Doctoral Dissertations in Chinese Interpreting Studies:

A Scientometric Survey Using Topic Modeling[1]

XU, Ziyun
Intercultural Studies Group
Universitat Rovira i Virgili

Abstract/Résumé
Tout en étant une discipline relativement nouvelle, la recherche sur l'interprétation en langue chinoise (RIC) a observé, depuis une vingtaine d'années, une croissance considérable du nombre de publications et de la diversité des thèmes étudiés. Le nombre de thèses de doctorat réalisées a également augmenté rapidement depuis la fin des années 1990. Puisque la RIC continue d'évoluer, il est important d'évaluer ses thèmes, tendances et institutions dominants de même que le développement professionnel des titulaires d'un doctorat dans ce sujet. Outre les techniques scientométriques traditionnelles, l'objectivité empirique de cette étude est renforcée par l'utilisation de la modélisation probabiliste par thèmes (PTM pour Probabilistic Topic Modeling), qui utilise l'allocation de Dirichlet latente (LDA pour Latent Dirichlet Allocation) pour analyser les thèmes abordés dans un corpus quasi exhaustif des thèses de la RIC. L'analyse révèle que les thèmes de la répartition des ressources cognitives, la déverbalisation et la modélisation du processus d'interprétation ont attiré beaucoup d'attention chez les doctorants. Des analyses supplémentaires ont été utilisées pour suivre la productivité de la recherche des institutions et les trajectoires professionnelles des titulaires de doctorat : une école s'est démarquée, représentant plus de la moitié du total des thèses produites et un doctorat en RIC s'est révélé être un atout précieux pour les nouveaux interprètes professionnels.

Keywords/Mots-Clés
Machine learning, scientometrics, Chinese interpreting studies, doctoral dissertations, topic modeling

I. Introduction

Doctoral dissertations are an important indicator of high-level research activities (Gile, 2013). They represent the final destination of an individual's journey through the world of formal education and, as well as demonstrating his or her mastery of the

existing literature in a given field, represent original contributions to that sum of knowledge (Kushkowski, Parsons, & Wiese, 2003). As a result they constitute an important component in the knowledge-creation process of any given discipline, and should be studied when assessing the evolution of that particular field. The object of this paper is to examine the areas of Chinese Interpreting Studies[2] (CIS) that its PhD students opt to study and the contributions they make to the field as a whole, as well as to ascertain which advisors and universities produce the most dissertations.

Given the academic value of doctoral dissertations in Translation and Interpreting Studies (TIS), a number of scholars have set out to study how they evolve and change over time. In 2013 Daniel Gile carried out a case study of major contributors to research in conference interpreting. His study revealed that over the span of four decades, L'École Supérieure d'Interprètes et de Traducteurs (ESIT), a leading TIS school, produced 11 dissertations, Spain 15, Italy one and China seven[3]. Despite ESIT's dominant position in generating the majority of TIS dissertations in the world in the 1970s, it accounted for only 2% of the total between 2000 and 2009.–Meanwhile, doctoral research productivity in Spain and China is on the rise. Strong leadership at major universities and the requirement for faculty members to have PhDs in order to obtain tenure may be the driving force in Spain. Mu, Zou and their doctoral students (2014) took a close look at the situation in China, examining 686 doctoral dissertations produced in TIS from 1992 to 2013, of which 39 dealt specifically with Interpreting Studies (IS). They found that most spotlighted interpreting strategies and cognition, and that experiments were the primary research method. They also observed that the topics themselves were focused and well-suited to the authors' capacities, but that few of the dissertations had solid theoretical underpinnings.

While Mu and Zou's study was pioneering in its analysis of Chinese doctoral dissertations, there were flaws in the way they classified research topics which resulted in a certain amount of overlapping. For example, it is difficult to understand why 'working memory, pedagogy, interpreting competence, interpreter's roles, and

---

[2] In this particular paper, CIS refers to research on Interpreting Studies with a specific focus on Chinese; it may be written in either Chinese or English. The focus of the present study is not exclusively on doctoral dissertations completed in China: any paper dealing with Chinese/English interpreting falls within its scope.
[3] The case study covered one university and three countries.

interpreting theories' were gathered together into one category, 'thinking process for interpreting, self-correction, and interpreting quality assessment' into another, and 'note-taking, norms, anxiety, communication, information processing and decision-making' a third. Elements from each ('working memory', 'thinking process', and 'information processing') might more profitably have been combined to form a category of their own ('cognitive issues'). To address these issues, the present author sought to produce the most comprehensive and rational system of categorization possible for the task at hand.

Also unlike Mu & Zou's study, which dealt only with papers written in China, the present paper explores CIS doctoral dissertations produced all over the world, examining their themes, theoretical influences, research methods, supervisors, and the modes of interpreting studied. A supplementary content-labeling method — probabilistic topic modeling (PTM) — was used to ensure empirical objectivity. One of the latest techniques used for document analysis and data-mining, PTM was originally developed in the field of machine learning: it consists of scouring large document archives to tease out hidden thematic information (Blei, 2012). Although the technique has been widely used by social media websites such as Twitter, Facebook and LinkedIn for detecting trending news, recommending targeted ads to users and establishing new connections between professionals of similar backgrounds or research interests, it had not previously been applied to scientometric research in Translation Studies (TS), which traditionally consumes a large amount of time and manpower. The present study therefore serves as a pilot project to explore its efficacy and limitations as a tool in TIS research.

II. Methodology

2.1 Research questions

Having only come into being as a stand-alone field of inquiry in the late 1970s (Li, 2007), CIS has nonetheless experienced massive growth in terms of the numbers of papers published and topics studied over the past three decades. Doctorate-level research, which began in the late 1990s, has to date resulted in more than 30 graduates at various institutions; dozens more are currently in training.

The aim of the present study was to answer questions in the following broad categories:

(1) An overview of CIS: What are the dominant theories influencing CIS doctoral research? What are the major topics that interest PhD students? What modes of interpreting have been thoroughly studied, and what research methods have been employed to investigate them?

(2) Institutional patterns: Which universities produce the most PhD graduates? Who are the most prolific advisors supervising dissertations?

(3) Career paths: In terms of publication counts, how productive are these student researchers? What career paths can be observed by studying CIS dissertations?

These three groups of questions together seek to provide distinct yet complementary perspectives on CIS as a whole. Those of the first group address the themes and theoretical underpinnings of the field, whereas the second and third shed light on the dynamics of the institutions and individuals that drive academic research. Unlike disciplines such as computer science and engineering, which often tend to be industry-led, it is fair to say that in TIS schools and researchers themselves are generally responsible for moving the field forward. The latter two groups of questions are intended to put those of the first into their broader context by examining the more 'businessy' aspects of doctoral education in CIS — asking "Who does it best?" and "How useful is it?" As the discipline matures it is important to go beyond the anecdotal to find data-driven answers to these questions, and to produce findings which the policy-makers of PhD-granting institutions may find useful for making informed decisions about doctoral instruction and long-term research planning. The aim of this study is to provide answers to these questions.

2.2 Data collection

While Mu and Zou (2014) examined a large number of TIS dissertations, for the present study the range was narrowed to focus exclusively on research into interpreting. A total of 32 doctoral dissertations on CIS were collected from China, the United States and the United Kingdom, using the CNKI database, Proquest and interlibrary loan requests. While the corpus was slightly smaller than that used for the aforementioned study, it provided good coverage of the most important studies on interpreting[4], and was representative enough for conclusions about doctoral research in CIS to be drawn from it[5]. In addition, the English abstracts of these dissertations were compiled for topic modeling analysis. In one instance a dissertation did not have an English abstract: the present author summarized its content and created an abstract from it.

2.3 Document labeling

The contents of all the dissertations were reviewed with a specific focus on their literature reviews and methodology sections. The author, title, academic affiliation, publication year, research methods, keywords, theoretical influences and modes of interpreting were identified for each dissertation and entered into an Excel spreadsheet for statistical analysis. Employing the same methodology described at length in earlier studies with MA theses and research papers on CIS (Xu, 2014 & 2015), the theoretical influences from each dissertation were consolidated into six categories: Cognition, Language, Communication Theory, Translation, Peoples and Cultures, and Miscellaneous. Drawing inspiration from Gile's coding scheme (2000), the keywords were grouped into six meme categories: Training, Professional, Language, Socio-cultural, Cognitive and Miscellaneous issues. Earlier scholars have proposed different ways of classifying topics (see for example Li, 2007; Liu & Wang, 2007); the present

---

[4] Of Mu and Zou's total of 686 dissertations on TIS, 39 dealt specifically with IS — 37 from mainland China and one each from Hong Kong and Macau; unlike the present study, their corpus featured no work from Anglophone countries.

