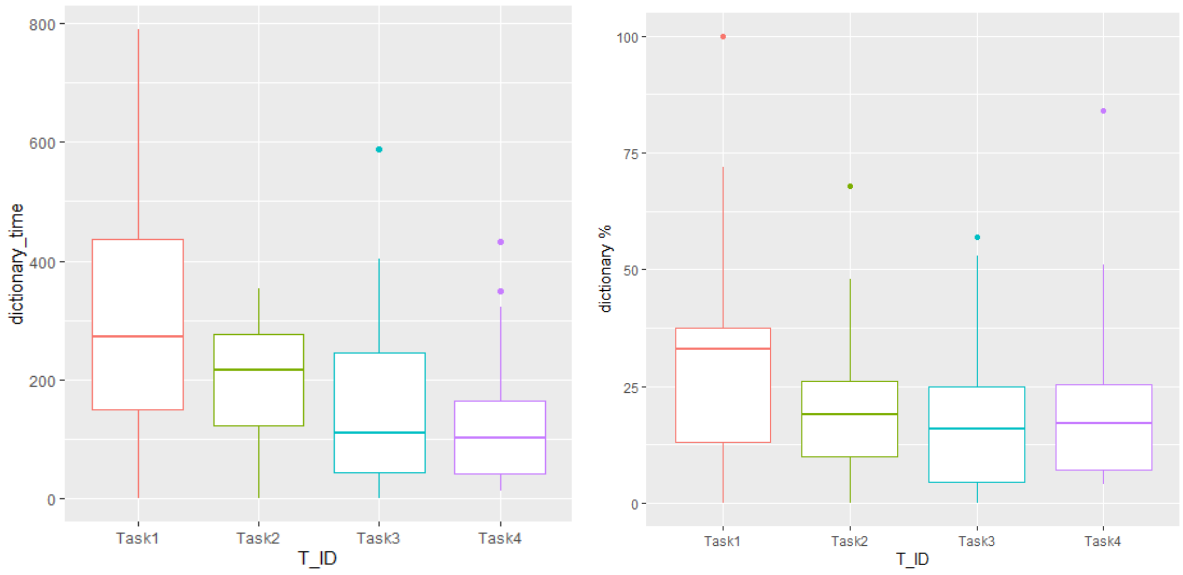

Figure 19 a) Time spent (time in seconds) and b) Percentage of time spent on dictionaries during the four tasks

In Figure 19, the horizontal line in the box interior denotes the median, and the upper and lower boundary of the box represent the 75% and the 25% quartiles respectively, also called the upper quartile (UQ) and lower quartile (LQ) (Murrell, 2005). The box's whiskers extend up to 1.5 times the interquartile range (IQR) beyond each quartile but equalling the smallest and largest values within that range. Potential outliers outside this range are represented by individual data points, as in Task 3 and Task 4 (represented by the blue dot and purple dots, respectively). To further investigate the three potential outliers, as mentioned at the beginning of this chapter, the original data (the Inputlog file and screen recording file) were examined for potential causes. Each outlier was checked with its corresponding Inputlog file and screen recording file. It turned out that on these three occasions the individual participants spent longer accessing dictionaries than the rest of the participants, and the outliers did not appear to have been caused by measurement error and were therefore considered valid data, hence, the three data points were retained. In Figure 19athe boxplot shows that time spent on consulting dictionaries decreased progressively from Task 2 to Task 4, and in Figure 19b it can be seen that the percentage of dictionary consultation time also decreases across the four tasks.

Table 16, below, presents the descriptive statistics of the time spent (dictionary_time) and percentage of time spent (dictionary_pct) consulting dictionaries, including the mean, median,

minimum, maximum, standard deviation (SD), and interquartile range (IQR) for the four tasks.

Table 16 Descriptive statistics of time spent (time in minutes:seconds and the percentage of time consulting dictionaries)

T_ID	dictionary_time					dictionary_pct				
	mean	sd	IQR	min	max	mean	sd	IQR	min	max
Task 1	5:12	3:57	4:48	0:00	13:10	30.21%	25.39%	24.50%	0.00%	100.00%
Task 2	3:13	1:56	2:34	0:00	05:53	22.63%	20.16%	16.00%	0.00%	68.00%
Task 3	2:40	2:37	3:22	0:00	09:49	18.21%	17.66%	20.50%	0.00%	57.00%
Task 4	2:20	2:03	2:02	0:13	07:12	21.26%	20.26%	18.50%	4.00%	84.00%

For the tasks, the mean time spent consulting dictionaries decreased noticeably from Task 1 (5min 12s) to Task 4 (2min 20s), with a reduction of 2min 52s. The change from Tasks 1--2 is worthy of note, representing a reduction of 1min 59s. The changes became less notable however from Task 2 to Task 3 (33s), and from Task 3 to Task 4 (20s). The SD, IQR, and maximum duration of dictionary usage all experienced a downward trend. In terms of the dictionary time percentage, the decrease from Task 1 to Task 2 amounted to 7.58% (30.21%-22.63%), and compared with Task 1, Task 3 and Task 4 decreased 12% (30.21%-18.21%) and 8.95% (30.21%-21.26%) respectively. One interesting finding is that the maximum for dictionary percentage is 100% which indicates that one of the participants used dictionaries exclusively in Task 1 while this did not happen in Task 2, Task 3, or Task 4.

Regression analysis was then used to examine whether the decreases observed in the sample were real and reflected true differences in the hypothetical population. In this case, this was whether the differences in dictionary access time or dictionary access percentage between the four tasks observed in Model 1 and Model 2 were statistically significantly different. The underlying null hypotheses test various possible relationships between the tasks and the response. As discussed in Section 4.9, linear mixed-effects models were used. Model 1 was built with Task ID (T_ID) as the explanatory variable and the time spent on dictionaries as the response variable. In addition, the Participant ID (P_ID) was the random-effect variable. Similarly, Model 2 was built with the percentage of time spent on dictionaries as the response variable.

Model 1 Regression model relating time spent on dictionaries

```
lmerTest::lmer(sqrt(dictionary_time) ~ T_ID + (1|P_ID) , data = websearch)
```

Model 2 Regression model relating the percentage of time spent on dictionaries

```
lmerTest::lmer(dictionary_pct ~ T_ID + (1|P_ID) , data = websearch)
```

The results for Model 1 and Model 2, are presented in below.

Table 17 Results of Model 1 and Model 2

	sqrt(**dictionary_time**)[18]				**dictionary_pct**		
Predictors	Estimates	CI[19]	p	Predictors	Estimates	CI	p
(Intercept)	15.73	12.78 – 18.67	<0.001	(Intercept)	0.3	0.21 – 0.40	<0.001
Task 2	-3.23	-6.32 – -0.14	0.041*	Task2	-0.08	-0.14 – -0.01	0.031*
Task 3	-4.61	-7.70 – -1.52	0.003**	Task3	-0.12	-0.19 – -0.05	<0.001***
Task 4	-4.92	-8.01 – -1.83	0.002**	Task4	-0.09	-0.16 – -0.02	0.011*
Marginal R2 / Conditional R2	0.082 / 0.495			Marginal R2 / Conditional R2	0.043 / 0.749		

*p < .05. **p < .01. ***p < .001.

Table 17 summarises the results for Model 1 which represents the differences in square-root transformed time spent consulting dictionaries for Tasks 2-4 relative to Task 1, and Model 2 which represents the differences in the percentage of dictionary time for Task 2-4 relative to Task 1. Taking the result of Model 1 (dictionary time) as an example, the Predictors column lists the explanatory variables. The "(Intercept)" refers to Task 1 as this is the baseline or reference group. For the p (p-value) column, the first p-value tests to what extent the true mean value is 0 for Task 1; this was not meaningful for the present situation. The remaining p-values compare Tasks 2-4 to Task 1. As an example, the p-value for Task 2 is 0.041, which indicates a significant difference between Tasks 1 and 2 (the significance criterion being $p<0.05$); this difference cannot be interpreted as Task 2 being significantly different from other tasks such as Task 3.

[18] Sqrt denotes a square-root transformation across this thesis.

[19] CI means "Confidence Interval'' across this thesis.

The differences in time spent consulting dictionaries during Task 2, Task 3, and Task 4 were all statistically significant compared to Task 1 (p=0.041, p=0.003, p=0.002). Specifically, as an example, the Estimates for Task 2 (-3.23) imply that participants in Task 2 spent approximately (3.23) ^2 = 24.2s less dictionary time than Task 1 on average. Since this is a mixed model with a random-effect variable, there were two R^2 quantities: the Marginal R^2 and Conditional R^2. The marginal R^2 only explains the fixed-effect variable; whereas the conditional R^2 explains both the fixed and random effects. In this case, the marginal R^2 explained the effect of the Task ID (T_ID), and the conditional R^2 explained both the Task ID (T_ID) and the Participant ID (P_ID). The marginal R^2 was 0.082, and the conditional R^2 was 0.495. The small marginal R^2 and the moderate conditional R^2 support the choice of entering Participant ID (P_ID) as the random-effect variable.

To address the first research question, the focus should be on the conditional R^2, which means that the whole model explains 49.5% of all the total variation. The effect plot can be seen in Figure 20. The bars are the confidence limits of the 95% CI; these have the form Estimate +/- 1.96 SE. In terms of the differences pertaining to the percentage of time spent on dictionary, the p-values were < 0.05 (p=.031, p=.001, p=.001) with an R^2 of .749, which formalised the significant difference.

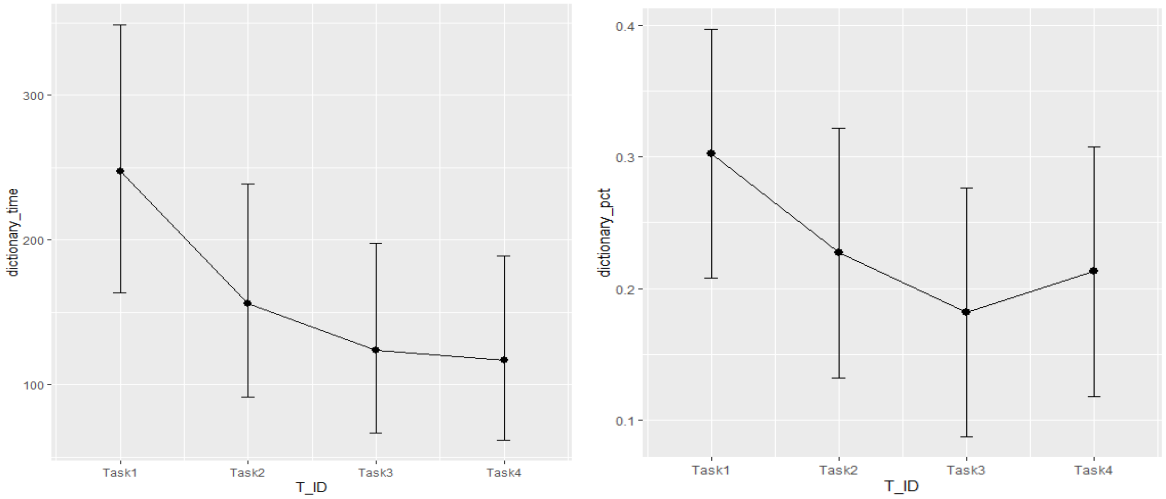

Figure 20 Effect plot of task on a) Time spent (time in seconds) and b) Percentage of time consulting dictionaries

Based on the results above, both the time spent consulting dictionaries and the percentage of

time spent on the dictionary consultation decreased significantly in Task 2, Task 3, and Task 4.

The results are in line with previous literature, including developmental (e.g., longitudinal) and non-developmental studies (e.g., novice vs. professional). For developmental studies, Chang's longitudinal study lasted one-year and included three data collection points in a postgraduate translation course. Over the three tasks, the frequency of the use of dictionaries decreased from Task 1 to Task 3. These results are also similar to a study by Paradowska (2020) which comprised two data collection points, before and after a four-month intervention (a general translation course). In this study ten participants were analysed together at group level, and more closely at the individual level. At group level, the mean percentage decreased from 23.8% to 15.0%, which inferential statistical analysis found to be significantly different. At individual level, eight out of the ten participants in the study significantly decreased their reliance on bilingual dictionaries.

In addition to the two studies mentioned above (Chang 2018 and Paradowska 2020) which took three and two measurements from the same cohort participants, previous literature comparing different groups (e.g., students and professionals) has also found decreasing reliance on dictionaries with more expertise group. In a study by Wang (2014), with three groups of participants, novices, semi-professionals and professionals, the percentage of the use of dictionaries decreased in relation to the level of expertise. Similarly in a study by Onishi and Yamada (2020), participants translated a 62 words text, and students were found to spend an average 1 min 30s longer on dictionary consultation, than professionals (2min 28 seconds and 1 min 48s respectively). The general consensus in the literature mentioned above therefore strongly suggests that novices rely more on dictionaries compared with professionals, and that pedagogical interventions related to web searching can also reduce this heavy reliance on dictionaries.

Since the use of dictionaries as a resource has been found to be so consistent in the literature, it is necessary to further investigate the usefulness of dictionaries and to determine whether the decreased use of dictionaries with experience impacts upon translation outcomes. A seminal empirical study of dictionary limitations was conducted by Livbjerg and Mees (2002), using think-aloud and Translog data collection methods. In this study, there were 115 verbalisations

related to dictionary consultation in total. Among them, participants felt only 66 dictionary consultations were profitable. Of the 66 consultations, the level of accuracy was retained in 40 (29 correct solutions were retained or changed into another correct solution, 11 errors were retained or changed into another error). Of the 26 consultations where the level of accuracy changed, 5 decreased and 21 increased the accuracy of the translation. Of the 21 improvements, nine were simple spelling corrections, one participant felt the dictionary helped him solve a problem, but the phrase he had a problem with was not found in the dictionary. Four other corrections were about the cultural transfer, which the dictionary could not solve. In all, there were, at best, seven "genuine quality-improving" (p.131) consultations among the 115 verbalisations. This study empirically indicates that, in this instance, the effectiveness of the dictionary for the improvement of translation was very limited.

Despite the fact that the effectiveness of dictionary consultations is questionable, research has found that dictionaries are often overused by students, for example Wang (2014) and Sycz-Opoń (2019). Two reasons for this were suggested in Livbjerg and Mees (2002). One was that students lack confidence in their own linguistic abilities, and the other that students were too focussed on lexical units and ignored other important factors, such as relevant background information which might help to understand the source text. Enríquez Raído (2011) and Sycz-Opoń (2019) both discussed dictionaries as being the student's "first port of call" (Enríquez Raído, 2011, p. 480) ignoring the nature of the problem they encounter. In addition, Sycz-Opon (2019) also mentioned that dictionaries are suitable for early linguistic education, but fall short of expectations in the more complex context of translation.

In this study, it would be too arbitrary to say that as the student's use of dictionaries decreased their resource use improved, however, it could be suggested that trainee translators become more confident with their linguistic abilities, realising the limitation of dictionaries, hence spent less time on dictionaries.

6.1.2 Time spent consulting different dictionaries

After concluding that the absolute time and the percentage of time spent on dictionaries decreased over the four tasks, in this section, the variable 'dictionary' was further investigated and categorised into 22 different dictionaries. A detailed catalogue can be seen in

Figure 21, presenting the total time spent consulting 22 different dictionaries for all four tasks combined.

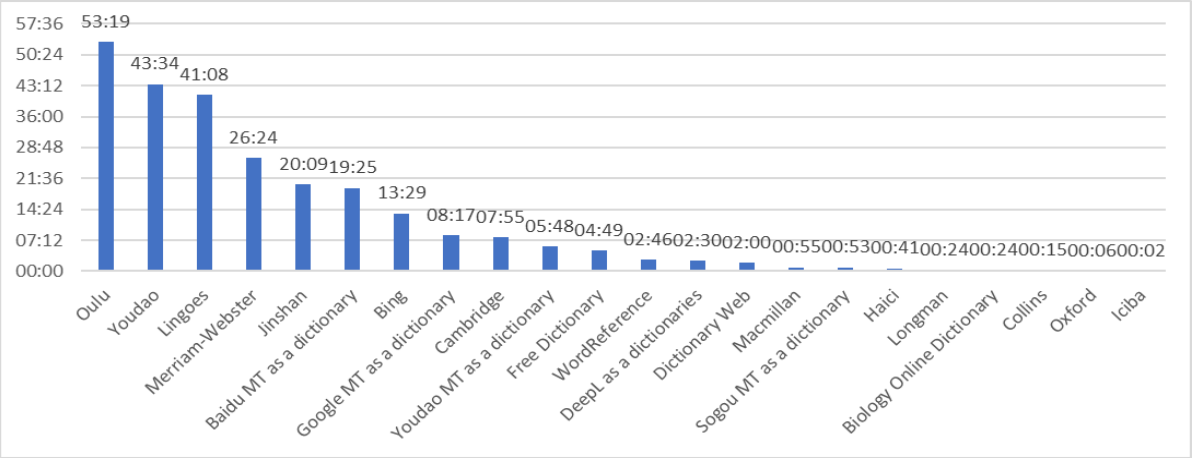

Figure 21 Total time spent consulting different dictionaries over the four tasks (time in minutes:seconds)

Participants spent a longer time consulting a few dictionaries, and quite a number of dictionaries are only sporadically used if at all. Specifically, Oulu, Youdao, and Lingoes, the top three dictionaries, were accessed for 59min 19s, 43min 34s and 41min 8s respectively. The subsequent four different dictionaries, Merriam-Webster, Jinshan, Baidu MT as a dictionary, and Bing, were used for between 26min 24s and 13min 29s. Following those, Google MT as a dictionary, Cambridge, Youdao MT as a dictionary, Free Dictionary, WordReference, DeepL as a dictionary, and Dictionary web were all used for less than a total of 10 minutes. The remaining eight were only used for less than one minute across the four tasks, and then only sporadically; they were Macmillan, Sogou MT as a dictionary, Haici, Longman, Biology Online, Collins, Oxford, and Iciba.

The time spent consulting different dictionaries was then broken down by the tasks:

Figure 22 and Table 18.

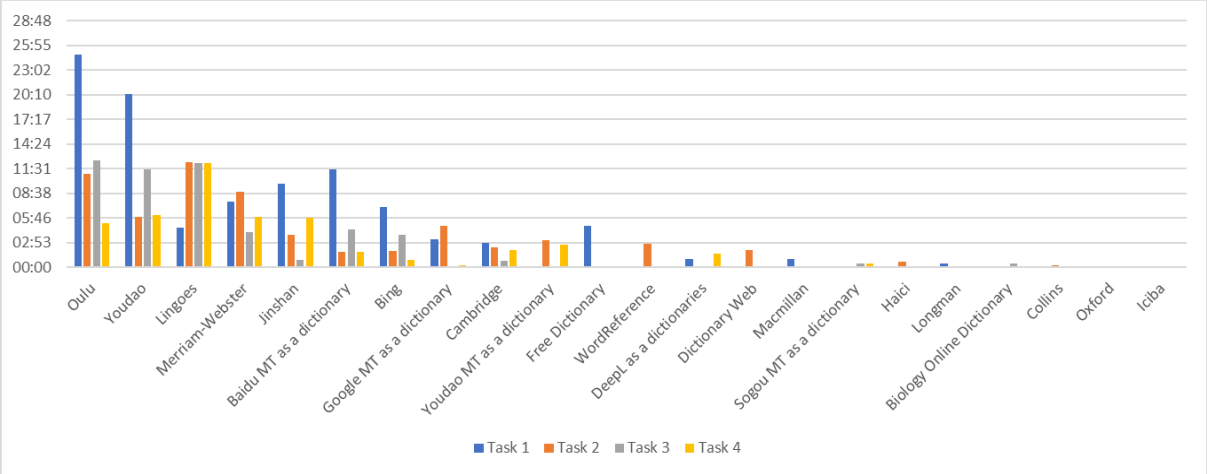

Figure 22 Time spent on different dictionaries by task (time in minutes:seconds)

Table 18 Time spent consulting different dictionaries (time in minutes:seconds)

Group	**Different dictionaries**	**T1**	**T2**	**T3**	**T4**
Group 1	Oulu	24:51	10:52	12:29	05:07
	Youdao	20:11	05:50	11:27	06:05
	Lingoes	04:34	12:17	12:11	12:06
	Merriam-Webster	07:39	08:47	04:04	05:54
	Jinshan	09:48	03:44	00:48	05:49
	Baidu MT as a dictionary	11:23	01:48	04:26	01:48
Group 2	Bing	07:02	01:54	03:44	00:49
	Cambridge	02:51	02:19	00:44	02:00
	Google MT as a dictionary	03:16	04:50	00:00	00:11
	Youdao MT as a dictionary	00:00	03:08	00:00	02:39
Group 3	DeepL as a dictionary	00:58	00:00	00:00	01:32
	Sogou MT as a dictionary	00:00	00:00	00:27	00:27
	Free Dictionary	04:49	00:00	00:00	00:00
	WordReference	00:00	02:46	00:00	00:00
	Dictionary Web	00:00	02:00	00:00	00:00
	Macmillan	00:55	00:00	00:00	00:00
Group 4	Haici	00:00	00:41	00:00	00:00
	Longman	00:24	00:00	00:00	00:00
	Biology Online Dictionary	00:00	00:00	00:24	00:00
	Collins	00:00	00:15	00:00	00:00
	Oxford	00:06	00:00	00:00	00:00
	Iciba	00:02	00:00	00:00	00:00

The 22 different dictionaries were then grouped according to the number of tasks in which they were used. Several considerations contributed to this grouping.

The aim of this section was to investigate the changes between the four tasks. For this reason, how many dictionaries were used or, more specifically, in which tasks they were used, or not used, gives a representative indication of the changes occurring over the four tasks.

According to Enríquez Raído (2011), there is a distinction between an individual information need and a common information need. Individual information needs refer to the needs owned by one participant, and common information needs refer to the needs common to at least two participants. Similarly, Gough (2016) adapted the taxonomy further to distinguish resources into individual resources and common resources. Based on this categorisation, for the purpose of the current study, different dictionaries were categorised according to whether they were used in one, two, three, or four tasks. This further grouping was also necessary due to the large number of variations, since 22 different dictionaries cannot be compared individually due to the time and space restraints in this thesis.

