A peer-reviewed version of this preprint was published in PeerJ on 17 September 2015.

View the peer-reviewed version (peerj.com/articles/1249), which is the preferred citable publication unless you specifically need to cite this preprint.

Xu Z, Pekelis L. 2015. Chinese Interpreting Studies: a data-driven analysis of a dynamic field of enquiry. PeerJ 3:e1249
https://doi.org/10.7717/peerj.1249

Chinese Interpreting Studies

A data-driven analysis of a dynamic field of enquiry

Ziyun Xu

Universitat Rovira i Virgili

xuziyun@gmail.com

Leonid Pekelis

Stanford University

lpekelis@gmail.com

Abstract:

Over the five decades since its beginnings, Chinese Interpreting Studies (CIS) has evolved into a dynamic field of academic enquiry with more than 3,500 scholars and 4,200 publications. Using quantitative and qualitative analysis, this scientometric study delves deep into CIS citation data to examine some of the noteworthy trends and patterns of behavior in the field: how can the field's progress be quantified by means of citation analysis? Do its authors tend repeatedly to cite 'classic' papers or are they more drawn to their colleagues' latest research? What different effects does the choice of empirical vs. theoretical research have on the use of citations in the various research brackets? The findings show that the field is steadily moving forward with new papers continuously being cited, although a number of influential papers stand out, having received a stream of citations in all the years examined. CIS scholars also have a tendency to cite much older English than Chinese publications across all document types, and empirical research has the greatest influence on the citation behavior of doctoral scholars, while theoretical studies have the largest impact on that of article authors. The goal of this study is to demonstrate the merits of blending quantitative and qualitative analyses to uncover hidden trends.

Keywords: scientometrics, Chinese Interpreting Studies, citation analysis, statistical modeling

PeerJ PrePrints

Biosketch

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.

Leonid Pekelis is a doctoral researcher in Statistics at Stanford University. He also works as a statistician for Optimizely. His research work focuses on correlations in problems of large scale inference. He is also passionate about bringing powerful statistical theories to everyday problems in data analysis.

PeerJ PrePrints

PeerJ PrePrints

1. Introduction

There are various channels through which scholars communicate with one another, easing the flow of knowledge and furthering the advance of science. One such important channel comes in the form of citations, which are the result of the duty incumbent upon all scholars to conduct comprehensive and critical reviews of existing literature before embarking on new research, to gain a deep understanding of the field and find the precise empty niche into which their own work will fit, referring to previous related work to bolster their arguments. Though citing other people's work did not become the norm in scientific writing until the early 1900s (Garfield 1979), it is now standard and required practice for authors to acknowledge the works of predecessors from which they have drawn inspiration, thereby maintaining the 'intellectual lineage' from one generation of academics to the next. Citation analysis has long attracted attention in the scientific community (see for example Garfield 1972; White & Mccain 1998; Baumgartner & Pieters 2003; Vallmitjana & Sabaté 2008). This is mostly as a consequence of Kuhn's 1970 ground-breaking work on the nature of science, in which he called on future scholars to recognize the crucial importance of adopting an empirical approach to studying the structure of the scientific community.

Such academic pursuits are particularly relevant in the Translation and Interpreting Studies (TIS) community, because it has experienced a significant growth in both quantitative and qualitative terms over the past two decades, and because hundreds of papers with diverse research methodologies and themes are produced on a yearly basis (Franco Aixelá 2013). During this period of significant growth, more empirical studies are needed if we are to fully appreciate the patterns of communication and trends in TIS. A number of earlier scholars have used citation data to trace the evolution of the field and

understand how scholars communicate with each other (see for example, Pöchhacker 1995; Gile 2005; Grbić & Pöllabauer 2009). However, despite its usefulness, there are limitations to a purely quantitative approach in analyzing TIS citation data, and qualitative analysis is called for in order to obtain a fuller picture of the discipline (Gile 2000). The purpose of this scientometric study is to marry quantitative and qualitative approaches to analyzing citation in order to obtain a panorama of CIS' evolution and reveal its hidden trends and predominant theoretical influences.

PeerJ PrePrints

2. Background

2.1 Major Questions

CIS has been developing rapidly since the 1990s, as evidenced by its increasing number of publications and researchers (Chen 2009). Using an all-but-exhaustive collection of citation data, three component strands of CIS (journal articles, MA theses, and doctoral dissertations) were studied with the aim of finding changes or differences in patterns of citation. In what ways is the citation network changing? Is the rate of change accelerating? Is it changing in new ways? Are authors still primarily influenced by older works or do more recent ones now hold the ascendancy? How do different research methods (theoretical, empirical, etc.) affect the use of citations in the works themselves? The three bodies of literature are generally produced by three distinct groups of authors: established researchers for journal articles and conference proceedings; graduate students for MA theses; and PhD students for doctoral dissertations (Xu 2014 & 2015). Examining these three strands individually is necessary if we are to fully understand how each contributes to advancing the field as a whole.

2.2 Literature Review

The study of research trends in Translation and Interpreting Studies (TIS) is currently dominated by citation analysis (see for example Gile 2005 & 2006; Gao 2008; Franco Aixelá 2004). There are various methods of carrying out citation analysis, but the overall

basic concept is always the same. First a sample of articles is selected; the researcher then counts the number of times each article is cited in other works. Citing (or 'source') works can be categorized according to type (conference proceedings, monographs, periodicals, etc.), and a weight assigned to each citation based on various factors: the type of publication in which it is being cited; the number of authors being cited; in the case of co-authorships an author's contribution to the work being cited (the 'target'); and others. Finally, a numerical score is calculated for each author, article, research institution, journal or whatever the researcher is focusing on; these scores can then be ranked to indicate each cited individual's or entity's relative impact (Lowry et al. 2007). The procedure is based on the premise that the number of times a work is cited is a measure of its influence in the academic world.

Citation analysis has increasingly been adopted to map out the historical evolution of a particular area of study, the impact of individual researchers, academic institutions or scientific publications, the extent of collaboration between these, or the influence of certain disciplines on others (Glänzel 2003; Kalaitzidakis et al. 2003). In their general study of the technique, Braun et al. (1985) found that articles cited between five and ten times each year during the period immediately following their publication tend to be assimilated into the relevant discipline's 'universal' stock of knowledge, and that conversely, if articles go uncited over the same period, there is little chance of such assimilation taking place. Citation analysis has been used in well-established disciplines such as linguistics (White 2004), psychology (Carr & Britton 2003; White & White 1977), and information science (White & McCain 1998), but has also been highly useful in assessing research patterns in fields with much shorter histories, such as TIS (Gile 2005).

Given the increasing popularity of citation analysis, Garfield's Institute of Scientific Information (ISI) produced the first citation index[1] for articles published in academic

[1] A citation index is a database that archives bibliographic information from publications: it allows users to trace the progress of a concept or subject of inquiry by sourcing published works that cite particular authors or articles.

PeerJ PrePrints

journals shortly after it was founded in 1960. The ISI has since produced numerous other indexes, which have grown to encompass more than 40 million records and 8,700 research journals (Meho 2006) and are now accessible online via Thomson Reuters' Web of Science. Although originally designed to facilitate access to information, the indexes are now widely recognized as an important source of empirical data for scientometric research (Ivancheva 2008).

Despite the growth in use of citation indexes, the exponential expansion of scientific research into new disciplines over the past four decades has resulted in numerous high-quality journals being excluded from the 'baskets' used by the leading indexes. To facilitate improved communication among researchers in the field of interpreting, in 1990 Daniel Gile set out to create an international network — the Conference Interpreting Research Information Network (CIRIN) — which publishes a biannual Bulletin. Since then several other searchable databases have been created for this discipline: the Bibliography of Interpreting and Translation (BITRA), for example, carries over 50,000 entries and is updated on a monthly basis, while the Translation Studies Bibliography (TSB) subscription service has 24,500 entries to date.

Gile (2005) surveyed citations from 47 papers on translator and interpreter training written by Western academics to find out which theories were most influential, the languages that target works were most often written in, and whether empirical or non-empirical research had more influence. The interpreter training material he sampled for the study revealed several interesting points: the model advocated by the Association Internationale des Interprètes de Conférence (AIIC) was the most frequently cited theory, while functional theories were dominant in translator training; the majority of the cited literature was written in English; and empirical research played very little part in the papers sampled. In another study (2006) he introduced a qualitative dimension to his analysis by grouping citations into different categories (concepts, methods, findings, etc.), on the assumption that such an approach would provide a more nuanced analysis of each category's impact on the evolution of Translation and Interpreting Studies (TIS). The study revealed that scholars were cited on their methods and findings in less than

PeerJ PrePrints

10% of the articles in the corpus. Adopting the same classification scheme, Nasr (2010) examined a corpus of 542 texts on translator training. Her study produced a similar result, indicating that empirical research was not influential in shaping research into that subject either.

By developing methodologies based on citation analysis, earlier researchers have laid the groundwork for assessing the impact of an individual's work and tracing the evolution of a field. In addition to quantitative analysis, qualitative approaches have been proposed to study how scholars cite one another. However, the application of these methodological techniques to investigating the evolution of CIS has to date been very limited. The goal of the present study was to adopt a blended approach with equal emphasis on both quantitative and qualitative considerations to explore how the CIS citation network changes over time and how different research methodologies have affected citation behaviors.

3. The Present Study

3.1 Research Questions

Expanding on the broad themes of enquiry outlined at the beginning of this paper, five more in-depth questions were drawn up to address some of the major issues unresolved by previous researchers. The rationale for each is summarized in the following section.

1. How do the average numbers of outgoing citations in Chinese papers, MA theses and doctoral dissertations change over time? How do the growing numbers of Chinese vs. English bibliographic references differ in the three categories? Are there any discernible differences between how English and Chinese citations are used in empirical and theoretical publications?

Compiling a bibliography at the end of a paper is essentially a Western scholarly tradition, but the inclusion of such information has, with the spread of intellectual

PeerJ PrePrints

enquiry, gradually come to be common practice in China since the mid-2000s. Citation habits in CIS have come a long way since the 1950s: the earliest paper in the data-set — Tang & Zhou 1958— had only two footnotes listing the sources the authors cited, but no bibliography at the end; by contrast, Huang's doctoral dissertation (2013) contained no fewer than 104 bibliographic references. While the number of references is not alone indicative of the quality of a paper, it does illustrate the author's breadth of knowledge of previous research conducted in the field (Gile 1999).

2. To what extent do the authors of the three different types of research cite recent and older work? Is recent work more commonly cited than older work, or vice versa?

Hargens (2000) observed that there are generally two major patterns for citation activities among researchers: some tend to follow the latest developments in research closely by citing more recent publications, whereas others focus on foundational work, disregarding the latest research. His study concluded that disciplines grouped within natural sciences fall into the first group, whereas social sciences and humanities fall into the second. For the present study the author set out to examine whether or not citation patterns in CIS tally with Hargens's observations.

3. How can the progress of CIS be quantified by means of citation analysis?

A number of scholars (Merton 1967; Lederberg 1972; Garfield 1977) have observed that at the same time as science constantly moves forward, there exists a phenomenon known as 'obliteration': the pace of scientific progress is so rapid, and new findings become so quickly and thoroughly absorbed into the 'general stock' of knowledge, that researchers are often unaware that they ought to be citing predecessors' work. For example, Darwin's 1859 Theory of Evolution which, as the name indicates, remains only a theory, is universally thought to be incontrovertible fact in the Western world: people referring to it nowadays may not even realize that they are actually citing one author's work.

PeerJ PrePrints

Franco Aixelá's study (2013) of the most cited works in Western Translation Studies (WTS) revealed a radically different situation: almost all the most frequently cited papers were 'classics' published well before the 2000s. His findings seemed to suggest that WTS scholars had a propensity for deepening and widening their understanding of the ages-old issues of translation and otherwise carrying on the intellectual lineage of 'classical' authors. The aim in this section was to discover whether or not the CIS community followed this academic tradition of WTS', and, more generally, to examine how the field's progress could be illustrated by means of citation analysis.

4. What are the most frequent citation types? Do they differ based on language of origin (Chinese vs. English) and document type (papers, MA theses and doctoral dissertations)?

While quantitative analysis of the academic influence of individual authors, institutions, geographical regions and publications may shed light on the entire CIS landscape, some qualitative analysis of citation types is necessary to provide insight into the interactions between the various schools of thought and research practice that constitute the field.

5. What different effects does the choice of empirical vs. theoretical research have on the use of citations in the three document types?

Citations illustrate a dynamic relationship between source and target authors; identifying whether a document is being cited for its methods, concepts or theories illustrates how researchers interact with and influence one another. As observed by Garfield (1979), a comprehensive survey of citation types could provide useful information on the structure and evolution of a science. When a source author is cited for his concepts, ideas or opinions, he is typically engaged in theoretical research, while citations of methodology and findings are typically taken from empirical research (Gile 2006b). An examination of the shares of citations that relate to empirical vs. theoretical research would shed light on the relative influences that the two methods have in the three categories of CIS publications.

PeerJ PrePrints

3.2 Data Organization

The author created a near-comprehensive database of 59,303 citations from the 1,289 Chinese MA theses, 32 doctoral dissertations and 2,909 research papers available to him. The CIS literature was collected from several sources: field trips to university libraries, interlibrary loans, book purchases, and academic databases such as CNKI, Wanfang and the National Digital Library of Theses and Dissertations in Taiwan. Publications with no bibliographic references were excluded from the analysis. Since these publications were obtained from multiple databases and different institutions, convenience sampling should not be an issue in this study: it would only be a problem if there were some inherent qualities among those one was able to sample that would not be present in the entire population. There is no good reason to believe that the publications found in the present sample would be different from ones found elsewhere. Once collected, the references were manually entered into a relational database which uses Structured Query Language (SQL) for managing data.

4. Methodology

4.1 How do the average numbers of outgoing citations in Chinese papers, MA theses and doctoral dissertations change over time? How do the growing numbers of Chinese vs. English bibliographic references differ in the three categories? Are there any discernible differences between how English and Chinese citations are used in empirical and theoretical publications?

For each category and individual year of CIS publications, the author divided the total number of outgoing citations by the total number of publications from which they came. This score was traced through time, allowing trends to be identified and differences according to document type or language to be detected if and when they occurred. Smoothing splines, determined by a penalty parameter of 0.6, were fit wherever necessary to facilitate the presentation of the data in a visual form and to trace the growth of outgoing citations; no splines were fit to the data for doctoral dissertations as the data-

PeerJ PrePrints

set for this group was too limited to fit reliable trends. To help in interpreting the data, the mean trend lines, with 95% confidence intervals, were fit for each document class and language type. Given that during the period under study Chinese scholars were becoming gradually more aware of the standards required for academic writing, including the importance of proper citations, it was reasonable to expect that the average number of outgoing citations in each category would have increased rapidly through the mid-2000s, but begun to stabilize from that point forward. Since a large proportion of Chinese researchers were not inclined to cite Chinese publications (Ren & Rousseau 2002), the average number of English references was expected to be significantly higher than that of Chinese.

A more nuanced analysis was conducted to determine whether there were any statistically significant differences in the average number of references between empirical and theoretical publications in the following categories:

(1) Chinese citations per paper in empirical vs. theoretical journal articles and conference proceedings
(2) English citations per paper *ibidem*
(3) Chinese citations per paper in empirical vs. theoretical MA theses
(4) English citations per paper *ibidem*
(5) Chinese citations per paper in empirical vs. theoretical doctoral dissertations; and
(6) English citations per paper *ibidem*

Welch two-sample two-sided t-tests were conducted to compare the number of Chinese and English references in articles and theses, but given the limited number of doctoral dissertations produced, Wilcoxon rank sum tests were chosen for this document type. The average numbers of both Chinese and English references in theoretical publications were expected to be higher in statistically significant terms than those in empirical ones. This predicted outcome was based on the differences in how these two different types of research are generally conducted: theoretical research is about developing new ideas through lengthy discussion of existing ones, whereas empirical studies deal in the

PeerJ PrePrints

collection and analysis of new data to confirm or refute them — it seems reasonable to hypothesize that there is more citation in the former than the latter.

4.2 To what extent do the authors of the three different types of research cite recent and older work? Is recent work more commonly cited than older work, or vice versa? And if there is such a tendency, does it depend on the language of origin (Chinese vs. English)?

The number of citations in any given publication year was counted and a plot with a smoothing spline generated for each set of data:

(1) All English citations vs. all Chinese ones;
(2) All Chinese citations split by category: papers, theses and dissertations;
(3) All English citations split by the same categories as in (2);
(4) English vs. Chinese citations in Chinese MA theses;
(5) English vs. Chinese citations in Chinese papers; and
(6) English vs. Chinese citations in Chinese doctoral dissertations.

In addition, the analysis was scaled, meaning that these yearly citation totals were divided by the total number of citations across all years. This approach compensated for the differences in numbers of citations in different years and allowed genuine patterns to be distinguished from random noise. A set of older, seminal works were expected to be cited more frequently than contemporary publications, and English citations were expected to be older than their Chinese counterparts across all three publication categories.

4.3 Do CIS authors tend repeatedly to cite 'classic' papers, or are they more drawn to the latest research within the field? How can the progress of CIS be quantified by means of citation analysis?

PeerJ PrePrints

Two null hypotheses were tested: the first was that of 'research stagnation'[2] — this tests whether new papers are not constantly being cited; and the second was that of 'perfect research flow' — this tests whether the citation process is stationary[3].

4.3.1 The Hypothesis of Research Stagnation

Research stagnation occurs when articles published after a given year (*t*) suddenly cease completely to be cited. One scenario which can lead to this state of affairs is when articles published before year *t* are so influential that they 'drown out' all citations from ones published after it. This hypothesis is rejected if new papers are being constantly cited.