[5] The dissertations were obtained from multiple databases and different institutions. Convenience sampling would only be a problem if the samples displayed characteristics not present in the entire population. There is no good reason to believe that the dissertations found in the present sample would be different from ones found elsewhere.

methodology was adopted because the proposed categories do not overlap and effectively cover all possible areas studied by CIS researchers. It should be noted here that each document may be assigned multiple labels for theoretical influences and memes, depending on the theories and topics that it touches on, but no two categories completely overlapped in which dissertations they applied to, and the distribution of dissertations was not highly skewed toward any particular category or categories.

2.4 Topic modeling

Another approach to the labeling of content was used for the present study, namely topic modeling, which had its origins in the field of Natural Language Processing (NLP). Using topic models one is able to determine, firstly, the subjects dealt with in a corpus of documents, and, secondly, the subject of each individual document in that corpus. The most well-known variant of topic modeling is Latent Dirichlet Allocation (LDA) (Blei, Ng, & Jordan, 2003); this is the method employed here.

The LDA model generates probability to describe the makeup of a corpus of documents. Each document is characterized by a multinomial distribution over topics, and each topic by a multinomial distribution over words. LDA is a simple statistical model, yet it is robust enough to deal with data with a large amount of variance (Blei, 2012); at the same time it is sensitive enough to detect hidden thematic structures when data is limited in size (Paul & Dredze, 2011).

Topic modeling has been successfully applied in a variety of different fields. They have, for example, been used to analyze the evolution of research in computational linguistics (Hall, Jurafsky, & Manning, 2008); to categorize press releases from the offices of US Senators (Grimmer, 2010); to examine the topics of papers published in the Proceedings of the National Academy of Sciences (PNAS) (Griffiths & Steyvers, 2004); and to investigate the evolution of topics published in papers in the journal *Science* (Blei & Lafferty, 2006 & 2007).

Two of the advantages that topic models–have over manual document-labeling methods are that they are, for the most part, objectively and empirically determined. With the exception of a few tuning parameters, the texts themselves determine the topics of a corpus and its constituent documents — there is no human intervention. This

objectivity adds value to the labeling process in that it keeps overt bias on the part of researchers to a minimum, preventing any preconceived notions they may have from influencing the choice of labels they assign to documents. Anything of importance to the authors of documents, as reflected in the amount they write about it, will be identified in the topic models, as opposed to potentially being missed under manual labeling schemes.

To fit an LDA model the documents under consideration must be featurized. First, the documents are stripped of so-called function words such as *the*, *a*, and *with*, which do not have any bearing on the topic under discussion. After this they are converted into 'bags of words': the number of occurrences of each word remaining after stripping is counted.

LDA assumes the following Bayesian model: first the topic weights are generated independently according to a prior Dirichlet distribution for each document. Then the weights of words within each topic are generated according to another prior Dirichlet distribution. Finally, for each word in a given document, a topic is chosen from its topic weights, and the words in each topic are selected according to that topic's word weights. The model is fit by obtaining a bag of words from the given set of documents.

For the present analysis the author used the LDA topic model toolkit in *Graphlab Create*, a leading software platform for data science (GraphLab Inc, 2014). The CIS dissertation abstracts were pre-processed to obtain tokens, which were English words. Once this was done, function words were removed to minimize the noise in the data analysis. Bags of words were obtained for each document, formatted as entries with frequencies in a dictionary. This information was used as training material for the *Graphlab* application to fit an LDA model on the data.

The result of an LDA fit was thus two quantities. First, for each topic we obtained an inferred distribution over words. Since function words were removed, the most likely words in each topic can be interpreted as keywords that represented the topic and can be used to assign a label to the topic for convenience. Secondly, an LDA fit returned, for each document, the estimated proportion of its words that came from each of the topics.

For training the LDA model, two parameters – number of topics and iterations – were adjusted to obtain better results, while the other parameters were set at default values. Given the modest size of the data-set, the number of topics was set at 10 and that of

iterations at 200. Though a higher number of iterations might have made it easier to correctly identify the keywords for each topic, it was found that increasing the number above 200 yielded no significant improvement in results in the present instance. Once the model training was completed with bags of words for each topic identified, the topics present in each author's paper were also determined.

III. Results and discussions

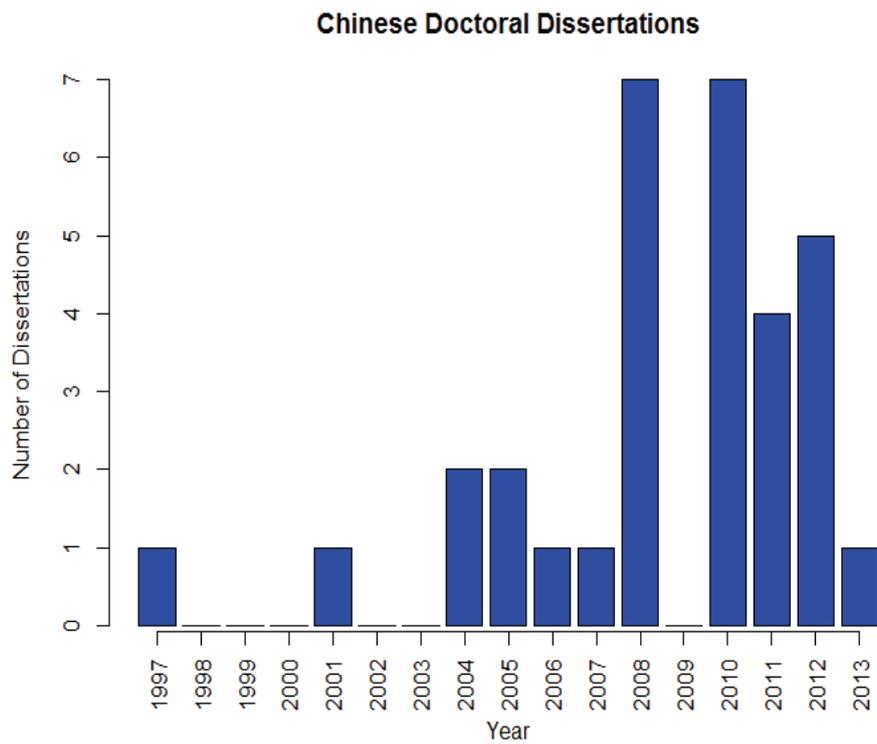3.1 Growth of doctoral dissertations

Chinese Doctoral Dissertations

Figure 1. Number of Chinese doctoral dissertations produced over time

The earliest dissertation in the data-set was written in 1997 by Robin Setton at the Chinese University of Hong Kong. Since then a number of papers have been completed by doctoral students around the world, but chiefly in mainland China over a twelve-year span (2001-2013)[6]. Figure 1 shows that their numbers increased appreciably after 2008, with an average of four dissertations produced each year. The total (31) may look small in comparison with that of MA theses on CIS completed during the same period (just under 1,200) (Xu, 2015), but is impressive when compared with the figures from Western countries such as Switzerland (3), Spain (12) and Italy (1) (Gile, 2013).

3.2 Theoretical influences

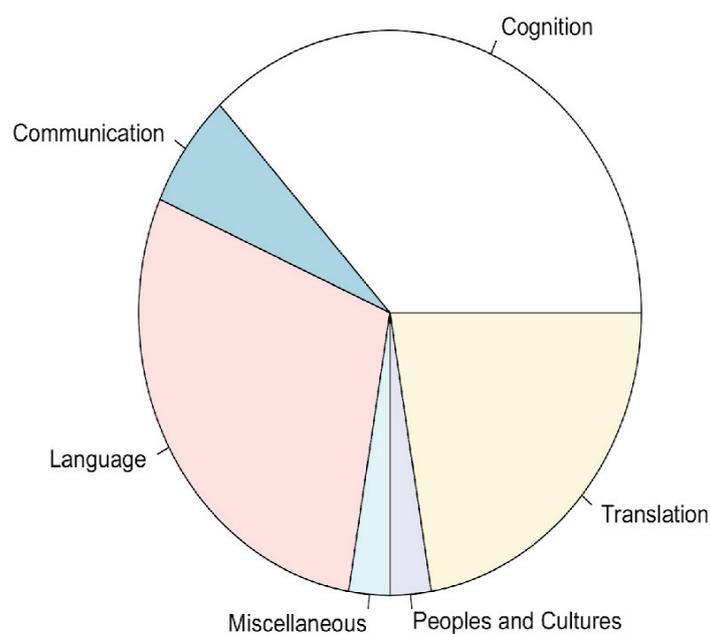Figure 2. Proportions of theoretical influences in doctoral dissertations