In the first group, seven different dictionaries were used by the participants in all four tasks (Oulu, Youdao, Merriam, Jinshan, Baidu MT as a dictionary, Bing, and Cambridge). Notably, the most used dictionary, Oulu, decreased by 19min 44s from Task 1 (24min 51s) to Task 4 (5min 7s). Apart from Lingoes, which was the only one that demonstrated an increasing trend among the different dictionaries used in all the four tasks, the time spent on the remaining six decreased dramatically.

According to Lingoes's official website,[20] it is a free customised dictionary and a text translation application. A screenshot of the Lingoes interface is attached in Figure 23. The red rectangle part includes various resources. A unique facility of Lingoes, compared with the majority of other dictionaries, is that it can incorporate various other online dictionaries (such as Longman Dictionary, Merriam-Webster Dictionary, Collins Dictionary) and encyclopaedia (such as Wikipedia and Baidu Baike) once they are installed by the users. The use of Lingoes can thus help participants search a larger scope of online resources more easily and quickly. The use of the Lingoes dictionary also attracted the attention of Shih (2017, 2019), who noted

[20] http://www.lingoes.net/index.html

that due to the convenience of being able to search a large number of resources within one portal, Lingoes could “alter the dynamics and even ergonomics” (p.5) of how participants use web resources.

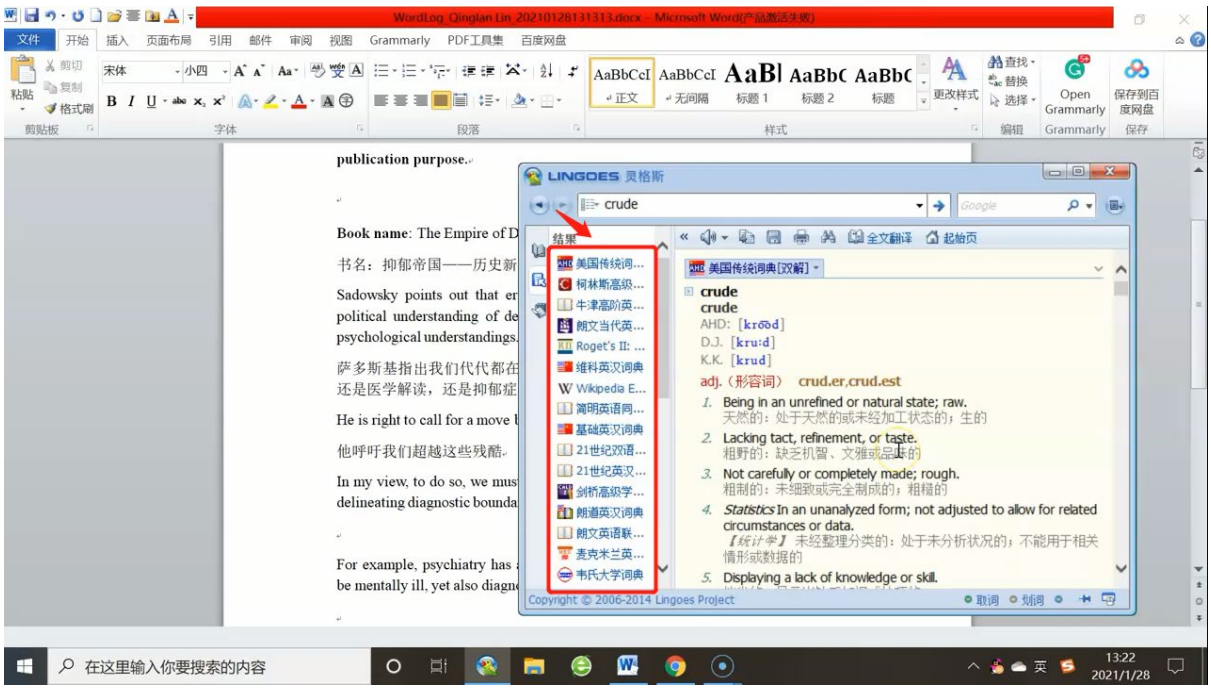

Figure 23 Screenshot of Lingoes interface

The second group comprised different dictionaries that were used in three of the tasks, and here Google MT, used as a dictionary, was the only one. The third group, comprised Youdao MT used as a dictionary, DeepL used as a dictionary, and Sogou MT used as a dictionary, which were used in only two of the tasks. All three are machine translation software acting as dictionaries. The remaining ten different dictionaries were only used in one of the four tasks. Interestingly, on closer inspection, five of the ten were used only in Task 1, four were only used in Task 2, and only one was used in Task 3, following a clear downward trend.

Another observation is that time using machine translation engines as dictionaries, i.e., entering lexical items as queries in machine translation engines, decreased. The total time of this type (Baidu MT, Google MT as a dictionary, Youdao MT as a dictionary, DeepL as a dictionary, Sogou MT as a dictionary) decreased from 15min 37s, 9min 46s, 4min 53s, and 6 min 38s in the four tasks respectively.

From the perspective of different dictionaries used, it is clear that the utilisation of dictionaries

diminished sequentially over the tasks.

6.2 Search engines

6.2.1 Time spent consulting search engines

Similar to the analysis of time spent consulting dictionaries, time spent accessing search engines was also first plotted as a boxplot in order to visually examine the general trends, see Figure 24.

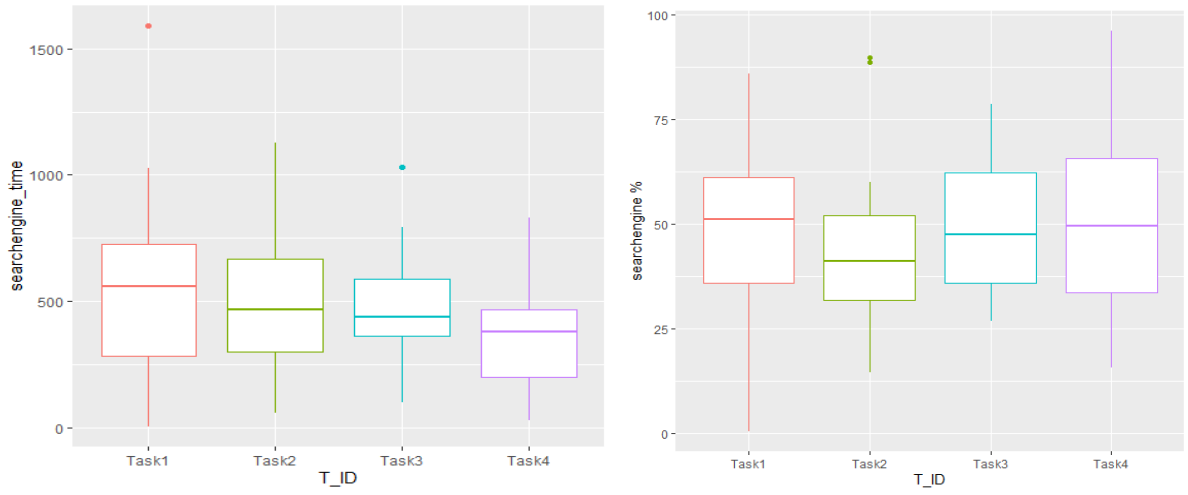

Figure 24 a) Time spent (time in seconds) and b) Percentage of time spent on search engines during the four tasks

The two dots representing potential outliers in Task 1 and 3 were retained after I carefully investigated the original files (see Section 4.8.1). Compared with the boxplot of dictionary usage, the general downward trend was relatively stable. In the box representing Task 4, it does not appear as stable as the other three tasks. More details are presented in Table 19. The search engine percentage seems quite stable for the four tasks.

Table 19 Descriptive statistics of time spent (time in minutes:seconds and the percentage of time consulting search engines)

T_ID	searchengine_time					searchengine_pct				
	mean	sd	IQR	min	max	mean	sd	IQR	min	max
Task 1	9:25	6:26	7:23	0:03	26:32	47.05%	20.62%	25.00%	0.00%	86.00%
Task 2	8:07	5:03	6:09	0:58	18:48	44.37%	19.91%	20.00%	14.00%	90.00%
Task 3	7:51	3:49	3:43	1:39	17:12	49.84%	16.67%	26.50%	27.00%	79.00%
Task 4	6:29	4:15	4:25	0:27	13:50	51.11%	21.34%	32.00%	16.00%	96.00%

Table 19 presents the descriptive statistics. The mean time spent on the four tasks shows a relatively gradual decrease (1min 18s, 15s, 1min 23s) compared with the decrease in time spent on dictionaries (1min 59s, 33s, 20s). There were no large reductions in time between tasks, although the difference from Task 1 (9min 25s) to Task 4 (6min 29s), was 2min 56s. The SD, IQR, and maxima all showed decreasing trends, and the longest access duration for Task 1 was 26min 32s, while for Task 4, it was only 13min 50s, only 49.7% of the time recorded for Task 1. The search engine usage percentage showed an increasing trend despite a 2.85% decrease at the beginning, from Task 1 to Task 2. After this, the change in percentage steadily grew, by 5.47% to 49.84% for Task 3, and by a total of 4.06% to 51.11% in Task 4.

Linear mixed-effects models, Model 3 and Model 4, were constructed to determine the significance of the differences in time spent and percentage of time spent consulting search engines during the four tasks. As for all the models in this thesis, Task ID (T_ID) was the explanatory variable, and Participant ID (P_ID) was the random effect variable, and in Model 3 and Model 4, the response variables were the time spent in consulting search engines and the percentage of time spent on search engines. The square root transformation was used.

Model 3 Regression model of time spent on search engines

```
lmerTest::lmer(sqrt(searchengine_time) ~ T_ID + (1|P_ID) , data = websearch)
```

Model 4 Regression model relating the percentage of time spent on search engines

```
lmerTest::lmer(searchengine_pct) ~ T_ID + (1|P_ID) , data = websearch)
```

The summary of the results for Model 3 and Model 4 is presented in Table 20.

Table 20 Results of Model 3 and Model 4

	sqrt(searchengine_time)				**searchengine_pct**		
Predictors	Estimates	CI	p	Predictors	Estimates	CI	p
(Intercept)	22.1	18.85 – 25.36	<0.001	(Intercept)	0.47	0.38 – 0.56	<0.001
Task 2	-1.22	-4.46 – 2.02	0.46	Task 2	-0.03	-0.10 – 0.05	0.465
Task 3	-1.07	-4.31 – 2.17	0.519	Task 3	0.03	-0.05 – 0.10	0.476
Task 4	-3.5	-6.74 – -0.26	0.034*	Task 4	0.04	-0.03 – 0.11	0.286
Marginal R2 / Conditional R2	0.031 / 0.518			Marginal R2 / Conditional R2	0.017 / 0.662		

*p < .05.

As Table 20 shows, Task 4 was significantly different from Task 1 (p=0.034) while the differences of Task 2 and Task 3 were insignificant (p=0.460, p=0.519), with a conditional R^2 of 0.518. Although the time spent consulting search engines decreased during the four tasks, only the difference in time spent on Task 4 compared with Task 1 was statistically significant. The percentage differences for search engines, were not significant with all the p-values >0.05 (p=.465, p=.476, p=.286) with an R^2 of .662.

The findings on search engines are quite interesting. The search engine percentage is quite stable across the four tasks as shown in Figure 24b. A slightly increasing trend was observed, but the regression model did not find any significance, which indicates that the participants distributed a similar proportion of time on search engines. Compared with the unanimous results for the use of dictionaries in previous literature, the findings of search engines are mixed. For example, in a longitudinal study by Paradowska's (2020) although descriptive findings suggested that the percentage of time spent on search engines was indicative of increased use of search engines, the inferential analysis results did not detect any significant increase or decrease. Similarly, students and professionals spent a similar percentage of time on search engines in Onishi and Yamada (2020), where a similar search engine percentage time (32.13% for students and 32.18% for professionals) was found when they translated a 62 words text. A slight increase in the frequency of search engine use was observed in Chang (2018)'s longitudinal study, during four tasks implemented in a one-year postgraduate translation course in the UK. Due to the small number of participants however, no inferential analyses were

conducted, except for Paradowska (2020).

In terms of the time spent on search engines (absolute time), although the mean time decreased gradually, statistically, only one decrease is detected in Task 4. It could suggest that search engine time did decrease during the four tasks, however, the amount of decrease did not reach a significant level, or it may have been influenced by the small number of participants (although the number of participants was quite large compared to other studies in translation studies) and lack of statistical power. The decreasing time of search engine time might be due to the participants' practice during the course; they might have got more familiar with the search engines.

As mentioned in Chapter 3, compared to dictionaries and knowledge-based resources, the use of search engines is quite unique due to the size of searchable information, and the function of linking into other types of resources (such as the links may contain dictionaries or Wikipedia). It is an interactive searching process, the essential part is submitting a query, receiving search results, and modifying a query (Choi, 2013). The most important component is search query, which is fully discussed in Chapter 7.

6.2.2 Time spent consulting different search engines

The types of search engines and their usage combined in the four tasks are plotted as a histogram in Figure 25. In total, there were six search engines used by all the participants across the four tasks. It is evident that Google Search was the dominant choice among all the search engines, totalling 6hr 13min 58s for the four tasks, followed by Baidu Search and Bing Search, which only totalled 2hr 6min 18s and 1hr 1min 14s respectively. The other three search engines, New Tab,[21] 360 Search, and Sogou Search, only made a small contribution with a combined total of only 42min 21s, being 7.3% of search engine usage.

[21] At this page, the detailed search engines are not shown.

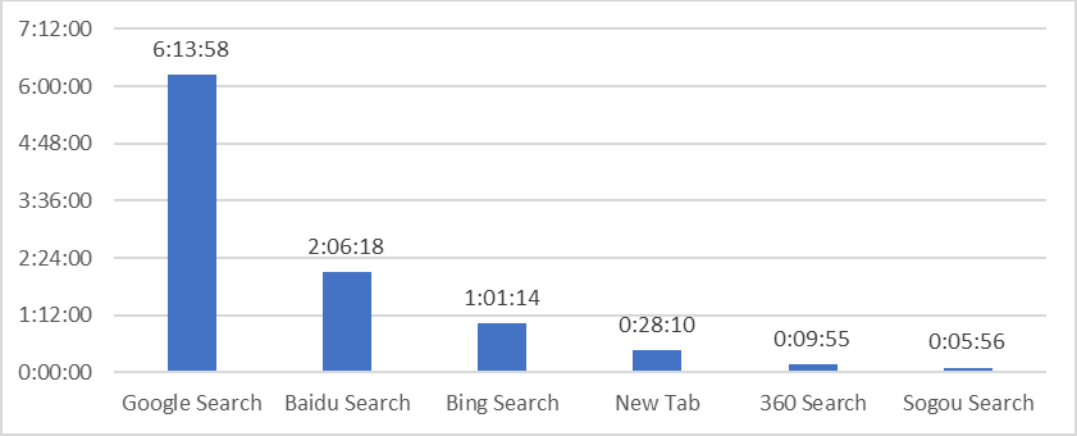

Figure 25 Total time spent in consulting different search engines for four tasks (time in hours:minutes:seconds)

To investigate further, it was necessary to break down the time spent on search engines per task and different search engines (see Figure 26 and Table 21). Since there were only six different search engines used in all four tasks by all of the 19 participants, it was feasible to compare them one by one. From Figure 26, there does not seem to be a noticeable difference between the four tasks graphically. A descriptive summary table is presented in Table 21.

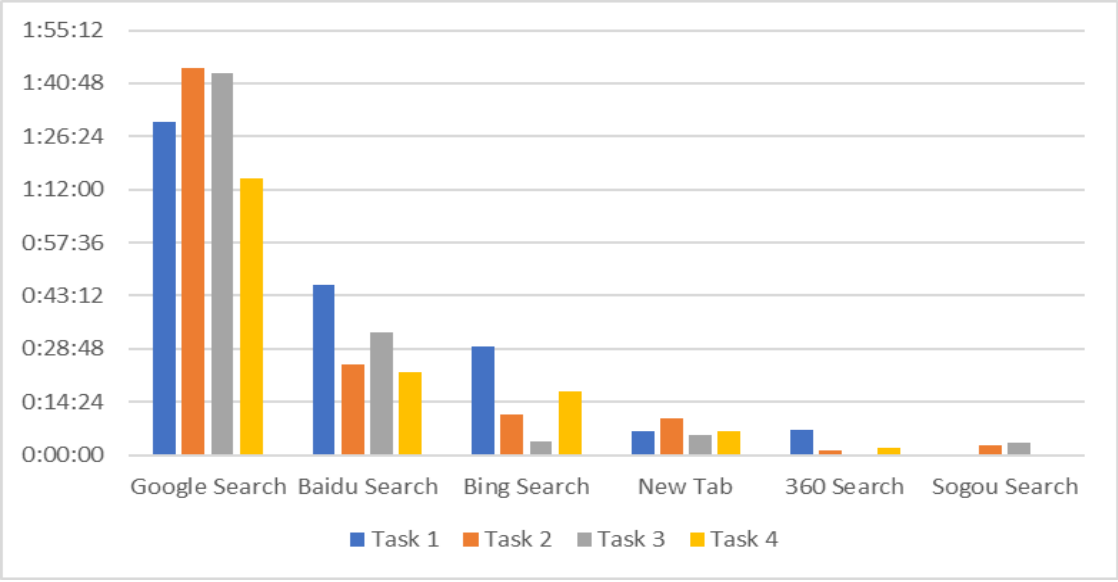

Figure 26 Time spent on different search engines by Task (time in hours:minutes:seconds)

In Table 21, Google Search was stable across the four tasks, ranging from 1hr 15min to 1hr 44min across the four tasks. Baidu Search presented a downward trend with some fluctuation,

dropping from 45min 59s in Task 1 to 22min 36s in Task 4. Bing Search started with 29min 18s in Task 1, then decreased by 18min 8s to 11min 10s in Task 2, and continued to decrease to 3min 34s in Task 3, however it then increased by 13min 38s to reach 17min 12s in Task 4. The New Tab search engine was the start page in the web browsers and indicated where the participants started to search. In these circumstances, the participants' default search engines did now show on the screen. New Tab was also stable across the four tasks, with a range from 5min 36s to 9min 55s. 360 Search and Sogou Search were only used sporadically compared with the others. Both search engines were designed by Chinese companies.

Table 21 Different search engines by task (time in hours:minutes:seconds)

Different search engines	**Task 1**	**Task2**	**Task3**	**Task4**
Google Search	1:30:26	1:44:58	1:43:33	1:15:01
Baidu Search	0:45:59	0:24:30	0:33:12	0:22:36
Bing Search	0:29:18	0:11:10	0:03:34	0:17:12
New Tab	0:06:18	0:09:55	0:05:36	0:06:21
360 Search	0:06:51	0:01:08	0:00:00	0:01:56
Sogou Search	0:00:00	0:02:34	0:03:22	0:00:00

Baidu and Google are the two most consistently used search engines in this study, while others are used less frequently. This result is consistent with Shih (2017), which also includes English and Chinese language combinations. In Shih (2017), Baidu and Google were also the most frequently used search engines by participants. Another result was that the first three different resources that the participants spent the most of their search engine time was Google Search, Baidu Search, and Bing Search. The time spent on Google Search was relatively stable over the four tasks, while the other two search engines (i.e., Baidu Search and Bing search) decreased. This was in agreement with Wang (2014). In his comparative study (novices, semi-professionals and professionals), where professional groups tend to use more Google Search than Baidu Search or Bing Search. In retrospective verbal reports, one of the participants mentioned the reasons for choosing Google instead of Bing in Task 3:

"从开学以来的一个体验，因为我之前我是真的没有怎么用谷歌，我一直以为百度和谷歌就差的不多，但是后来发现谷歌是真的比国内的必应要好用太多了，基本上在国内查不到的东西，一到谷歌上查就全都有了。"

[Translation] "One of my lessons learned in this semester is Baidu and Google are very different. Before this semester, I thought the two are almost the same, but recently I found out that Google is so much better than Baidu. The contents are much richer, what I cannot find in Chinese search engine, such as Baidu, is available in Google."

From here, it can be seen that the use of different search engines does have an impact on search results, however few studies take into account the different search engines used in translation studies. This may be due to the fact that in some languages, one search engine is almost exclusively used. For example, in Polish, Google accounted for of 97.85% in 2020 (Paradowska 2020).[22] In China, the situation was different: in 2020, the dominant one is Baidu which is 84.27%, followed by Bing comprised 6.73%, Sogou 3.05%, Google only took up 2.51%, while the rest (i.e., Haosou and Shenma) only comprised 3.44%.[23] The main reason for the low utilisation rate is the limited accessibility of Google in China, as people will need Virtual Private Network (VPN) service to access Google. In this research, all the participants were capable of accessing Google, however, the position of Google and Baidu was reversed, and Google was the most frequently accessed search engine accounting for 61.8%, while Baidu only took up 20.9%. This indicated the uniqueness of translation-oriented web searching, and the need for web searching studies to be contextualised in translation studies.

6.3 Knowledge-based resources

6.3.1 Time spent consulting knowledge-based resources

Figure 27 is a boxplot of the time and percentage participants spent consulting knowledge-based resources during each of the tasks.

[22] Data originated from https://gs.statcounter.com/search-engine-market-share/all/poland.

[23] Data comes from https://gs.statcounter.com/search-engine-market-share/all/china/2020.