4.3.2 The Hypothesis of Perfect Research Flow

Perfect research flow occurs when the citation process is stationary. The following example illustrates a case of perfect research flow: for articles published in a given year *t*, let us suppose that no citations come from year *t-4* or earlier, and that most citations come from papers published in year *t-3*, with half as many for each successive year down to *t* itself. Perfect research flow comes about when this distribution of citations is true for all the years *t* examined in the study.

A typical scenario that would cause this hypothesis to be rejected would be if a few very influential ('classic') articles were published in a given year *t_0* and cited more than the average article, even ten years later: in this case the citation process would indeed not be

[2] 'Research stagnation' in this context is shorthand for the stagnation of the citation process, whereby new articles are not cited and therefore after a given year the distribution of citations falls to zero. As we shall see in the Results and Discussions section (5.3.1), this hypothesis was later rejected. Of course, numerous other factors need to be taken into consideration to determine whether or not a field of inquiry is moving forward. Unfortunately the analysis of these is outside the scope of the present study.

[3] A process is said to be stationary if its distribution remains unchanged over time. In the example given in section 4.3.2 the distribution of papers cited in year *t* is said to be stationary if its relation to the previous years (*t, t-1, t-2*, etc) does not depend on *t*.

PeerJ PrePrints

stationary, because in year *t_0+10* citations of this article published ten years previously would still be being produced! We therefore would not be dealing with a case of perfect research flow.

4.3.3 Hypothesis Testing

The aforementioned hypotheses concern the distribution of the citation process. To test them, all the papers published in year *t* and the years of all citations contained in those papers were identified. The distribution of papers cited in year *t* was estimated as the average number of citations per paper published in year *t* coming from each previous year: *t-1*, *t-2*, *t-3* and so on. The same methodology was applied to all publication years between 1990 and 2013. Once the distribution of cited papers for each year *t* was established, it was possible to test whether the figure was stagnant, and, by measuring how it changed from year to year, whether it was stationary.

These two hypotheses were tested by comparing the performance of a model corresponding to each to that of a third, namely a varying coefficient model (VCM). Please refer to the Appendix for a more detailed description of the VCM and the hypothesis testing methodology.

4.4 What are the most frequent citation types? Do they differ based on language of origin (Chinese vs. English) and document type (papers, MA theses and doctoral dissertations)?

The citations were labeled according to the way each cited paper was used by the referring paper. After an initial pilot study,[4] a citation classification system was developed to evaluate how authors were cited in the CIS literature (see Table 1).

[4] Building on the studies of Gile (2006) and Nasr (2010), the author conducted a pilot study of citation types in CIS by randomly sampling a total of 239 in-text citations from Chinese papers, theses and doctoral dissertations. After labeling these according to the methods described in Gile's pilot research project

PeerJ PrePrints

Table 1: Citation classification system

Citation types	Definitions
Prescriptive opinion	The cited author explicitly expresses his position on an issue or issues, and directs readers. Statements of this nature often rely on modals such as *ought*, *should* and *must*.
Non-prescriptive opinion (Claim)	The cited author expresses a personal view but *without* directing the reader.
Assessment	The cited author gives an evaluation of an issue or issues he deems important.
Concept	The cited author puts forward a detailed idea.
Rules/standards	The cited author talks about principles of conduct or codified regulations.
Theory/model	The cited author creates a group of propositions that are used to explain or predict certain phenomena.
Theoretical analysis	The cited author examines a phenomenon, concept or behavior in abstract terms, basing his reasoning on existing theoretical frameworks.
Idea	The cited author's thought is non-technical and lacks the detail found in 'concepts', 'theories' and 'theoretical analysis'.
Tangential Research	The cited author has made a detailed study of a particular subject in the hope of obtaining new information or deepening understanding.
Research method	The cited author adopted a particular approach to uncovering new information or advancing understanding.

(2006), a collaborator who was familiar with the topic was asked to give a 'second opinion' by labeling them again himself, with the aim of ensuring a greater measure of objectivity and reliability. These labelling activities were completed in four installments. The sequential analysis was restricted to 239 citations because by that point the author had sufficient knowledge of regularly occurring citation types in CIS, and the differences in labeling between the author and the collaborator were minimal.

PeerJ PrePrints

Research finding	The cited author draws factual or empirical findings from a study.
Non-TS factual citation	The cited author covers factual information outside the scope of Translation Studies.
TS-related factual citation	The cited author alludes to factual information that falls within the scope of Translation Studies.
General Principle	The cited author talks of fundamental 'truths' which fall short of being absolute.
General Report	The cited author observes and describes a phenomenon or behavior.
Qualification	The cited author sets parameters or defines the limitations of a phenomenon.
Definition	The cited author explains a phenomenon, behavior or concept in terms of its unique characteristics.
Characterization	The cited author describes the features of a particular issue or concept.
Textual	The citation is of an author's textbook or other instruction manual.
Non-theoretical analysis	The cited author makes a face-value examination of certain issues or phenomena without the use of theories.
Famous quotes	The cited author quotes a well-known person's words to emphasize a point.
Miscellaneous	Any citations that do not fit into the aforementioned categories.
Not available	A bibiliographic reference does not correspond to any of the in-text citations in a paper.

A random sampling without replacement was conducted on each of the six citation databases of CIS: English citations in MA theses, research papers, and doctoral dissertations; and Chinese citations *ibidem*. This form of sampling was used because it leads to more accurate results than sampling with replacement, thanks to an effect known in the simulation literature as 'variance reduction' (Rao 1963). The minimum sample size was fixed at one large enough for detecting any statistically significant difference in the

PeerJ PrePrints

proportions of each citation type between Chinese and English. Given that there was no prior knowledge about whether the number of Chinese citations would be larger than that of the English ones, or vice versa, a power calculation for a two-sided two-sample proportion test was conducted for each of the six citation databases in order to give a 95% chance of detecting a difference of 10%[5] between two proportions at the 0.05 significance level. The minimum number of samples required for each database (648 as determined by the power calculation) was collected, analyzed and labeled.

The majority of the in-text citations were expected to belong to a single citation type, but some belonged to more than one — this occurred primarily when an author cited a particular source at different places in his text. If one of these multi-category observations was selected for inclusion in the random sample, that citation needed to be assigned to a *single* category to maintain consistency. To address this, one of the assigned categories with equal probability was selected at random to decide which category the citation should have been assigned to. Though it may appear crude, assuming equally probable categories as a first approach yields good results, and is frequently used (Kempthorne 1952; Freedman 1997; Schulz & Grimes 2002) when there is no prior knowledge of the distribution of the data.

Working from the results of the power calculation, which gave an effective way of sampling; confidence intervals of 95% were constructed for the proportion of each citation type. It was expected that N/A would be the most frequently used citation type across all three categories of CIS publication in both Chinese and English. Citation practice in the Chinese academic community is distinctly different from that of the Western world: large numbers of scholars list in their bibliographies the literature they consult while conducting research, even if it is not directly cited in the body of their texts. However, according to most Western style guides, such as that of the American Psychological Association (APA), authors are required to cite the works of those who

[5] If the true difference in proportions were, for example, 1%, no difference would be detectable unless the sample sizes were at least 64,974! The entire population for the present study comes nowhere near this size. A 10% difference was chosen because it allowed the author to work with a reasonably sized sample.

PeerJ PrePrints

have directly influenced their research, and every resource in the bibliography must have a corresponding in-text citation, with the exception of some classics such as the Bible (APA 2010). Because of the lack of textual references to these N/A citations it is not easy to code them. There may be limits to the space available to authors for recording references, meaning that not all the works that have influenced them will make it to the final list, so it is reasonable to assume that they are painstaking in their choice of what to include. Those citations that are listed, including those in the N/A category, must have had a major influence on the author, be it factual, theoretical or inspirational. Because of the different impacts that Western and Chinese literatures have had on CIS, the proportions of N/A for English and Chinese citations were expected to differ — analyzing these proportional differences might help to illustrate the ways in which Eastern and Western thought have influenced CIS.

Another expected finding was that authors would be frequently cited for their Ideas and Prescriptive and Non-prescriptive Opinions; this might come as a surprise to those from disciplines where opinion-based citations are not common. These citations of Ideas and Opinions are prevalent because practicing interpreters are highly respected in the CIS community. This is corroborated by the fact that numerous professionals with no background in research are regularly invited to serve as keynote speakers at the biannual National Conference and International Forum on Interpreting, the most important research conference in China. Tangential Research was another citation type expected to be in frequent use, because scholars may often feel obligated to make 'ceremonial citations', i.e. referencing the leading experts in the field without actually having read their research (Meho 2006). At the other end of the spectrum, research methodology and findings were expected to be in much less frequent use, because non-empirical research still accounts for the overwhelming majority of published works in CIS.

4.5 What effect does the choice of empirical vs. theoretical research have on the use of citations in journal articles, MA theses and doctoral dissertations?

PeerJ PrePrints

The analysis for this research question proceeds from the assumption described in section 3.1 that certain citation types (Research Methodology and Finding) are typically associated with empirical research, while others (Concepts, Ideas and Opinions) are linked to theoretical studies.

It was expected that doctoral dissertations might contain a greater proportion of empirical citations than MA theses and research papers. Furthermore, it was predicted that citations relating to theoretical research would be most frequently found in papers, followed by theses and dissertations. These predicted outcomes were based on the fact that 80% of dissertations were empirical (compared with 50% of theses and 20% of papers), indicating that empirical methodologies were the preferred research approach among doctoral students, whereas paper authors, who were mostly established academics, preferred theoretical research.

First it was necessary to determine how many data points would be required to detect a 10% difference between two proportions at the 0.05 significance level. Given that there is reason to believe that the differences in the proportions of empirical and theoretical citations are directional, a power calculation for a one-sided two-sample proportion test was conducted to determine the minimum number of citations that would be needed as samples from each of the three categories of CIS publications to guarantee enough statistical power of the test. Since there was found to be no relationship between the hypotheses in Questions 4 and 5, re-using the same samples is not an issue. The minimum sample size for Question 4 was 648 for every possible combination of language and document type, whereas the same number for Question 5 was 755 for each category of document. The actual sample of 1,296 (combining the sampled Chinese and English citations in each document type) easily exceeded the necessary minimum for Question 5, further increasing the power without increasing a type one error. Once all the required citations were labeled, a two-proportion z-test was performed, yielding p-values. The z-test examined whether the proportions of citations associated with theoretical research were equal to those associated with empirical research.

PeerJ PrePrints

5. Results and discussions

5.1 How do the average numbers of outgoing citations in Chinese papers, MA theses and doctoral dissertations change over time? How do the growing numbers of Chinese vs. English bibliographic references differ in the three categories? Are there any discernible differences between how English and Chinese citations are used in empirical and theoretical publications?

Examination of the citation data for journal articles revealed that the average number of Chinese references grew from around three per paper in the early 1990s to six in 2005, stabilizing at that figure from then onward (see Figure 1). At the same time the average number of English citations rose from slightly above two in 1995 to just over four in 2012; the available data gives no sign of this trend slowing (see Figure 2).

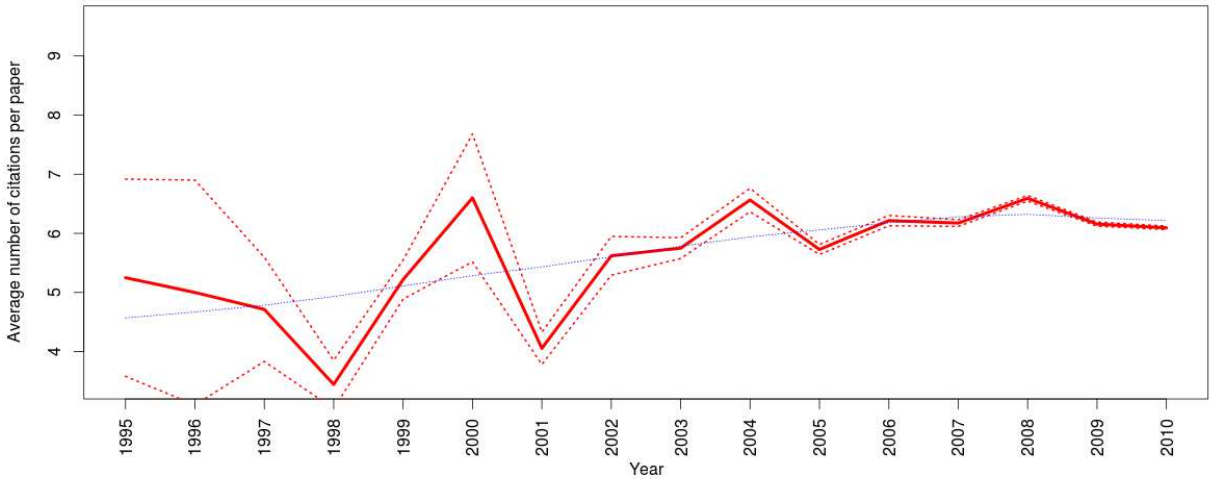


Figure 1: Average number of Chinese citations per paper ($R^2 = 0.4820229$)

PeerJ PrePrints

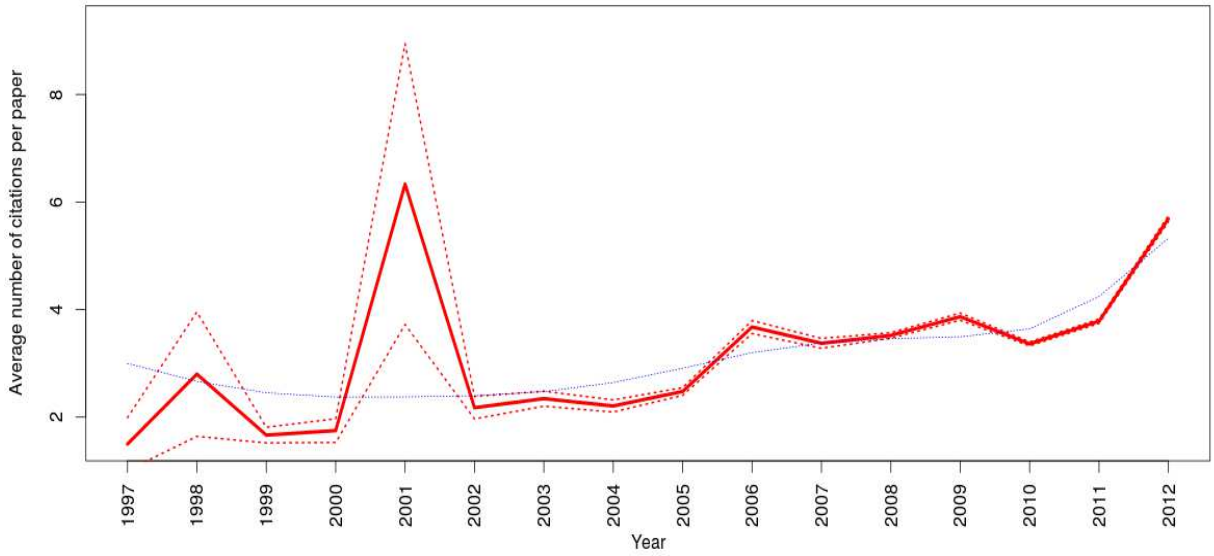


Figure 2: Average number of English citations per paper (R^2 = 0.2783813)

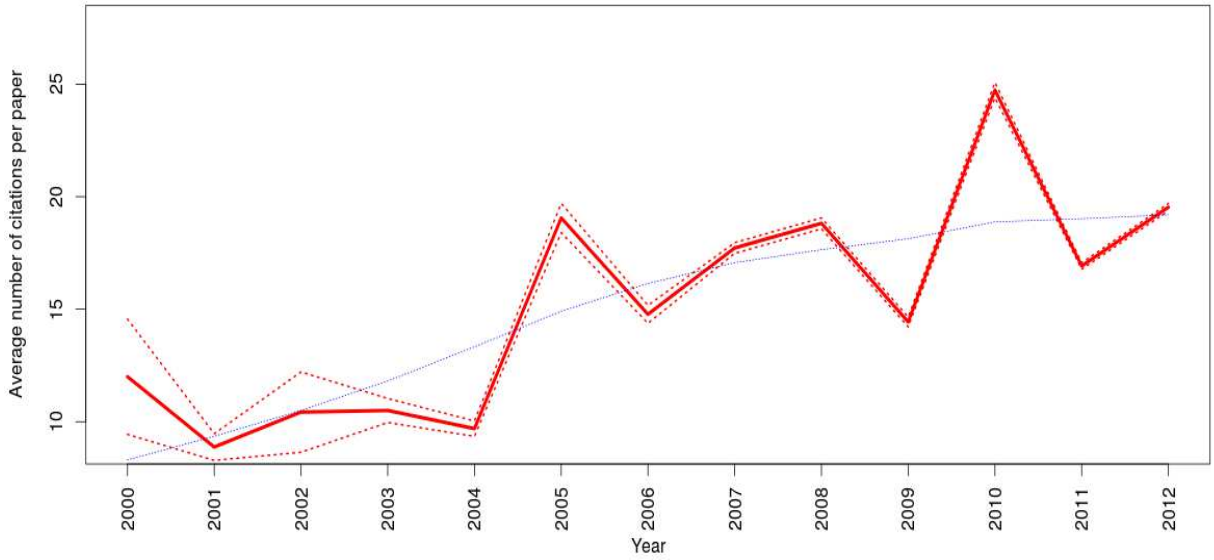


Figure 3: Average number of Chinese citations per thesis (R^2 =0.6259455)