In this section we examine the theoretical influences which served as foundations for each CIS doctoral research project, to discover their underlying trends. Cognition was

---

[6] In the data-set three CIS dissertations were completed in the United States, one in the United Kingdom, and one in Hong Kong, during the same period.

the dominant influence with a share of 36.8%, followed by Language-related disciplines (28.9%) and Translation (22.4%). Communication Theory (6.6%), Peoples and Cultures (2.6%), and Miscellaneous (2.6%) made up the remainder. Given that Interpreting and Translation Studies are closely related, it is interesting to observe that rather than Translation theories, Cognition claimed the top spot. A possible explanation for its popularity is that over recent years there has been an increasingly vocal call for interdisciplinary research in the international IS community (Baker & Saldanha eds., 1998), and within CIS itself scholars are encouraged to conduct research using methodologies from other disciplines (Cai, 2001; Zhang, 2012). Of the 32 authors, 17 recruited participants for their research, and employed Cognition-related theories such as Effort Models (Gile, 1995), Baddeley's working memory model (Baddeley, 1992) and the Probability-predict model (Chernov, 2004), to help them explain why interpreters behave in particular ways. However, similarly to the author's earlier finding with regard to CIS research papers (Xu, 2014), there were only very limited influences on doctoral papers from the Communication Theory and Peoples and Cultures categories, as can be seen from their low rankings — a finding which serves to highlight that those two categories are under-researched in CIS. A quick survey revealed that nearly all PhDs in CIS are either offered through a university's foreign language department, as in the case of Sichuan University, or granted directly by the Graduate School of Translation and Interpreting, such as is the case at the Shanghai International Studies University (SISU); none are offered by departments of Communication or Intercultural Studies. This may explain why the aforementioned categories are unable to gain traction.

3.2.1 Most popular theories in Chinese doctoral dissertations

With the aim of identifying the dominant theories to be found in CIS doctoral studies, a count of the frequency of all those mentioned in the data-set was carried out. Analysis of the data revealed that the 32 dissertations made use of 72 different theories as their foundations. Of those 72, only five received more than three mentions:

- The Interpretive Theory of Translation (12 mentions): Seleskovitch's theory (Seleskovitch, 1978), which emphasizes the importance of unpacking meaning

from its linguistic wrappings and is often used as the benchmark for interpreter training, enjoyed a reasonable degree of popularity with Chinese doctoral students. Zhang Jiliang even dedicated his entire dissertation (2008) to examining the evolution of this particular theory in the context of contemporary interpreting research.

- Effort Models (12): Wan Hongyu (2006), Sun Xu (2010), and Xu Qilu (2012) all used this theory (Gile, 1995) to explain how their research participants allocated limited cognitive resources to different tasks during their performances.

- Working memory (5): Liu Minhua (2001), Tzou Yeh-zu (2008), Zhang Wei (2005), and Xu Qilu (2012) explored the relationship between working memory and SI performance, while Sun Xu (2010) used theories concerning working memory to study the connections between language proficiency and CI competence.

- Psycholinguistics (4): While the theory of working memory focuses on the limited brain capacity available for storing information, psycholinguistics studies language comprehension, processing and production. Some doctoral researchers, such as Hu Lingque (2008) and Huang Yi (2013), turned to this specialism (Warren, 2012) to examine the psychological and neurobiological factors which contribute to the success of an interpreting assignment.

- The expert-novice paradigm (4): This is widely used in the study of music, sports, and aviation to describe the differences between how experts and novices perform. In recent years scholars have attempted to develop instructional activities based on research methods originally used to capture the differences between the two grades of performer (Fadde, 2009). Moser-Mercer (1997) launched the multi-university project on language and communication, which was based on this paradigm. Since then a number of Chinese doctoral researchers, such as Liu Minhua (2001), Sun Xu (2010), Zhu Jinping (2010) and Huang Yi (2013), have adopted it to study skills acquisition in CI and SI.
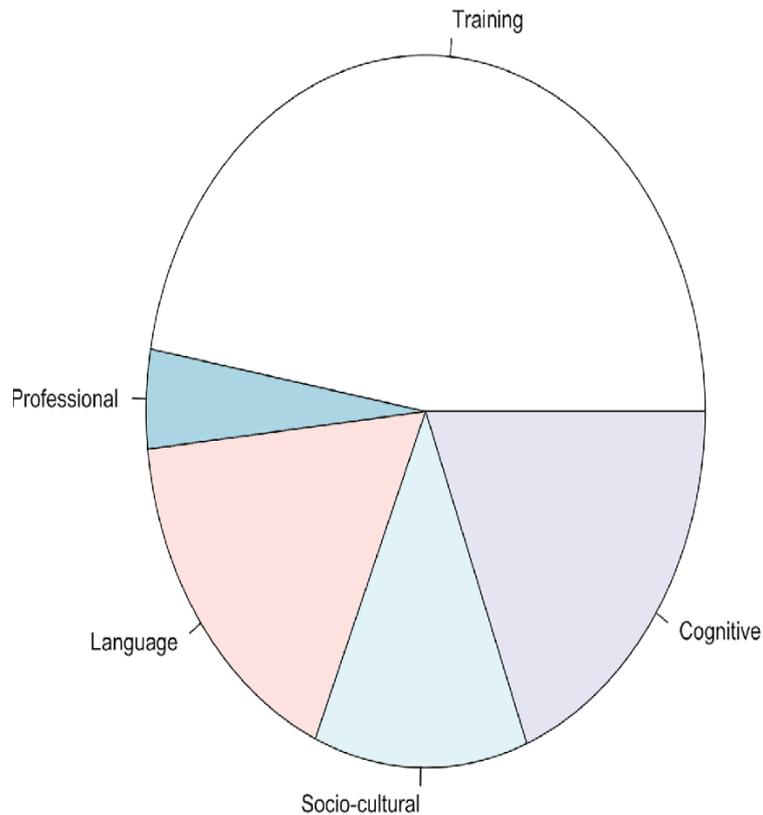
3.3 Themes of CIS

Figure 3. Memes in doctoral dissertations

Theoretical influences can be considered the 'input' of doctoral research activities and memes their 'output'. When it came to examining the latter, it was observed that Training was the most studied among PhD students, with an enormous 47.2% of the total, far outstripping all the others. According to Mu and Zou (2014), the majority of Chinese doctoral students are interpreter trainers, so may find Training issues convenient to tackle and relevant to their jobs. Cognitive (19.1%) and Language-related (16.9%) issues received a moderate amount of attention, while Socio-cultural (12.4%) and Professional (4.5%) issues were the least favored. These findings are broadly in line with those for CIS research papers (Xu, 2014).

3.4 Keywords in Chinese doctoral dissertations

A more detailed examination of the data-set was performed to gauge how often the keywords — the 'building blocks' of the memes from the previous section — were used. The keywords used by the original authors were taken into consideration when generating comprehensive keyword sets for each paper, in order to capture the topics addressed as fully as possible. This process revealed that a wide range of topics were covered: 46 keywords received only one mention, 8 received two, 3 received three, and another 3 received four. The following is a list of the keywords which received more than four mentions:

- Assessment (9 mentions): A widespread interest in creating systems for evaluating interpreters' performances made Assessment the most studied keyword.
- Interpreting performance (7): This keyword is closely related to assessment. Numerous researchers observed performance either in laboratory conditions or by using recordings of real-life assignments.
- Interpreting process (6): A good number of authors sought to offer explanations of the different stages an interpreter passes through when handling incoming information and rendering it into another language.
- Interpreting strategies (5): Some students investigated when and how interpreters employ particular solutions for dealing with difficulties in interpreting.

3.5 Topic models

3.5.1 Content analysis for the corpus

As a supplementary measure to manual labeling, PTM was used to identify the topics covered in the corpus. To perform this analysis an LDA model with ten topics was fit to the doctoral dissertation abstracts. The resulting topics are listed in Table 1: the left-hand column contains the names assigned to the topics after the keywords had been evaluated, the middle shows the percentage of doctoral dissertations that deal

substantially with the given topic (topic weight of at least 5%), and the right-hand contains the top five keywords associated with each estimated topic. Note that the percentages in the second column do not add up to 100% because a large number of dissertations covered more than one topic.