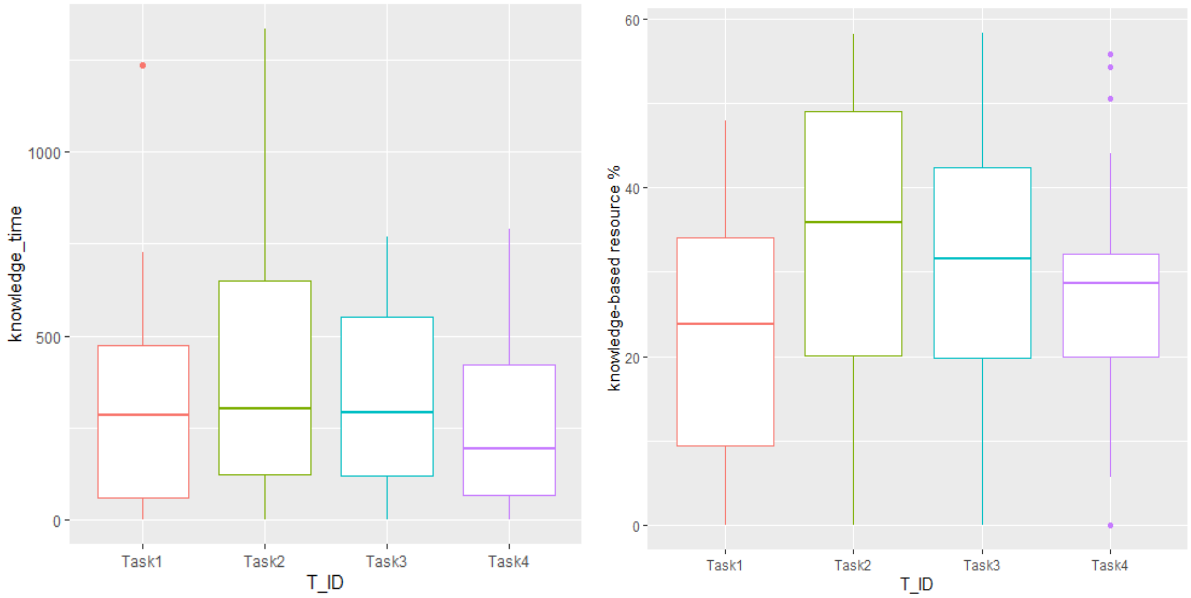

Figure 27 a) Time spent (time in seconds) and b) Percentage of time spent on knowledge-based resources during the four tasks

Figure 27 a) indicates that the trend of the time spent on knowledge-based resources was relatively stable, although with a small increase from Task 1 to Task 2, and a slight decrease from Task 3 to Task 4. The data in Task 2 above its median exhibited more variability compared to the other three tasks. This can be seen more clearly below in Table 22.

Figure 27 b) shows the percentage of time spent on knowledge-based resources shows an increasing trend from Task 1 to Task 4, which is particularly true for Task 2.

Table 22 Descriptive statistics of time spent (time in minutes:seconds and the percentage of time consulting knowledge-based resources)

T_ID	**knowledge_time**					**knowledge_pct**				
	mean	sd	IQR	min	max	mean	sd	IQR	min	max
Task 1	5:22	5:14	6:53	0:00	20:37	22.63%	16.03%	24.50%	0.00%	48.00%
Task 2	7:21	6:37	8:49	0:00	22:15	32.95%	18.55%	28.50%	0.00%	58.00%
Task 3	5:47	4:04	7:13	0:00	12:49	32.05%	16.23%	22.50%	0.00%	58.00%
Task 4	4:18	3:54	5:56	0:00	13:11	27.58%	16.41%	12.50%	0.00%	56.00%

From Table 22, it is clear that the average time spent on knowledge-based resources fluctuated: specifically it increased and then decreased. The increase in the mean access time from Task 1 to Task 2 is particularly noticeable, peaking at 7 minutes and 21 seconds from 5 minutes 22 seconds. In terms of range, the SD (6min 37s), the maximum (22min 15s) were all highest in Task 2. From Task 2 to Task 4, it decreased by 2min 57s, from Task 2 (7min 21s) to Task 3 (5min 47s), and Task 4 (4min 18s).

The overall trend of percentage of time spent on knowledge-based resources was, however, upward although there was a slight decrease in Task 4. The 10.3% increase from Task 1 to Task 2 was notable, however there was only a 0.9% difference between Task 2 and Task 3, and a 4.47% decrease between Task 3 and Task 4. The increase between Task 1 and Task 4 of 4.95% suggests a consistent trend, although the difference between sequential tasks was relatively small.

In the following section, regression was used to check whether the differences were significant. As for the time spent on dictionaries and search engines (Models 1 and 2), Model 3 was also built with Task ID (T_ID) as the explanatory variable, and Participant ID (P_ID) as the random effect variable. The response variable was time spent in consulting knowledge-based resources. This dataset was also transformed using a square-root transformation to improve its performance.

Model 5 Regression model of time spent on knowledge-based resources

```
lmerTest::lmer(sqrt(knowledge_time) ~ T_ID + (1|P_ID) , data = websearch)
```

Model 6 Regression model relating the percentage of time spent on knowledge-based resources

```
lmerTest::lmer(knowledge_pct ~ T_ID + (1|P_ID) , data = websearch)
```

The summary of the results is presented below in Table 23.

Table 23 Results of Model 5 and Model 6

	sqrt(knowledge_time)				**knowledge_pct**		
Predictors	*Estimates*	*CI*	*p*	Predictors	*Estimates*	*CI*	*p*
(Intercept)	5.8	4.59 – 7.00	<0.001	(Intercept)	0.23	0.15 – 0.30	<0.001
Task 2	0.88	-0.25 – 2.00	0.126	Task 2	0.1	0.02 – 0.18	0.012*
Task 3	0.64	-0.48 – 1.77	0.263	Task 3	0.09	0.01 – 0.17	0.023*
Task 4	-0.37	-1.49 – 0.76	0.524	Task 4	0.05	-0.03 – 0.13	0.228
Marginal R2 / Conditional R2	0.034 / 0.578			Marginal R2 / Conditional R2	0.056 / 0.467		

*p < .05.

According to Table 23, the results for time spent on knowledge-based resources did not yield any significant differences, p>0.05 (p=0.119, p=0.347, p=0.456), although some variation in the means are observed in the descriptive summary. The p-values for the percentage difference for knowledge-based resources were significant for Task 2 (p=.012) and Task 3 (p=.023), while Task 4 (p=.228) was insignificant, and the model R^2 value was 0.467. The effect plots are shown below in Figure 28.

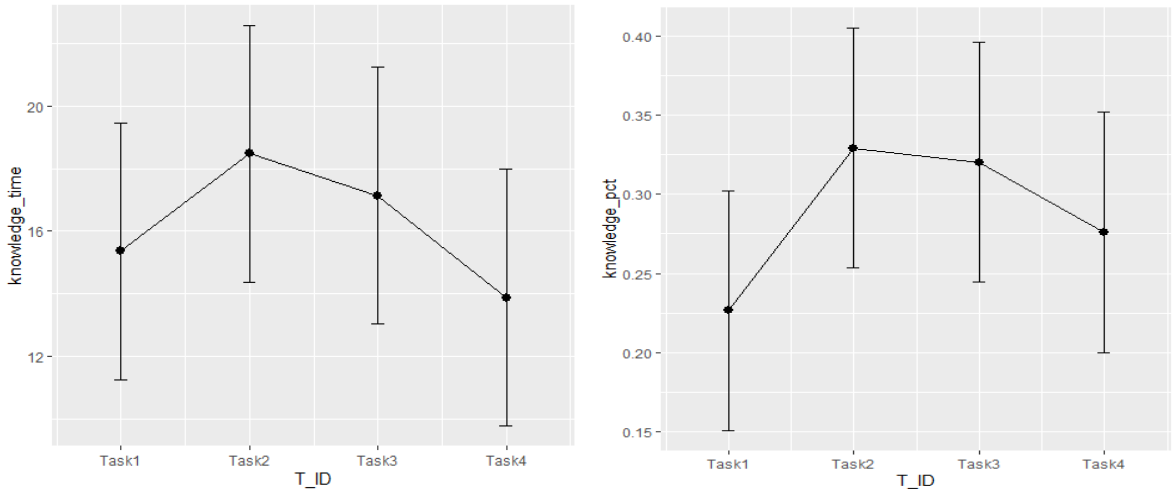

Figure 28 Effect plot of task on a) Time spent (time in seconds) and b) Percentage of time consulting knowledge-based resources

To sum up, , the mean time spent on knowledge-based resources increased from Task 1 to Task 2, and then decreased, but there were no significant differences detected. For the percentage of

time spent on those resources, compared with Task 1, the remaining three tasks all experienced an increase, and statistically significant increases were detected in Task 2 and Task 3. Compared to the time spent on the other two main resources (i.e., dictionaries and search engines), time spent on knowledge-based resources was the only one that did not decrease. For the percentage of knowledge-based resources time, compared to Task 1, the remaining three tasks all seemed to increase, and a statistical significance for Task 2 and Task 3 increases were found.

In the study by Chang (2018), the participants accessed online encyclopaedias, travel websites, and non-travel websites, whereas in the present study, the resources were divided into dictionaries, search engines, machine translation engines, online encyclopaedias, travel websites, online images and online maps. The addition of the latter four types of resources was the same as for knowledge-based resources. Compared with Onishi and Yamada (2020), students spent 36.9% while professionals spent 56.43% on the non-dictionary resource. Although they used different terminology, the scope of non-dictionary resources in their study was the same as this study, representing resources other than search engines and dictionaries. The increasing use of knowledge-based resources may imply that participants pay more attention to the information that can be searched in those resources, such as background information (compared to superficial lexical information) only.

6.3.2 Time spent consulting different types of knowledge-based resources

After investigating knowledge-based resources as a whole, it would be useful to investigate the constituent components in greater detail. As previously explained, the categorisation evolved in a bottom-up design, to be more specific, it was based on the resources the participants consulted to complete the four translation tasks.

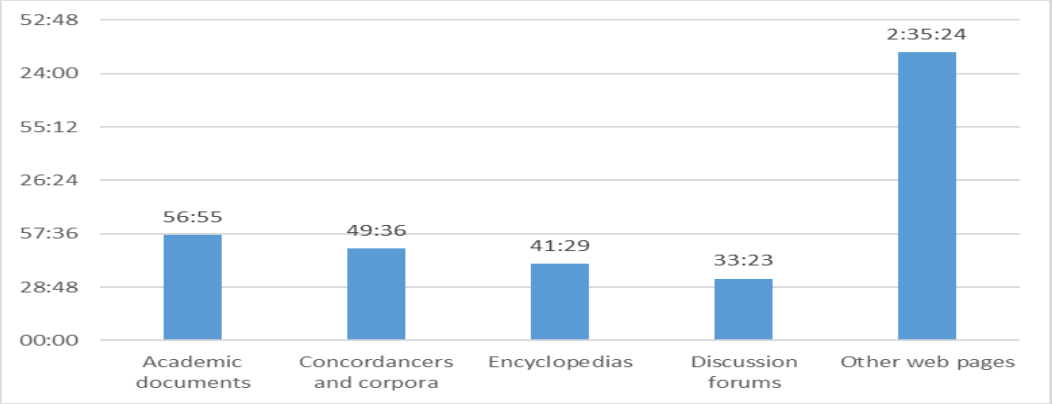

Figure 29 Total time spent consulting different knowledge-based resources during four tasks (time in hours:minutes:seconds)

In Figure 29, five types of different knowledge-based resources were identified across the four tasks: academic documents, concordancers and corpora, encyclopaedias, discussion forums, and other web pages. Among them, academic documents included Google scholar, Baidu Scholar, China National Knowledge Infrastructure (CNKI), Academic web pages (e.g., library page), and academic PDF file (e.g., formal reports). Concordancers and Corpora contained Linguee, [24] British National Corpus (BNC), Corpus of Contemporary American English (COCA), and English-Corpora. Encyclopaedias were comprised of Wikipedia, Baidu Baike, and Sogou Baike. Discussion forums included Questia, Zhihu, Douban, Baidu Zhidao, and Sci-Hub. The category, Web pages, encompassed all other web pages, excluding the aforementioned types of knowledge-based resources. For example, the most frequently used "other web page" was the Amazon webpage, which introduces the books about which the book reviews had been written that they were assigned to translate.

To ensure data transparency, I have also provided data on the detailed resources inside the five types of knowledge-based resources in Appendix J. Figure 30 shows that in addition to other web pages, academic documents were the most used resources, which accounted for 56min 55s. After this, concordancers and corpora, and encyclopaedias were used for 49min 36s and 41min 29s, respectively. The least used were discussion forums, which were only consulted for 33min

[24] Previous literature did not reach an agreement on the categorisation of Linguee, such as Zetzsche (2015) regarded it as a dictionary, Zapata (2016) treated it as a sperate resource different from any others, and Gough (2016) classified it as a concordancer. In this thesis, it is allocated to "concordancers and corpora" category.

23s. In the last bar in Figure 30, other web pages were the dominant knowledge-based resource (> 2hr 35min). The use of different knowledge-based resources for the four tasks is therefore further explored in Figure 30 and Table 24.

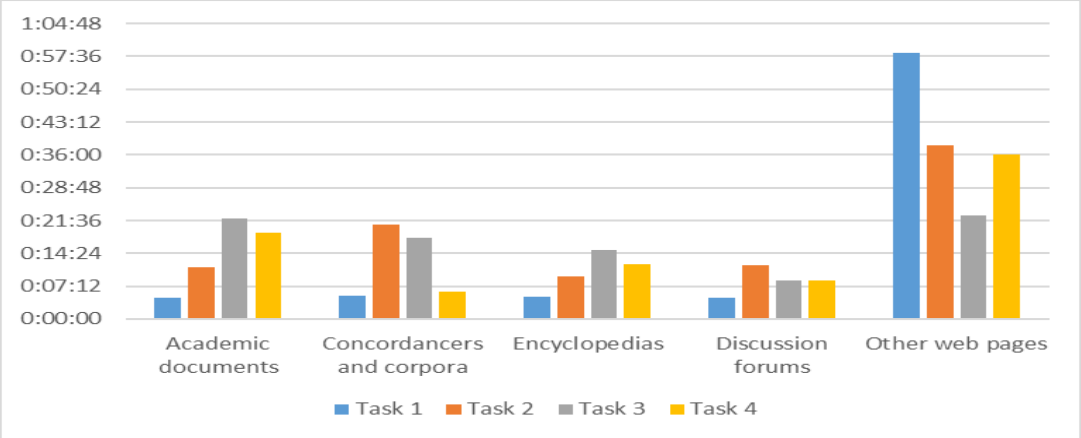

Figure 30 Time spent on different knowledge-based resources by task (time in minutes:seconds)

Table 24 Different types of knowledge-based resources by task (time in minutes:seconds)

Different types of knowledge-based resources	Task 1	Task 2	Task 3	Task 4
Academic documents	04:36	11:22	21:59	18:58
Concordancers and corpora	05:07	20:43	17:42	06:04
Encyclopaedias	04:57	09:17	15:09	12:06
Discussion forums	04:42	11:53	08:22	08:26
Other web pages	58:28	38:11	22:43	36:02

From Table 24, it can be seen that there were five types of knowledge-based resources used in all four tasks. Academic documents were consulted for 4min 36s in Task 1, then the time increased steadily in Task 2, Task 3, and Task 4, with the highest consultation time reaching 21min 59s in Task 3. For concordancers and corpora, participants spent an average of 5min 7s in Task 1, and then increased to 20min 43s in Task 2 and 17min 42s in Task 3. The concordancers and corpora time only reached 6min 4s in Task 4; but Task 4 to Task 1, still increased by 57s. Time spent on encyclopaedias increased from 4min 57s in Task 1 to 12min 6s in Task 4, with the longest time spent on Task 3 (15min 9s). Participants spent 4min 42s accessing discussion forums in Task 1, this increased to the maximum of 11min and 53s in Task 2, then decreased slightly in Task 3 and Task 4 with 8min 22s and 8min 26s respectively. For these resources, academic documents, concordancers and corpora, encyclopaedias, and

discussion forums, the trends were all positive, peaking at Task 2 or Task 3, and decreasing in Task 4. Compared to the time spent on each type of knowledge-based resource in Task 1, the average time spent in Task 4 was longer. As for the last knowledge-based resources, i.e., other web pages used, the time spent on this type was the only one with a negative trend. As mentioned, other web pages largely include general content in a simple design compared with other knowledge-based resources. It is possible that the high reliance on simple and general diminished, and participants moved away to more specialised or more authoritative resources, seeking information at a deeper level.

The above results suggest that other web pages were the most dominant and consistently consulted during all the four tasks, although the length of time spent in consultation decreased over the four tasks. Participants may have explored academic documents, concordancers and corpora, encyclopaedias, and discussion forums in Task 2 and Task 3, and due to a learning effect gravitated to the resources they found most useful over time.

6.4 The dynamics of resource use

Here I investigate the interplay of the three main types of resources. After summarising the relationship between the three types I compare the results with previous literature and interpret the findings. As the exact number and inferential analysis results about the percentage of each main resource have been analysed and reported earlier, this section focus on the relationship between the three, and the changes that occurred over the four tasks. Figure 31 presents descriptively the percentage change of the three main resources throughout the four tasks.

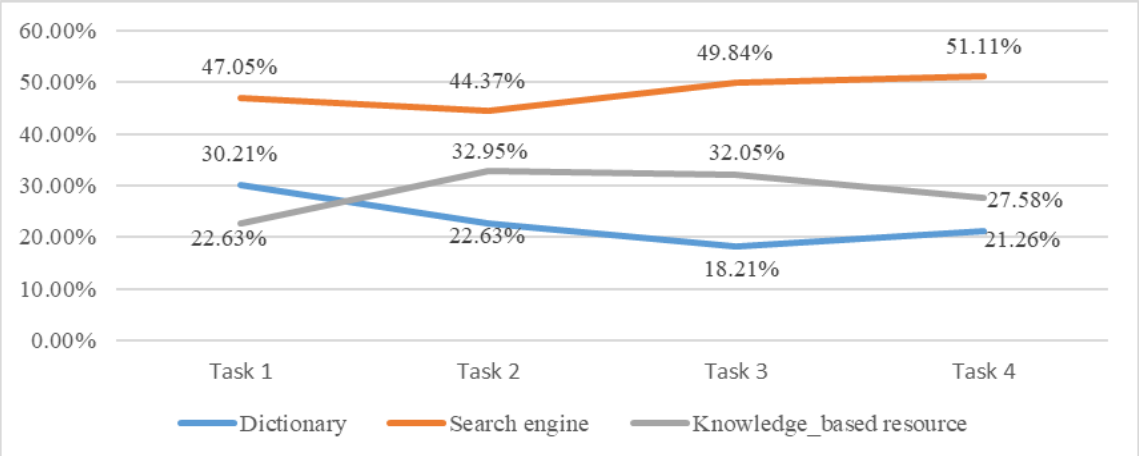

Figure 31 Percentage change of resource usage during the four tasks

According to Figure 31, the search engine usage showed an increasing trend, dictionary usage showed a general downward trend, while the overall trend of knowledge-based resources was positive (albeit there was a slight decrease in Task 4). In Task 1, the percentage of dictionary usage was larger than the percentage of knowledge-based resource use, but from Task 2 onwards, the percentage of dictionary usage became less than the percentage of knowledge-based resources accessed, and although this difference was reduced in Task 4, the percentage of knowledge-based resource access still exceeded dictionary usage.

The findings of this study are in line with previous translation-oriented web searching studies. One of the earliest studies was Enríquez Raído (2011), where during two tasks in week three and week four of an introductory course on technical and scientific translation, the students used mostly dictionaries in searching for information in Task 1 while in Task 2 the diversity of information increased, although some other factors impacted the resources use, such as text difficulty. In Chang (2018), during four tasks with similar difficulty levels implemented in one year, students also gradually moved away from dictionary-based resources towards other resources. More recent studies conducted by Paradowska (2020) also confirmed that the reliance on dictionaries decreased (statistically confirmed), the percentage of search engine increased (not statistically confirmed), while the percentage of other resources increased, although no inferential analysis was conducted to support these observations. In addition, Onishi and Yamada (2020) also examined resource use differences by recruiting students and professionals, and found that the differences between the two groups lay mainly in the percentage use of dictionary and non-dictionary resources, to be more specific, students used more dictionary resources while professionals used more non-dictionary resources.

There are at least two possible explanations for the finding of the current study: one is that trainee translators become aware of the importance of non-lexical information, such as background information, which cannot be found in dictionaries. For example, all the four source texts assigned to the participants were book reviews, and since these were not classic reviews, which could not be found in dictionaries. This required the participants to access alternative sources, such as knowledge-based resources, to access the required information. A second possible explanation is that, compared to the limited information in dictionaries, the

totality of knowledge-based resources gives participants the room to explore different types of resources. One thing to be emphasised here is that there are no more advanced or less advanced resources, for as also mentioned by Gough, there is no "correct way of obtaining certain types of information" (2019, p.354), however deeper and more diversified resources are needed to solve translation-related searching problems (Enríquez Raído 2011, Wang 2014, Shih 2017, Gough 2016, Cheng 2017, Chang 2018).

6.5 A comparison between observed resource use and self-declared resource use

In this section I compare the findings from questionnaires with the findings from observed resource accessing behaviours. In Section 5.3 the self-declared frequency use of dictionaries from the questionnaires is presented. Participants reported that the dictionary was the most frequently used resources, for which the ordinal mean was 4.58/5.[25] The self-rated search engines mean was 4.37/5. The remaining resource types rated in the questionnaires were web pages, encyclopaedias, machine translations, documents, discussion forums, and concordancers, which comprise the sum total of the knowledge-based resources. The ordinal mean of the remaining knowledge-based resource types was 3.02/5, which is lower than dictionary and search engine access.

As described in Section 4.6, the participants filled in the questionnaire in Task 1. The findings from this can be compared with observed behaviours from Task 1. From Figure 18, the distribution of the three main types of resources in Task 1 are presented, where: search engines took up most of their time (47.05%), dictionaries were the least used (30.21%), and knowledge-based resources percentage accounted for 22.63% of their enquiries.

The self-declared data from the questionnaire and the process observations initially seems quite different, for dictionaries were self-rated as the most frequently used resources while, in fact, they only comprised 30.21% of resource usage time. One of the reasons is, as recalled, that this questionnaire data is collected at the beginning of this course, it only reflected the baseline data students' resource behaviour. This is also probably due to dictionaries being considered the

[25] Participants were asked to rate from 1 (low), 2, 3, 4, 5 (high).