PeerJ PrePrints

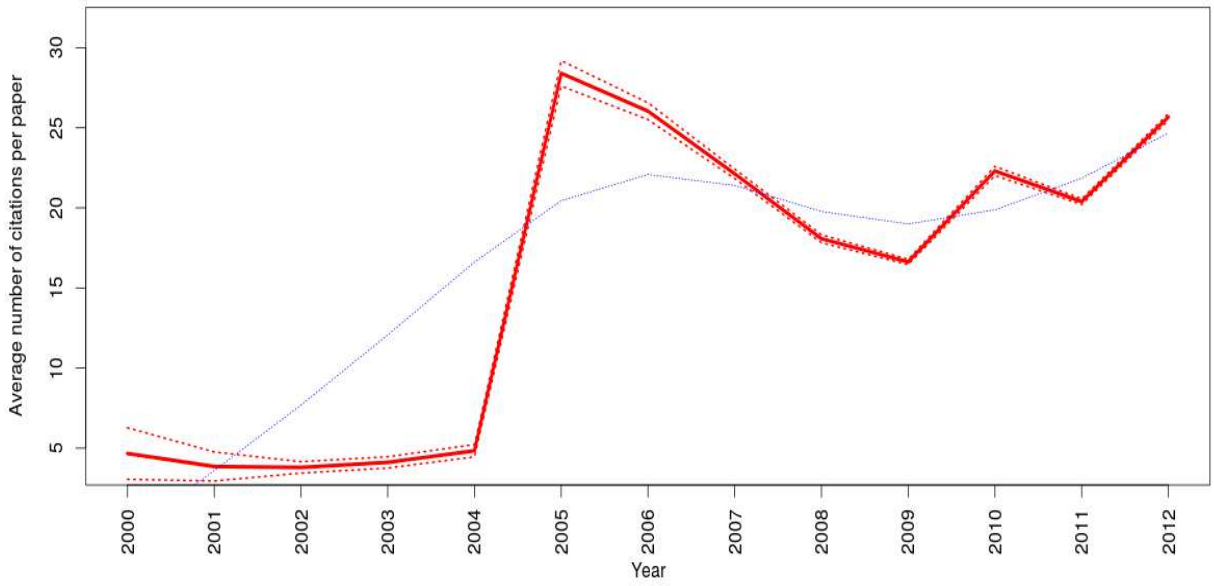


Figure 4: Average number of English citations per thesis ($R^2 = 0.7015041$)

When it came to MA theses, the number of Chinese citations per thesis rapidly increased from around 12 in 2000, peaked at 25 in 2010, and stabilized at around 20 in 2012 (see Figure 3). The overall trend for Chinese citations was upward, but the speed of the increase significantly slowed approaching the late 2000s. As the use of bibliographic references is regarded as indicative of a scholarly community's increasing maturity (Gile 2006a), the significant rise in the average number of Chinese citations in the 2000s would appear to indicate and reflect the ongoing development of MA-level CIS research: students seem to take their work more seriously. There is no fixed limit for how long – or short – a list of references can be, so the plateau reached in the early 2010s may suggest that 20 Chinese citations per thesis is the 'happy medium' number for an MA thesis.

At the same time, the number of English citations experienced a similar trend, though the increase was much more noticeable, from fewer than 5 citations per thesis in 2004 to nearly 30 in 2005 (see Figure 4). The English pattern was also somewhat different from that of the Chinese: from 2005 onward English citations per thesis experienced a downward trend, the figure dropping to as low as 18 in 2009, but there followed an upturn, the number bouncing back to around 25 in 2012.

PeerJ PrePrints

Prior to 2004, Beijing Foreign Studies University dominated the production of MA theses, which were written exclusively in Chinese, but starting from 2005 a number of universities, among them Shanghai International Studies University (SISU), Xiamen University and Guangdong Foreign Studies University, came onto the scene, requiring students to complete their theses in English. This change almost certainly explains the sudden and rapid increase in the number of English citations per thesis between 2004 and 2005. In the late 2000s numerous schools restructured their graduate programs, allowing their students to complete a master's in two rather than two and a half years. Given these six fewer months spent on researching and writing their theses coupled with increasing pressure on them to seek gainful employment, one can imagine students will have had less time to read, review and cite previous research, which may explain the decline in the number of English citations per thesis from 2005 onwards. It might further be speculated that the relatively static number of Chinese citations during the same period (2005-2012) is due to students' continuing to feel obligated to acknowledge the influence of their Chinese predecessors; a corroborating piece of evidence for this argument is that the number of Chinese citations labelled as Tangential Research in MA theses is greater than those of English ones in statistically significant terms (see section 5.4 for more detail). As is the case for Chinese citations, the data appear to suggest that most students settle on an average of 25 English citations per thesis.

PeerJ PrePrints

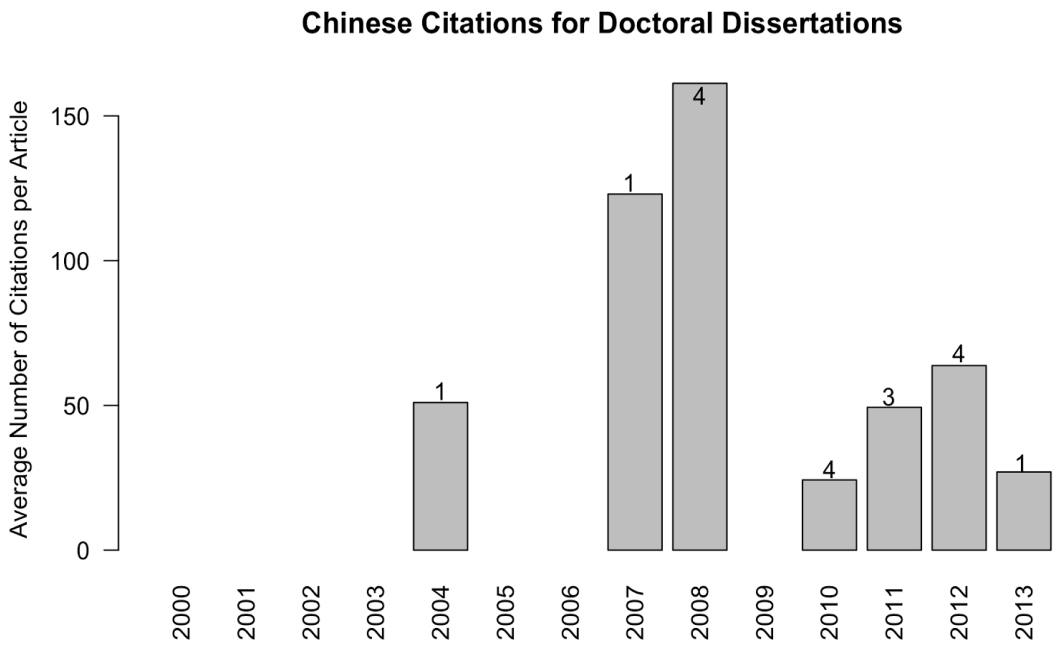


Figure 5: Average number of Chinese citations per dissertation[6]. The numbers on the top of each bar represent the number of dissertations produced in that year.

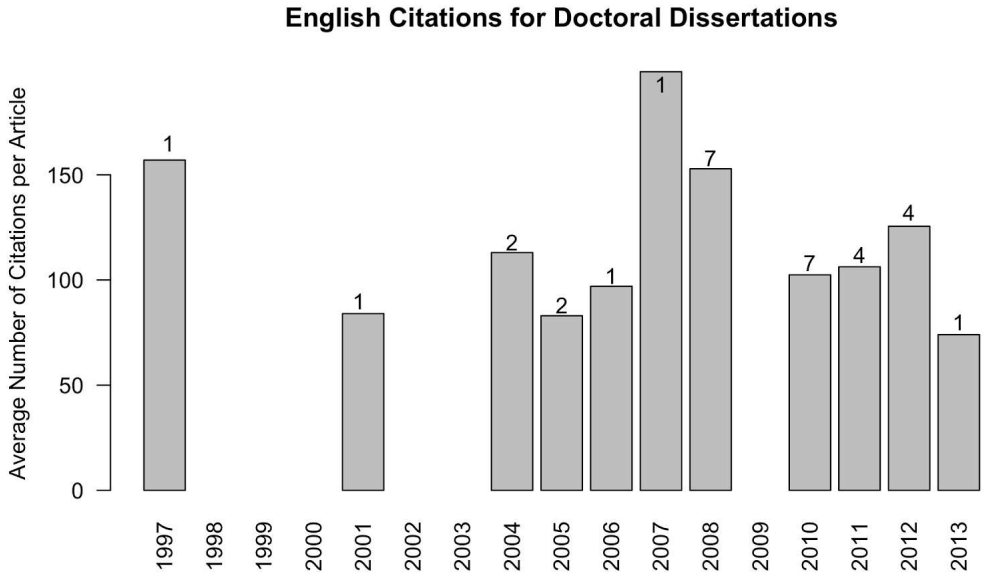


Figure 6: Average number of English citations per dissertation

[6]No smoothing splines were generated for the citations in doctoral dissertations because of their limited numbers.

PeerJ PrePrints

While no trends for citations from doctoral dissertations were observable because of the small sample size (only 32 dissertations were available compared with 2,909 research papers and 1,289 theses), it was apparent that on average the number of English citations per dissertation far exceeded those in Chinese (see Figures 5 and 6), suggesting that Western literature has a marked influence on Chinese doctoral research.

Given that the number of citations per paper was rather low, the research methodology employed in each type of document was examined further to see whether the number of citations differed between empirical and non-empirical papers.

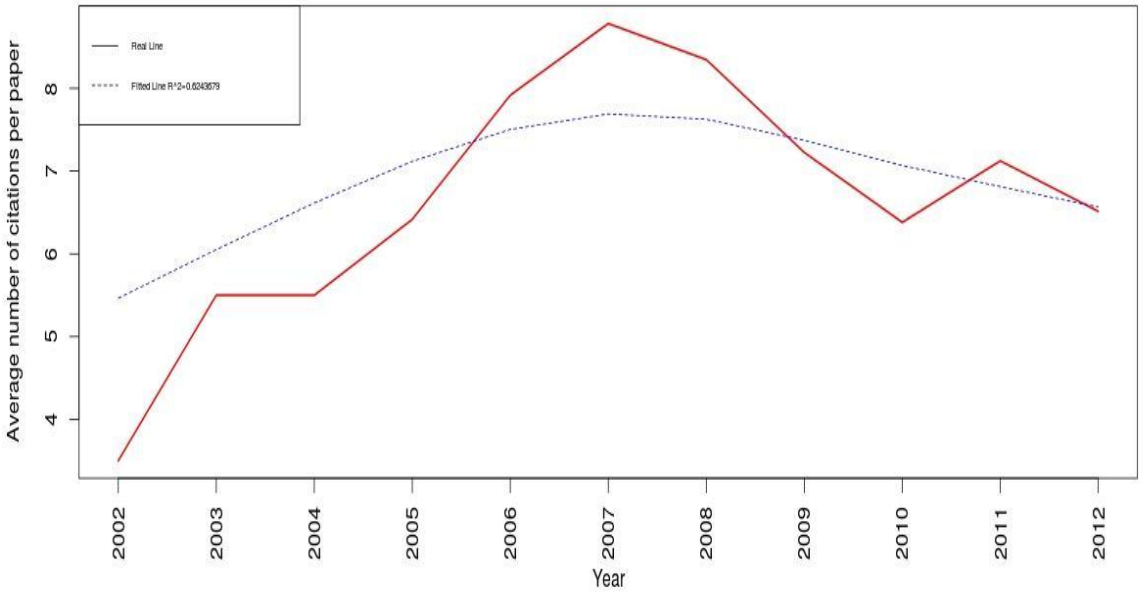


Figure 7: Average number of Chinese citations per paper for empirical journal articles and conference proceedings

PeerJ PrePrints

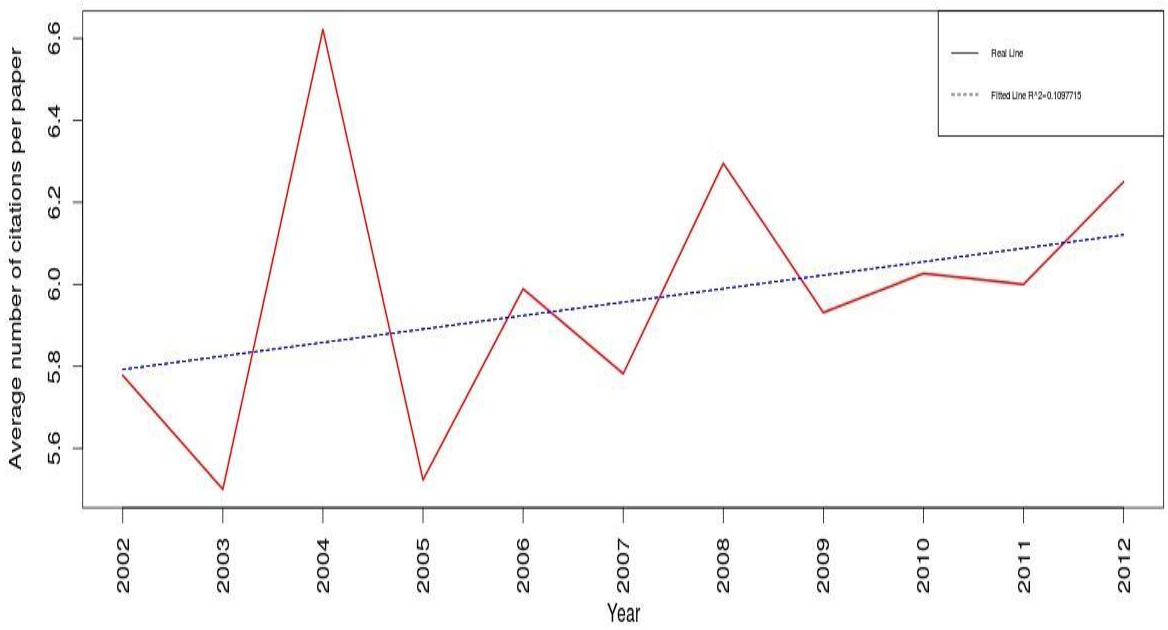


Figure 8: Average number of Chinese citations per paper for non-empirical journal articles and conference proceedings

It can be seen that the average number of Chinese citations for empirical papers increased significantly from two per paper in 2002 to a peak of ten in 2007, but subsequently experienced a slight downward trend to around six in 2012 (see Figure 7). A possible explanation for this stabilization is that journals typically set a word limit for the papers they publish, so the number of citations per paper cannot grow indefinitely. By contrast, the number of Chinese citations per non-empirical paper witnessed a slight upward trend with cyclical patterns from 5.8 to slightly over six in the same period (see Figure 8). To examine whether there was a statistically significant difference in the number of Chinese citations between empirical and non-empirical papers, a Welch two-sample two-sided t-test was conducted. This test was chosen because it performs better than a simple Student's t-test when the variances of the two samples are different, something which we might expect to happen given that the empirical and non-empirical samples were non-overlapping. The results revealed that there was a significant difference in the numbers of Chinese citations for empirical (Median = 7.076923, Standard Deviation = 5.602727) and non-empirical (M = 6.008817, SD = 4.310034) papers. The p-value was 0.002535<0.01; there is therefore strong evidence to suggest that

PeerJ PrePrints

the number of Chinese citations in empirical papers was significantly higher than in non-empirical ones.