Topic Name	Document Frequency	Top Five Keywords
Allocation of cognitive resources	14.60%	language, performance, cognitive, information, allocation
Deverbalization[7] for surmounting intercultural barriers	12.50%	deverbalization, speech, intercultural, eco-translatology, comprehension
Modeling the interpreting process	12.50%	theory, interpreting, model, process, mechanism
Compression strategies in SI	12.50%	interpreting, simultaneous, strategies, compression, cultural
Interpreting competence	10.40%	interpreting, competence, development, psychological, verbal
Corpus research on conference interpreting	10.40%	interpreters, professional, sentences, analysis, corpus
Verbal fluency	8.30%	communication, interpreter, syntactic, strategy, fluency
Coherence issues in interpreting	7.30%	source, coherence, Chinese, trainees, role
A discourse analysis approach to assessing interpreting quality	6.25%	process, quality, discourse, knowledge, structure
Contextual awareness for going beyond surface meaning	5.20%	processing, linguistic, form, surface, awareness

Table 1. Most frequently mentioned LDA topics

---

[7] Deverbalization refers to the ability of an interpreter to render the meaning of the source into the target language without relying on finding direct linguistic correspondences.

Note that not all of the top keywords associated with a given topic will necessarily appear in an abstract about that topic. There are two complementary reasons for this. Firstly, the keywords are merely the top few words associated with the particular topic, and so do not completely represent it. Secondly, the LDA framework classifies a document as having a particular topic if it contains *many* of the words prominently associated with that topic, not *all* of them.

Indeed, it would be unreasonable to expect every document classified as belonging to a given topic to contain all the words associated with that topic, because authors may discuss differing aspects of the same topic or use different synonyms to refer to the same concepts. For a document to be classified as belonging to a particular topic, then, it must contain words associated with the topic, which need not encompass all of the top keywords most associated with that topic.

The results of the analysis show that Chinese doctoral researchers were particularly interested in issues connected with the allocation of mental resources and in modeling the interpreting process, a finding broadly in line with the earlier meme classification that Cognitive issues were a key focus for PhD students. LDA was able to detect these two specific subtopics of Cognitive that were entirely missed by the traditional method of keyword analysis. With 80% of all Chinese interpreting students having never spent more than a month or two in overseas English-speaking environments (Setton and Guo, 2009), it is fair to say that they do not have enough exposure to how English is used naturally, so have to expend significant mental resources on language production when interpreting into their target tongue. This may explain why mental resource allocation is a subject of such particular interest to them. As for process modeling, certain theoretical frameworks are popular in the Chinese academic world (the Interpretive Theory of Translation, Effort Models, etc.). However, there are a limited number of studies testing the empirical validity of these claims, so dissertation authors such as Wan Hongyu (2006), Zhang Wei (2007) and Chen Liwen (2011) set out to offer their own models for interpreting processes with support from empirical data.

Table 1 indicates that topics such as deverbalization, interpreting competence, and compression strategies are also popular among Chinese doctoral students. This finding is in line with the results for meme classification, which revealed that Training was the

most studied topic. With regard to deverbalization, 12 of the 32 dissertations cited the Interpretive Theory of Translation as part of their theoretical foundation — a cornerstone of this theory is the concept of deverbalization. Over the years the concept has come in for a great deal of criticism in the West; however, numerous Chinese interpreter training programs, such as SISU's, have actively incorporated it into their teaching philosophy and practice. The logic behind this is that the syntactic differences between Chinese and English make it difficult to find linguistic equivalents when interpreting. To examine the effectiveness of deverbalization, authors such as Zhang Qunxing (2010), Sun Haiqin (2012) and Huang Yi (2013) conducted experimental studies the aim of which was to use empirical data in support of the concept.

One hot topic of interest completely missed by the meme and keyword analyses but captured by LDA topic modeling was corpus research. Close examination of the meta-data revealed that the corpora created by CIS researchers can be divided into two types: (1) those that employ observational data obtained in natural contexts where there is little interference from the experimenters (see for example Gao Bin, 2008; Wang Yongqiu, 2008; Zhan Cheng, 2011); and (2) those which collect data from experiments in which the experimenters intervene directly (see for example, Robin Setton, 1997; Yu Wenting, 2012).

3.5.2 Analysis of individual researchers' topics: LDA vs. manual keyword labeling

Once the LDA model fitting was complete, the top three topics associated with each doctoral researcher were identified; these are listed in Table 2[8]. It should be stressed that only the top three topics are presented here, since they have the highest probability scores and thus should be the primary foci of each author, though in reality each author is likely to have covered more than three topics. For the purposes of comparison, Table 2 also lists the keywords identified by the present author using the manual approach, to highlight the strengths of each topic-labeling approach.

Author's	Topic	Topic Name	Keywords

---

[8] The complete references for all the dissertations are given at the end of this paper.

Name	No.		
CHANG Chia-chien	Topic 10	Allocation of cognitive resources	interpreting performance, directionality, interpreting strategies
	Topic 7	Interpreting competence	
	Topic 8	Corpus research on conference interpreting	
CHEN Jing	Topic 2	Verbal fluency	interpreting process, interpreter's roles, noise
	Topic 4	Deverbalization for surmounting intercultural barriers	
	Topic 7	Interpreting competence	
CHEN Liwen	Topic 5	Contextual awareness for going beyond surface meaning	interpreting process, meaning negotiation, subjectivity
	Topic 6	Modeling the interpreting process	
	Topic 9	Compression strategies in SI	
CHEN Shengbai	Topic 1	Coherence issues in interpreting	press conferences, interpreting process, interpreting strategies
	Topic 4	Deverbalization for surmounting intercultural barriers	
	Topic 9	Compression strategies in SI	
GAO Bin	Topic 10	Allocation of cognitive resources	evolution, trends, themes, history
	Topic 4	Deverbalization for surmounting intercultural barriers	
	Topic 7	Interpreting competence	
GONG Longsheng	Topic 5	Contextual awareness for going beyond surface meaning	interpreting process, context

	Topic 8	Corpus research on conference interpreting	
	Topic 9	Compression strategies in SI	
GUO Liangliang	Topic 10	Allocation of cognitive resources	word order, EVS, information density, interpreting strategies
	Topic 8	Corpus research on conference interpreting	
	Topic 9	Compression strategies in SI	
HU Lingque	Topic 1	Coherence issues in interpreting	note-taking, domain-specific knowledge, cognitive load, information retrieval
	Topic 6	Modeling the interpreting process	
	Topic 9	Compression strategies in SI	
HUANG Yi	Topic 4	Deverbalization for surmounting intercultural barriers	recognition memory, deverbalization
	Topic 6	Modeling the interpreting process	
	Topic 8	Corpus research on conference interpreting	
JIA Dan	Topic 4	Deverbalization for surmounting intercultural barriers	interpreter training, interpreting performance
	Topic 6	Modeling the interpreting process	
	Topic 7	Interpreting competence	
KANG Zhifeng	Topic 2	Verbal fluency	anxiety, interpreting performance
	Topic 6	Modeling the interpreting process	
	Topic 7	Interpreting competence	

LIU Qingxue	Topic 10	Allocation of cognitive resources	interpreting strategies, cross-cultural communication
	Topic 2	Verbal fluency	
	Topic 9	Compression strategies in SI	
LIU Menglian	Topic 10	Allocation of cognitive resources	interpreter training, online learning, self-evaluation
	Topic 3	A discourse analysis approach to assessing interpreting quality	
	Topic 6	Modeling the interpreting process	
LIU Minhua	Topic 10	Allocation of cognitive resources	expertise, working memory, assessment, interpreting performance
	Topic 6	Modeling the interpreting process	
	Topic 9	Compression strategies in SI	
PENG Kuei-Chuan	Topic 1	Coherence issues in interpreting	coherence, interpreting quality, self-correction, self-evaluation, peer feedback
	Topic 3	A discourse analysis approach to assessing interpreting quality	
	Topic 8	Corpus research on conference interpreting	
REN Wen	Topic 4	Deverbalization for surmounting intercultural barriers	interpreter's subjectivity, power, ethics, visibility, empowerment, neutrality
	Topic 5	Contextual awareness for going beyond surface meaning	
	Topic 6	Modeling the interpreting process	
SETTON	Topic 1	Coherence issues in interpreting	interpreting process,