“first port of call” (Enríquez Raído 2011, p.480), for when finding information participants, ignoring the type of information need, address their information goals by checking in the dictionary which of the many meanings of a word or phrase was being used in contexts. This initial use of dictionaries perhaps gave the participants the false impression that dictionaries were the resources they used most frequently.

Although search engines were rated as being accessed less often compared to dictionaries, they were the most dominant resources accessed in this study. This might be due to search engines’ multiple functions and applications. For example, for one query, the search engine result pages (SERPs) may provide enough information to solve their question straight away, or it may serve as a bridge to enable the participants to access a diversity of other resources. For instance, in this study, when the participants entered the author of the books as a query concerning the book review, after browsing the SERPs, they usually further clicked on the Wikipedia page about the author: as one participant mentioned in the retrospective verbal report,

> “我把这书名放到 Google 里面去搜的，搜到这个作者是谁，然后再去找这个作者的那个 Wikipedia,这才大概能知道这个作者他主要是干吗了写什么的”
>
> [Translation] “I put the book title in Google, then I can immediately know who the author is, and by only clicking into Wikipedia, I will know who he is and his background.”

In this situation, the participant's focus was on Wikipedia and the use of search engine seemed neglected.

Compared to the rating of dictionaries and search engines however, the knowledge-based resources were rated very low. This indicated that there was room for improvement in the use of diversified resources, and this was reflected in Tasks 2 to 4 where the percentage of knowledge-based resources gradually increased and surpassed the percentage of time spent in accessing dictionaries.

This result is in line with previous questionnaire findings, such as the snapshot survey conducted by Massey and Ehrensberger-Dow (2011), and the questionnaire results in Gough (2016). In the snapshot survey, the participants were asked to recall and rank their resource usage during a recent representative translation assignment. Of the 16 resource types, the four

types of dictionaries (multilingual online dictionaries, multilingual print /CD/DVD dictionaries, and monolingual online dictionaries) were ranked one, two, three, and seven. In Gough's study the participants also claimed that dictionaries were their most frequently used resources.

To sum up, there is a mismatch between self-declared resource use and the observed resource use. As Enríquez Raído (2011) and Chang (2018) advised, the use of retrospection might therefore help students better understand their resource behaviours and effectively and efficiently find the resources they need.

6.6 Summary

In conclusion, the resource use did change during the four tasks over the semester. Such changes in dictionary usage occurred quickly, and from Task 2 participants spent significantly less time looking up words in dictionaries (both time spent and percentage), and the diversity of resources accessed also fell. Time spent accessing search engines also gradually decreased, but this decrease was only significant for Task 4. The percentage however remained relatively stable, and the diversity of search engines accessed remained relatively consistent, with only some small changes in search engines access observed. The actual time spent consulting knowledge-based resources did not seem to change significantly, but the percentage increased, and participants did explore a broader range of resources, particularly during Tasks 2 and 3. The move away from dictionaries to knowledge-based resources, and the massive use of knowledge-based resources indicate that the participants became more aware of the importance of non-lexical information and realized that knowledge-based resources give them a greater possibility to explore the information needed.

7. Query Behaviour

In this chapter, the second research question concerning query behaviours is addressed. Search queries were analysed in terms of query time (Section 3.2.1), query complexity (Section 3.2.2), and query language (Section 3.2.3).

7.1 Query time

As explained in Section 3.2, the search process can be regarded as a process connected by consecutively formulated queries. Query time represents the time a searcher spends on a query, including formulating queries, browsing results, as well as further clicking on other resources. It is calculated as total resource time divided by the number of queries per task and participants. From an analysis of Figure 32, it appears the general trend is flat. Table 25 provides more detailed information.

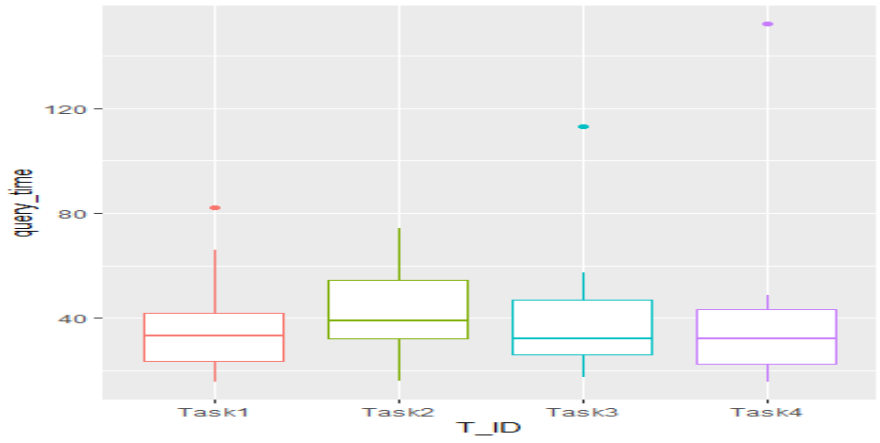

Figure 32 Boxplot of average query time during the four tasks (time in seconds)

From Table 25, the query time fluctuated during the four tasks. It increased from Task 1 to Task 2, summiting at 42.9 seconds per query. Task 2 to Task 4 decreased by 3.7 seconds, from 42.9 seconds to 39.2 seconds, and then to 37.8 seconds.

Table 25 Descriptive statistics for query time (time in seconds)

Task_ID	mean	median	min	max	SD	IQR
Task 1	36.12	33.26	15.60	82.01	17.04	18.33
Task 2	42.92	39.10	15.78	74.07	17.57	22.41
Task 3	39.22	32.16	17.47	112.92	21.76	20.87
Task 4	37.81	32.11	15.57	152.40	29.60	20.81

Whether the increase and decrease observed could be inferred from the whole population is investigated by building a linear mixed-effect model, Model 7. Task ID (T_ID) was the explanatory variable, Participant ID (P_ID) was the random effect variable, and the response variable was the average query time (query_time). The log-transformed dataset was entered into the model.

The increase and decrease observed could be explained by a flat trend and this was investigated by fitting a linear mixed-effect model, Model 7. Task ID (T_ID) was the explanatory variable, Participant ID (P_ID) was the random effect variable, and the response variable was the average query time (query_time). The log-transformed response was used to make the analysis more robust against outliers.

Model 7 Regression model of the effect of task on query time

```
lmerTest::lmer(log(query_time) ~ T_ID + (1|P_ID) , data = websearch)
```

The results are presented in Table 26 below.

Table 26 Regression results of the effect of task on average query time

	log(query_time)		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	3.49	3.29 – 3.70	<0.001
T_ID [Task2]	0.18	-0.05 – 0.41	0.134
T_ID [Task3]	0.07	-0.16 – 0.30	0.565
T_ID [Task4]	-0.02	-0.25 – 0.21	0.894
Marginal R^2 / Conditional R^2	0.027 / 0.399		

In Table 26, all the p-values (p=0.13, p=0.57, p=0.89) were insignificant, with conditional R^2 of 0.399. The data shows no evidence that any of Task 2, Task 3, or Task 4 differed from Task 1.

If the query time decreased, it is reasonable to assume that the participants became more familiar with queries, thus they became faster; if the query time increased, it could be that the participants acquired some query-related knowledge, such as the advanced use of queries, and then they need more time to practice, and thus became slower. The results showed that the query time was quite stable during the four tasks, however, and this may be attributed to the participants not being new to constructing search queries to solve tranlsation-related problems, or that they were postgraduate students (although first year), and already had some translation experience (although not full-time professional translation experience).

As it was concluded that the query time did not change, it is necessary to investigate the query per se (product), in order to see whether there were any changes in terms of query complexity and the langauges of the queries. This is explored in the following two sections.

7.2 Query complexity

As discussed in Section 3.2.2, according to Enríquez Raído (2011), an advanced query is defined as "the use of advanced search options, operators and/or term modifiers entered by the user" (p.374). In the present study, no participants used advanced search options, and thus the search operators were the features of advanced queries. In this section, query complexity was analysed in terms of quantity, the number of features, and where the search operators were submitted.

There were, in total, 2439 queries constructed during the four tasks by all the 19 participants combined. Of these, only 18 contained search operators, representing 0.73% of all queries.

To compare the differences between the four tasks, the number of search operators was broken down per task and participant in Table 27.

Table 27 Distribution of search operators per participant and per task

Participant ID	Task 1	Task 2	Task 3	Task 4	Total for per participant
P1	0	0	0	0	0
P2	0	0	1	1	2
P3	0	0	0	1	1
P4	0	0	0	0	0
P5	0	0	0	0	0
P6	0	0	0	0	0
P7	0	0	0	0	0
P8	0	0	0	0	0
P9	0	11	0	0	11
P10	0	0	0	0	0
P11	0	0	0	0	0
P12	0	0	0	0	0
P13	0	0	0	0	0
P14	0	0	0	0	0
P15	0	0	0	0	0
P16	2	1	0	0	3
P17	0	0	0	0	0
P18	0	0	0	0	0
P19	0	0	0	0	0
Total search operators	2	12	1	2	17
The number of queries	652	494	512	434	N.A
Search operator percentage	0.31%	2.43%	0.20%	0.46%	N.A

In Table 27, the total advanced queries row shows the total number of advanced queries constructed for each task: Tasks 1--4 contained a total of 2, 12, 1, 2 search operators, i.e., regarded as advanced queries, respectively. The last row shows the percentages of search operator were 0.31% (2/652), 2.43% (12/494), 0.2% (1/512), and 0.46% (2/434), respectively. It appears that the 12 advanced queries were used in Task 2 more than in the other three tasks, however 92% (11/12) of the search operators were constructed solely by one participant (P9), and the remaining one was formulated by P16. P9's search operator profile was then investigated across the four tasks, and it was shown that P9 only used search operators during Task 2 and none in the other tasks. It could be said that a particular participant caused the difference of Task 2 compared with the other three tasks, and whether there were any differences between the four tasks cannot be confirmed.

Table 28 presents more detailed information for the 17 advanced queries which contained search operators, each in a separate row. The information provided for each query includes: which participant constructed the advanced query and during which task, what advanced features were used, and where the advanced queries were submitted.

Table 28 Detailed advanced queries

Task ID	Participant ID	Advanced Query	Feature	location
Task 1	P16	“移民接收国经济”	quotation	Google Search
	P16	“非人类迁徙”	quotation	Google Search
Task 2	P9	“subsidiarity hypothesisa”	quotation	Google Search
	P9	“subsidiarity hypothesis”辅助性原则	quotation	Google Search
	P9	“subsidiarity hypothesis”辅助性	quotation	Google Search
	P9	“subsidiarity hypothesis”辅助性假说	quotation	Google Search
	P9	“wild commodity”野生商品	quotation	Google Search
	P9	“wild commodity”野生	quotation	Google Search
	P9	“wild commodity”	quotation	Google Search
	P9	“野生商品”	quotation	Linguee
	P9	“the Namibian case study”	quotation	Google Search
	P9	“proprietorship–price–subsidiarity hypothesis”价格	quotation	Google Search
	P9	“proprietorship–price–subsidiarity hypothesis”	quotation	Google Search
	P16	“所有权-价格-补助”假说	quotation	Google Search
Task 3	P2	set the ^ perspective	caret	Google Search
Task 4	P2	“listen to patients”	quotation	DeepL as a dictionary
	P3	“black power movement”	quotation	Wikipedia

Overall, two types of advanced queries were constructed using search operators (see Feature Column in Table 28). One was the use of quotation marks, and the other was the use of caret. Quotation marks were the most commonly used, occurring 16 times. These serve to locate the exact phrase enclosed between the quotation marks, thus filtering unnecessary results on the search engine results pages. Caret was only used once. According to the etymology, the Latin word caret means "it lacks",[26] and when used by writers or proofreaders it means that something needs to be inserted there.

[26] https://en.wiktionary.org/wiki/caret

On taking a closer look at the caret, which P2 uses in Task 3, it seems that the source item the participant was attempting to translate was "set the evo-devo perspective". P2 formulated four queries in total to address this problematic phrase: 1) set a perspective, 2) set perspective, 3) set the ^ perspective, and 4) set the evo-devo perspective. The participant verbalised this search after completing the translation as follows:

> Set the evo-devo perspective, evo-devo 是在维基百科的时候看到的，是一个固定表达的查证。set the perspective 我结合上下文，也包括在搜词的意思，也并不是很能确定它到底是什么意思我觉得难度是 4.
>
> [Translation] Set the evo-devo perspective. I found the equivalence of evo-devo in Wikipedia, and it is a collocation. To translate “set the perspective”, I put it in the context and searched online, but am still not sure what it means. I rate the difficulty as four.

According to the actual queries the participant constructed and the verbalisation, this item was therefore perceived as quite difficult to translate (rated four),[27] as the participant could not determine the meaning of “set the perspective”. It is assumed that the caret was used here to replace “evo-devo”, and the participant was initially trying to search the meaning of “set the perspective”. Quality assessment of the translated texts is out the scope of this thesis, however, based on the above verbalisation, it is suggested that the participant did not successfully solve this problem, and the use of this search operator did not achieve the intended outcome. Based on the fact that only two types of advanced features were used, and one feature was only used once (and might have been used inappropriately), the changes in the search operators used during the four tasks could not be determined.

As shown in Table 28, 14 out of 17 of the search operators were used in search engine, specifically Google Search. The remaining three were used in Linguee, DeepL as a dictionary, and Wikipedia. For each task, advanced queries in Task 1 and Task 3 were all submitted in the search engine (100%), 92% (11/12) were employed for Task 2, but none were employed in Task 4 (0%). It seems that the advanced queries were more frequently raised while using resources other than search engines, but lacking enough evidence to confirm. On taking a closer

[27] As recalled, the difficulty rating from 1 (very easy) to 5 (very difficult).

look at those queries constructed in resources other in search engines. “野生商品” was constructed in Linguee, in Task 2 by P3. The “Listen to patients" query formulated in Task 3 by P2 is quite interesting. It was entered in DeepL, a machine translation engine, but because it is a phrase rather than a paragraph or text, the use of the machine translation engine here can be regarded as a dictionary, as mentioned in Section 3.1. This query was copied and pasted from the source text, and the quotation marks were as in the original source text. The use of quotation marks in dictionary or machine translation engines does not impact the results in DeepL, for example, “black power movement” was constructed in Wikipedia in Task 4 by P3. Due to the scarcity of advanced queries, however, it is inappropriate to hypothesise whether the location of the search operators changed.

Another interesting observation is that none of the participants mentioned the use of advanced queries in their retrospective verbal reports. This may indicate that 1) they were very familiar with the search operators already and use them unconsciously, however according to the rare use (only 0.73%), this cannot be true; 2) they were not familiar with the use of advanced queries; or 3) they strategically choose to use simple rather than complex queries in search engines.

From the perspective of the number of advanced queries, and their features, and locations, the differences between the four tasks are hard to detect due to the scarcity of advanced queries used in the data-collection sessions. All the evidence is compromised by the limited number of advanced queries.

Most of the previous studies relate advanced queries to information competence, search style, or web search expertise. One of the latest studies to have included search operators is Paradowska (2020), where search operators were directly employed as one of the indicators of the improvement of information competence. Ten search operators were included in the analysis: quotation marks, asterisk, minus, related, near, site, plus, define, location, and tilde. After a four-month intervention, there was a significant increase in the use of two of the search operators: quotation marks and site. In Enríquez Raído (2011), in the first task, 59.36% of advanced queries were performed by one professional participant alone. It was deduced that advanced queries were linked with a more in-depth search style. In Wang (2014), most of the professional groups used search operators, and in response to the questionnaire in Gough (2016) participants (all are professional translators) claimed they frequently used at least one advanced

feature, although 38% of them did not use any advanced search features in the study, and three of the four frequent advanced queries users were classified only as "shallow searchers". From this, I would deduce that using advanced queries is not only a reflection of web search expertise, but is also related to personal research style and preferences.

From a pedagogical perspective, Enríquez Raído (2014) emphasised the training need for the use of advanced queries to efficiently locate relevant information, and this was supported by Paradowska (2020) and implemented as an intervention, however no increase in the use of advanced queries over four sequential tasks was found.

7.3 Query language direction

The language direction of a query can, to some extent, indicate the participant's search directions. In this section, the queries were divided into three categories according to the languages used: source language (English), target language (Chinese), and a mixture of the source and target languages.

Figure 33 summarises the average number of entries in the three query direction categories. Across the four tasks, source language (English) was dominant, with 22.59 queries per task and 71.53%. The second most used target language (Chinese) accrued 6.41 queries per task and comprised 18.12%. The least used were the mixed source and target language, consisting only of 3.08 queries and 10.32%.

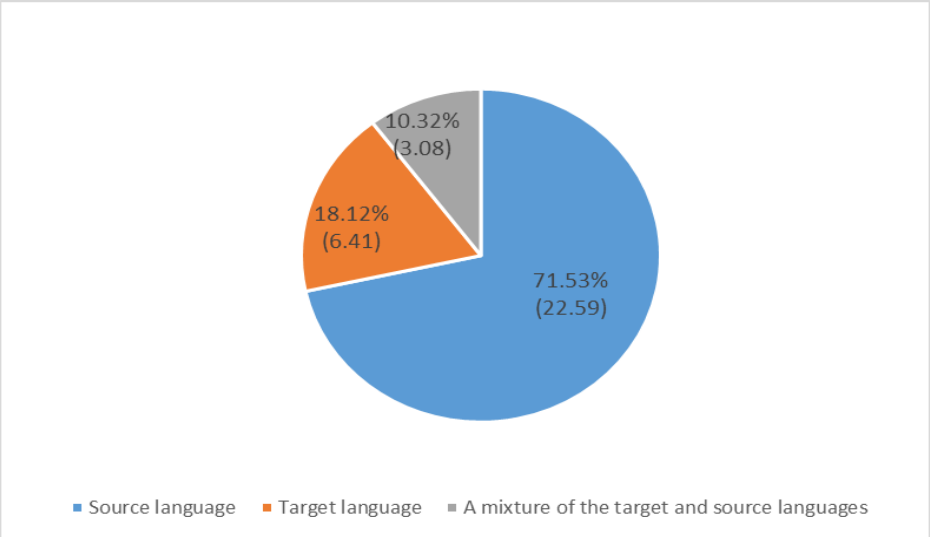

Figure 33 Average number of the source language, target language, and mixed queries

This distribution is in line with previous studies, such as in Shih (2017), where constructed queries predominantly consisted of the source language with only occasionally the target language, and in Gough (2016) where the prevalent queries were ST-oriented and TT-oriented queries were a small minority.

7.3.1 Search queries consist of the source language

The number of source language queries formulated during each task is presented in box and whisker plots in Figure 34 below.

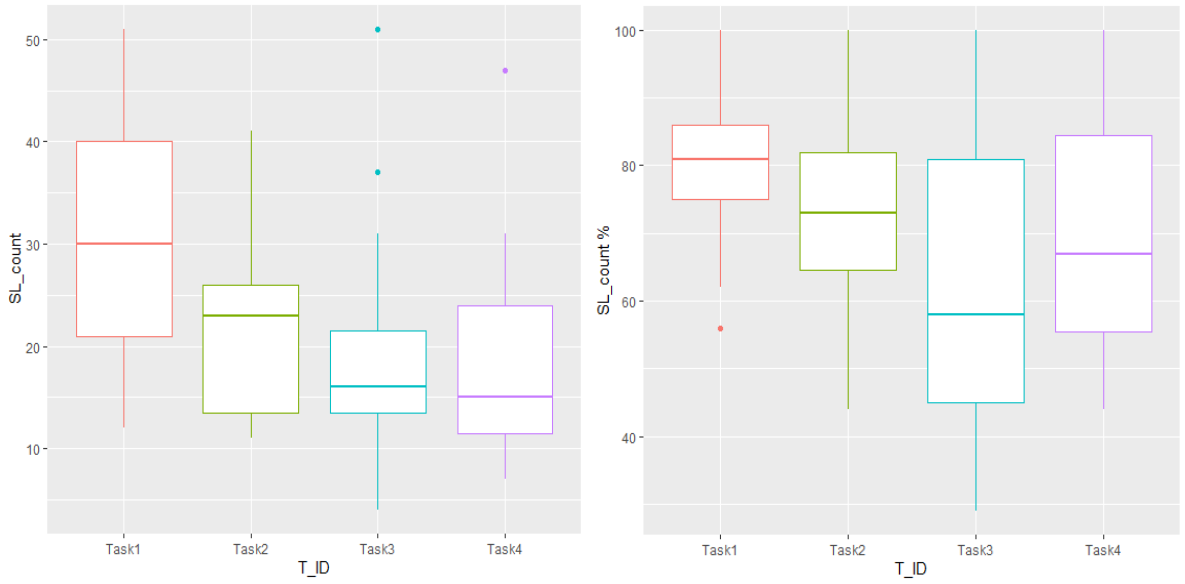

Figure 34 a) The number of source language queries and b) Percentage of source language queries

In Figure 34a, it is evident that in terms of the number of source language queries, there was a downward trend from Task 1 to Task 4 for both the median (the horizontal line inside the box) and the range (the vertical line extending from the box to indicate the extent of the lower and upper limit). In Figure 34b, the percentage of source language queries also decreased, while the range expanded to the lower side of the boxplots. Table 29, below, shows the descriptive summary of both the number of source language queries and the percentage of source language queries for each task.

Table 29 Descriptive statistics of the number of source language queries

T_ID	SL_count					SL_pct				
	mean	sd	IQR	min	max	mean	sd	IQR	min	max
Task1	30.63	12.07	19.00	12.00	51.00	80.00%	11.18%	11.00%	56.00%	100.00%
Task2	21.89	8.90	12.50	11.00	41.00	72.63%	14.33%	17.50%	44.00%	100.00%
Task3	19.37	11.10	8.00	4.00	51.00	62.05%	22.09%	36.00%	29.00%	100.00%
Task4	18.47	10.23	12.50	7.00	47.00	71.42%	18.52%	29.00%	44.00%	100.00%

Table 29 shows that the average number of source language queries for the four tasks decreased by 12.16, from Task 1 (with 30.63 queries) to Task 4 (18.47 queries). A downward trend can be observed from Task 1 to Task 2, with less than 8.71 source language queries constructed. The trend then levelled out with 2.52 and 0.9 fewer queries for the differences between Task 2 and Task 3, and Task 3 and Task 4, respectively. The range, and both the SD and IQR, also decreased. The average percentage of source language presents an overall decrease of 8.58% from Task 1 (80.00%) to Task 4 (71.42%), although there was an increase of 9.37% from Task 3 (62.05%) to Task 4 (71.42%), compared Task 4 to Task 1, the mean percentage still decreased. To confirm whether the participants used significantly fewer source language queries or smaller percentage as observed in descriptive statistics, Model 8 and Model 9, were built with the four tasks (T_ID) as the explanatory variable, the number of source language queries (SL_count) and the percentage of source language queries (SL_pct) as the response variables, and participants (P_ID) as the random effect.