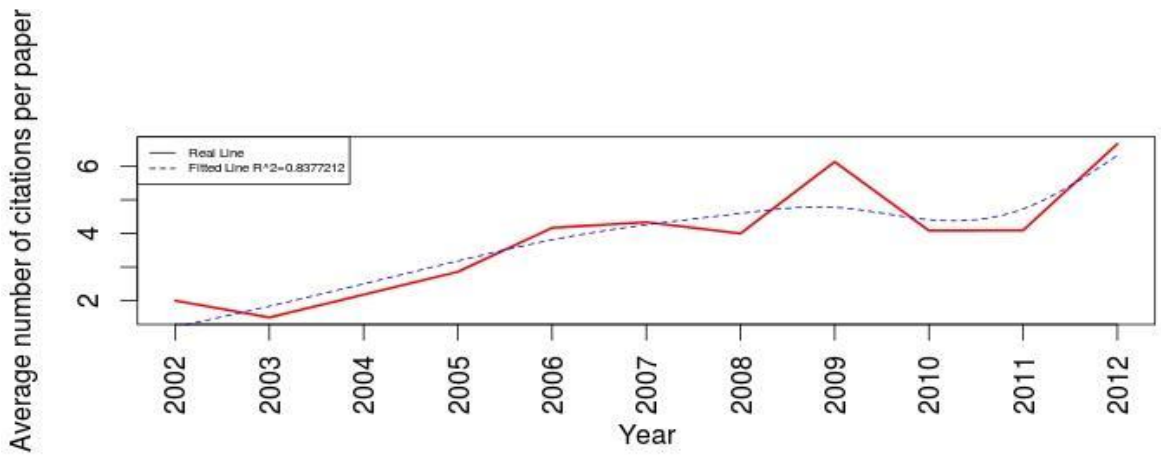


Figure 9: Average number of English citations per paper in empirical journal articles and conference proceedings

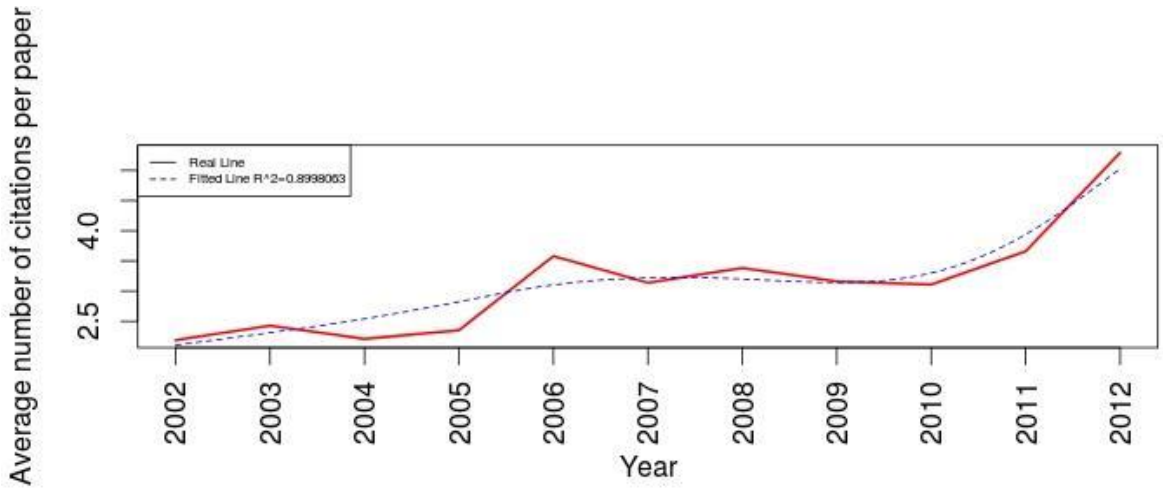


Figure 10: Average number of English citations per paper in non-empirical journal articles and conference proceedings

The data revealed that empirical papers experienced an increase in the number of English citations per paper from two in 2002 to six in 2012, and that there was no sign of the number stabilizing (see Figure 9). A similar trend was observed for the average per non-empirical paper (see Figure 10). To test whether the numbers were similar, a Welch

PeerJ PrePrints

two-sample two-sided t-test was conducted. The results revealed that there was very strong evidence of a significant difference in the numbers–for empirical (M = 4.903346, SD = 5.334164) and non-empirical (M = 3.585233, SD = 4.664803) papers. The p-value was 0.0002819<0,001, giving very strong grounds to believe that there was a statistically significant difference between English citations in empirical and non-empirical journal articles. The fact that citations in empirical papers outnumbered those in non-empirical ones is statistically in line with the earlier finding regarding Chinese citations in empirical vs. non-empirical papers. Irrespective of the language in which the cited document was written, authors of empirical papers used more citations on average than those of non-empirical ones. It was interesting to see that not only were the two findings consistent, but the differences they pointed to were of the same magnitude: the mean number of citations was 27% lower in non-empirical than empirical papers in Chinese, and 15% lower in English.

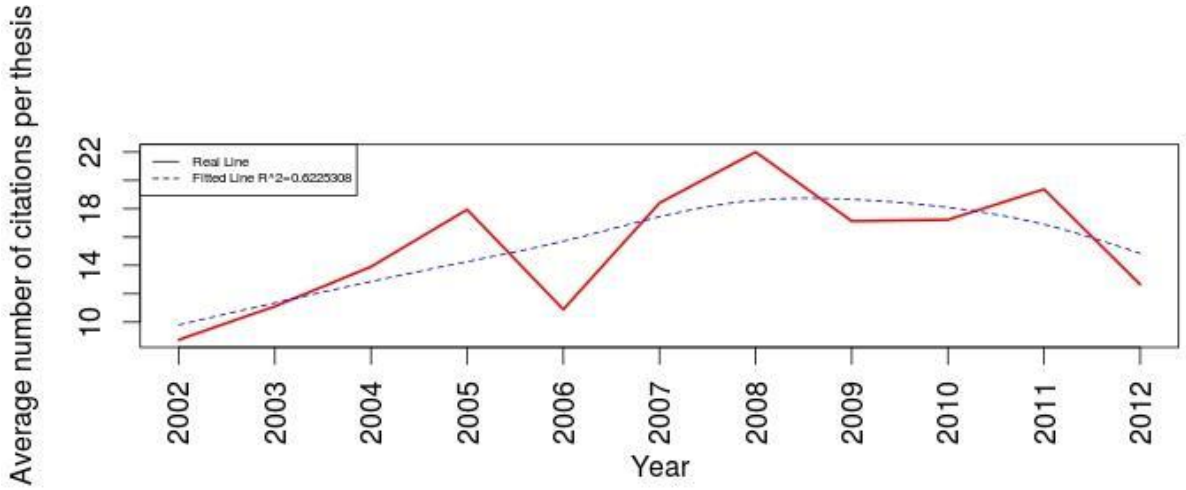


Figure 11: Average number of Chinese citations per empirical thesis

PeerJ PrePrints

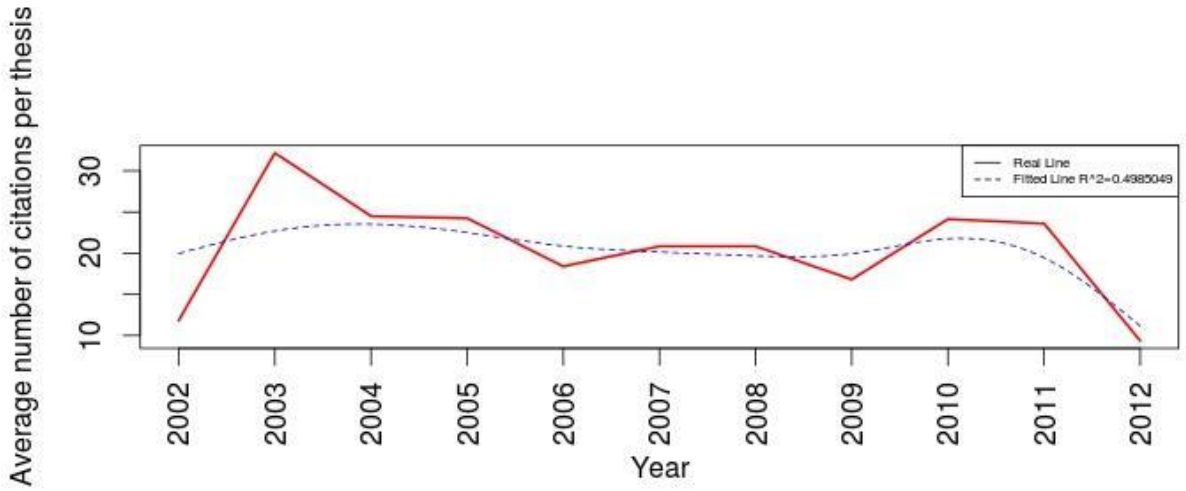


Figure 12: Average number of Chinese citations per non-empirical thesis

A similar approach was used to examine the trends for Chinese citations in empirical vs. non-empirical theses. It was observed that the number in the empirical type increased significantly from three in 2002 to 21 in 2008, and from that point forward there was a slight downward trend to around 12 per thesis in 2012 (see Figure 11). When it came to non-empirical theses, the number experienced an upward trend from 12 in 2002 to 32 in 2003, but the average has been declining ever since (see Figure 12). To examine whether there was a statistically significant difference between empirical and non-empirical theses, a Welch two-sample two-sided t-test was also conducted: there was a significant difference in the numbers of Chinese citations for empirical (M = 16.82697, SD = 12.14141) and non-empirical (M = 18.75378, SD = 13.00549) theses; $t(846) = -2.2392$, $p = 0.0254$. This finding suggests that the number of Chinese citations in empirical theses was statistically lower than that in non-empirical ones.

PeerJ PrePrints

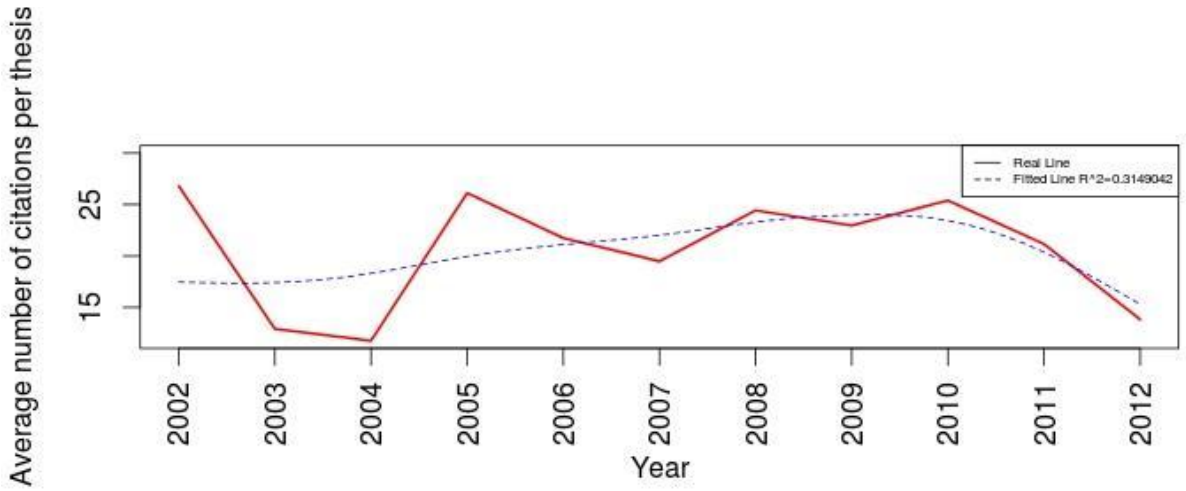


Figure 13: Average number of English citations per empirical thesis

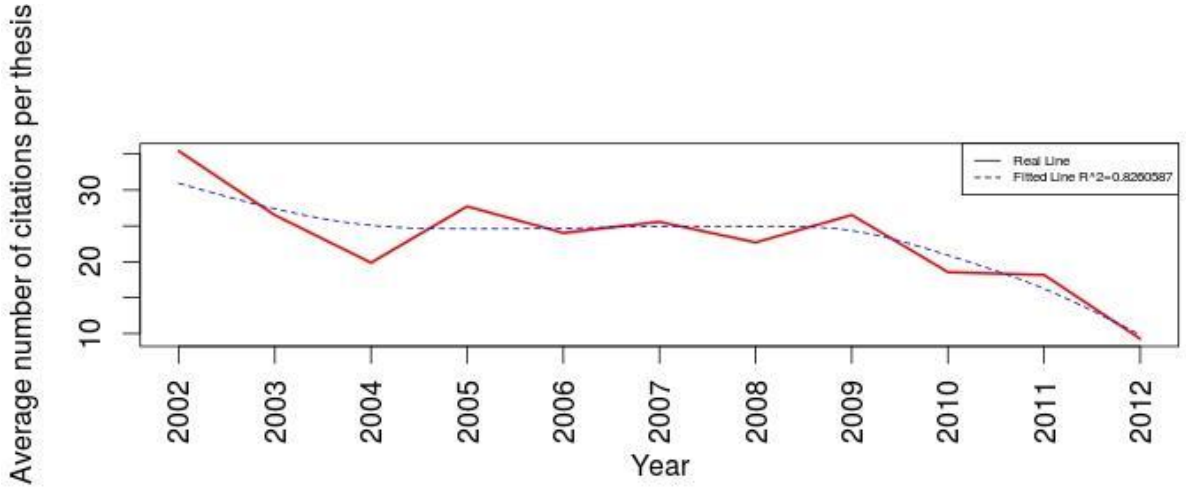


Figure 14: Average number of English citations per non-empirical thesis

As can be seen in Figure 13, despite the initial fluctuation in numbers between 2002 and 2005, the average number of English citations per empirical thesis remained stable at around 20 until 2011. Starting from 2012, the average dropped rather sharply to slightly less than 15 per thesis. In the case of English citations per non-empirical thesis (see Figure 14), the average number experienced a significant decline from 35 in 2002 to 20 in 2004, and from 2009 to 2012 there was another noticeable decline of nearly 20 per thesis. A Welch two-sample two-sided t-test indicated that there was no difference in the number of English citations between empirical (M=21.21875, SD=17.98875) and non-

PeerJ PrePrints

empirical (M=20.38214, SD=16.17692) theses, because the p-value was 0.4338. Earlier examination of journal articles showed that more citations — both Chinese and English — were used in empirical than non-empirical papers. The data for theses, however, was insufficient to permit confirmation that the same applied there.

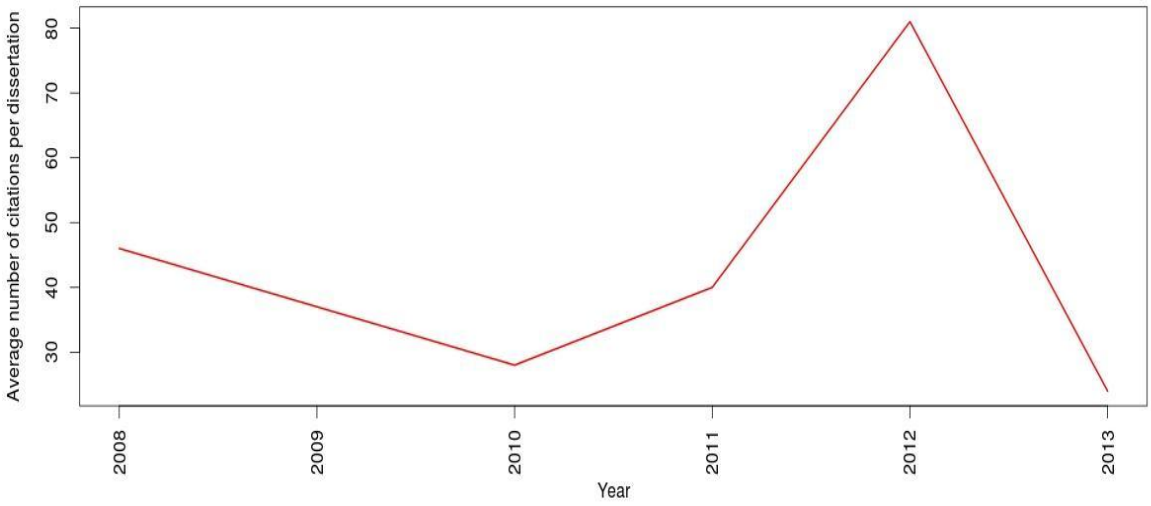


Figure 15: Average number of Chinese citations per empirical dissertation

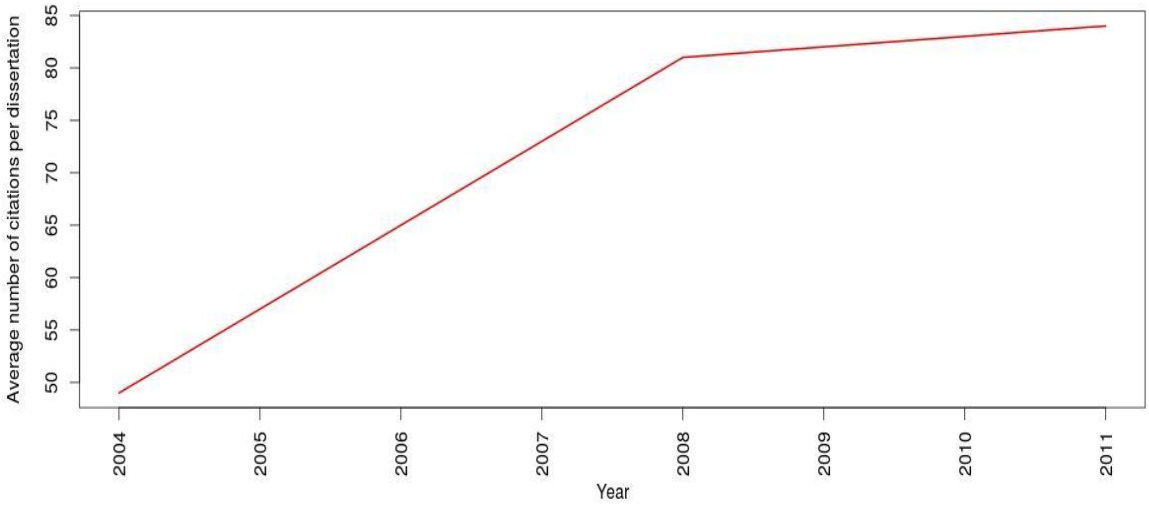


Figure 16: Average number of Chinese citations per non-empirical dissertation

Because of the limited availability of data, no smoothing splines were generated to analyze the trends for the number of Chinese citations in doctoral theses. The average

PeerJ PrePrints

number in empirical dissertations varied greatly from year to year (see Figure 15). However, it was immediately apparent that there was a gradual increase in the number of Chinese citations in non-empirical, from fewer than 40 per paper in 2004 to around 80 in 2008, and that the average seemed to stabilize at around that number from there through to 2011 (see Figure 16). To determine whether empirical and non-empirical dissertations differed in the number of Chinese citations they contained, a Wilcoxon rank sum test was conducted. This particular test was chosen because it is distribution-free: it can be used when the data set is small and no distribution assumption can be made (Rice 1995). This test outperforms the t-test in non-normal distributions[7]. The p value was 0.09902, indicating that there was no evidence of any significant difference between the two types.