Robin	Topic 3	A discourse analysis approach to assessing interpreting quality	modeling, word order differences
	Topic 6	Modeling the interpreting process	
SHA Liewen	Topic 2	Verbal fluency	translation units, interpreting techniques, linguistic redundancy, anticipation
	Topic 4	Deverbalization for surmounting intercultural barriers	
	Topic 9	Compression strategies in SI	
SU Wei	Topic 3	A discourse analysis approach to assessing interpreting quality	interpreting performance, assessment, competence, verbal fluency, curriculum design
	Topic 4	Deverbalization for surmounting intercultural barriers	
	Topic 7	Interpreting competence	
SUN Haiqin	Topic 2	Verbal fluency	information density, deverbalization, interpreting quality, assessment
	Topic 4	Deverbalization for surmounting intercultural barriers	
	Topic 8	Corpus research on conference interpreting	
SUN Xu	Topic 10	Allocation of cognitive resources	interpreting competence, interpreting quality, assessment, information density
	Topic 7	Interpreting competence	
	Topic 8	Corpus research on conference interpreting	
TZOU Yeh-zu	Topic 10	Allocation of cognitive resources	working memory, language proficiency, interpreting performance,
	Topic 2	Verbal fluency	
	Topic 9	Compression strategies in SI	

			assessment
WAN Hongyu	Topic 10	Allocation of cognitive resources	interpreting process, segmentation, modeling
	Topic 4	Deverbalization for surmounting intercultural barriers	
	Topic 6	Modeling the interpreting process	
WANG Yongqiu	Topic 10	Allocation of cognitive resources	interpreting quality, assessment, condensation, interpreting strategies
	Topic 8	Corpus research on conference interpreting	
	Topic 9	Compression strategies in SI	
XU Qilu	Topic 1	Coherence issues in interpreting	working memory, attention, concurrent speaking and listening
	Topic 10	Allocation of cognitive resources	
	Topic 6	Modeling the interpreting process	
YANG Liuyan	Topic 2	Verbal fluency	interpreter training, verbal fluency, assessment
	Topic 7	Interpreting competence	
	Topic 9	Compression strategies in SI	
YU Wenting	Topic 4	Deverbalization for surmounting intercultural barriers	self-correction, assessment, interpreting performance, self-monitoring
	Topic 6	Modeling the interpreting process	
	Topic 8	Corpus research on conference interpreting	
ZHAN Cheng	Topic 1	Coherence issues in interpreting	shifts, interpreter's

	Topic 5	Contextual awareness for going beyond surface meaning	roles, assessment, turn-taking
	Topic 8	Corpus research on conference interpreting	
ZHANG Wei	Topic 10	Allocation of cognitive resources	working memory, omissions, directionality, modeling
	Topic 5	Contextual awareness for going beyond surface meaning	
	Topic 7	Interpreting competence	
ZHANG Jiliang	Topic 1	Coherence issues in interpreting	history, theories
	Topic 5	Contextual awareness for going beyond surface meaning	
	Topic 9	Compression strategies in SI	
ZHANG Qunxing	Topic 2	Verbal fluency	deverbalization, free recall, memory capacity
	Topic 3	A discourse analysis approach to assessing interpreting quality	
	Topic 4	Deverbalization for surmounting intercultural barriers	
ZHU Jinping	Topic 10	Allocation of cognitive resources	interpreting competence, b language, language proficiency, interpreter training
	Topic 3	A discourse analysis approach to assessing interpreting quality	
	Topic 7	Interpreting competence	

Table 2. Research foci of each doctoral dissertation

Table 2 contrasts the top three topics identified by LDA with all the keywords generated via the manual labeling approach. For example, LDA identified that Chang addressed 'allocation of cognitive resources', 'interpreting competence' and 'corpus research on conference interpreting' in her dissertation, whereas the foci of her research were identified as 'interpreting performance', 'directionality' and 'interpreting strategies' by manual scanning. The results demonstrate that LDA is effective at picking out themes that are commonly discussed by multiple people, whereas the manual approach can pick out unique topics focused on by one or two individual authors. One intriguing finding is that the LDA method enables one to uncover hidden themes that the human eye alone may not immediately notice. Let us take Chang Chia-chien's dissertation (2005) as an example. While the manual method enables us to correctly identify 'directionality' as a focus of her work, despite its not being among the top three LDA topics associated with her, it failed to identify 'cognitive resource allocation'. On closer scrutiny it was revealed that Chang only made an oblique reference to resource allocation near the end of her abstract: "The difference in their performances seems [to be] a result of their metacognitive awareness of the limits of their language abilities …". Obviously this would be rather difficult for a researcher to detect; one cannot afford to dwell for long over a single document when conducting scientometric research. LDA is effective in uncovering hidden thematic topics like this one because it calculates the probability of words occurring together in a document, identifies a 'bag' of keywords associated with a certain topic, and, when a large number of those words appear in a given document, it enables the researcher to determine whether that particular document contains that topic. Given that scientometric research in TIS requires a tremendous amount of manual data-mining, the promising results from the present analysis strongly suggest that LDA can serve as a powerful supplementary procedure to manual labeling.

3.6 Modes of interpreting

Interpreting is far from a homogenous discipline, covering as it does a range of culturally and socially distinct activities from conference to community interpreting, all of which require somewhat different skill-sets. For the present study Pöchhacker's classification of interpreting (2004) was adapted to investigate the working modes,

social contexts and various other forms of interpreting ('miscellaneous') as they occur in PhD theses.

On examining the working modes it was observed that 15 PhD theses addressed consecutive interpreting (CI), 12 the simultaneous mode (SI) and one sight translation (ST). Again, this was in line with the findings for MA theses (Xu, 2015). Only Wan Hongyu (2006) took on the subject of ST, seeking to improve upon Gile's model (1995) by stressing the importance of working memory and effort coordination in that particular mode.

Analysis of the dissertations' social contexts revealed that three explicitly addressed conference interpreting, two diplomatic and one escort. Zhan Cheng (2011) studied diplomatic interpreting by analyzing the audio recordings of six staff interpreters for the Foreign Affairs Office of Guangdong Province. Chen Shengbai (2012) also approached this social context in her analysis of the press conference interpreting for China's Premier between 2010 and 2012. None of the authors studied business interpreting, which was the third most studied social context in MA theses (Xu, 2015). Technical, court and other forms of community interpreting received no attention from Chinese doctoral researchers.

In the miscellaneous category, of the 32 dissertations only two addressed TV interpreting. Chen Shengbai's study (mentioned above) touched on TV interpreting in that the Premier's conferences were televised. Wang Yongqiu (2008) analyzed the use of compression strategies by the Cantonese simultaneous interpreters working on three televised meetings of Hong Kong's Legislative Council. As of the time of writing (January, 2015) no study has addressed telephone interpreting. Despite its wide application in developed countries such as the United States, the United Kingdom, and Australia, telephone interpreting has only come into use for major international sporting events in China since 2010 (Zhan & Suo, 2012). As it becomes gradually more mainstream and popular in China, one might expect this mode to receive increasing attention from PhD students.

3.7 Empirical research in doctoral dissertations
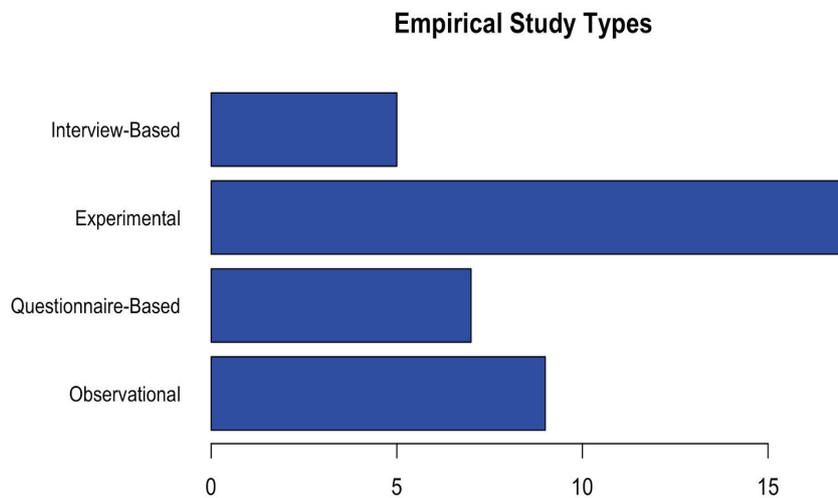
Empirical Study Types

Figure 4. Frequency of different types of empirical research methods used in PhD dissertations

Empirical research can help support or refute theories and hypotheses, a process vital for the robust development of a discipline such as IS in which results and findings display a high degree of variability (Gile, 2013). The data for the present study revealed that 26 of the 32 dissertations (81.3%) were empirical in nature, in comparison with the data-set for MA theses, where the proportion was just under 50% (Xu, 2015). This finding suggests that doctoral researchers are perhaps more inclined or have more time to use a data-driven approach to finding answers to their research questions. It is clear from Figure 4 that experiments were the primary research method for empirical studies: 17 took that form, followed by observational (9), questionnaire-based (7) and interview-based (5) studies. (It should be noted that some studies employed a number of research methods, so the total does not add up to 26.) Gile (1994) points out that in comparison with other disciplines in behavioral sciences, experiments in IS face fewer technical difficulties such as the creation of test environments similar to those of actual working conditions. Simultaneous interpreters work in booths, similar to those found in language labs, and consecutive interpreters typically sit at the conference room table or stand at the podium; both can easily be simulated in a regular classroom. The relative ease of

replicating ecologically valid conditions may help to explain why so many doctoral researchers favor experiments.