Model 8 Regression model relating the number of source language queries

```
glmernb(SL_count ~ T_ID + (1|P_ID), data = websearch)
```

Model 9 Regression model relating the percentage of source language queries

```
lmerTest::lmer(SL_pct ~ T_ID + (1|P_ID) , data = websearch)
```

The results of Model 8 and Model 9 are shown in Table 30.

Table 30 Results of Model 8 and Model 9 (differences in the number and the percentage of source language queries during the four tasks)

Predictors	*Incidence Rate Ratios*	*CI*	*p*	*Predictors*	*Estimates*	*CI*	*p*
(Intercept)	29.06	23.71 – 35.63	<0.001	(Intercept)	80	72.34 – 87.66	<0.001
Task 2	0.71	0.58 – 0.88	<0.001***	Task 2	-7.37	-15.53 – 0.79	0.077
Task 3	0.63	0.51 – 0.77	<0.001***	Task 3	-17.95	-26.11 – -9.78	<0.001***
Task 4	0.59	0.48 – 0.73	<0.001***	Task 4	-8.58	-16.74 – -0.42	0.039*
Marginal R^2 / Conditional R^2	0.164 / 0.587			Marginal R^2 / Conditional R^2	0.124 / 0.503		

*p < .05. ***p < .001.

For the number of source language queries, all the p-values were significant, <0.05 (p<.001, p<.001, p<.001), with the model explaining 58.7% of the data. For the percentage of source language queries (SL_pct), differences in Task 3 (p<0.001) and Task 4 (p=0.039) are statistically supported, with the model explaining 50.3% of the data. In summary, it is reasonable to say that the participants constructed progressively fewer source language queries during the four tasks and the percentage of source language queries also occupied a smaller amount in Task 3 and Task 4. The effect plot can be seen in Figure 35.

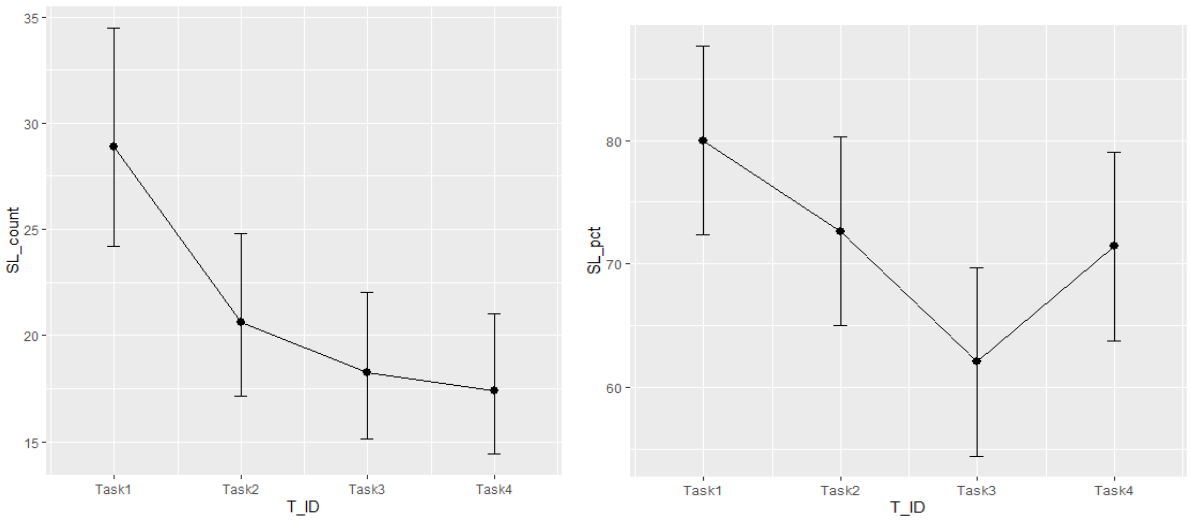

Figure 35 Effect plot of task on a) the number of source language queries and b) Percentage of

the number of source language queries

There are two possible reasons for the decreasing use of source language queries. One potential reason is that during the four tasks the participants' internal support could have been more robust: an improvement in their comprehension of the source language would lead to a decrease in the number of source language queries. The other possible reason is that the source language queries were transformed into the other two language-related types, i.e., target language queries and queries consisting of both source language and target language. As there is an intertwined relationship between source language queries and target language queries, the analysis of the decrease of source language queries will be discussed with the target language queries in the next section, after presenting results about the target language queries.

7.3.2 Search queries consist of the target language

In this section, the number of target language queries is portrayed in Figure 36 and summarised in Table 31.

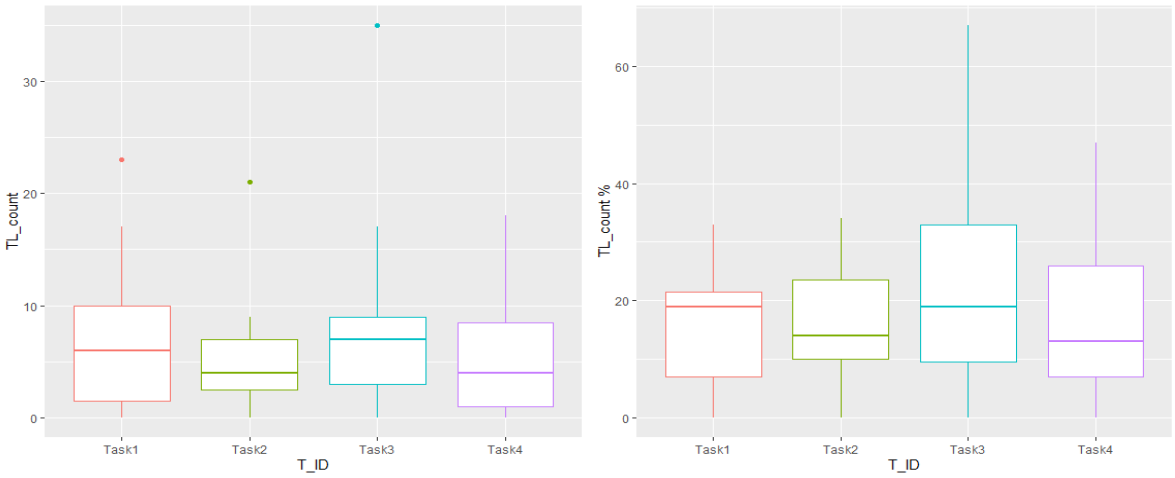

Figure 36 a) The number of target language queries and b) Percentage of target language queries

In Figure 36a, the general trend seems stable within certain limits. Except for a slightly smaller quartile in Task 2, the other three tasks present a similar pattern. In Figure 36b, the trend is also stable overall except for the target language query percentage in Task 3 with a notably greater range. Table 31 presents the descriptive statistical breakdown of the target language queries.

Table 31 Descriptive statistics of the number of target language queries

T_ID	TL_count					TL_pct				
	mean	sd	IQR	min	max	mean	sd	IQR	min	max
Task1	6.79	6.11	8.50	0.00	23.00	15.53%	10.21%	14.50%	0.00%	33.00%
Task2	5.37	4.67	4.50	0.00	21.00	16.42%	9.36%	13.50%	0.00%	34.00%
Task3	7.95	8.23	6.00	0.00	35.00	22.47%	17.69%	23.50%	0.00%	67.00%
Task4	5.53	5.75	7.50	0.00	18.00	18.05%	15.48%	19.00%	0.00%	47.00%

From Table 31, it can be seen that the mean value for the number of target language queries first decreased by 1.42% from Task 1 to Task 2, then increased by 2.58 to Task 3 before decreasing by 2.42% to Task 4. The mean percentage of target language queries increased by 6.94% from Task 1 (15.53%) to Task 3 (22.47%), and then decreased by 4.42% from Task 3 to Task 4 (18.05%). These differences were, however, relatively small. Two further regression models were built to check whether these differences are significant.

The linear mixed-effects model, Model 10 and Model 11 were built with Task ID (T_ID) as the explanatory variable, the number of target language queries (TL_count) and the percentage of target language queries (TL_pct) as the response variables, and Participant ID (P_ID) as the random effect variable.

Model 10 Regression model relating the number of target language queries

```
glmernb(TL_count ~ T_ID + (1|P_ID), data = websearch)
```

Model 11 Regression model relating the percentage of target language query

```
lmerTest::lmer(TL_pct ~ T_ID + (1|P_ID) , data = websearch)
```

The results are presented in Table 32, below.

Table 32 Results of Model 10 and Model 11 (differences in the number and the percentage of target language queries during the four tasks)

	TL_count				sqrt(TL_pct)		
Predictors	*Incidence Rate Ratios*	*CI*	*p*	*Predictors*	*Estimates*	*CI*	*p*
(Intercept)	4.79	2.89 – 7.96	<0.001	(Intercept)	3.58	2.71 – 4.45	<0.001
Task 2	0.87	0.58 – 1.32	0.523	Task 2	0.24	-0.57 – 1.05	0.559
Task 3	1.15	0.78 – 1.72	0.481	Task 3	0.58	-0.23 – 1.39	0.162
Task 4	0.78	0.52 – 1.18	0.237	Task 4	0.1	-0.71 – 0.91	0.805
Marginal R^2 / Conditional R^2	0.018 / 0.705				0.013 / 0.572		

As Table 32 shows, none of the differences was significant for both the number of target language queries and the percentage of target language queries, with all the p-values greater than .05 (p=0.523, 0. 481, 0. 237) respectively for three tasks for TL_count, and p=0.559, 0.162, 0.805 respective for the three tasks for TL_pct, therefore participants did not construct different numbers and percentage of target language queries.

In previous research, Enríquez Raído (2011) found that students consulted predominantly with source-text queries while professionals used target language as queries more often than students. Other studies, although using different terminologies, rather than explicitly using 'source language' or 'target language', reached similar conclusions. Zheng (2014) used "reverse lookup" (p.2) to refer to the use of the target language as queries. In three groups of participants, i.e., professionals, semi-professionals, and novices, the professional group used the "reverse lookup" most frequently, and "reverse lookup" was found to be correlated with "consultation profiency" (Zheng 2014, p.2). Wang (2014), used "主动查证"[Translation:active lookup] (p.25) and defined it as "输入事先想到的英文对等表达" [Translation: entering English (target language) equivalent], in this study the professional group was also empirically proven to use "主动查证" [Translation:active lookup] the most frequently. Onishi and Yamada (2020), used "target language entry" and "source language entry". In this study, similar to other studies mentioned above, trainee translators constructed more "source language entry" queries, while professional translators performed more "target language entry" queries. This could be attributed to the participants' shift from source language comprehension problems to target language production problems. Gough (2016) coded the "research direction" (p.180) as ST-oriented research (initiated with source language queries) and TT-oriented research (initiated with target language queries). However, this study concluded differently to the studies

mentioned above, finding no relationship between the length of experience and the research direction, and hypothesised that research direction might be related to familiarity with the source-text domain and individual preference or habit.

From retrospective verbalisation, participants frequently mentioned using source language queries to find a meaning, and target language queries to confirm the provisional translation.

Example:

> Rambunctious Garden, Rambunctious 它的直译指的是喧闹的，指的是这位作者他写的一本书，我们把这个名字输入到谷歌中，输入到谷歌中可以看到是喧闹的花园，把喧闹的花园反向输入到谷歌中，也会看到 rambunctious garden 这个词。
>
> [Translation]: Rambunctious Garden. The literal meaning of Rambunctious is noise and excitement, and (Rambunctious Garden) refers to the book the writer wrote. As I put Rambunctious Garden in Google, 喧闹的花园 is appeared in Google. Then, I put 喧闹的花园 in Google, Rambunctious Garden appeared.

This example indicates that the participant constructed a source language query "Rambunctious Garden" to find the Chinese meaning, and then the participant formulated a target language query "喧闹的花园" to confirm whether the provisional translation was correct.

7.3.3 Search queries consist of mixed languages

In this section I report the findings of an analysis concerning queries consisting of both source and target languages.

Figure 37 presents boxplots representing the number and the percentage of mixed queries formulated by participants during the four tasks.

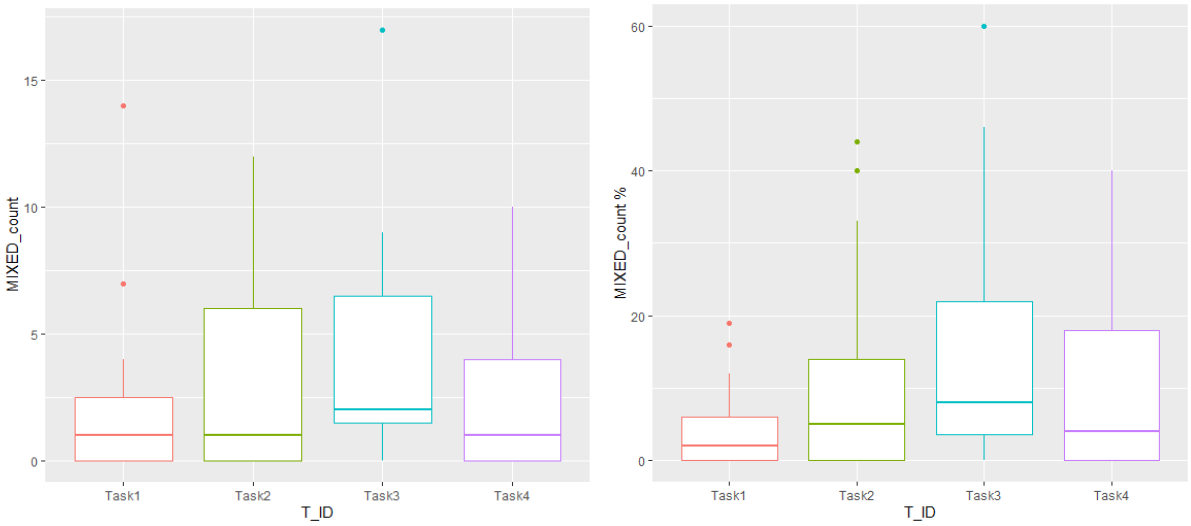

Figure 37 a) The number of mixed language queries and b) Percentage of mixed language queries

In Figure 37a, it can be seen that, in terms of the number of mixed language queries, there was a general upward trend, and the distributions for Task 2, Task 3, and Task 4 were noticeably skewed to the lower end. The most distinctive feature was that the upper 50% of the data above the median had a greater range, with the upper 50% of Tasks 2, 3, and 4 noticeably greater than Task 1. In Figure 37b, in terms of the percentage of mixed language queries, an upward trend can also be observed. The descriptive statistics for mixed queries during the four tasks are presented in Table 33.

Table 33 Descriptive statistics of the number of mixed queries

T_ID	**MIXED_count**					**MIXED _pct**				
	mean	sd	IQR	min	max	mean	sd	IQR	min	max
Task1	2.05	3.42	2.50	0.00	14.00	4.37%	5.70%	6.00%	0.00%	19.00%
Task2	3.42	3.92	6.00	0.00	12.00	11.05%	13.83%	14.00%	0.00%	44.00%
Task3	4.05	4.26	5.00	0.00	17.00	15.32%	16.97%	18.50%	0.00%	60.00%
Task4	2.79	3.51	4.00	0.00	10.00	10.53%	14.09%	18.00%	0.00%	40.00%

The average number of mixed queries in Tasks 2, 3, and 4 were all greater than in Task 1. An increase of 1.37 was observed between Task 1 (with 2.05 mixed queries) and Task 2 (with 3.42 mixed queries), and again between Task 3 which reached a peak with 4.05 for the mixed queries. After the early increases, a decrease of 1.26 was observed from Task 3 (with 4.05 mixed queries)

to Task 4 (with 2.79 mixed queries). The range, SD, and IQRs of Task 2, Task 3, and Task 4 were all greater than for Task 1. In terms of percentage, the mixed queries increased by 10.95% from Task 1 (4.37%) to Task 3 (15.32%), with a slight decrease of 4.79% from Task 3 to Task 4 (10.53%).

Model 12 and Model 13 were then built to test the significance of these differences, with Task ID (T_ID) as the explanatory variable, the number of mixed queries (MIXED_count) and the percentage of mixed queries (MIXED_pct) as the response variables, and Participant ID (P_ID) as the random effect.

Model 12 Regression model relating the number of mixed queries

```
glmernb(MIXED_count ~ T_ID + (1|P_ID), data = websearch)
```

Model 13 Regression model relating the percentage of mixed query

```
lmerTest::lmer(sqrt(MIXED_pct) ~ T_ID + (1|P_ID) , data = websearch)
```

Model 12 and Model 13 results are summarised in Table 34.

Table 34 Results of Model 12 and Model 13 (differences in the number and the percentage of mixed language queries during the four tasks)

	MIXED_count				**sqrt(MIXED_pct)**		
Predictors	*Incidence Rate Ratios*	*CI*	*p*	*Predictors*	*Estimates*	*CI*	*p*
(Intercept)	1.20	0.59 – 2.47	0.611	(Intercept)	-0.91	-2.41 – 0.58	0.232
Task 2	1.76	0.92 – 3.34	0.086	Task 2	0.84	-0.68 – 2.35	0.279
Task 3	2.08	1.11 – 3.88	0.022*	Task 3	2.25	0.74 – 3.77	0.003**
Task 4	1.31	0.69 – 2.49	0.416	Task 4	0.7	-0.82 – 2.21	0.366
Marginal R^2 / Conditional R^2	0.040 / 0.660				0.058 / 0.518		

*p < .05. **p < .01.

Table 34 shows that Task 3 was significant (p=0.02), while Task 2 and Task 4 were insignificant (p=0.086, p=0.416), with an R^2 of 0.66. This suggests that the number of mixed queries was significantly greater in Task 2 and Task 3 compared with Task 1. The differences

of the mixed percentage (MIXED_pct) in Task 3 (p=0.003) was confirmed. The effect plot is presented in Figure 38.

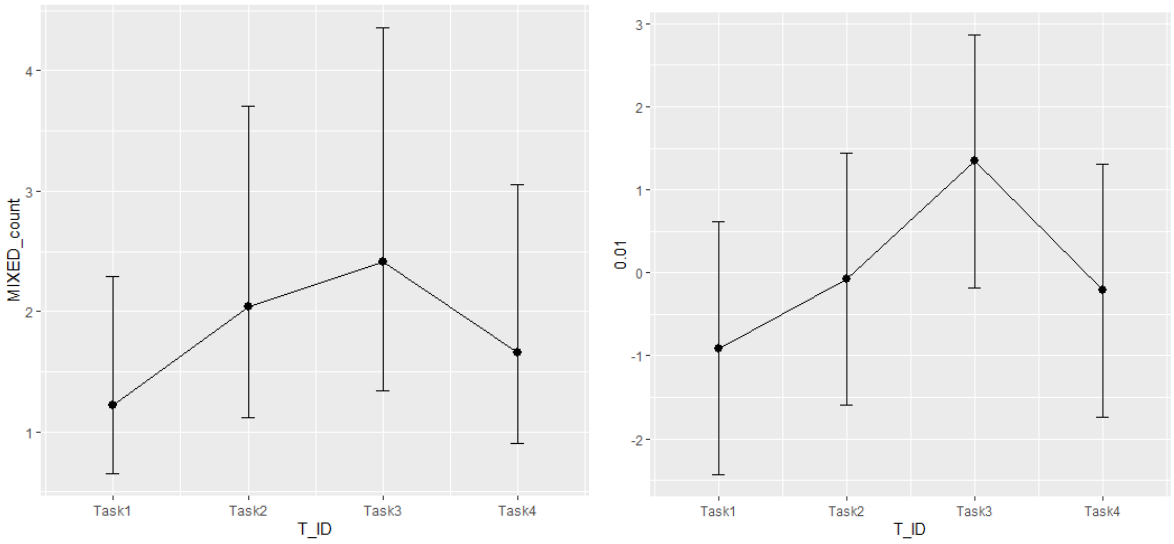

Figure 38 Effect plot of task on a) The number of mixed queries and b) Percentage of the number of mixed queries

As mentioned, the mixed use of both source language and target language queries is often ignored, and there are only a few studies that have mentioned it. Of these, Wang (2014) found that of the three groups, the semi-professionals used the most mixed language queries. Wang and Lim (2016) directly regarded the use of source language and target language as "sophisticated search techniques" (p.68), however the junior group were found to employ mixed language queries more often than the senior group. It was reasoned that this might have been because the senior group had better internal support, so the sophisticated search techniques seemed unnecessary. In a study by Shih (2017) the mixed language queries were further categorised into 1) a ST term and its corresponding TT equivalent, 2) a ST term and some question words in the target language, 3) a ST term and the name of the TL. The findings suggested that the mixed language queries were very effective.

The mixture use of source language and target language was mentioned frequently in the retrospective verbal reports transcriptions. For example, when participants explained their search process:

Example 1:

接下来的一个是 invasion biological，这个在词典里直接输入词组，并没有查到，通过输入全部的英文以及加入生物学家这几个汉字，在必应网页里面可以查到有相关的文章里边写到是入侵生物学家，于是我采用了这个译法,难度是 2。

[Translation] The next one I searched is invasion biological. First, I entered “invasion biological” in the dictionary, but I didn't get any results. Then, I entered “invasion biological 生物学家” in the search engine, then I found a parallel text, and it contains the equivalence of 入侵生物学家, then I adopted this translation. I rated the difficulty as two.