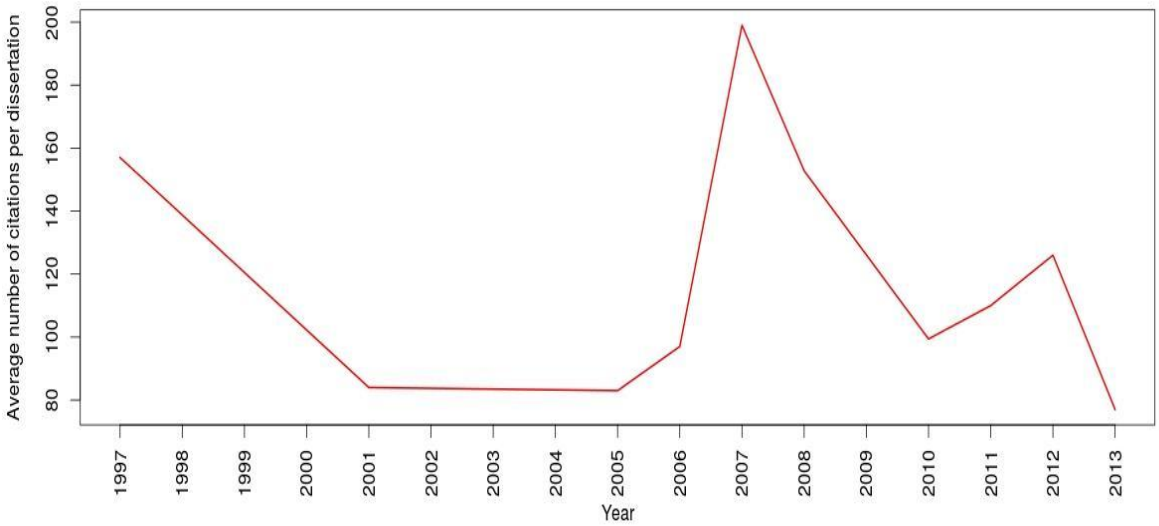


Figure 17: Average number of English citations per empirical dissertation

[7] For sufficiently large data sets (e.g. those used for the paper and thesis data), one could use a test which assumed a normal distribution. But the limited size of the present dissertation sample did not allow that assumption to be maintained, hence the switch to Wilcoxon rank sum.

PeerJ PrePrints

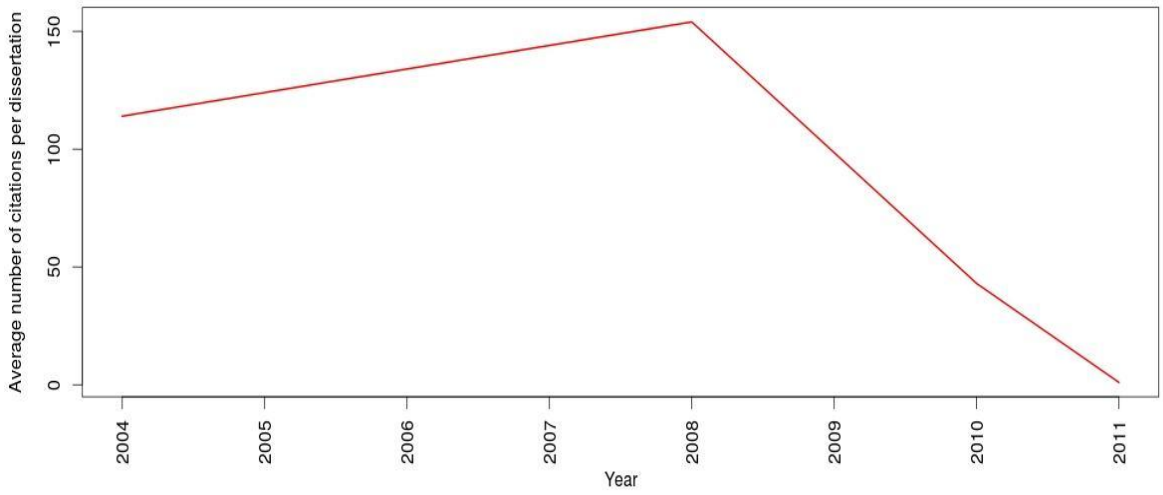


Figure 18: Average number of English citations per non-empirical dissertation

An examination was also made of the trends in the average number of English citations per empirical vs. non-empirical dissertation. While the average for the former hovered just above 80 from 1997 to 2013 (see Figure 17), the figure for the latter-increased from around 120 in 2004 to nearly 150 in 2008 and then declined sharply to around ten in 2011 (see Figure 18). A Wilcoxon rank sum test was also conducted to examine whether there were any differences in the average number of English citations between the two types of dissertation. The result revealed that the p-value was 0.3467, suggesting that there was no evidence of a significant difference. The size of the current data-set does not permit any definite conclusions to be drawn about the difference, but as CIS doctoral research continues to grow, some differences may become observable in the future. In sum, at the time of writing (January, 2015) doctoral students conducting empirical and non-empirical research exhibited no difference when it came to citing English or Chinese works.

PeerJ PrePrints

5.2 To what extent do the authors of the three different types of research cite recent and older work? Is recent work more commonly cited than older work, or vice versa? And if there is such a tendency, does it depend on the language of origin (Chinese vs. English)?

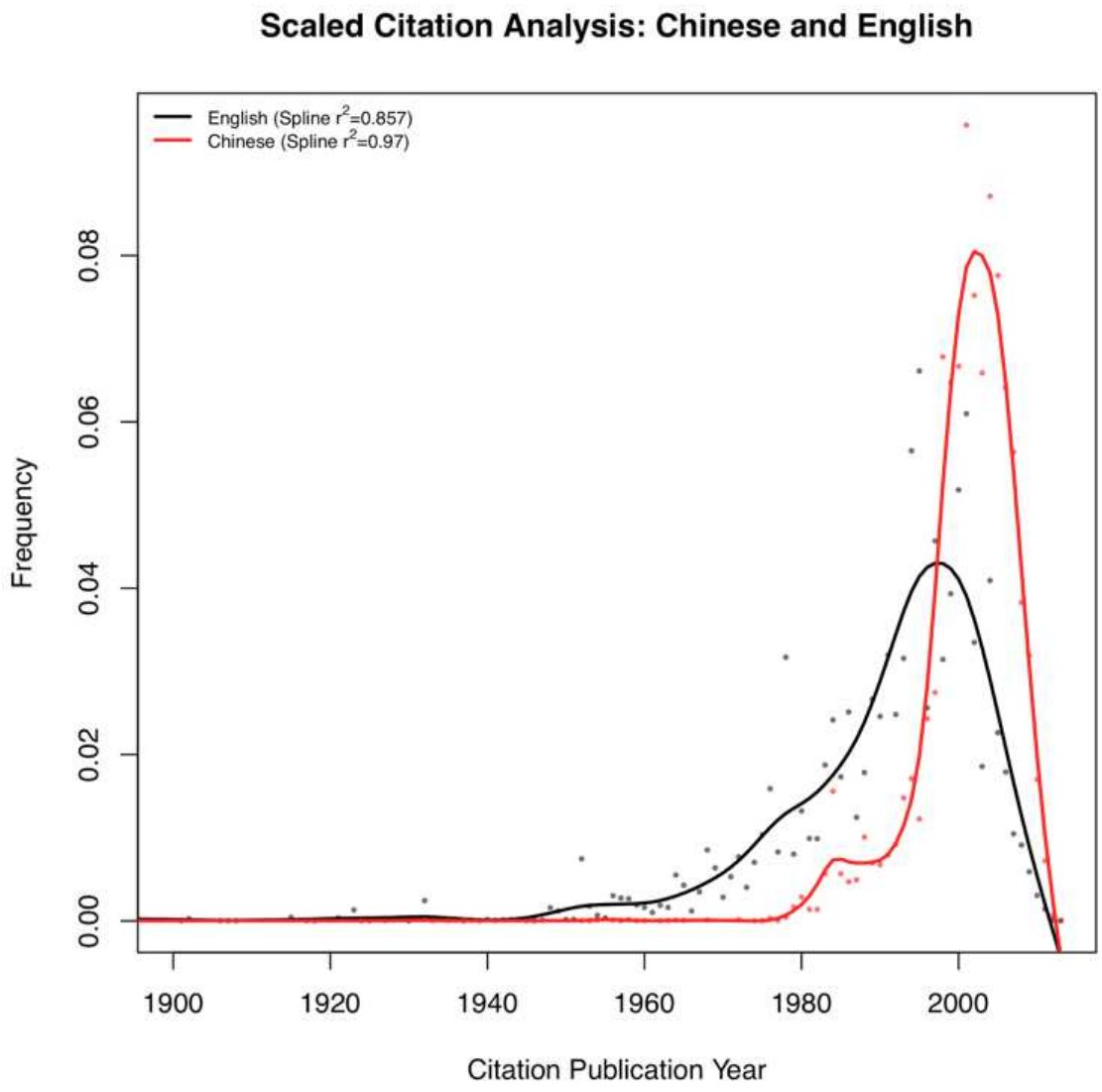


Figure 19: All English citations over time compared to all Chinese ones

PeerJ PrePrints

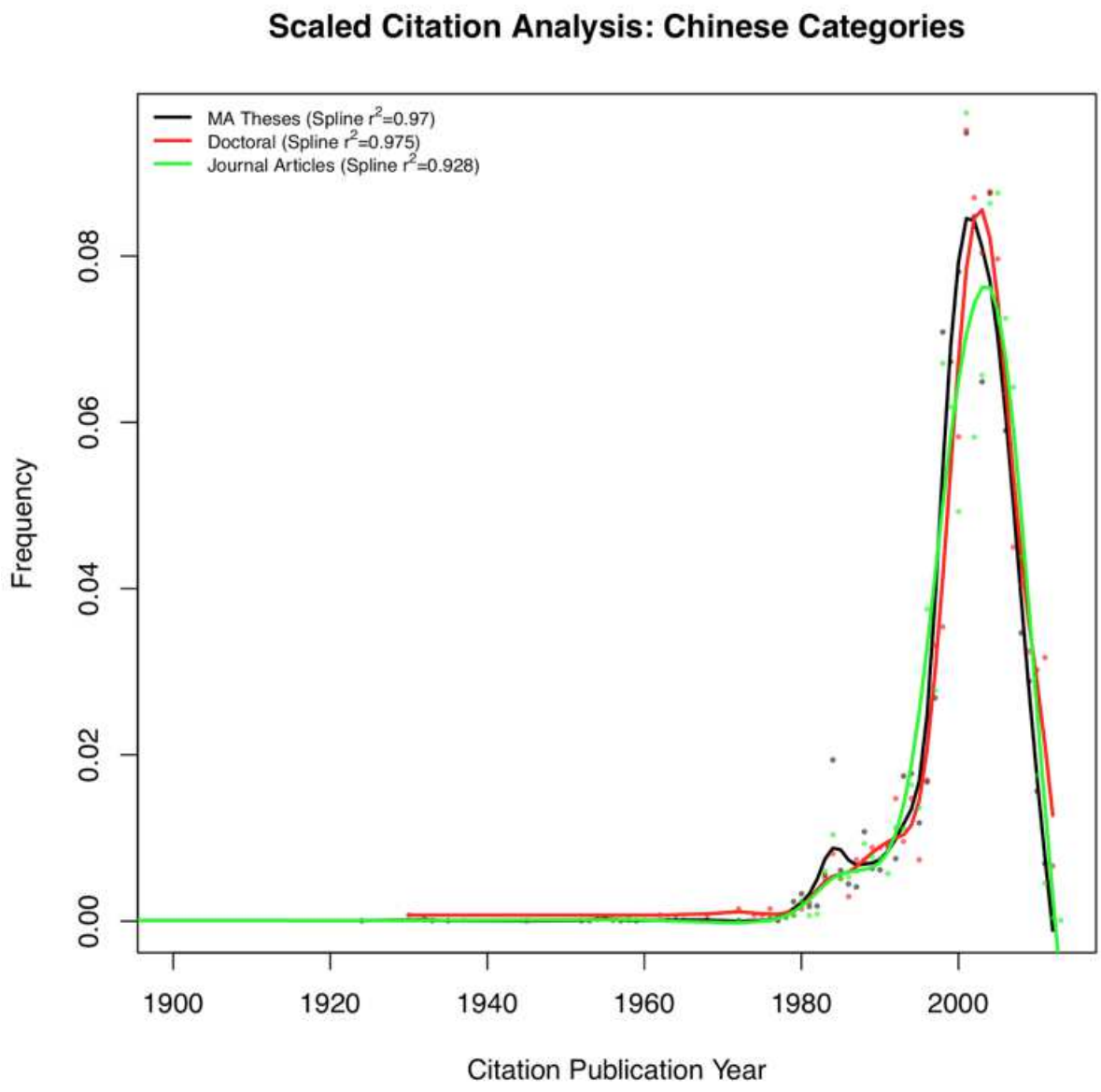


Figure 20: All Chinese citations over time, split by category: papers, theses and dissertations

PeerJ PrePrints

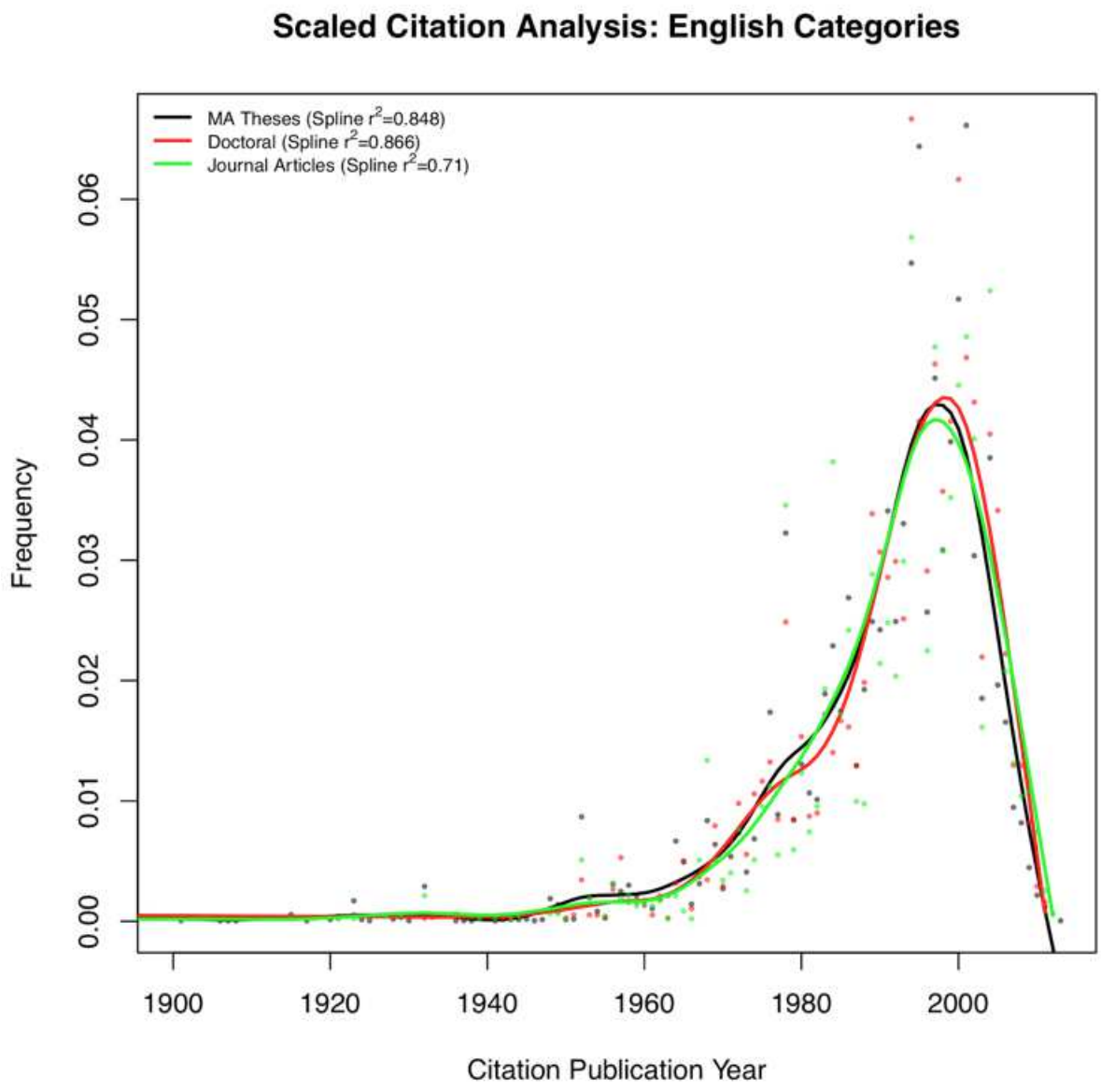


Figure 21: All English citations over time, split by category: papers, theses and dissertations