However, in the same article Gile also points out that the serious lack of uniformity in test subjects and the difficulty of recruiting competent interpreters are issues that the authors of experimental studies need to be particularly chary of. These issues did indeed prove to be especially problematic in the case of Chinese MA theses (see Xu, 2015), so the author set out to examine whether the same was true of doctoral dissertations. Examining the sample sizes and types of participants used in the 17 experimental studies revealed that:

- 7 used both professionals and students
- 3 used professionals only
- 7 used students only

The sample sizes for those using professionals (with or without students) were between 3 and 20 (3, 5, 6, 10, 10, 12, 13, 13, 15, 20). Incidentally, this range is similar to the use of professionals in MA theses (Xu, 2015). In the papers for which student interpreters (with or without professionals) were recruited, the sample sizes ranged from 8 to 115 (8, 8, 10, 16, 23, 23, 30, 36, 38, 45, 50, 53, 57 and 115). A particular examination was made of Zhang Wei's dissertation of 2007, to date the only one in IS to have received the accolade of Best Chinese Dissertation in Social Sciences. He recruited a total of 13 interpreters and 115 students, of whom 45 were majoring in subjects other than interpreting; of the interpreting students 35 were first-year and 35 second-year. To control for variability in working memory capacity between participants in each sub-group of interpreting students, Zhang emphasized that he used random sampling to select his participants.

None of the researchers conducted multiple experiments in their studies. This may have limited their ability to draw valid conclusions that were representative of the overall population under scrutiny. The variability of test subjects is a notable feature of Interpreting Studies (Gile, 1994). Even the same individual cannot produce identical interpretations when asked to render the same material twice, not to mention the significant differences between participants in terms of language proficiency, training and experience. Multiple experiments with different participants and under various working conditions are required before generalizations can be made.

3.8 Dissertation-producing universities

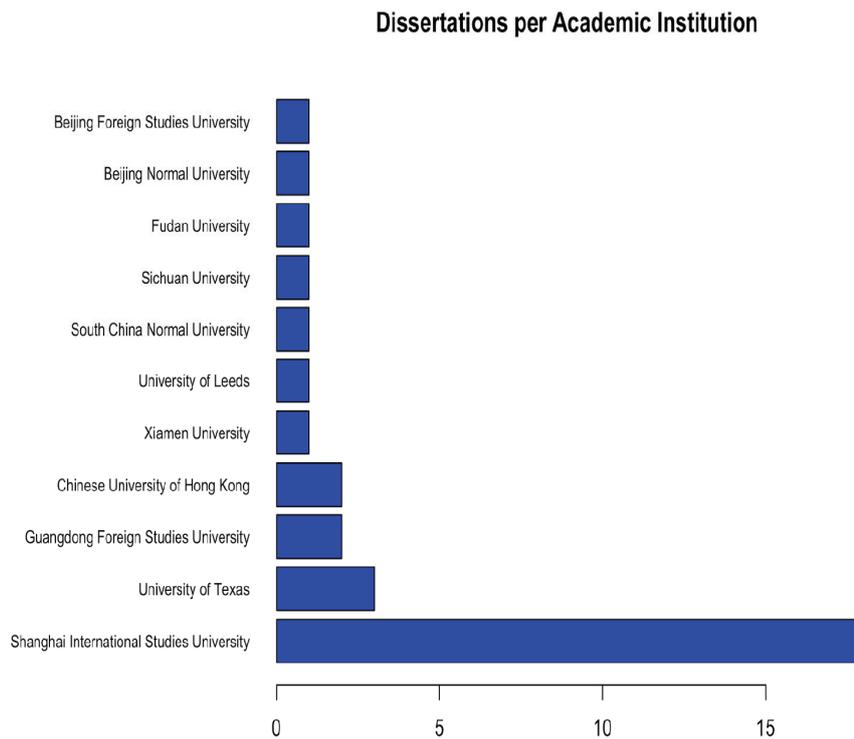**Dissertations per Academic Institution**

Figure 5. Dissertation-producing universities

The rapid development of CIS doctoral research has not occurred spontaneously in a vacuum: the universities at which it is produced have obviously played their role in the discipline's boom. The present analysis shows that the Shanghai International Studies University (SISU) was the most prolific in generating dissertations, with a total of 18. The University of Texas came in second with three, followed by the Chinese University of Hong Kong (CUHK) and Guangdong Foreign Studies University (GFSU) with two each. Only one dissertation was recorded for each of the remaining universities. In 2004 SISU was the first school in mainland China to be granted the authority to award PhDs in Translation Studies — its first PhD class graduated in 2008. It has actively

contributed to the development of CIS since the early 2000s. In tandem with its prolific generation of dissertations, in 2014 it launched China's first TIS doctoral summer school, welcoming 120 participants from all over the country (SISU, 2014). While this cannot be considered a means of *completing* a dissertation, it nonetheless acts as a supplement to PhD education and creates an opportunity for young researchers to network with experts and colleagues in the field.

Some other universities also produced dissertations on CIS, though the authors were not necessarily TIS doctoral students. For example, both Sha Liewen's and Chen Jing's dissertations on IS were completed in 2004; the former's doctorate was in applied linguistics and the latter's in English language and literature. That said, institutional recognition of the discipline at SISU has undoubtedly contributed to its total number of dissertations far exceeding that produced at any other school.

3.9 PhD Advisorship

Doctoral Supervisors

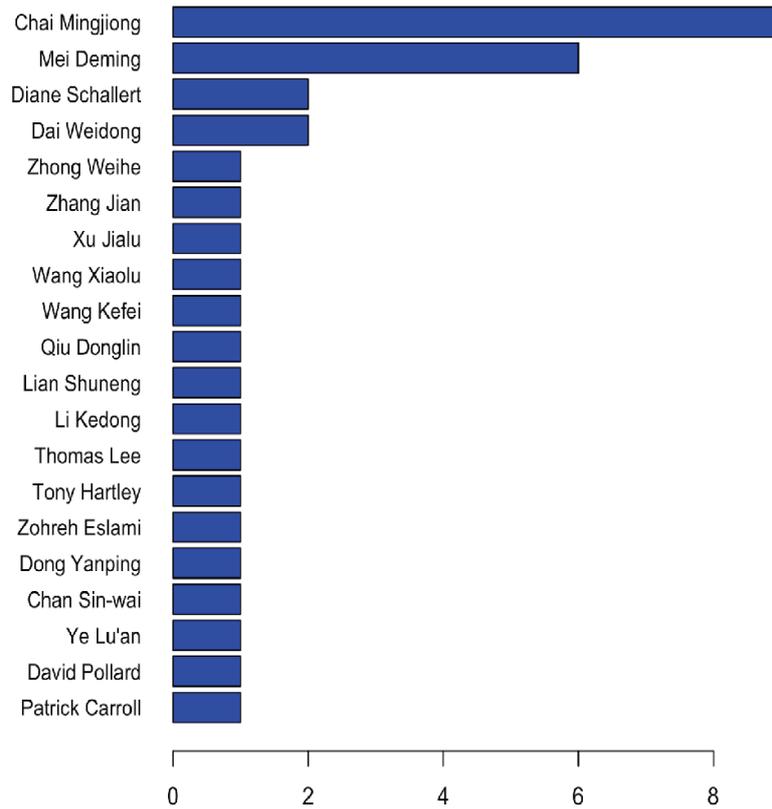| Supervisor	Number
Chai Mingjiong	9
Mei Deming	6
Diane Schallert	2
Dai Weidong	2
Zhong Weihe	1
Zhang Jian	1
Xu Jialu	1
Wang Xiaolu	1
Wang Kefei	1
Qiu Donglin	1
Lian Shuneng	1
Li Kedong	1
Thomas Lee	1
Tony Hartley	1
Zohreh Eslami	1
Dong Yanping	1
Chan Sin-wai	1
Ye Lu'an	1
David Pollard	1
Patrick Carroll	1

Figure 6. Number of dissertations supervised by different advisors

Figure 6 shows that Chai and Mei, both affiliated to SISU, supervised the most dissertations; the majority supervised only one student each. Gile (1994) recommended that PhD students ideally have two supervisors who can complement one another's strengths, e.g. one well-versed in TIS and the other with expertise in linguistics or psychology. From the data it would seem, however, that Chinese doctoral advisors are more inclined to take on supervision single-handed. Of the 32 doctoral dissertations in the data-set, only Liu Minhua, Robin Setton, and Yu Wenting had two supervisors each. Diane Schallert, an expert in learning motivation, and Patrick Carroll, a specialist in

psycholinguistics and cognition, formed the team advising Liu, who studied skills acquisition and its relationship with working memory. Thomas Lee and David Pollard teamed up to advise Setton on his doctoral studies, the former bringing to the table his knowledge of language acquisition and the latter his experience in translating classical Chinese. Mei Deming partnered with Jeroen van de Weijer to supervise Yu's doctoral project on self-correction in consecutive interpreting. While both Mei and van de Weijer have expertise in linguistics, the latter specializes in phonology and psycholinguistics.