Example 2:

institutional economics。这个我先是在 Linguee 上查，他给出的是他给的是机构经济学，但是又不对，在百度上查到有机构经济学，但是并不是这个说法，跟一些对称性或是不对称性的经济学有关。 这个我又在谷歌上结合它的原文，以及输入了一个经济学的关键词联合搜索，搜索出来它是制度经济学，那么反向在中文的网站上进行查证，也证明了这个说法，所以我给他打 3 分.

[Translation] Institutional economics. Firstly, I started to search on Linguee, entered “institutional economics”, and got the result “机构经济学”。Then, I entered “机构经济学” in Baidu, and found out that it was not the thing that I wanted to translate. After this, I entered “institutional economics 经济学” in Google, and found out “制度经济学” could be the candidate. Then, I entered 制度经济学 in a Chinese website (Baidu), it proved that 制度经济学 is the right one. I rated the difficulty as three.

From the two examples above, the participants seemed satisfied with the results after using mixed language queries.

To sum up, in terms of the query language, the dominant direction of search queries was the source language, followed by the target language, and then mixed. The source language queries, both the number and percentage, present a significant overall decrease over sequential tasks (except for the percentage in Task 2 which was insignificance). The number of target language

queries (both the number and percentage) remained unchanged; no significant difference was found. The mixed language queries increased, with a significant number of mixed languages queries detected in Task 2 and Task 3 and the percentage of mixed language queries found in Task 3.

7.4 Summary

In conclusion, the translator trainee's query behaviour is investigated in terms of query time, query complexity, and query language direction over the Chinese-English translation course. According to the data collected and the analysis made, query time did not present much change. The changes in query complexity were hard to confirm since the advanced queries were only sporadically constructed, only 0.73% (18/2439) for all the four tasks combined. Lastly, in terms of query language direction, it is evident that participants constructed fewer source language queries in four tasks, a similar amount of target language queries, and more mixed language queries. These findings speculate/demonstrate that over the four tasks which took place over the translation course, the participants acquired more robust internal support and became more effective at constructing queries to solve their translation problems.

8. Conclusion

In this study I have investigated the changes in trainee translators' resource use and query behaviour. The empirical data for this purpose were gathered over four months in a translation practice course with the English and Chinese languages for Master-degree students, and four data collection sessions were implemented in a remote setting. The data collection tools included a questionnaire, four source texts and their translations, a key logger (Inputlog), a screen recorder (Screencast-O-Matic), and retrospective verbal reports. Nineteen participants each performed four tasks. In this chapter, I summarise the main findings of this study, and then consider the limitations. Finally, the contributions of the research and future avenues for further research are discussed.

8.1 Main findings

The findings of this thesis are summarized around the two research questions concerning the changes in the participants' **resource use** and **query behaviour**.

8.1.1 Resource use

The first research question concerned the changes in the use of online resources by the participants when completing the four translation tasks over the course of a semester. The analysis was based on the categorisation proposed in Section 3.1 and the results of the models is summarized in Table 35 below.

Table 35 Summary of the findings of resource use

Categories	Measurements	Task 2			Task 3			Task 4		
		Estimates	CI	p	Estimates	CI	p	Estimates	CI	p
Dictionaries	Time	-3.23	-6.32–-0.14	0.041*	-4.61	-7.70–-1.52	0.003**	-4.92	-8.01–-1.83	0.002**
	Percentage	-0.08	-0.14–-0.01	0.031*	-0.12	-0.19–-0.05	<0.001	-0.09	-0.16–-0.02	0.011*
Search engines	Time	-1.22	-4.46–2.02	0.46	-1.07	-4.31 – 2.17	0.519	-3.5	-6.74–-0.26	0.034*
	Percentage	-0.03	-0.10–0.05	0.465	0.03	-0.05 – 0.10	0.476	0.04	-0.03–0.11	0.286
Knowledge-based resources	Time	0.88	-0.25–2.00	0.126	0.64	-0.48 – 1.77	0.263	-0.37	-1.49–0.76	0.524
	Percentage	0.1	0.02–0.18	0.012*	0.09	0.01 – 0.17	0.023*	0.05	-0.03–0.13	0.228

*p < .05. **p < .01. ***p < .001.

Resources were first divided into three categories: dictionaries, search engines, and knowledge-based resources. For dictionaries, compared to Task 1, the actual time spent and percentage time spent in Task 2, Task 3, and Task 4 all showed significant differences. More specifically, time spent consulting dictionaries decreased substantially and quickly (as early as Task 2). This result is consistent with almost all previous empirical studies, which compared differences between different levels of expertise, such as Enríquez Raído (2011), Wang (2014), and between the same group of participants tested at different points of time, such as Chang (2018) and Paradowska (2020). Other related findings include the time spent on machine translation engines as bilingual dictionaries (defined and justified in Section 4.8.3.3) as dictionary time decreased sequentially, and the time spent in Task 4 only amounted to half of the time spent in Task 1. Time spent on each specific dictionary decreased, except for the only increase which was for the use of Lingoes. This could be explained by the fact that Lingoes comprises customized dictionaries, which can incorporate other online dictionaries (such as Merriam-Webster Dictionaries, or Collins Dictionaries), presenting a larger scope of information.

With regard to search engines, except for the significant reduction in time spent on Task 4, the time and percentage time spent on other tasks remained stable. This was to some extent comparable with previous studies where both the magnitude of the time differences between participants, and the changes in the use of dictionaries due to specific pedagogical interventions, were evident. For example, Onishi and Yamada (2020) found a similar proportion of search engine percentage time between novice and professional translators, however Paradowska

(2020) did not find any significance after a 4-month intervention which aimed to develop participants' information competence. Only Chang (2018) found a slight increase in the frequency of search engine usage after a one-year postgraduate translation course.

Baidu, Google, and Bing were the three most frequently used search engines across the four tasks. Time spent on Google searches was relatively stable while time spent on the other two (Baidu and Bing) decreased. From the retrospective verbal reports data, participants claimed that choosing Google instead of Bing or Baidu helped them gain access to richer content. Another finding was that when different search engines used by the participants in this research (Google 61.8% Baidu 20.9%) were compared to the general usage of search engines by users at large in China overall in 2020 (Google 2.51% Baidu 84.27%), the popularity rating of the two search engines were reversed.

Differences in the time spent on knowledge-based resources for the three tasks were not significant. Differences in the percentage of time spent on knowledge-based resources was significant however, and the significant differences were found in both Task 2 and Task 3. This is similar to the general literature, for example Chang (2018) and Onishi and Yamada (2020) also found differences, although they used different terms to refer to the resources accessed other than dictionaries and search engines. General web pages were the dominant, alternative type of knowledge-based resource used across the four tasks, although time spent on them decreased sequentially over the four tasks. The use of other specialized or specific resources, such as academic documents, however increased.

Taking those three resources together into consideration, a clear trend away from dictionaries to knowledge-based resources can be observed. These results are to be expected since dictionaries mainly provide lexical equivalent information, whereas knowledge-based resources provide more in-depth, text-based information. Compared to dictionaries and knowledge-based resources, search engine usage remained relatively stable over the four tasks. As the most important element in search engines is search queries, the second research question specifically related to search query behaviour is fully summarized in Section 8.1.2 below.

8.1.2 Query behaviour

The findings pertinent to the second research question concerning the changes in query behaviour observed over the semester, are summarized in Table 36.

Table 36 Summary of the findings of query behaviour

Operationalization		Task 2			Task 3			Task 4		
		Estimates	CI	p	Estimates	CI	p	Estimates	CI	p
Query_time		0.18	-0.05 – 0.41	0.134	0.07	-0.16 – 0.30	0.565	-0.02	-0.25 – 0.21	0.894
Query_complexity		cannot firm with 0.73% (18/2439)								
Query language direction	SL_count	0.71	0.58 – 0.88	<0.001***	0.63	0.51 – 0.77	<0.001***	0.59	0.48 – 0.73	<0.001***
	SL_pct	-7.37	-15.53 – 0.79	0.077	-17.95	-26.11 – -9.78	<0.001***	-8.58	-16.74 – -0.42	0.039*
	TL_count	0.87	0.58 – 1.32	0.523	1.15	0.78 – 1.72	0.481	0.78	0.52 – 1.18	0.237
	TL_pct	0.24	-0.57 – 1.05	0.559	0.58	-0.23 – 1.39	0.162	0.1	-0.71 – 0.91	0.805
	MIXED_count	1.76	0.92 – 3.34	0.086	2.08	1.11 – 3.88	0.022*	1.31	0.69 – 2.49	0.416
	MIXED _pct	0.84	-0.68 – 2.35	0.279	2.25	0.74 – 3.77	0.003**	0.7	-0.82 – 2.21	0.366

*p < .05. **p < .01. ***p < .001.

As can be seen in Table 36, query time did not change significantly. As explained earlier, this might be due to the fact that the graduate level participants were already familiar with how to construct queries in order to solve their search problems. They have therefore already formed their own working routines and habits.

The changes in query complexity were hard to confirm since the advanced queries were only sporadically constructed and only comprised 0.73% (18/2439) on all the four tasks combined. This result contrasts with apparent differences found between novice and professional translators (Wang 2014), and the clear development (formulating more advanced queries) found by Paradowska (2020) after a pedagogical intervention.

The number and percentage of source language queries decreased for query language direction over the four tasks (except for the percentage of source language queries in Task 2), while the number and percentage of mixed language queries increased in Task 3. Both the number and percentage of target language queries remained stable across all four tasks. The dynamic among different query language directions is in line with previous studies, such as Wang (2014) and Onishi and Yamada (2020), but both these two studies only categorized the queries into source

language and target language. The additional category in this thesis, i.e., the combination of source and target language queries, which points to a more refined categorisation.

As presented above, the changes in resource use, are clear for both the first and second levels, however, for query behaviour, the only change confirmed from the inferential analysis relates to the number of the three types of language-direction queries, rather than query time and query complexity. According to the data, the two main variables (resource use and query behaviour) in this thesis seem to have not developed at the same pace. It could be speculated that resources are WHAT the participants used while query behaviours are essentially about HOW the participants searched. The former change is therefore easier to acquire while the latter could be regarded as habits, hence taking longer to change. For example, one of the resources used by the participants was UN terms.[28] When the participants realized it was a useful website, they used it more frequently in the following tasks. The use of advanced queries seemed to be more challenging however, perhaps because they needed more practice and more time and confidence to become familiarised and assimilate advanced queries into their customary skill set.

Despite every effort, there were still some limitations in this study which are inherent in other empirical research in translation studies. In the following section, I introduce the limitations first and then provide some future avenues for research.

8.2 Limitations

To start, one of the limitations of this thesis is that, although this empirical research conducted from a developmental perspective, the four data collections were conducted in a relatively short period of time, which was only four months, and this program lasted only two to three years. This may be a reason that this could not be counted as a complete process for their web search development.

A primary concern is the limited number of participants in this research. In total, there were 76 observations (19 participants*four tasks) in the final dataset, which although compared to other translation-oriented web search studies with the use of inferential analysis, such as Onishi and

[28] https://unterm.un.org/unterm/portal/welcome

Yamada (2020) with five students and four professionals, and Paradowska (2020) with only ten participants tested at two time points, is a comparably large cohort. The small number does however limit the use of statistical analyses, and the results are therefore difficult to generalize to a wider population, such as to all trainee translators working with an English-Chinese translation task, or to all trainee translators worldwide.

Sound statistical methodology and practice are important components when examining any research hypothesis, and consequently, there are limitations which should be acknowledged in the present thesis. All the linear mixed models built in this thesis are additive, meaning no interactions were assumed. This is justified in part due to the participants' background information, collected by questionnaire as presented in Chapter 5, which is all very similar and suggests that all the participants came from a homogenous cohort. In addition, no additional data were collected to adjust for potential confounding variables that might have compromised the validity of the explanatory variable (the four tasks). For example, variables such as age and gender could have been collected and incorporated into the analyses.

8.3 Contributions

In this section, I summarize the theoretical, methodological and pedagogical contributions of this study.

8.3.1 Theoretical contributions

In this thesis a two-dimensional theoretical framework derived from translation studies and LIS was introduced, and the two dimensions were resource use and query.

The term 'resource' in the present study refers to the online resources the translators (or trainee translators) used for solving translation problems. Due to the rapid development of technology and the emergence of different resources, a greater number and variety of resources related to different translation language pairs are now available. In addition, with different research purposes and a wide variety of potential source texts to be translated, the comparability of resources is hard to achieve. The two-level categorisation proposed in the present study enables comparability between different studies and ensures the flexibility to investigate changing

resource use since it can potentially include any new emerging resources, which is especially useful during the rapid development of available online resources. The first-level categories comprised dictionaries, search engines, and knowledge-based resources, which included all resources that could not be categorized into dictionaries and search engines. This was developed from previous literature and based on the uniqueness of each of the three resource categories. The second level comprised further categories within each of the three first-level categories determined by a bottom-up approach, and informed by the collected data.

Search queries in this investigation refer to items typed into search engines, dictionaries, or any other searchable resources, excluding those designed to post-edit a whole sentence, paragraph, or text in machine translation engines (explained in Section 4.7.3). The analytical framework of query behaviour is composed of three indicators. 'Query time' represent the average time a participant spent on a query; 'Query complexity' is represented by the number of advanced queries, including advanced search operations, operators, and term modifiers; and 'query language direction', categorized as source language queries, target language queries, and queries composed of the source language and target language. The query language direction could to some extent imply the search direction of the searcher. These are the three important indicators for evaluating query behaviour in translation studies and LIS.

The proposed theoretical framework could be used in other translation-oriented web search studies, to observe the comparison between different levels of expertise or the changes in the same group of participants at different times.

8.3.2 Methodological contributions

From a methodological perspective a research design was developed and refined with the simultaneous use of a key logger and a screen recorder in a remote setting. More specifically, the participants recorded their translation processes independently and remotely, without the researcher at hand. This methodology was developed through three pilot studies followed by the main study comprising four sequential data collection points. The data collection process was integrated and highly replicable, including the participant recruitment and informed consent procedures, the instruction email, and the step-by-step video tutorial, which can all be adapted and used as a template for similar studies.

Applications of remote settings in translation-oriented web search studies are emerging, as exampled in studies such as Gough (2016) and Shih (2017, 2019). The remote setting adds the "quality and credibility" (Saldanha & O'Brien, 2014, p. 33) of the research compared with an onsite setting, which is often criticized for compromised ecological validity (Mellinger & Hanson, 2017). In addition, the remote setting increases the number of potential participants involved since it had no time constraints, for the participants could complete the translation in their own time, and there are no location constraints as they did not need to travel. This was particularly pertinent in the pandemic period. In the above studies, however, the data collection tool used to record the translation process was only a screen recorder. The applied methodology in the present thesis added the use of a key logger (Inputlog), which generated a timeline with actions (such as mouse movements and clicks) carried out by the participants. This allowed a semi-automatic way of processing and analyzing collected data, which greatly reduced the workload of researchers, further enabling a larger number of participants.

8.3.3 Pedagogical contributions

In this section, I first present the findings directly retrieved from Section 8.1. Then, I consider the wider perspective to discuss the pedagogical benefits of web search instructions. I conclude by presenting pedagogical implications and applications in Chinese settings.

The findings show that the participants moved away from dictionaries to knowledge-based resources, and the diversity of knowledge-based resources accessed expanded while the time spent in accessing different dictionaries diminished. Time spent on different dictionaries sequentially decreased, except for Lingoes, which provides more information, compared to other traditional dictionaries. The use of knowledge-based resources shifted from general web pages to more specialized or specific resources, such as academic documents. These changes indicate that the participants shifted their focus from lexical problems to larger and more contextual problems. For example, less time was spent on finding word equivalents in dictionaries, while more time was devoted to checking background information from knowledge-based resources, such as encyclopaedias. This trend provides a confirmed perspective to enable the translation trainers to understand how trainee translators' web search processes change and to evaluate their web search competence. Based on this empirical evidence, future course designs can take into account what could be improved and adjust

accordingly to match the students' learning needs (Hao & Pym, 2021), which ultimately helps students solve translation problems more efficiently and effectively.

The differences found in the search engines accessed by participants in this study (Google 61.8% Baidu 20.9%), compared with the use of search engines accessed in China overall in 2020 (Google 2.51% Baidu 84.27%), calls for the need for web search training in the context of translation compared to web search training in general. For example, some universities offer information literacy courses which are led or taught by librarians for students without identifying their subject areas. This highlights the need for contextualized information literacy training for translator education.

The change in query complexity among the four tasks was difficult to determine due to the limited use (0.73%) combined in the four tasks, however, one thing that can be inferred is that the usage can, and should, be improved, as some studies have empirically shown that the advance use of queries can efficiently and effectively help searchers to solve their problems (Wang 2014, Paradowska 2020). Translator training should include, but not be confined to the use of search queries. Other contents, such as the basic knowledge of how search engines work (Enríquez Raído 2011, Wang 2014, Shih 2019), could help trainee translators better understand the use of advanced queries and encourage them to use them appropriately, also, as mentioned earlier, more practice and time are needed to develop student confidence and familiarity with search queries.

The mismatch between self-declared resource use and the resource use evident in the data collected and analysed in this thesis suggests that self-reflection, or self-awareness, needs to be incorporated into web search instruction. This would benefit both students and teachers. It would help students better understand their web search processes, and help trainers to identify trainee weaknesses, and the reasons behind this mismatch could provide insights into training pedagogy. LIS scholars have also suggested in higher education that "reflective thinking" (Sanches et al., 2022, p. 489) is beneficial for enhancing students' information literacy skills and for promoting "lifelong learning" (Webber & Johnston, 2014, p. 15).

In China 2022, there are translator training programs at both bachelor level (Bachelors of Translation and Interpreting, BTI) and master's level (Masters of Translation and Interpreting

MTI) reaching 316 and 282 programs, respectively.[29]

Formal training in translation-oriented web searching seems to be only a component in module design, and web search-related instruction is often embedded in other courses, such as Translation Practice or Computer-aided Translation courses (Yu & Que, 2019), but not offered as a stand-alone course. Compared with other countries, such as Switzerland, which initiated formal training as early as 2002, in a course entitled "Tools for Translators" and another similar course run by the Zurich University of Applied Science, named "Internet per mediatori linguistici" (Massey et al., 2007). In Spain, there is also a compulsory course entitled Documentación aplicada a la traducción y la interpretación (translated as "Information literacy for translators and interpreters") at the undergraduate level (Sales, 2022, p. 2). The pedagogical contribution of this study could inform the curriculum design and implementation of a dedicated course to address this training gap in China. For instance, it provides empirical evidence for the need to propose some of web-search instructions as early as possible. As already discussed in Section 3.2.2 and summarized in Section 8.1.2, the use of advanced queries probably takes a longer time to acquire, therefore it is better to incorporate this type of instruction in the early stages of the program.

8.4 Future work

In this section, possible future avenues for related research are discussed in terms of data collection tools, statistical methods, as well as for researching the web search process in general.

The main data collection tool used in this research, Inputlog, was initially designed for writing process research. Translation is sometimes regarded as a type of writing, and this opens up various exciting research possibilities (Muñoz 2022). As discussed, Inputlog can only be used with a Windows operation system (compared to the ISO operation system), which also greatly limited the number of potential participants in the present study. This limitation is understandable since Inputlog was originally developed for researchers as compared to the participants. It would be great to have different versions for different types of Inputlog users, for example, for researchers in strict experimental settings and for participants who collect their

[29] https://cnbti.gdufs.edu.cn/index.html

own data remotely.

Statistically, in future empirical web search research, more possible confounding variables could be explored to extend the additive models, as well as allowing for interaction effects. In addition, when analysing collected data, some categorical variables could be used as a numeric variable for predictive purposes. For example, in this study, the four tasks (Task_ID) were regarded as categorical variables to investigate the changes among the four tasks. These four tasks could also be used as a numeric variable, and then it would be possible to reasonably model what might happen in a fifth task, sixth task, and so on.

Translation process research is moving forward and there is a need for more rigorous research to test for repeatability and validity (Mellinger, 2020). Center for Research and Innovation in Translation and Translation Technology (CRITT) Database (Carl, 2021) is currently the most influential one which consists of more than 500 hours translation processes. To the best of my knowledge, however, although this includes some web search process data recorded by Inputlog and screen recorder, the ecological validity could be considered dubious since: 1) in most datasets, the participants were asked not to spent too much time on searches; 2) the Translog-II requires the screen layout for translation and searching to be fixed throughout the recorded process (see Section 4.7.3). There is therefore an urgent need for a similar, but ecologically valid, database for the web search process, for as observed by Enríquez Raído (2011), “one must cut down a whole forest to create a single toothpick” in translation process research, and this is still true today.

In conclusion, however the novel semi-automatic method of processing and integrating Inputlog data and screen recording data will enormously alleviate the researchers’ burden during data processing and analysis. It should now also be possible, by implementing the comprehensive methodology developed in this thesis, to recruit a larger number of participants in future studies, who can join the research at any time and remote setting that suits them best.

References

Aharony, N., Limberg, L., Julien, H., Albright, K., Fourie, I., & Bronstein, J. (2017). Information literacy in an era of information uncertainty. Proceedings of the Association for Information Science and Technology,

Alabau, V., Carl, M., Casacuberta, F., GarcíaMartínez, M., González-Rubio, J., BartoloméMesa-Lao, Ortiz-Martínez, D., Schaeffer, M., & Sanchis-Trilles, G. (2016). Learning Advanced Post-editing. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*. Springer.

Alcina, A. (2008). Translation technologies. *Target. International Journal of Translation Studies, 20*(1), 79-102. https://doi.org/10.1075/target.20.1.05alc

Alves, F. (2015). Translation process research at the interface. In *Psycholinguistic and Cognitive Inquiries into Translation and Interpreting* (pp. 17-40). https://doi.org/10.1075/btl.115.02alv

Balling, L. W. (2008). A Brief Introduction to Regression Designs and Mixed-effects Modelling by a Recent Convert.

Balling, L. W., & Hvelplund, K. T. (2015). Design and Statistics in Quantitative Translation (Process) Research. *Translation Spaces, 4*(1), 170-187. https://doi.org/10.1075/ts.4.1.08bal

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval. *Journal of Documentation*.