PeerJ PrePrints

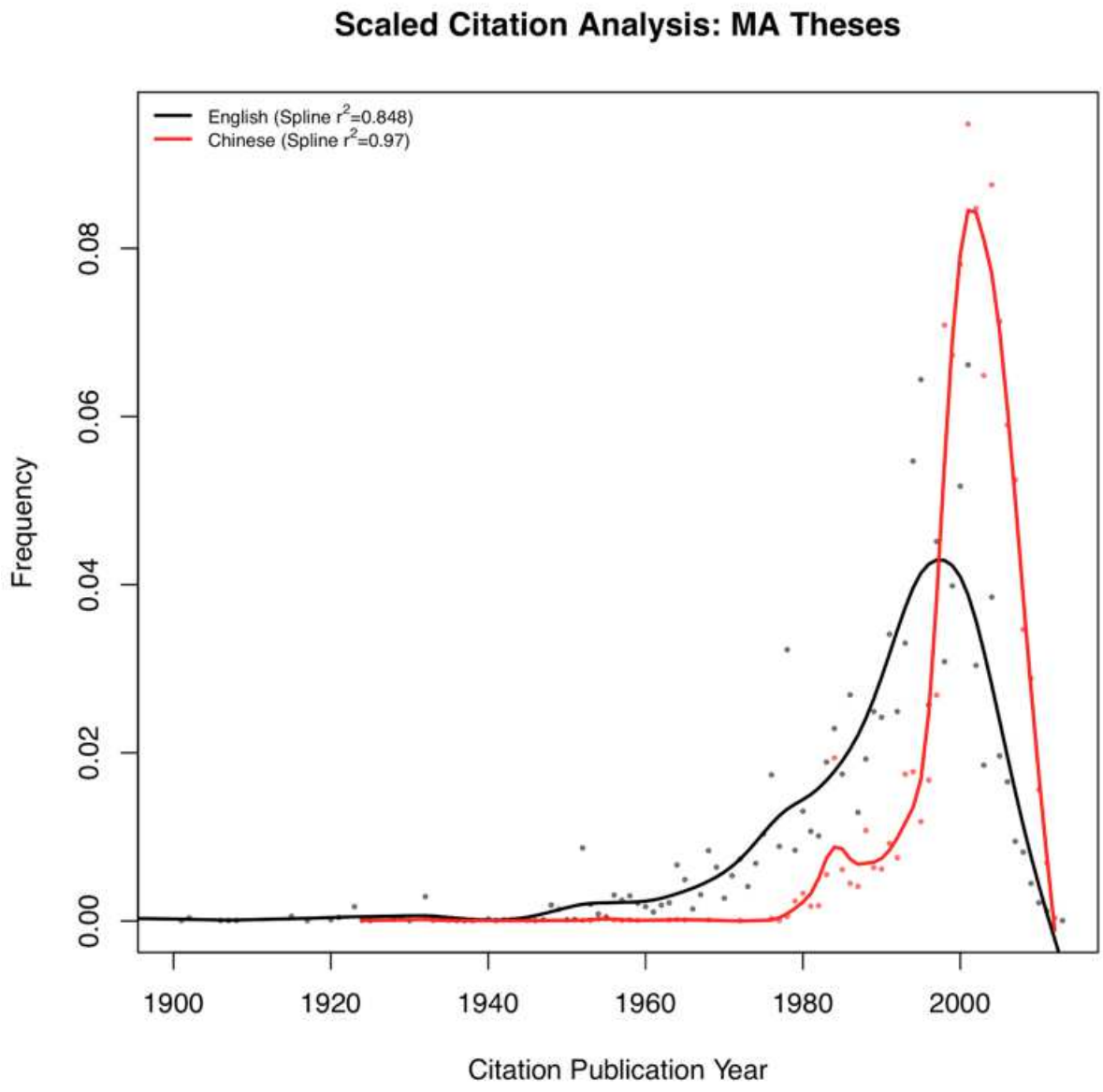


Figure 22: English compared to Chinese citations over time in Chinese MA theses

PeerJ PrePrints

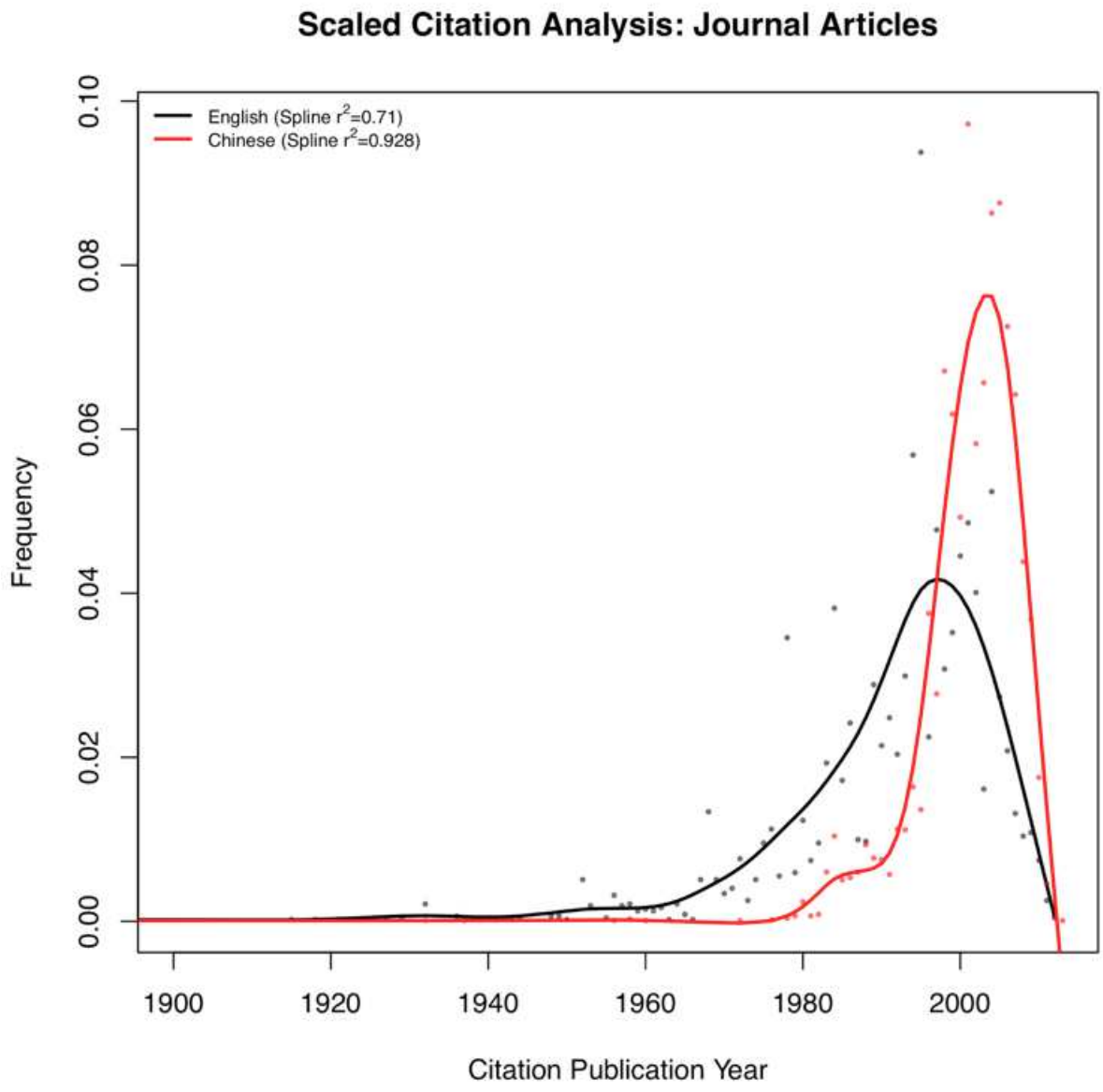


Figure 23: English compared to Chinese citations over time in Chinese papers

PeerJ PrePrints

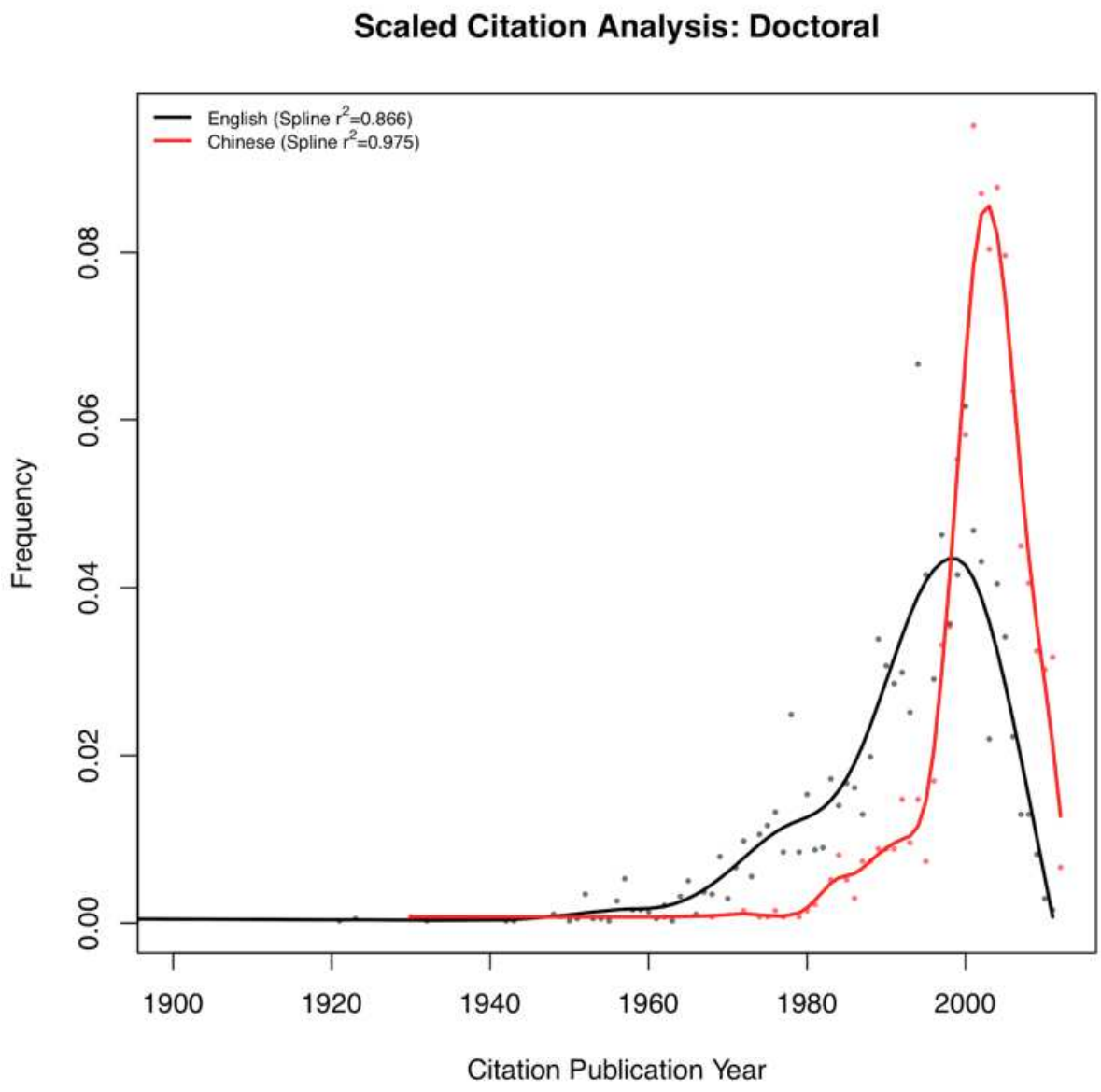


Figure 24: English compared to Chinese citations in Chinese doctoral dissertations

The data revealed that the oldest citation in the entire sample was from 1758 and the newest from 2013, but all distributions are extremely negative-skewed, i.e. they have a very large left tail; this long tail was cut off in all of the plots because there was very little data between 1758 and 1900 — doing so made the patterns easier to see. Chinese papers, MA theses and doctoral dissertations all had similar citation tendencies: the green, red and blue lines on the plot all look very similar. The Chinese curves exhibit a steeper

PeerJ PrePrints

‘spike’ centered on a later date than the more sloping distributions of English documents (see Figures 19), which indicates that Chinese authors tended to cite much older English publications than Chinese ones.

It was observed that 90% of the Chinese citations in theses were published in the period 1993-2007, compared to 1974-2004 for English citations (see Figure 22). In research papers 90% of Chinese citations were from 1994-2008, compared to 1977-2004 for English ones (see Figure 23). In the case of doctoral dissertations 90% of Chinese citations dated from 1994-2009, whereas the same percentage of English ones dated from 1976-2004 (see Figure 24). For all the categories combined, 90% of the Chinese citations came from the period 1993-2007 (see Figure 20), compared to 1975-2004 for the English ones (see Figure 21).

A possible explanation for this variance is that Western Interpreting Studies (WIS) have a much longer history than their Chinese counterpart (CIS), so it goes without saying that the English works cited by Chinese scholars were much older than any Chinese ones. Gile’s research (1994) revealed that interpreting research took off in the West in the 1950s when a number of experience-based essays and manuals on principles and techniques were published; as shown in an earlier paper (Xu 2014), CIS only began to grow rapidly after 2000; in addition, copyright clearance and translation issues may also have played their part in delaying the introduction of the latest Western theories and research to Chinese readers. That said, it is somewhat surprising to see that the majority of the English works cited were published before 2004. One possible interpretation is that Chinese researchers have failed to capitalize on the latest Western research and theories, especially in the area of cognition, one of the top two influences on MA theses and research papers (Xu 2014 & 2015) – research into cognition has developed rapidly over the past decade, with major discoveries and findings being published regularly in journals all over the world. Another probable compounding factor is that Chinese scholars do not have easy access to relevant English publications, especially in TS. Unlike other mainstream disciplines such as cognitive science and linguistics, papers on which have been comprehensively archived in academic databases such as Proquest and EBCSO, TS

PeerJ PrePrints

literature is decidedly more scattered and in many cases difficult to access unless the researcher's library has a subscription service to the journal publishers.

5.3 Do CIS authors tend repeatedly to cite 'classic' papers, or are they more drawn to the latest research within the field? How can the progress of CIS be quantified by means of citation analysis?

Figures 39 - 42 are the graphs representing the distribution of citation processes for MA theses, doctoral dissertations and journal articles in different years.