Despite the fact that only a handful of doctoral students had co-supervisors, it should be noted that all had access to a large support network, as can be seen in the number of people mentioned in each dissertation's acknowledgements section. For example, Zhan Cheng, who examined the interpreter's role as a mediator in diplomatic meetings, acknowledged receiving research assistance from 26 people during the writing of his thesis. In addition, faculty members from the same program may provide consistent research guidance and support to students who are not their official advisees. Setton is a case in point: he spent three to four months every year teaching seminars on cognitive science, research methodology and interpreter training during his tenure at SISU, though he never took on any doctoral students himself. He even co-authored a paper with Guo Liangliang (2009) on the professional identity issues surrounding Chinese translators and interpreters, even though Guo's official supervisor was Chai Mingjiong.

3.10 Academic publications and career paths of doctoral students

The PhD generally represents the highest qualification one can obtain in an academic discipline[9]: it requires candidates not only to master their given subject but to add to the 'stock' of knowledge by publishing academic papers on it. The data shows that of 32 successful PhD scholars, 28 have gone on to publish papers. On average each has produced five, a lower number than the eight per person in Mu and Zou's study of PhD holders in TIS, indicating that on average doctoral students in CIS produce slightly less

---

[9] In a number of countries, such as Germany, France, Austria and Switzerland, habilitation is the highest academic degree: this qualifies scholars for full professorship. The qualification does not exist in China, North America, or the United Kingdom.

research than their peers in TS. That said, regardless of the number of papers they went on to publish, all 32 scholars secured permanent academic positions with various universities immediately upon graduation, suggesting that doctorates in interpreting are in high demand. Unlike other disciplines within humanities such as linguistics or musicology, where even PhD holders from elite universities are sometimes obliged to take on non-permanent posts during their first few years after graduation, before being offered secure permanent positions, 93.8% of the CIS scholars in the data-set stayed on with their first employer. The remaining 6.2% switched employers by choice, and there was no gap between their various academic appointments – in fact, they often held multiple faculty positions with various universities at the same time. These statistics perhaps suggest that the CIS academic market has not yet aged to the point of saturation: that, compared with long-established disciplines, a relatively small number of PhD-holders compete for a growing number of teaching positions worldwide, and that as a consequence doctors in CIS have less difficulty finding permanent work. It is likely that as the field continues to mature and produce highly qualified interpreter trainers, the competition for teaching posts will intensify; in the meantime a PhD in Chinese Interpreting Studies remains a highly beneficial acquisition which, by setting the holder apart from hundreds of MA-holders, unlocks numerous doors.

IV. Conclusion

Doctoral-level research in CIS has been developing rapidly over the past decade or so. Cognition and translation-related disciplines are the dominant theoretical influences among researchers, and training and cognitive issues are the most popular choices for study. Over four fifths of dissertations are empirical, experiments being the preferred research method. Nearly half of all students focus on CI, and slightly over a third on SI.

One of the aims of the present study was to explore the use of PTM as an objective method of content labeling, complementary to the manual approach. The results revealed that LDA modeling was successful in detecting hidden themes in CIS doctoral research that are missed by the traditional labeling method. As increasing numbers of dissertations, theses and papers are digitized, topic modeling looks likely to be much

used in future research into larger corpora. The mass OCR-ing of texts will increase the numbers of available keywords relating to given topics, which will in turn greatly improve the accuracy of topic prediction. When data on a number of papers published by a certain author are available, PTM will help to identify how that author's research interests change over time.

Though the current study uses LDA to analyze texts only in English, it can also be applied to analyzing Chinese texts, albeit with certain technical challenges. In an earlier study Zhang and Qin (2010) demonstrated the effectiveness of using Chinese characters as the basic units of data. Unlike English and the majority of alphabetic languages, the basic structural unit of written Chinese is the character or morpheme, not the word, and the fact that characters are written without spaces between them makes it difficult for computers to parse units of meaning. Most previous research using topic models for Chinese documents did not take into consideration the relationship between characters and units of meaning, but simply treated characters as the linguistic 'building blocks' of documents. In a more recent study (Zhao, Qin & Wen, 2011), a new model was proposed which took that relationship into account by placing an asymmetric prior to the topic-meaning unit distribution of the standard Latent Dirichlet Allocation (LDA) model. Zhao et al. concluded that, in comparison with LDA, the revised model can improve performance in document classification especially when the test data contains a considerable number of Chinese meaning units which do not appear in the training data. Future scientometric research with Chinese texts may be greatly facilitated by this latest development in machine learning.

As the competition for academic positions gets tougher, a large majority of Chinese universities now require PhDs for new openings in interpreter training. This has provided an incentive for interpreting students to enroll in PhD programs. The American Council of Graduate Schools carried out a survey (Sowell, 2008) on the completion rate of doctoral dissertations: their findings revealed that only 56% of students in social sciences were able to complete their dissertation within ten years of beginning their doctoral studies. The high attrition rate has become a major concern in the US, because failure to complete often leaves students with debts and limited opportunities for career advancement (Lovitts, 2001). While no such data is available for the completion rates of Chinese IS doctoral students, there is no doubt that pursuing

a PhD in Interpreting Studies in mainland China is less time-intensive and follows a far more predictable path: all the students in the present study's data-set managed to complete their studies and successfully defend their dissertations within three years. In addition, they all went on to pursue careers as academics, a fact which undoubtedly spurs on aspiring students. Its strong social and institutional network of support has helped to transform CIS from a state of virtual non-existence into the formidable presence it is today — and all within the space of a decade. Nobody knows what the next ten years may have in store for the discipline, but this author hopes that it will be as remarkable as the last.

REFERENCES

Baddeley, A. (1992). Working memory. *Science, 255*(5044), 556-559. doi:10.1126/science.1736359

Baker, M., & Saldanha, G. (Eds.). (1998). *Routledge encyclopedia of translation studies*. London: Routledge.

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. *Proceedings of the 23rd International Conference on Machine Learning*. Retrieved from http://www.cs.columbia.edu/~blei/papers/BleiLafferty2006a.pdf

Blei, D. M., & Lafferty, J. D. (2007). A correlated topic model of Science. *Annals of Applied Statistics, 1*(1), 17-35. doi:10.1214/07-AOAS114

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84. doi:10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3*, 993-1022.

Cai, X. (2001). Interpretation studies with an interdisciplinary perspective (In Chinese). *Chinese Translators Journal, 22*(2), 26-29.

Chang, C. (2005). *Directionality in Chinese-English Simultaneous Interpreting: Impact on performance and strategy use* (Unpublished doctoral dissertation). University of Texas.

Chen, J. (2004). *Intercultural noise in interpreting as a communication process* (In Chinese) (Unpublished doctoral dissertation). Xiamen University.

Chen, L. (2011). *A conceptual integration study on meaning negotiation in the Interpreting process* (Unpublished doctoral dissertation). Shanghai International Studies University.

Chen, S. (2012). *An ecological approach to interpretation studies* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Chernov, G. V. (2004). *Inference and anticipation in simultaneous interpreting: A probability-prediction model*. Amsterdam: John Benjamins.

Fadde, P. J. (2009). Training of expertise and expert performance. *Technology, Instruction, Cognition and Learning, 7*(2), 77-81.

Gao, B. (2008). *Conjectures and refutations* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Gile, D. (1994). Methodological aspects of interpretation and translation research. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 39-56). Amsterdam: John Benjamins.

Gile, D. (1995). *Basic concepts and models for interpreter and translator training.* Amsterdam: John Benjamins.

Gile, D. (2000). The history of research into conference interpreting: A scientometric approach. *Target, 12*(2), 297-321.

Gile, D. (2013). Institutional, social and policy aspects of research into conference interpreting. In O. G. Becerra, E. M. Pradas Macías, & R. Barranco-Droege (Eds.), *Quality in interpreting: Widening the scope* (Vol. 1, pp. 9-31). Editorial Comares.