Bowker, L., & Pearson, J. (2002). *Working with Specialized Language A Practical Guide to Using Corpora*. Routledge.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological), 26*(2), 211-243.

Brown, V. A. (2021). An Introduction to Linear Mixed-Effects Modeling in R. *Advances in Methods and Practices in Psychological Science, 4*(1). https://doi.org/10.1177/2515245920960351

Byström, K. (2002). Information and information sources in tasks of varying complexity. *Journal of the American Society for Information Science and Technology, 53*(7), 581-591. https://doi.org/10.1002/asi.10064

Carl, M. (2021). *Explorations in Empirical Translation Process Research*. Springer.

Case, D. O. (2016). *Looking for Information A Survey of Research on Information Seeking, Needs, and Behavior* (4 ed.).

Champagne, G. (2004). *The economic value of terminology: An exploratory study*.

Chan, S. W. (2015). *The Routledge Encyclopedia of Translation technology*. Routledge.

Chang, C.-Y. (2009). *Testing Applicability of Eye-tracking and fMRI to Translation and Interpreting Studies: An Investigation into Directionality* University of London].

Chang, L. (2018). *A Longitudinal Study on the Formation of Chinese Students' Translation Competence: with a particular focus on metacognitive reflection and web searching* University College London].

Cheng, S. (2017). *Translation Competence Development among Learners: A Problem-Solving Perspective* The University of Manchester].

Choi, Y. (2013). Analysis of image search queries on the web: Query modification patterns and semantic attributes. *Journal of the American Society for Information Science and Technology, 64*(7), 1423-1441. https://doi.org/10.1002/asi.22831

Cui, Y., & Zheng, B. (2020). Consultation behaviour with online resources in English-Chinese translation: an eye-tracking, screen-recording and retrospective study. *Perspectives, 29*(5), 740-760. https://doi.org/10.1080/0907676x.2020.1760899

Daems, J. (2016). *A translation robot for each translator? A comparative study of manual translation and post-editing of machine translations: process, quality and translator*

attitude Ghent University].

Daems, J., Carl, M., Vandepitte, S., Hartsuiker, R., & Macken, L. (2016). The Effectiveness of Consulting External Resources During Translation and Post-editing of General Text Types. In *New Directions in Empirical Translation Process Research* (pp. 111-133). https://doi.org/10.1007/978-3-319-20358-4_6

Dahl, Ö. (2004). *The Growth andMaintenance of Linguistic Complexity*. John Benjamins Publishing Company.

Dam-Jensen, H., & Heine, C. (2009). Process research methods and their application in the didactics of text production and translation. *Trans-kom, 2*(1), 1-25.

Dragsted, B. (2005). Target. International Journal of Translation Studies. *Segmentation in translation: Differences across levels of expertise and difficulty, 17*(1), 45-70.

Ehrensberger-Dow, M. (2014). Challenges of translation process research at the workplace. *MonTI. Monografías de Traducción e Interpretación*, 355-383. https://doi.org/10.6035/MonTI.2014.ne1.12

Ehrensberger-Dow, M., & Perrin, D. (2009). Capturing translation processes to access metalinguistic awareness. *Across Languages and Cultures, 10*(2), 275-288. https://doi.org/10.1556/Acr.10.2009.2.6

EMT Board. (2017). *EMT Competence Framework 2017*. Retrieved 23 August from https://ec.europa.eu/info/sites/default/files/emt_competence_fwk_2017_en_web.pdf

EMT Expert Group. (2009). *Competences for professional translators, experts in multilingual and multimedia communication*. Retrieved 23 August from https://ec.europa.eu/info/sites/default/files/emt_competences_translators_en.pdf

Enríquez Raído, V. (2011). *Investigating the web search behaviors of translation studnets* [PhD thesis, Universitat Ramon Llull, Barcelona, Spain), https://www.tdx.cat/bitstream/handle/10803/21793/Enriquez_PhD_Thesis_Final.pdf

Enríquez Raído, V., Angelone, E., & and Olalla-Soler, C. (Forthcoming). Screen Recording.

In *Data Collection Methods in Cognitive Translation and Interpreting Studies*. John Benjamins.

Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. MIT Press.

Eyelink. (2009). *EyeLink 1000 User Manual 1.5.0*.

Fernández, O. (2015). *Exploratory Research into the Use of Web Resources of Students Enrolled in an Introductory University-level Medical Translation Course* Arizona State University].

Göpferich, S. (2009). Towards a model of translation competence and its acquisition: the longitudinal study TransComp. In S. Göpferich, L. J. Arnt, & I. M. Mees (Eds.), *Behind the mind: Methods, models and results in translation process research* (pp. 11-37). Samfundslitteratur.

Göpferich, S. (2011). From multidisciplinarity to transdisciplinarity: the investigation of competence development as a case in point. *MikaEL electronic proceedings of the KäTu Symposium on Translation and Interpreting Studies, 5*.

Gough, J. (2016). *The patterns of interaction between professional translators and online resources* [Unpublished Doctoral dissertation, University of Surrey), https://core.ac.uk/download/pdf/78911493.pdf

Gough, J. (2019). Developing translation-oriented research competence: what can we learn from professional translators? *The Interpreter and Translator Trainer, 13*(3), 342-359. https://doi.org/10.1080/1750399x.2019.1656404

Hansen, G. (1999). *Probing the Process in Translation: Methods and Results*. Samfundslitteratur.

Hansen, G. (2003). Controlling the Process. Theoretical and Methodological Reflections. In F. Alves (Ed.), *Triangulating Translation: Perspectives in process oriented research*. John Benjamins Publishing.

Hansen, G. (2006). Retrospection methods in translator training and translation research.

Journal of specialised Translation, 5(1), 2-41.

Hao, Y., & Pym, A. (2021). Translation skills required by Master's graduates for employment: Which are needed, which are not? *Across Languages and Cultures, 22*(2), 158-175. https://doi.org/10.1556/084.2021.00012

Hirci, N. (2013). Changing Trends in the Use of Translation Resources: The Case of Trainee Translators in Slovenia. *ELOPE: English Language Overseas Perspectives and Enquiries, 10*(2), 149-165. https://doi.org/10.4312/elope.10.2.149-165

Hurtado Albir, A. (2017). *Researching Translation Competence by PACTE Group*. John Benjamins Publishing Company.

Hvelplund, K. T. (2011). *Allocation of cognitive resources in translation an eye-tracking and key-logging study*. Frederiksberg: Copenhagen Business School.

Hvelplund, K. T. (2014). Eye tracking and the translation process: reflections on the analysis and interpretation of eye-tracking data. *MonTI. Monografías de Traducción e Interpretación*, 201-223. https://doi.org/10.6035/MonTI.2014.ne1.6

Hvelplund, K. T. (2017). Translators' Use of Digital Resources during Translation. *Journal of Language and Communication in Business*(56).

Hvelplund, K. T. (2017a). Translators' Use of Digital Resources during Translation. *Journal of Language and Communication in Business*(56).

Hvelplund, K. T. (2017b). Four fundamental types of reading during translation. In A. L. Jakobsen & B. Mesa-Lao (Eds.), *Translation in transition. Between cognition, computing and technology*. John Benjamins Publishing Company.

Hvelplund, K. T. (2019). Digital resources in the translation process – attention, cognitive effort and processing flow. *Perspectives, 27*(4), 510-524. https://doi.org/10.1080/0907676x.2019.1575883

Hvelplund, K. T., & Dragsted, B. (2018). Genre familiarity and translation processing Differences and similarities between literary and LSP translators. In I. Lacruz & R.

Jääskeläinen (Eds.), *Innovation and Expansion in Translation Process Research* (pp. 55-76). John Benjamins Publishing Company. https://doi.org/10.1075/ata.18.04tan

Jääskeläinen, R. (2011). Studying the Translation Process. In K. Malmkjær & K. Windle (Eds.), *The Oxford Handbook of Translation Studies*. Oxford. https://doi.org/10.1093/oxfordhb/9780199239306.013.0010

Jakobsen, A. L. (2014). The development and current state of translation process research. In E. Brems, R. Meylaerts, & L. v. Doorslaer (Eds.), *The Known Unknowns of Translation Studies*. John Benjamins Publishing Company.

Jakobsen, A. L. (2020). Translation technology research with eye tracking. In M. O'Hagan (Ed.), *The Routledge Handbook of Translation and Technology*. Routledge.

Jansen, B. J. (2000). The effect of query complexity on Web searching results.

Jensen, K. T. H. (2009). Indicators of Text Complexity. *Copenhagen Studies in Language, 37*, 61-80.

Kiraly, D. C. (2013). Towards a view of translator competence as an emergent phenomenon. In D. C. Kiraly, S. Hansen-Schirra, & K. Maksymski (Eds.), *New Prospects and Perspectives for Educating Language Mediators* (pp. 197-224). Gunter Narr Verlag.

Kurbanoğlu, S., Boustany, J., Špiranec, S., Grassian, E., Mizrachi, D., & Roy, L. (2018). *Information Literacy in the Workplace*.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). *lmerTest Package: Tests in Linear Mixed Effects Models.* In *Journal of Statistical Software* (Version 3.1-3) https://cran.r-project.org/web/packages/lmerTest/index.html

Kuznik, A. (2017). Use of Instrumental Resources. In A. Hurtado Albir (Ed.), *Researching Translation Competence by PACTE Group* (pp. 219-241). John Benjamins Publishing Company.

Kuznik, A., & Hurtado Albir, A. (2015). How to define good professional translators and interpreters: Applying the behavioural approach to studying competences in the field

of translation studies1. *Across Languages and Cultures, 16*(1), 1-27. https://doi.org/10.1556/084.2015.16.1.1

Kuznik, A., & Olalla-Soler, C. (2018). Results of pacte group's experimental research on translation competence acquisition. The acquisition of the instrumental sub-competence. *Across Languages and Cultures, 19*(1), 19-51. https://doi.org/10.1556/084.2018.19.1.2

Larson-Hall, J. (2016). *A Guide to Doing Statistics in Second Language Research Using SPSS and R*. Routledge Group.

Lau, J. (2006). *Guidelines on Information Literacy for Lifelong learning*. https://www.ifla.org/wp-content/uploads/2019/05/assets/information-literacy/publications/ifla-guidelines-en.pdf

Leijten, M., & Van Waes, L. (2013). Keystroke Logging in Writing Research. *Written Communication, 30*(3), 358-392. https://doi.org/10.1177/0741088313491692

Leijten, M., Van Waes, L., Schrijver, I., Bernolet, S., & Vangehuchten, L. (2019). Mapping Master's Students' Use of External Sources in Source-Based Writing in L1 and L2. *Studies in Second Language Acquisition, 41*(3), 555-582. https://doi.org/10.1017/s0272263119000251

Leijten, M., Waes, L. V., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research, 5*(3), 285-337.

Lew, R. (2013). Online dictionary skills. Proceedings of eLex,

Lewandowski, D. (2006). Query types and search topics of German Web search engine users. *Information Services & Use*(26), 261–269.

Li, C., & Wang, S. (2020). *An English-Chinese Translation Course in Beiwai*. China Translation and Publishing House.

Livbjerg, I., & Mees, I. M. (2002). Patterns of dictionary use in non-domain-specific translation.

In F. Alves (Ed.), *Triangulating Translation: Perspectives in process oriented research* (pp. 123-136). Benjamins.

Lörscher, W. (1992). Process-Oriented Research into Translation and Implications for Translation Teaching. *TTR : traduction, terminologie, rédaction, 5*(1). https://doi.org/10.7202/037110ar

Massey, G., & Ehrensberger-Dow, M. (2011a). Investigating information literacy: A growing priority in translation studies. *Across Languages and Cultures, 12*(2), 193-211. https://doi.org/10.1556/Acr.12.2011.2.4

Massey, G., & Ehrensberger-Dow, M. (2011b). Technical and Instrumental Competence in the Translator's Workplace: Using Process Research to Identify Educational and Ergonomic Needs. *ILCEA Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*.

Massey, G., Riediger, H., & Lenz, S. (2007). Teaching instrumental competence in an e-learning environment: A Swiss perspective. eCoLoTrain Workshop, Lubljana, Slovenia.

McCulloch, C. E., R.Searle, S., & Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models* (2 ed.). John Wiley & Sons.

Mellinger, C. D. (2020). Methodological considerations for survey research Validity, reliability, and quantitative analysis. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19*, 172–190.

Mellinger, C. D., & Hanson, T. A. (2017). *Quantitative Research Methods in Translation and Interpreting Studies*. Routledge.

Muñoz Martín, R. (2010). On paradigms and cognitive translatology. In G. M. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 169-187). Benjamins.

Muñoz Martín, R., & Apfelthaler, M. (2022). A task segment framework to study keylogged translation processes. *The International Journal of Translation and Interpreting Research, 14*(2), 8-31. https://doi.org/10.12807/ti.114202.2022.a02

Mutta, M., Pelttari, S., SalM, l., Chevalier, a., & johanSSon, M. (2014). Digital literacy in academic language learning Contexts: Developing information-Seeking Competence. In *Digital Literacies in Foreign and Second Language Education*. CALICO.

Newton, R. R., & Rudestam, K. E. (1999). *Your Statistical Consultant : Answers to Your Data Analysis Questions*. Sage Publications.

Nitzke, J. (2019). *Problem solving activities in post-editing and translation from scratch*.

Nitzke, J., & Hansen-Schirra, S. (2021). *A short guide to post-editing*. Language Science Press.

Nunes Vieira, L., Ragni, V., & Alonso, E. (2021). Translator autonomy in the age of behavioural data. *Translation, Cognition & Behavior*. https://doi.org/10.1075/tcb.00052.nun

O'Brien, S., & Rodríguez Vázquez, S. (2020). Translation and technology. In S. Laviosa & M. González-Davies (Eds.), *The Routledge Handbook of Translation and Education*. Routledge Press.

Olalla-Soler, C. (2018). Using electronic information resources to solve cultural translation problems. *Journal of Documentation, 74*(6), 1293-1317. https://doi.org/10.1108/jd-02-2018-0033

Onishi, N., & Yamada, M. (2020). Why translator competence in information searching matters: An empirical investigation into differences in searching behavior between professionals and novice translators. *Invitation to Interpreting and Translation Studies*(22), 1-23.

Osborne, J. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research, and Evaluation, 15*(1), 1-12. https://doi.org/10.7275/qbpc-gk17

PACTE Group. (2011). Results of the validation of the PACTE translation competence model. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 317-343). John Benjamins Publishing.

Pakkala-Weckström, M. (2015). Student's Data Mining Skills in Second-Year Undergraduate Translation. *Current Trends in Translation Teaching and Learning E*(2), 139-170.

Paradowska, U. (2020). *Developing information competence in translator training – a mixed-method longitudinal study*

Pastor, V., & Alcina, A. (2010). Search Techniques in Electronic Dictionaries: A Classification for Translators. *International Journal of Lexicography, 23*(3), 307-354. https://doi.org/10.1093/ijl/ecq015

Pastor, V., & Alcina, A. (2013). Researching the use of electronic dictionaries. In H. Jackson (Ed.), *The Bloomsbury Companion To Lexicography*.

Pinto, M., Fernandez-Pascual, R., & Puertas, S. (2016). Undergraduates' information literacy competency: A pilot study of assessment tools based on a latent trait model. *Library & Information Science Research, 38*(2), 180-189. https://doi.org/10.1016/j.lisr.2016.05.004

Pinto, M., & Sales, D. (2007). A research case study for user-centred information literacy instruction: information behaviour of translation trainees. *Journal of Information Science, 33*(5), 531-550. https://doi.org/10.1177/0165551506076404

Pinto, M., & Sales, D. (2008a). Knowledge transfer and information skills for student-centered learning: Some academic experiencies in Spain. *Libraries and the Academy, 8*(1).

Pinto, M., & Sales, D. (2008b). INFOLITRANS: A model for the development of information competence for translators. *Journal of Documentation, 64*(3), 413-437.

Pinto, M., & Sales, D. (2008c). Towards User-centred Information Literacy Instruction in Translation. *The Interpreter and Translator Trainer, 2*(1), 47-74. https://doi.org/10.1080/1750399x.2008.10798766

Pokorn, N., Blake, J., Reindl, D., & Pisanski Peterlin, A. (2020). The influence of directionality on the quality of translation output in educational settings. *The Interpreter and Translator Trainer, 14*(1), 58-78. https://doi.org/10.1080/1750399x.2019.1594563

Quinci, C. (2013). *Translators in the Making: An Empirical Longitudinal Study on Translation Competence and its Development* University of Trieste].

Read, J. (2000). *Assessing Vocabulary*. Cambridge University Press.

RStudio Team. (2020). *RStudio: Integrated Development for R. RStudio, PBC, Boston.* In (Version 1.4.1106) http://www.rstudio.com/

Saastamoinen, M., & Järvelin, K. (2017). Relationships between work task types, complexity and dwell time of information resources. *Journal of Information Science, 44*(2), 265-284. https://doi.org/10.1177/0165551516687726

Saldanha, G., & O'Brien, S. (2014). *Research Methodologies in Translation Studies*. Routledge.

Sales, D. (2022). Threading metaliteracy into Translation and Interpreting undergraduates information literacy training: a reflective active learning approach. *Anales de Documentación, 5*. https://doi.org/10.6018/analesdoc.504691

Sales, D., & Pinto, M. (2011). The professional translator and information literacy: Perceptions and needs. *Journal of Librarianship and Information Science, 43*(4), 246-260.

Sanches, T., Lopes, C., & Antunes, M. L. (2022). *Critical Thinking in Information Literacy Pedagogical Strategies: new dynamics for Higher Education throughout librarians' vision* 8th International Conference on Higher Education Advances (HEAd'22),

Sarkar, D. (2008). *Lattice Multivariate Data Visualization with R.* In (Version 0.20-44) Springer. https://cran.r-project.org/web/packages/lattice/index.html

Schmitz, K.-D., & Gornostay, T. (2014). Terminology as a Service – a model for collaborative terminology management. EAFT Terminology Summit,

Shih, C. Y. (2017). Web search for translation: an exploratory study on six Chinese trainee translators' behaviour. *Asia Pacific Translation and Intercultural Studies, 4*(1), 50-66. https://doi.org/10.1080/23306343.2017.1284641

Shih, C. Y. (2019). A quest for web search optimisation: an evidence-based approach to trainee translators' behaviour. *Perspectives, 27*(6), 908-923.

https://doi.org/10.1080/0907676x.2019.1579847

Shih, C. Y. (2021). Navigating the Web A Study on Professional Translators' Behaviour. In *Empirical Studies of Translation and Interpreting* (pp. 74-92). Routledge.

Shreve, G. M., & Angelone, E. (2010). *Translation and Cognition*. John Benjamins Publishing Company.

Spinner, P., Gass, S. M., & Behney, J. (2013). Ecological Validity in Eye-Tracking. *Studies in Second Language Acquisition, 35*(2), 389-415. https://doi.org/10.1017/s0272263112000927

Sycz-Opoń, J. (2019). Information-seeking behaviour of translation students at the University of Silesia during legal translation – an empirical investigation. *The Interpreter and Translator Trainer, 13*(2), 152-176. https://doi.org/10.1080/1750399x.2019.1565076

Sycz-Opoń, J. E. (2021). Trainee translators' research styles: A taxonomy based on an observation study at the University of Silesia, Poland. *The International Journal of Translation and Interpreting Research, 13*(2). https://doi.org/10.12807/ti.113202.2021.a08

[Record #963 is using a reference type undefined in this output style.]

Taylor, R. S. (2015). Question-Negotiation and Information Seeking in Libraries. *College & Research Libraries, 76*(3), 251-267. https://doi.org/10.5860/crl.76.3.251

Teixeira, C. S. C., & O'Brien, S. (2017). Investigating the cognitive ergonomic aspects of translation tools in a workplace setting. *Translation Spaces, 6*(1), 79-103. https://doi.org/10.1075/ts.6.1.05tei

Tirkkonen-Condit, S., & Jääskeläinen, R. (2000). *Tapping and Mapping the Processes of Translation and Interpreting*. John Benjamins Publishing Company.

Tukey, J. W. (1977). *Exploratory data analysis*.

Vakkari, P. (1999). Task complexity, problem structure and information actions Integrating studies on information seeking and retrieval. *Information Processing and Management*.

Varantola, K. (1998). Translators and their Use of Dictionaries: User Needs and User Habits. In B. T. S. Atkins (Ed.), *Using Dictionaries Studies of Dictionary Use by Language Learners and Translators*. Max Niemeyer Verlag.

Varian, H. (2006). The economics of internet search. *Rivista di politica economica 96, 12*(11).

Venables, W. N., & Ripley, B. D. (2004). *Modern applied statistics with S.* In (Version 7.3-54) Springer.

Wang, V. X., & Lim, L. (2017). How do translators use web resources Evidence from the performance of English–Chinese translators. In D. Kenny (Ed.), *Human Issues in Translation Technology*. Routledge.

Wang, Y. (2014). *An Empirical Study of Chinese Translators' Search Behavior In Chinese-To-English Translation*

Webber, S., & Johnston, B. (2014). Transforming information literacy for higher education in the 21st century: A lifelong learning approach. In M. Hepworth & G. Walton (Eds.), *Developing people's information capabilities: Fostering information literacy in educational, workplace and community contexts*. Emerald Group Publishing Limited.

White, M. D., Matteson, M., & Abels, E. G. (2008). Beyond dictionaries Understanding information behavior of professional translators. *Journal of Documentation, 64*(4), 576-601. https://doi.org/10.1108/00220410810884084

White, R. W. (2016). *Interactions with Search Systems*. https://doi.org/10.1017/cbo9781139525305

Whyatt, B., Witczak, O., & Tomczak, E. (2021). Information behaviour in bidirectional translators: focus on online resources. *The Interpreter and Translator Trainer*, 1-18. https://doi.org/10.1080/1750399x.2020.1856023

Wickham, H. (2021). *dplyr: A Grammar of Data Manipulation.* In (Version R package version 1.0.7) https://CRAN.R-project.org/package=dplyr

Witczak, O. (2021). *Information searching in the post-editing and translation process*

[Unpublished doctoral dissertation, Adam Mickiewicz University),

Xiao, K., & Muñoz, R. M. (2020). Cognitive Translation Studies: Models and methods at the cutting edge. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*(19).