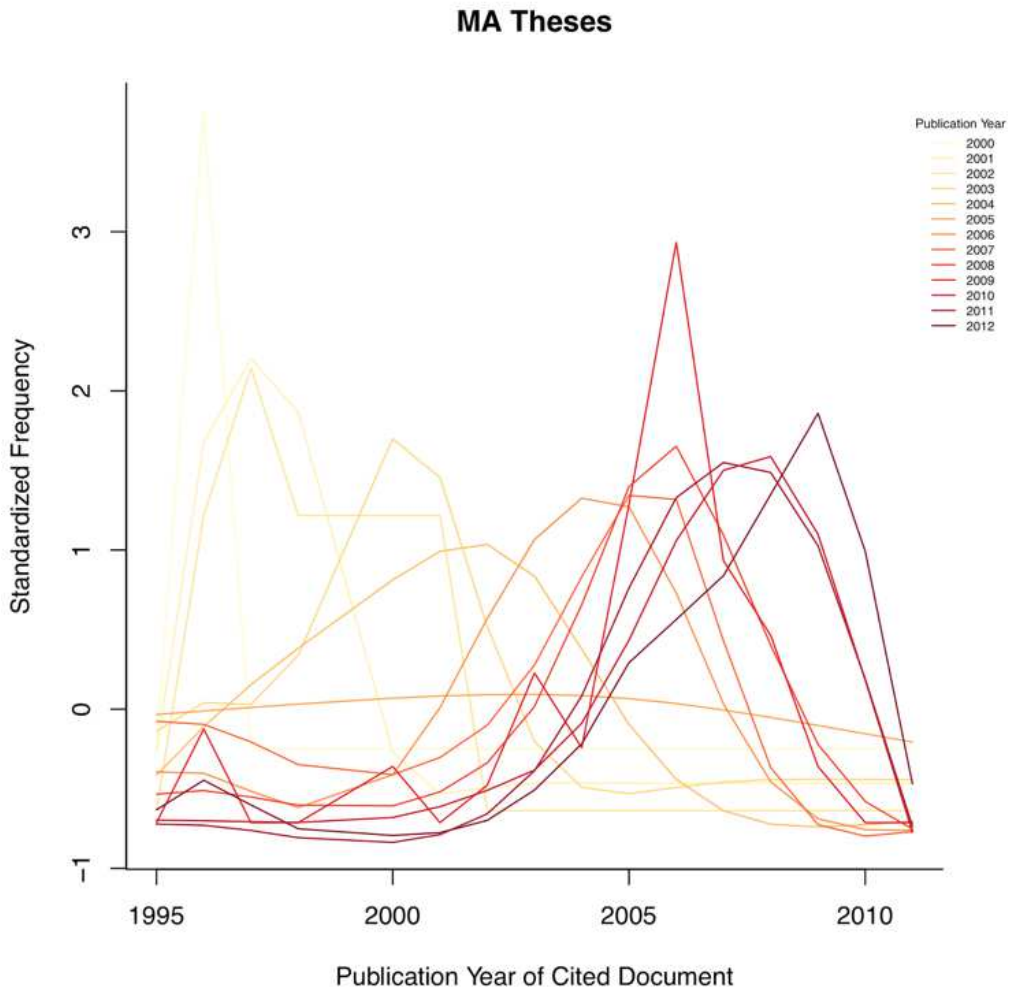


Figure 25: Normalized incoming citations for MA theses over time

PeerJ PrePrints

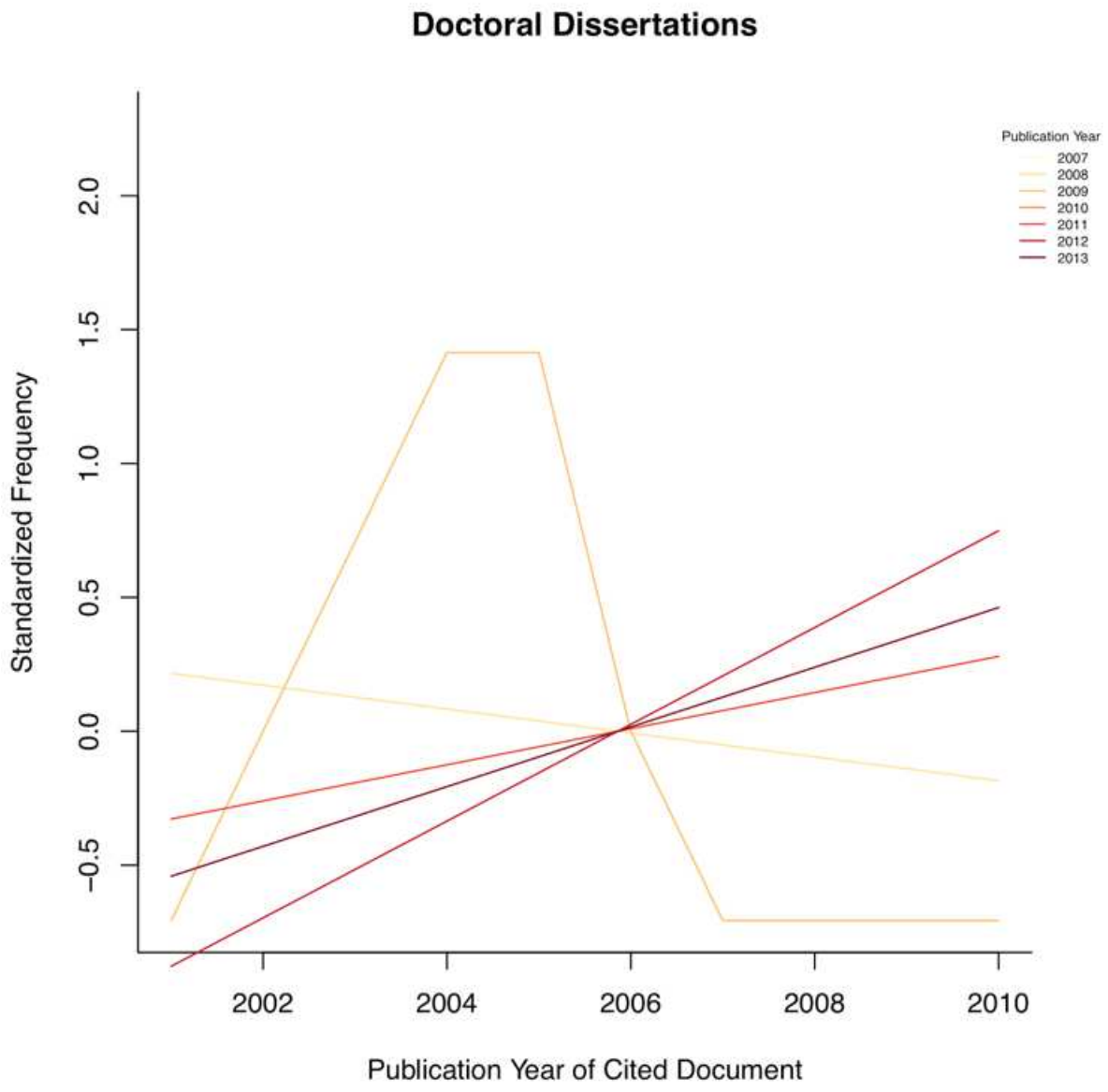


Figure 26: Normalized incoming citations for doctoral dissertations over time

PeerJ PrePrints

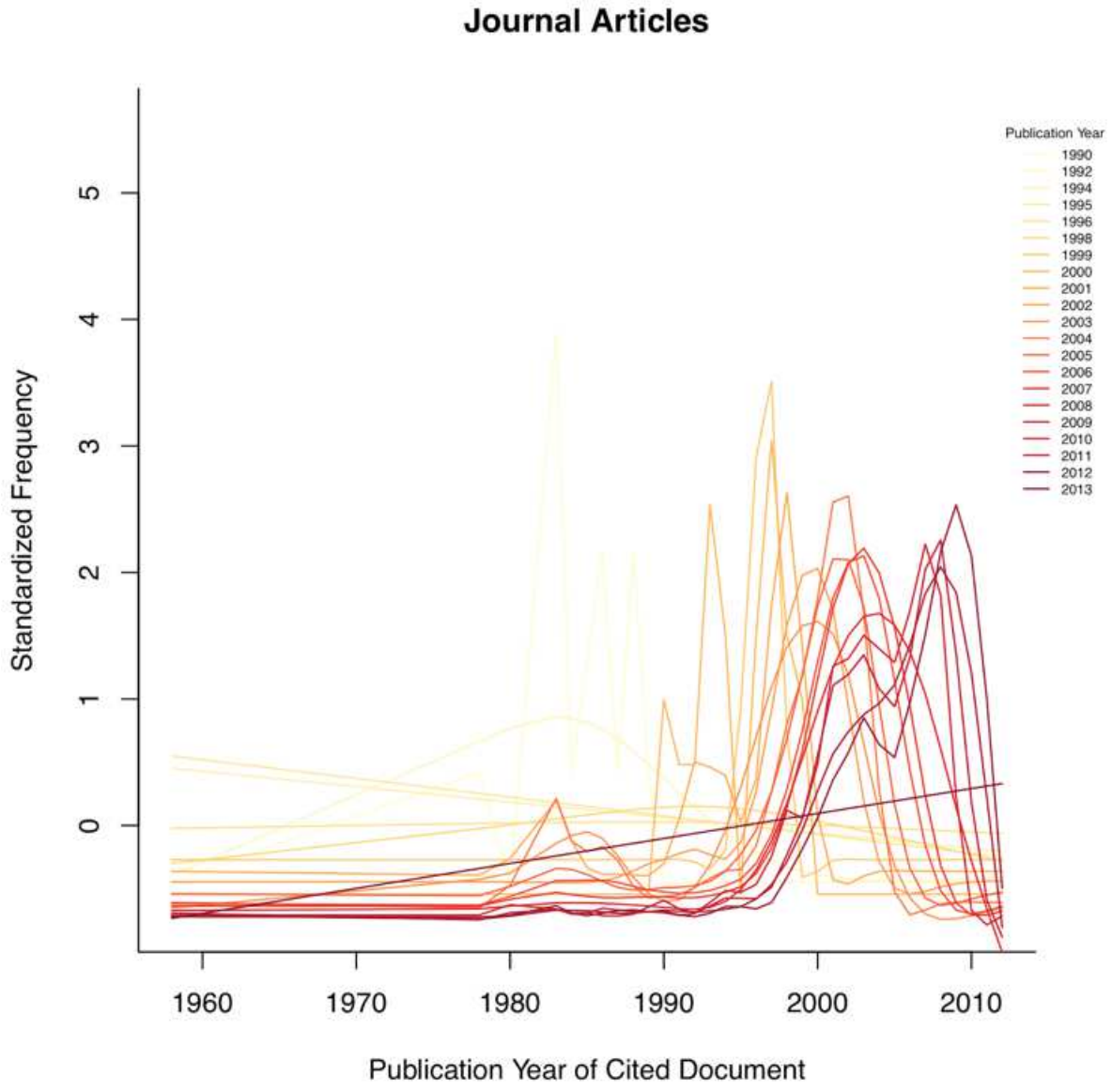


Figure 27: Normalized incoming citations for research papers over time

PeerJ PrePrints

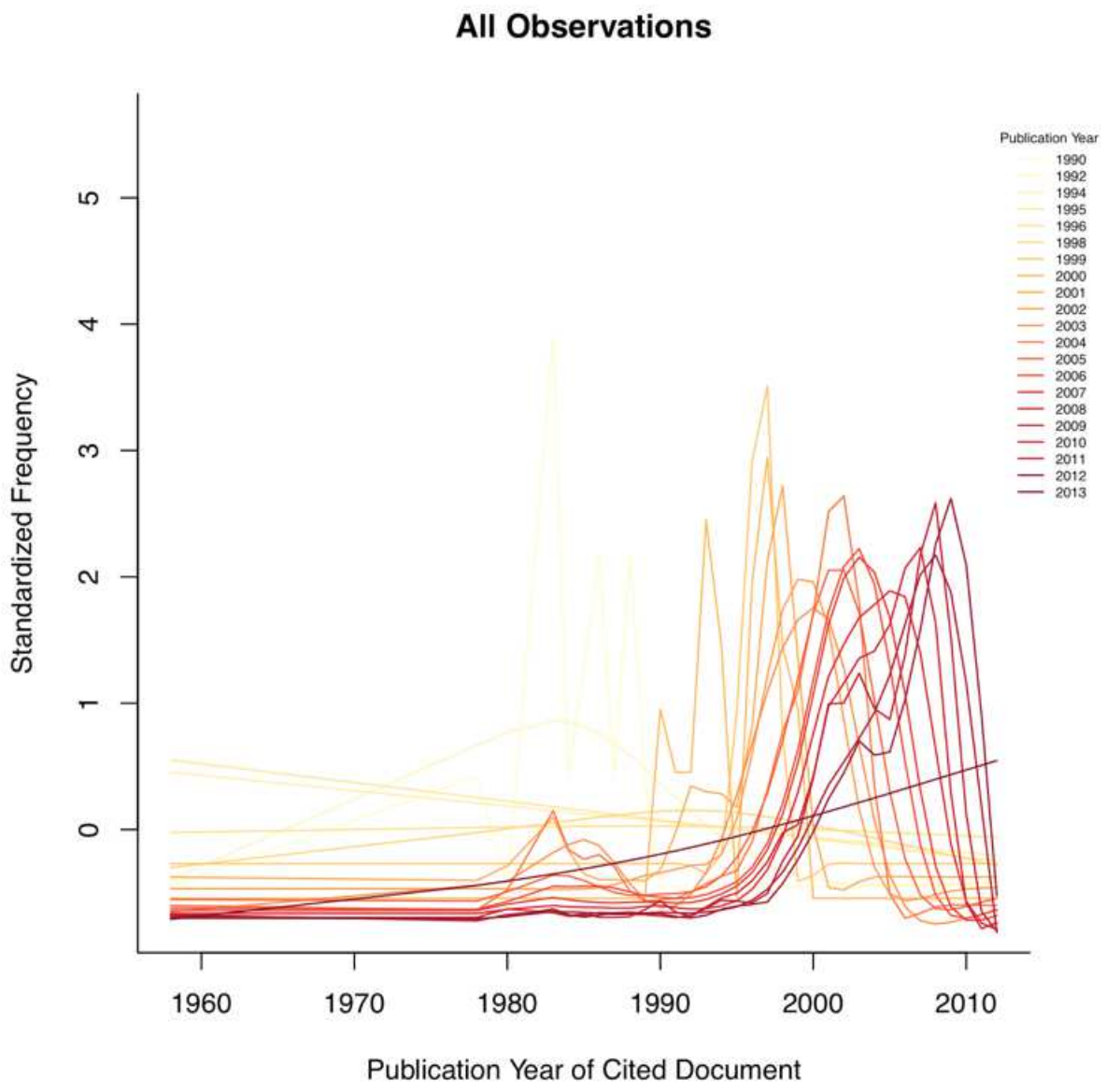


Figure 28: Normalized incoming citations for all CIS publications over time

5.3.1 Hypothesis of Research Stagnation

Figures 25 to 28 indicate definite movement over time for the incoming citation curves. In Figure 28 concerning all CIS publications, if all the curves had looked the same, this would have supported the hypothesis that the field of CIS is static. This is not the case

PeerJ PrePrints

here: the 'peaks' in the curves move forward from year to year and do not 'stagnate' at a given year. In sum, the figures suggest that CIS research is moving forward.

In addition, the hypothesis of research stagnation was rejected on statistical grounds: more recent CIS publications were constantly being cited, as opposed to classic papers receiving the majority of citations as time went by, and that caused the model corresponding to research stagnation to fit less accurately the data than did the VCM model, as demonstrated by the very low p-values for the corresponding tests[8] (see the Appendix for more information).

While newer citations may not necessarily contain innovation — they may simply restate the positions found in classic works — there is assuredly some foundation to Zuckerman's argument (1987) that the use of more recent citations nonetheless indicates that academic inquiry is moving forward. The argument is as follows: a cited paper (Paper A) gains influence when it is cited by multiple authors; however, authors may sometimes be inclined to cite other more recent papers that specifically refer to Paper A, as opposed to citing it directly. While these more recent publications may or may not generate new findings or innovative material, they effectively serve as an intellectual conduit connecting contemporary researchers with past foundational work. Paper A has become so thoroughly incorporated into the field's stock of knowledge, has become so fundamental to it, that authors feel no need to make explicit reference to it. Therefore, the rejection of the research stagnation hypothesis indicates that contemporary researchers build on more recent work and that academic enquiry is moving forward.

Even though it was both visually and statistically confirmed that CIS is moving forward, whether it has been doing so at a steady pace remains an open question. The rejection of the research stagnation hypothesis says nothing about how research evolves, and notably if the flow of research is 'perfect' in the sense where the distribution of citations is

[8] 'The corresponding tests' refers to those that compare the research stagnation model against the VCM model to see if the former fit data better than the latter. This hypothesis was rejected for all document types: MA theses, journal articles and PhD dissertations.

PeerJ PrePrints

always the same from year to year, and for that reason the hypothesis of perfect research flow needs to be tested.

5.3.2 Hypothesis of Perfect Research Flow

Perfect research flow is the opposite extreme of stagnation, and it means that papers are cited in exactly the same fashion every year. Figures 25 to 28 also enable us to grasp visually the rejection of the stationarity hypothesis. If this hypothesis were true, this would imply that the lines shown in the plots did not change with source year. This is clearly not the case here.

Moreover, a statistical analysis was conducted to validate visual intuition concerning the hypothesis of perfect research flow. Indeed, this hypothesis was statistically rejected, because the Varying Coefficient Model fit better to the citation data than did the model corresponding to perfect research flow (described in the Appendix). Once again, this is demonstrated by the very low p-values of the corresponding tests in Supplemental Table 1 of the Appendix.

5.3.3 Hypothesis Testing and Graphical Interpretations

To test both of the previously mentioned hypotheses a VCM model was used first to describe the data as accurately as possible, then this model's performance was tested to compare it with those of the models corresponding to each hypothesis.

For each year t, a spline was fit to incoming citations as a function of $|t-i|$, where i was the year of publication of the cited article.. The VCM model was constructed so that it would be easy to control the variation of the coefficients over time.

The resulting graphs (see Figures 29-31) can be likened to a frame-by-frame movie of the evolution of incoming citations over time.