Gong, L. (2008). *An analytical study of the application of Adaptation Theory in Interpreting* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

GraphLab Inc. (2014). GraphLab Create (Version 1.2) [Computer software]. Retrieved from http://graphlab.com/index.html

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences, 101*, 5228-5235. doi:10.1073/pnas.0307752101

Grimmer, J. (2010). A Bayesian Hierarchical Topic Model for political texts: Measuring expressed agendas in senate press releases. *Political Analysis, 18*(1), 1-35. doi:10.1093/pan/mpp034

Guo, L. (2011). *An analysis of the word order pattern in the SI target language and its underlying reasons in the language combination of English and Chinese* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Hall, D., Jurafsky, D., & Manning, C. D. (2008). Studying the history of ideas using topic models. *Proceedings of the conference on empirical methods in natural language processing* (pp. 363-371).

Hu, L. (2008). *Note-taking in Consecutive Interpreting: A psycholinguistic perspective* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Huang, Y. (2013). *Sentence memory and deverbalization: An experimental study of recognition memory of sentence surface structure of professional interpreters in English-Chinese simultaneous interpretation* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Jia, D. (2010). *A reflective approach to teacher development in interpretation classroom* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Kang, Z. (2011). *A study of AA mode and IA mode in interpretation from the perspective of cognitive psychology* (Unpublished doctoral dissertation). Fudan University.

Kushkowski, J. D., Parsons, K. A., & Wiese, W. H. (2003, January 1). *Master's and doctoral thesis citations: Analysis and trends of a longitudinal study*. Ames: Iowa State University.

Li, X. (2007). A report on the thematic trends of Chinese interpreting research (In Chinese). *New West, 2*, 206-208.

Liu, M. (2001). *Expertise in simultaneous interpreting: A working memory analysis* (Unpublished doctoral dissertation). University of Texas.

Liu, M. (2010). *A research on a computer assisted interpreter self-training* (In Chinese) (Unpublished doctoral dissertation). South China Normal University.

Liu, Q. (2011). *Communication strategies in interpretation from cross-disciplinary perspectives* (Unpublished doctoral dissertation). Shanghai International Studies University.

Liu, S., & Wang, L. (2007). A survey of the Chinese interpreting research over the past decade (In Chinese). *Journal of Guangdong University of Foreign Studies, 1*, 37-40.

Lovitts, B. E. (2001). *Leaving the ivory tower: The causes and consequences of departure from doctoral study*. Lanham, MD: Rowman & Littlefield.

Moser-Mercer, B. (1997). The expert-novice paradigm in interpreting research. In E. Fleischmann, W. Kutz, & P. A. Schmitt (Eds.), *Translationsdidaktik:*

Grundfragen der Übersetzungswissenschaft (pp. 255-261). Tübingen: Gunter Narr.

Mu, L., & Zou, B. (2014). A bibliometric analysis of two decades of China's production of doctorate dissertations on translation: What the findings say about the development of TS in the country (In Chinese). *Chinese Translators Journal, 2*, 14-20.

Paul, M. J., & Dredze, M. (2011). You are what you tweet: Analyzing Twitter for public health. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 265-272).

Peng, K. (2006). *The development of coherence and quality of performance in conference interpreter training* (Unpublished doctoral dissertation). University of Leeds.

Pöchhacker, F. (2004). *Introducing interpreting studies*. London: Routledge.

Ren, W. (2008). *A study of liaison interpreters' subjective consciousness* (In Chinese) (Unpublished doctoral dissertation). Sichuan University.

Seleskovitch, D. (1978). *Interpreting for international conferences: Problems of language and communication*. Washington: Pen and Booth.

Setton, R., & Guo, L. (2009). Attitudes to role, status and professional identity in interpreters and translators with Chinese in Shanghai and Taipei. *Translation and Interpreting Studies, 4*(2), 210-238. http://dx.doi.org/10.1075/tis.4.2.05set

Setton, R. (1999). *Simultaneous interpretation: A cognitive-pragmatic analysis* (Doctoral dissertation, Chinese University of Hong Kong, 1997) (pp. 1-397). Amsterdam: John Benjamins. doi:10.1075/btl.28

Sha, L. (2004). *Research on interpreting theories* (In Chinese) (Unpublished doctoral dissertation). Beijing Normal University.

SISU. (2014, August). Graduation ceremony for the 2014 Shanghai summer school for Translation Studies (In Chinese). Retrieved September 5, 2014, from http://www.giit.shisu.edu.cn/xueyuandongtai/note74.htm

Sowell, R. (2008, March 31). *Ph.D. completion and attrition: Analysis of baseline data*. Lecture presented at NSF Workshop: A fresh look at Ph.D. education. Retrieved August 13, 2014, from www.phdcompletion.org/resources/CGSNSF2008_Sowell.pdf

Su, W. (2011). *Interpreting competence in the apprentice stage* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Sun, H. (2012). *Influence of density of special knowledge in the speech content on interpreter's deverbalization performance* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Sun, X. (2010). *The relation between linguistic competence and interpreting competence during information processing in consecutive interpreting* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Tzou, Y. (2008). *The roles of working memory, language proficiency, and training in simultaneous interpretation performance: Evidence from Chinese-English bilinguals* (Unpublished doctoral dissertation). University of Texas. Retrieved September 5, 2014, from http://hdl.handle.net/1969.1/ETD-TAMU-2943

Wan, H. (2006). *A cognitive study of Sight Translation* (Unpublished doctoral dissertation). Shanghai International Studies University.

Wang, Y. (2008). *A corpus-based study on the compression strategy in Chinese (Cantonese) - English simultaneous interpreting* (Unpublished doctoral dissertation). Chinese University of Hong Kong.

Warren, P. (2012). *Introducing psycholinguistics* (1st ed.). Cambridge University Press.

Xu, Q. (2012). *The relationship between Simultaneous Interpreting and working memory in the case of concurrent listening and speaking* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Xu, Z. (2014). Chinese Interpreting Studies: Genesis of a discipline. *Forum: International Journal of Interpretation and Translation, 12*(2), 159-190. Retrieved from http://interpretrainer.com/Final_for_FORUM_October_2014.pdf.

Xu, Z. (2015). The past, present and future of Chinese MA theses in Interpreting Studies: A scientometric survey. *Perspectives*. (Forthcoming) doi:10.1080/0907676X.2015.1011175

Yang, L. (2012). *Fluency development in English production of Chinese consecutive interpreting learners* (Unpublished doctoral dissertation). Shanghai International Studies University.

Yu, W. (2012). *Self-monitoring and self-repair patterns in consecutive interpreting* (Unpublished doctoral dissertation). Shanghai International Studies University.

Zhan, C., & Suo, R. (2012). Telephone interpreting for the Guangzhou Asian Games (In Chinese). *Chinese Translators Journal, 1*, 107-110.

Zhan, C. (2011). *The interpreter's role as mediator in political settings: A corpus-based analysis of shifts in interpretation* (Unpublished doctoral dissertation). Guangdong University of Foreign Studies.

Zhang, J. (2008). *The Interpretive Theory of Translation in the context of modern Interpreting Studies* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

Zhang, Q. (2010). *Mechanisms of deverbalization in consecutive interpreting* (Unpublished doctoral dissertation). Guangdong Foreign Studies University.

Zhang, W. (2011). *Cognitive research into interpreting: Relationship between simultaneous interpreting and working memory* (Doctoral dissertation, Beijing Foreign Studies University, 2007) (pp. 1-357). Beijing: Foreign Language Teaching and Research Press.

Zhang, W. (2012). Interdisciplinary approaches to interpreting studies: Puzzles and solutions (In Chinese). *Chinese Translators Journal, 3*, 13-19.

Zhang, Y., & Qin, Z. (2010). A topic model of observing Chinese characters. *Intelligent Human-Machine Systems and Cybernetics, 2*, 7-10.

Zhao, Q., Qin, Z., & Wen, T. (2011). Topic modeling of Chinese language using character-word relations. *Neural Information Processing, 7064*, 139-147.

Zhu, J. (2010). *Tracking and comparison of interpreting competence and foreign language competence in conference interpreter training* (In Chinese) (Unpublished doctoral dissertation). Shanghai International Studies University.

XU, Ziyun

Ziyun Xu is a doctoral researcher at the Intercultural Studies Group of the Universitat Rovira i Virgili. He currently works as Chief Interpreter for the US-China Exchange Council in the United States. In this role he interprets for Chinese and American political leaders, business people and

academics, supervises and trains a roster of interpreters and translators, and helps to develop executive training programs in collaboration with Stanford University and the University of California, Berkeley. He also works as a business consultant, facilitating complex negotiations for Chinese start-up companies interested in acquiring clean technologies from US businesses.

xuziyun@gmail.com