Yu, J., & Que, Y. (2019). Fostering Comprehensive Competence in Translation-related Search and Verification. *Chinese Translators Journal*(2), 79-85.

Appendices

Appendix A Source text for the first pilot study

2015 年乐视广告交易平台框架协议

如有甲方广告链接地址被计算机病毒感染的情况发生，乙方有权先暂停该广告发布，同时通知甲方进行杀毒，待甲方对服务器进行杀毒且经乙方确认广告链接安全后方可恢复广告发布。在此期间出现的广告发布之暂停，不应视为乙方违约，该广告暂停损失由甲方自行承担。乙方不予以补充发布，甲方仍需按照广告合同向乙方支付全额广告费用。

乙方负责网络广告内部发布工作，包括：流量规则设置、广告位设置、广告播放数据监测。

Reference

2015 LeTV Advertising Trading Platform Framework Agreement

If Party A’s advertising link address is infected by computer virus, Party B has the right to suspend the advertisement and notify Party A to perform anti-virus. After Party A disinfects the server and Party B confirms the advertisement link, the advertisement can be resumed. The suspension of advertising during this period shall not be regarded as a breach of contract by Party B, and Party A shall bear the loss of the suspension of the advertisement. Party B shall not supplement it and Party A shall still pay Party B the full amount of advertising fees in accordance with the advertising contract.

Party B is responsible for the internal release of online advertising, including: traffic rule settings, ad slot settings, and ad play data monitoring.

Appendix B Source text for the second pilot study

Translation brief: Please translate the following text from English into Chinese for publication purpose

The materialization of several key downside risks could prolong the period of weak global growth

Global economic prospects remain subject to significant uncertainties and risks that are weighted on the downside, with the potential to obstruct the modest acceleration in growth that is currently forecast for 2017-2018. Some of these risks stem from monetary policy actions in major developed economies. The impact of introducing untested monetary policy instruments — such as the negative interest rate policies in Japan and Europe — remains unclear. There is a risk that such measures could lead to a deterioration of bank balance sheets, causing credit conditions to tighten, with the potential to destabilize fragile and undercapitalized banks. The timing of interest rate rises in the United States is another area of uncertainty. As interest rate differentials relative to other developed economies widen, this has the potential to trigger financial volatility, reversal of capital inflows to developing economies, and abrupt adjustments in exchange rates. Such volatility would exacerbate vulnerabilities associated with high levels of debt and rising default rates in a number of developing countries, with the potential to push up borrowing costs, raise deleveraging pressures, and increase banking sector stress. (194 words)

Appendix C Questionnaire in the second pilot study

This questionnaire will take you approximately 6 minutes to complete. The first section contains questions about personal information and the second section includes questions for information online. No right or wrong answers in this questionnaire, and your answers will be kept confidential.

Section 1

1. Full Name:

2. Please select the test(s) you have taken from the list below.

A. TEM-8 (Test for English Major-8)

B. IELTS (International English Language Testing System)

C. TOEFL (Test of English as a Foreign Language)

D. Other, please specify:

3. What is the highest score you have achieved in TEM 8? (If the participants choose A in Question 2)

4. What is the highest score you have achieved in IELTS? (If the participants choose B in Question 2)

5. What is the highest score you have achieved in TOEFL? (If the participants choose C in Question 2)

6. Have you passed any of the following test(s)? (Please tick all the boxes that apply)

A. CATTI-I in written translation[30]

B. CATTI-II in written translation

C. CATTI-III in written translation

D. None of the above

7. Have you ever worked as a full-time professional translator?

A. Yes

B. No

8. How many years have you worked as a full-time professional translator? (If the participants choose A in Question 7)

A. Less than 1 year

B. 1-3 years

C. More than 3 years

Section 2:

[30] CATTI is the abbreviation for China Accreditation Test for Translators and Interpreters.

9. How often do you use the following resources to look for translation information online?

	Never	Rarely	Sometimes	Often	Very Often
Search Engines (e.g. Google search)	○	○	○	○	○
Web Pages (e.g., parallel texts)	○	○	○	○	○
Dictionaries (e.g., monolingual, bilingual, or specialized dictionaries)	○	○	○	○	○
Machine Translation (e.g., Google/Youdao Translate)	○	○	○	○	○
Concordancers (e.g. Linguee)	○	○	○	○	○
Encyclopedias (e.g. Wikipedia)	○	○	○	○	○
Document	○	○	○	○	○
Discussion Forums (e.g. blog or Zhihu)	○	○	○	○	○

10. Are you familiar with any search engine operators (e.g., the operator “OR” to find one or more words in a search query) used to find an exact phrase (i.e., containing all the words in a search query)? Pease list a maximum of 5 and describe their function, i.e. what they are used for. If you do not know any search engine operators, you can skip this question

Search engine operator 1

Search engine operator 2

Search engine operator 3

Search engine operator 4

Search engine operator 5

11. Please select your perceived sense of knowledge to each of the questions below:

	not at all knowledgeable	slightly knowledgeable	moderately knowledgeable	very knowledgeable	extremely knowledgeable
what information you can find on the web:	○	○	○	○	○
where you can find the information you need:	○	○	○	○	○
when to change search strategy, including stopping searching:	○	○	○	○	○
how to evaluate the information you find:	○	○	○	○	○

Thanks for completing this questionnaire!

Appendix D Participant Information Sheet

Faculty of Arts

18 Symonds St

The University of Auckland

Private Bag 92019

Auckland, New Zealand

Participant Information Sheet

Title: Translator Trainee's Online Search Behaviours: From Current Practice to Desirable Practice

Principal Investigator (PI): Dr Vanessa Enríquez Raído

Researcher introduction

Yuxing Cai, a PhD candidate in the Faculty of Arts at the University of Auckland, is conducting this research project. The supervisors of this project are Dr Vanessa Enríquez Raído and Professor Changshuan Li.

Project description and invitation

This project aims to explore the development of students' use of information sources online. It involves students like you, who will complete the non-literary translation course in the Graduate School of Translation and Interpretation at the Beijing Foreign Studies University. There is no exclusion criterion, i.e. all the students who will take this course are potential participants. Professor Bin Yao, the Executive Dean of the Graduate School of Translation and Interpretation at the Beijing Foreign Studies University, has guaranteed that participation or non-participation is voluntary, and that your participation or non-participation in this project has no effect on your study, grades or relationship with the University.

Project Procedures

1. Prepare

- Fill out the questionnaire (https://www.wjx.cn/jq/90964461.aspx)
- Watch the video tutorial with information on how to install the software applications and how to proceed with your translation.
- Install Inputlog (Password: IL7558)
- Install Screencast (URL: https://screencast-o-matic.com/translation) and log into it: (username: ycai699@aucklanduni.ac.nz; password: translation)

2. Translate

Start translating the source text attached.

3. Describe (in Chinese)

For the whole text:

- Verbally rate your familiarity with the topic of this text. (1 not familiar at all, 5 very familiar)
- Verbally rate how difficult of translating the text. (1 not difficult at all, 5

very difficult)

For each individual search:

- Highlight the items in the source text that you searched for during translation.
- Verbally describe the type of information you searched for to translate each of the highlighted items (e.g., a source-term definition, a cultural aspect, an acronym, a specific collocation, a fluent/ idiomatic expression in Chinese, an equivalent term, etc.).
- Verbally rate how difficult did you find this search (1 not difficult at all, 5 very difficult).

4. Upload recording files

After you complete the task, please upload the Screencast file, and email the Inputlog file to me (ycai699@aucklanduni.ac.nz).

If you choose to participate in this project, you will receive a gift of 400 China Yuan (100 NZ dollars) to thank you for your time. You will also benefit from the results of this project from a translation and information search learning perspective by choosing to receive a summary of the results obtained.

Data storage/retention/destruction/future use

The data will be stored in a University of Auckland computer for 6 years and will be accessed and analysed only by me. After 6 years, all the data will be destroyed securely and permanently.

Right to Withdraw from Participation

You have the right to withdraw from participating in the research project at any time without giving a reason. You also have the right to withdraw your data from the research up to a period of 2 weeks after your initial participation by emailing me.

Anonymity and Confidentiality

Confidentiality is guaranteed. All the data collected will only be analysed by me. The information you share with me will remain confidential to us. Anonymity, however, cannot be guaranteed, as I will obtain some personal information from you, e.g., educational background and information on English scores. However, in my research I will assign pseudonyms to all participants. If the information you provide is reported or published, this will be done in a way that does not identify you as its source.

A summary of results will be e-mailed to you if you wish to learn about the findings of the project.

Thank you very much for your time and consideration.

Yuxing Cai

PhD candidate in Translation Studies

Faculty of Arts

The University of Auckland

ycai699@aucklanduni.ac.nz

Private Bag 92019

Auckland 1142

Contact Details

Head of Department

Bernadette Luciano

School of Cultures, Languages, Languages, and Linguistics

Faculty of Arts

The University of Auckland

b.luciano@auckland.ac.nz

+64 9 923 7917

Principle Investigator

Vanessa Enríquez Raído

School of Cultures, Languages, and Linguistics

Faculty of Arts

The University of Auckland

v.enriquez@auckland.ac.nz

+64 9 923 7108

Co-Investigator

Changshuan Li

Graduate School of Translation and Interpretation

Beijing Foreign Studies University

lichangshuan@gmail.com

+86 13641197147

For any concerns regarding ethical issues you may contact the Chair of the University of Auckland Human Participants Ethics Committee at the University of Auckland Research Office, Private Bag 92019, Auckland 1142. Telephone 09 373-7599 ext. 83711.

Email: ro-ethics@auckland.ac.nz

Approved by the University of Auckland Human Participants Ethics Committee on 11-Oct-2018 for three years. Reference Number 022125.

Appendix E Consent Form

Faculty of Arts

18 Symonds St

The University of Auckland

Private Bag 92019

Auckland, New Zealand

CONSENT FORM

THIS FORM WILL BE HELD FOR A PERIOD OF 6 YEARS

Project title: Translator Trainee's Online Search Behaviours: From Current Practice to Desirable Practice

Name of Investigator/Supervisor: Dr Vanessa Enríquez Raído

Name of Student Researcher: Yuxing Cai

I have read the Participant Information Sheet, have understood the nature of the research and why I have been selected. I have had the opportunity to ask questions and have had them answered to my satisfaction.

- I agree to take part in this research.

- I understand that I am free to withdraw from participating in the research at any time without giving a reason, and withdraw my data from the research up to a period of 2 weeks after participation.

- I agree to be screen-and-keystroke recorded.

- I wish to receive the summary of findings.

  I do not wish to receive the summary of findings.

  Name:_____ Email:

  Signature:_____ Date:_____d

Approved by the University of Auckland Human Participants Ethics Committee on11-Oct-2018 for three years. Reference Number 022125.

Appendix F Instruction Email

操作步骤

亲爱的同学，

非常感谢你愿意参加我的项目。

这封邮件包括实验所需要的材料及操作步骤。附件里是软件安装程序、需要翻译的文本、同意书和知情书。请在同意书上签名，如果不方便签名，可以直接邮件回复我愿意参加。首先需要下载附件里的文件，然后按照以下步骤进行操作：

1. 准备

• 填写问卷，链接为：https://www.wjx.cn/jq/90964461.aspx

• 观看视频教程，了解操作过程，链接：https://www.bilibili.com/video/BV13541187Ef/

• 安装键盘记录软件 Inputlog（安装密码：IL7558）

• 安装并登录屏幕录制软件 Screencast（URL: https://screencast-o-matic.com/translation，用户名：ycai699@aucklanduni.ac.nz，密码：translation）

2. 翻译

• 翻译附件里的文章（文件名为 Source Text）

3. 口述

对于整篇文章：

a. 你对这篇文章的主题是否熟悉（1 是一点也不熟悉，5 是非常熟悉）

b. 你觉得翻译这篇文章的难度如何（1 是非常简单，5 是非常难）

对于每一个单独的查证：

a. 在原文（英文）文本中高亮**所有**进行了查证的部分

b. 口述查证的类型（例如中文的习惯表达、英文的术语定义、词语搭配、缩略语等）

c. 对查证难度进行打分并解释原因（1 是非常简单，5 是非常难）

4. 上传及发送邮件

上传 Screencast 文件，并将 Inputlog 文件发送至 ycai699@aucklanduni.ac.nz 。

如果实验遇到任何问题，请随时与我联系（邮箱 ycai699@aucklanduni.ac.nz；微信 15600581477）。

期待你的回复！

非常感谢

蔡宇星

Appendix G Questionnaire in the main study

This questionnaire will take you approximately 6 minutes to complete. The first section contains questions about personal information and the second section includes questions for information online. No right or wrong answers in this questionnaire, and your answers will be kept confidential.

Section 1

1. Full Name:

2. Please select the test(s) you have taken from the list below.

A. TEM-8 (Test for English Major-8)

B. IELTS (International English Language Testing System)

C. TOEFL (Test of English as a Foreign Language)

D. Other, please specify:

3. What is the highest score you have achieved in TEM 8? (If the participants choose A in Question 2)

4. What is the highest score you have achieved in IELTS? (If the participants choose B in Question 2)

5. What is the highest score you have achieved in TOEFL? (If the participants choose C in Question 2)

6. Have you passed any of the following test(s)? (Please tick all the boxes that apply)

A. CATTI-I in written translation[31]

B. CATTI-II in written translation

C. CATTI-III in written translation

D. None of the above

7. Have you ever worked as a full-time professional translator?

A. Yes

B. No

8. How many years have you worked as a full-time professional translator? (If the participants choose A in Question 7)

A. Less than 1 year

B. 1-3 years

C. More than 3 years

Section 2:

[31] CATTI is the abbreviation for China Accreditation Test for Translators and Interpreters.

9. How often do you use the following resources to look for translation information online?

	Never	Rarely	Sometimes	Often	Very Often
Search Engines (e.g. Google search)	○	○	○	○	○
Web Pages (e.g., parallel texts)	○	○	○	○	○
Dictionaries (e.g., monolingual, bilingual, or specialized dictionaries)	○	○	○	○	○
Machine Translation (e.g., Google/Youdao Translate)	○	○	○	○	○
Concordancers (e.g. Linguee)	○	○	○	○	○
Encyclopedias (e.g. Wikipedia)	○	○	○	○	○
Document	○	○	○	○	○
Discussion Forums (e.g. blog or Zhihu)	○	○	○	○	○

10. Please select your perceived sense of knowledge to each of the questions below:

	not at all knowledgeable	slightly knowledgeable	moderately knowledgeable	very knowledgeable	extremely knowledgeable
what information you can find on the web:	○	○	○	○	○
where you can find the information you need:	○	○	○	○	○
when to change search strategy, including stopping searching:	○	○	○	○	○
how to evaluate the information you find:	○	○	○	○	○

Thanks for completing this questionnaire!

Appendix H Four source texts for the main study

Source Text 1

Please translate the following book review from English into Chinese for publication purpose.

Book name: The Next Great Migration: The Beauty and Terror of Life on the Move

Shah convincingly argues that politicians against immigration distort and misuse data to create unnecessary and cruel barriers. She tells gut-wrenching stories of struggling families on the move, and presents evidence that migrants are generally healthier and less apt to commit violent crimes than are the residents of the country they move to. And studies show that immigrants benefit host economies.

Turning to plants and animals, she takes biologists to task for abusing data to make introduced species look worse than they are. She charges that renowned ecologist Charles Elton "cherry-picked" case studies of the most disruptive introduced species (such as the sea lamprey (Petromyzon marinus), which originated in the Atlantic Ocean and devastated trout populations in the Great Lakes). She levels that later biologists exaggerated the economic impact of non-natives by including the costs of removing them in calculations. Shah concludes that invasion biologists' predictions of "ecological Armageddon" have failed to transpire.

As a writer on ecology and conservation (for example, my 2011 book *Rambunctious Garden*), my expertise lies with the non-human migrations that Shah covers. I found it odd that she tends to group together many kinds of biological movement that specialists think of as quite different.

Source Text 2

Please translate the following book review from English into Chinese for publication purpose.

Book name: Biodiversity, Sustainability and Human Communities: Protecting Beyond the Protected

The book introduces concepts that promise better, more equitable resource allocation and

governance — property rights, free choice and markets, access to justice and equitable local control — but does not build these into a guiding framework. There is an emerging proprietorship–price–subsidiarity hypothesis, which says that if rights to natural resources are clearly delineated, if the value of wild commodities exceeds other land uses, and if people closest to the land capture most of this value, then renewable resources are likely to be husbanded better.This hypothesis underpins the Namibian case study (by Markus Nuding), the book's only example of entrusting landholders with wild resources. The strategy shows promising results and presents a strong moral justification for landholder control. Passing resource control to millions of rural people at the periphery of power requires new systems of governance that maximize value while curbing overuse. Prohibitions, policemen and fines are increasingly unacceptable. Making the necessary shifts in power relations is the greatest challenge, however. Political and institutional economics are important tools for understanding this transition, and conservationists need to arm themselves with these. As our demands on the planet increase, our ability to govern and add value to biodiversity will determine how much is conserved and for how long, as well as its contribution to livelihoods.

Source Text 3

Please translate the following book review from English into Chinese for publication purpose.

Book name: The Evolution of Developmental Pathways

Plato’s Republic contains an allegory of people chained in a cave, who see only the flickering shadows of reality on the wall, until some are able to break away and see the underlying truth. This is similar to Wilkins' message on the role of developmental pathways in the evolution of development. Wikins takes on the paradox that morphologically disparate creatures use conserved regulatory genes, and asks how different developmental processes can have evolved using much the same molecular machinery. He argues that the regulatory gene pathways are the reality under the morphology that we see and want to explain. His perspective is powerful and compelling. Although it may be too early to decide what constitutes the fundamental reality in the interactions between evolution and development, the book provides the framework for a coherent research programme that should be highly effective. The book contains some familiar

topics to set the evo-devo perspective. These include a historical summary, a bit about the fossil record and a discussion of phylogeny. It then progresses rapidly to the core issues by using a few concrete examples of changes in gene function in evolution, introduces the reader to developmental pathways and their functions, and then moves on to conserved genes and their evolutionary roles.

Source Text 4

Please translate the following book review from English into Chinese for publication purpose.

Book name: The Empire of Depression: A New History

Sadowsky points out that era after era grapples with false choices — between a political understanding of depression and a medical one, or between physical and psychological understandings. He is right to call for a move beyond these crude binaries. In my view, to do so, we must face the central roles of racism, sexism and ableism in delineating diagnostic boundaries over the years — not dismiss them as unfortunate. For example, psychiatry has a history of labelling some people as too uncivilized to be mentally ill, yet also diagnosing anticolonialism as mental illness.

In fact, many forms of resistance have been deemed symptoms of mental illness, from enslaved Africans fleeing brutality in the nineteenth-century United States to the Black Power movement of the 1960s. It is alarming, then, that apart from using “empire” as an analogy for the global dominance of Western psychiatry in interpreting distress, Sadowsky devotes little attention here to power and politics — especially given his previous work on colonialism (in the 1999 book *Imperial Bedlam*).

The book ends with the wise injunction: “listen to patients”. Yet, apart from illness memoirs, the voices (and research) of people who experience depression, those who become patients, those who refuse to become patients, and service users or psychiatric survivors are almost completely absent from the book.

Appendix I MS Excel Code for calculating time spent on each type of resources

```
Sub hz()
Dim i&, m&, arr, x, rng As Range
Dim d, k, t, Sht As Worksheet
l2 = Sheets.Count
 If Cells(1, l2 + 1) = "123" Then MsgBox "123456": Exit Sub
 [a1].CurrentRegion.ClearContents
Set d = CreateObject("Scripting.Dictionary")
ReDim brr(1 To 65000, 1 To l2)
l = 1: n = 1
brr(1, l) = "need"
For Each Sht In Sheets
    If Sht.Name <> "hz" Then
     m = Sht.[e65536].End(3).Row
      For x = 333333 To m Step 2
        If rng Is Nothing Then Set rng = Sht.Cells(x, 1) Else Set rng = Union(rng, Sht.Cells(x, 1))
      Next
      If Not rng Is Nothing Then rng.EntireRow.Delete Shift:=xlUp
```

```
        Set rng = Nothing
        l = l + 1
        brr(1, l) = Sht.Name
        arr = Sht.Range("e2:f" & m)
        For i = 1 To UBound(arr)
            ss = arr(i, 1)
            If Not d.Exists(ss) Then
                n = n + 1
                d.Add ss, n
                brr(n, 1) = arr(i, 1): brr(n, l) = arr(i, 2)
            Else
                brr(d(ss), l) = brr(d(ss), l) + arr(i, 2)
            End If
        Next
    End If
Next
[a1].Resize(n, l2) = brr
Cells(1, l2 + 1) = "123"
Set d = Nothing
End Sub
```

Appendix J Knowledge-based resources

Table 37 Specific knowledge-based resources by task (time in minutes:seconds)

Specific resources	Task 1	Task 2	Task 3	Task 4
Web Pages	58:28	38:11	22:43	36:02
Wikipedia	15:04	07:09	25:21	05:55
Linguee	05:07	16:55	17:38	06:04
Baidu Baike	13:53	02:08	13:18	06:11
Academic	02:53	20:27	07:10	01:59
PDF File	01:43	16:51	05:10	00:08
Douban	04:33	00:28	01:01	04:46
Google Books	00:00	18:53	04:06	14:46
CNKI	00:00	00:21	04:21	01:53
Google Scholar	00:00	02:38	01:06	00:12
Zhihu	00:00	00:00	05:23	02:52
Baidu Zhidao	00:09	00:00	01:58	00:48
BNC	00:00	02:33	00:04	00:00
Baidu Scholar	00:00	00:12	00:06	00:00
Sci-Hub	00:00	07:44	00:00	00:00
Questia	00:00	03:41	00:00	00:00
COCA	00:00	01:03	00:00	00:00
Sogou Baike	00:00	00:00	00:30	00:00
English-Corpora	00:00	00:12	00:00	00:00