PeerJ PrePrints

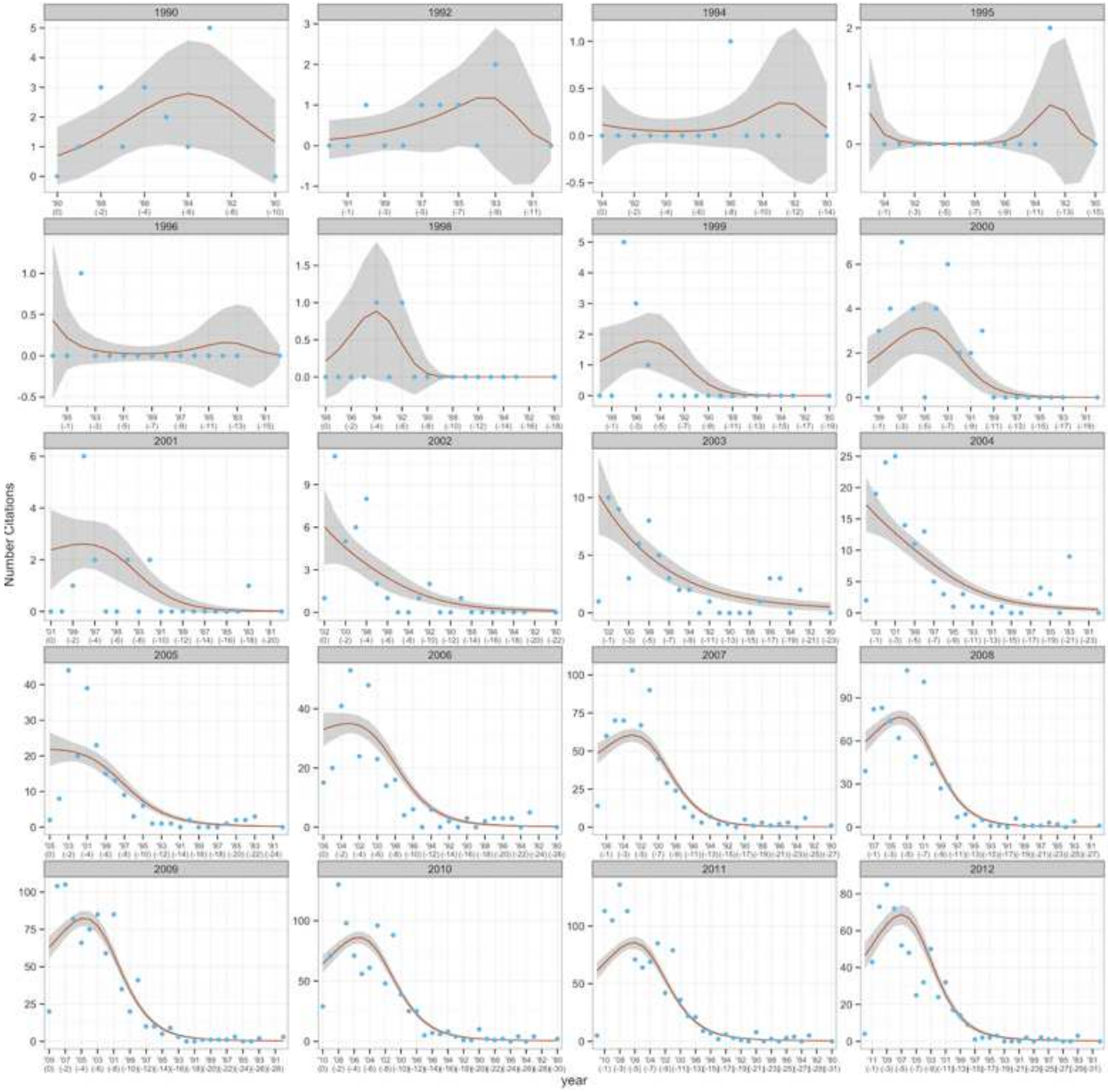


Figure 29: Trends in citations for research papers

PeerJ PrePrints

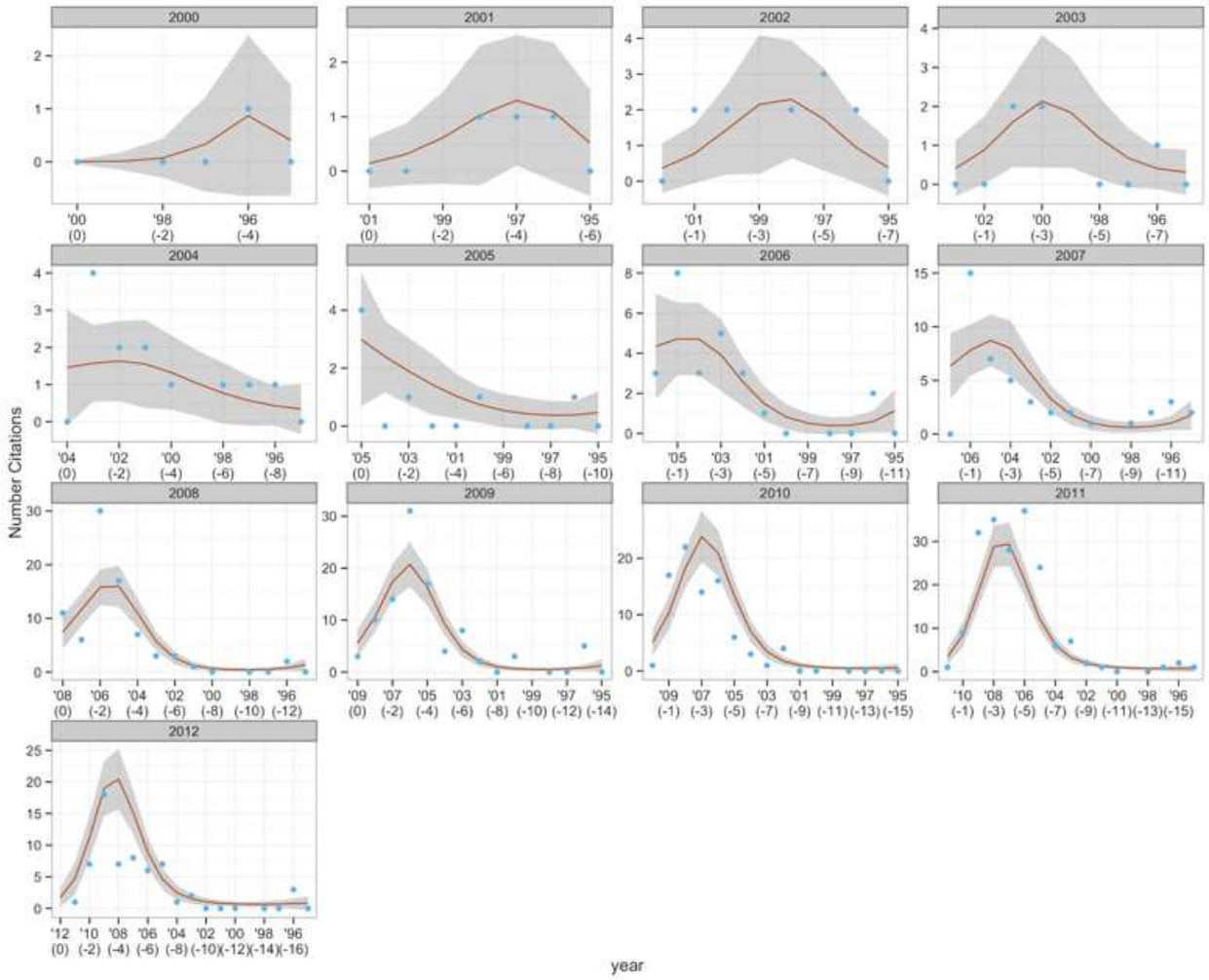


Figure 30: Trends in citations for MA theses

PeerJ PrePrints

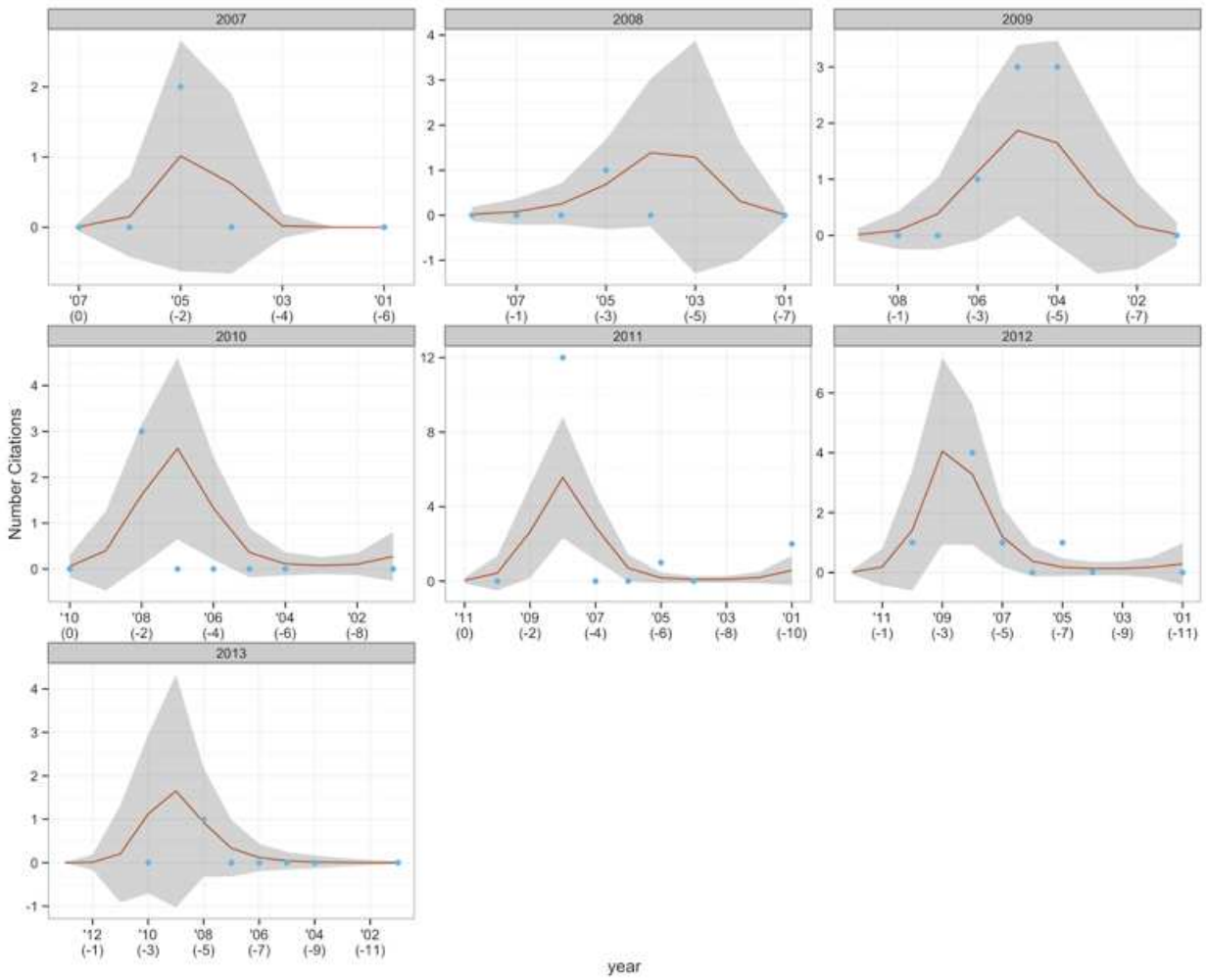


Figure 31: Trends in citations for doctoral dissertations

The red line is the fit for the VCM model and can be considered the average citations count for that year; the blue dots are the actual number of citations produced in each year; and the gray shaded areas represent a 95% confidence interval for the red line. The gray headers show the year under consideration — for example, '2000' means that all the papers written in 2000 were examined to ascertain the number of citations in them dating from 2000 (*t*), 1999 (*t-1*), 1998 (*t-2*), and so forth.

On examining the incoming citation data it was observed that recent papers were regularly cited within an interval of a year or two — this trend was particularly obvious from 2009 to 2012. Moed (2005) has argued that an author might include a certain

PeerJ PrePrints

reference not only because its content fits the flow of an argument, but because he believes the scholar he is citing has gained a certain stature in the field and will lend credibility to his own ideas. For example, it would be more credible to cite the definition of empirical research by a scholar who has conducted extensive studies of that type than by one whose focus is purely theoretical. The finding that recent papers are cited so soon indicates that newer research has a more or less instant impact on the latest studies and that CIS research is in a state of continuous progression. It was also remarked that, in disregard of the 1-2 year rule mentioned above, citations from material published in 1990 were made in CIS papers throughout the period under study, suggesting that that year may have seen the publication of particularly influential material, whose impact on research has been especially long-lasting. On further examination of the incoming citations, Hu Gengshen's *An overview of interpreting research in China* stood out as the aforementioned material. Hu's paper took a scientometric approach to assessing the themes and trends in interpreting research. From the Y axis it was also clear that many more citations were being made in later years, probably because the number of CIS papers being written was increasing year on year.

The situation for MA theses was slightly different from that of journal articles, though research was moving forward here too. These authors were somewhat hesitant to cite recently completed theses, preferring those produced at least three years previously, which they could be sure, had been adopted by the academic community and become established. It was also noticeable that material produced in 1996 was cited by numerous MA authors in all subsequent years, suggesting that some very influential work was produced in that year. Detailed analysis revealed that work to be Ru Mingli's thesis *Interpreting quality and the role of the interpreter from the perspective of users*, which was produced under the supervision of Chen Yongyu. It should be noted, however, that MA thesis authors cited their predecessors' work far less often than the authors of research papers did theirs: in 2010, for example, research papers produced in 2008 were cited no fewer than 148 times; the same figure for MA theses was a mere 22. There are two possible reasons for this phenomenon: (1) a number of researchers (Lawrence 2001; Harnad & Brody 2004; Hajjem et al. 2005) have identified that open-access articles

PeerJ PrePrints

receive a substantially higher number of citations than those that require a subscription — this is true across many disciplines including computer science, physics, sociology and psychology. Proceeding from their findings, it is reasonable to speculate that the difficulty — and expense — of obtaining access has contributed to the significantly lower number of CIS theses being cited in comparison to research papers. (2) In addition, in the academic world MA theses are generally considered to be of lower quality than research papers, which have gone through rigorous peer review.

Given that the total number of doctoral dissertations was only 32, little in the way of trends was observable. It should be noted, however, that a particular doctoral dissertation produced in 2008 was consistently quoted by later PhD authors in the period 2010-13 — this was Gong Longsheng's *An analytical study of the application of Adaptation Theory in interpreting*, written under the supervision of Dai Weidong. Gong is such a well-established and visible academic within the CIS community[9], that it is hardly surprising that his work might attract a large number of incoming citations.

To conclude, two null hypotheses were both visually and statistically rejected: research stagnation and perfect research flow. To perform those tests two models corresponding to each of the hypotheses, and a third, the Varying Coefficient Model, were constructed. The three were tested to see how well they fit the CIS citation data. Both hypotheses were rejected, because the first two models performed poorly in comparison with the VCM model. Analysis of the citations yielded enough evidence to say that this field is going forward, though not at a constant pace.

5.4 What are the most frequent citation types? Do they differ based on language of origin (Chinese vs. English) and document type (papers, MA theses and doctoral dissertations)?

[9] Gong served as an associate dean of the Graduate School of Business at SISU, has supervised nearly 30 MA interpreting students, and was involved in developing the Shanghai Interpretation Accreditation Test.

PeerJ PrePrints

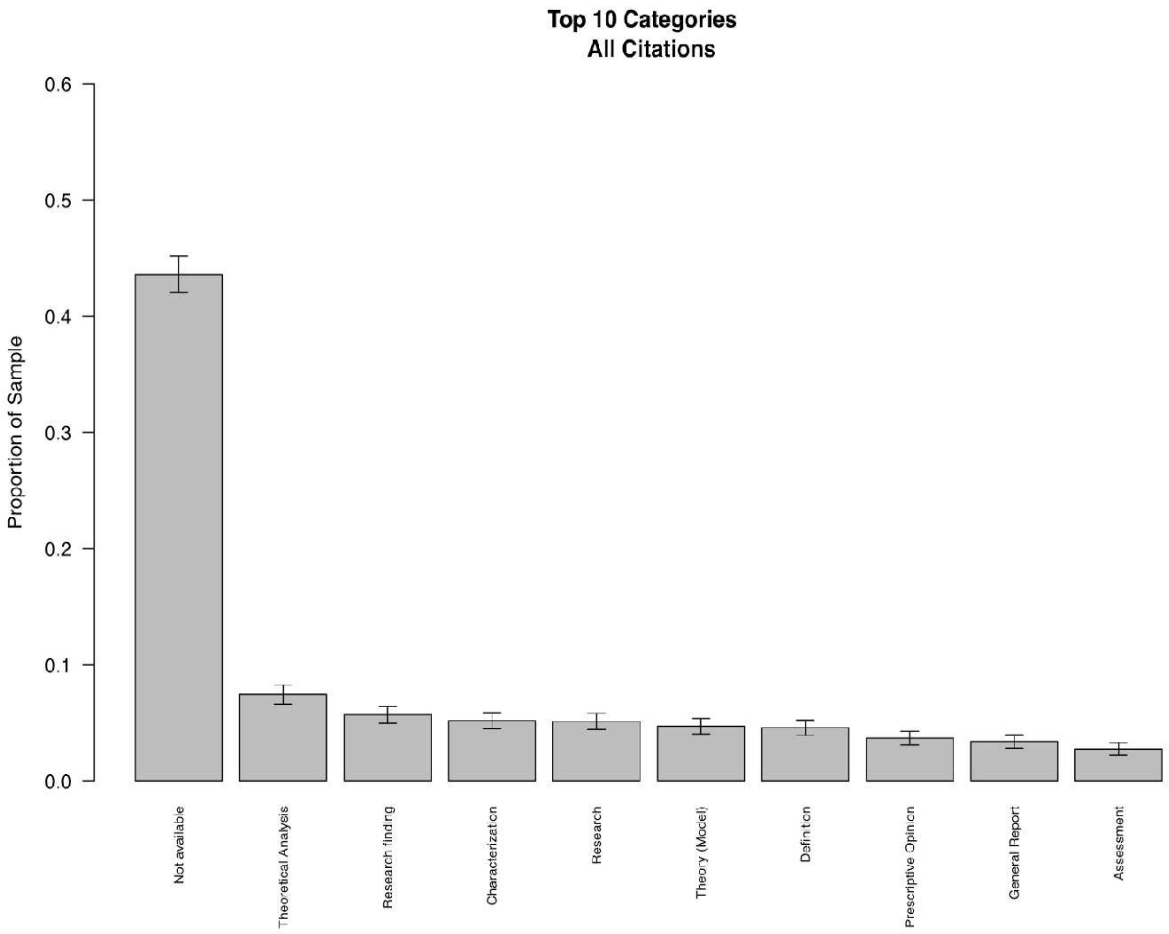


Figure 32: Proportion of the top 10 citation categories with confidence bars across all databases

When a fair and unbiased random sample is taken from a large population, the law of large numbers ensures that the sample average for any category ought to be close to the true value of the mean in the total population. Confidence intervals represent intervals where the real average proportion of each category will be with 95% probability. For example, in the sample for the present study (3,888 citations), nearly 45% were labeled N/A. Therefore the 95% confidence interval for all the N/A citations was [42% ; 45.1%], a finding which appears to suggest that almost half of the works listed in the bibliographies of CIS authors did not appear in the body of the text. The prevalence of N/A references indicates that certain cited authors may not necessarily be linked directly to the research of those citing them, despite having played an instrumental role in shaping their outlook on interpreting or influencing their professional training. For example, a

PeerJ PrePrints

further analysis revealed that 75.3% of the references to Mei Deming and 85.6% of those to Zhang Weiwei belonged to the N/A category; their cited works are the leading interpreting textbooks in China, though not regarded as theoretical or empirical contributions to CIS research. The second most popular citation type (theoretical analysis) stood at 7.5%; the corresponding 95% confidence interval was [6.6% ; 8.3%]. From this finding it is reasonable to speculate that theoretical research has played a crucial role in shaping CIS. It was also observed that 5.8% of citations belonged to the research finding category (95% confidence interval: [5.1; 6.5%]). This is interesting because that citation type is generally associated with empirical research. Its being the third most frequently used type seems to indicate that CIS authors were also keenly aware of the importance of empirical research, and consciously analyzed how their research could amplify the findings of previous scholars.

An examination was also made of the distribution of citation types by the following methods:

(1) Citation type distribution for all Chinese citations (see Figure 33)
(2) Citation type distribution for all English citations *ibidem*
(3) Top ten citation types for all Chinese citations in theses, dissertations and papers (see Figure 34)
(4) Top ten citation types for all English citations in theses, dissertations and papers *ibidem*
(5) Citation types for MA theses, dissertations and papers (see Figure 35)

PeerJ PrePrints

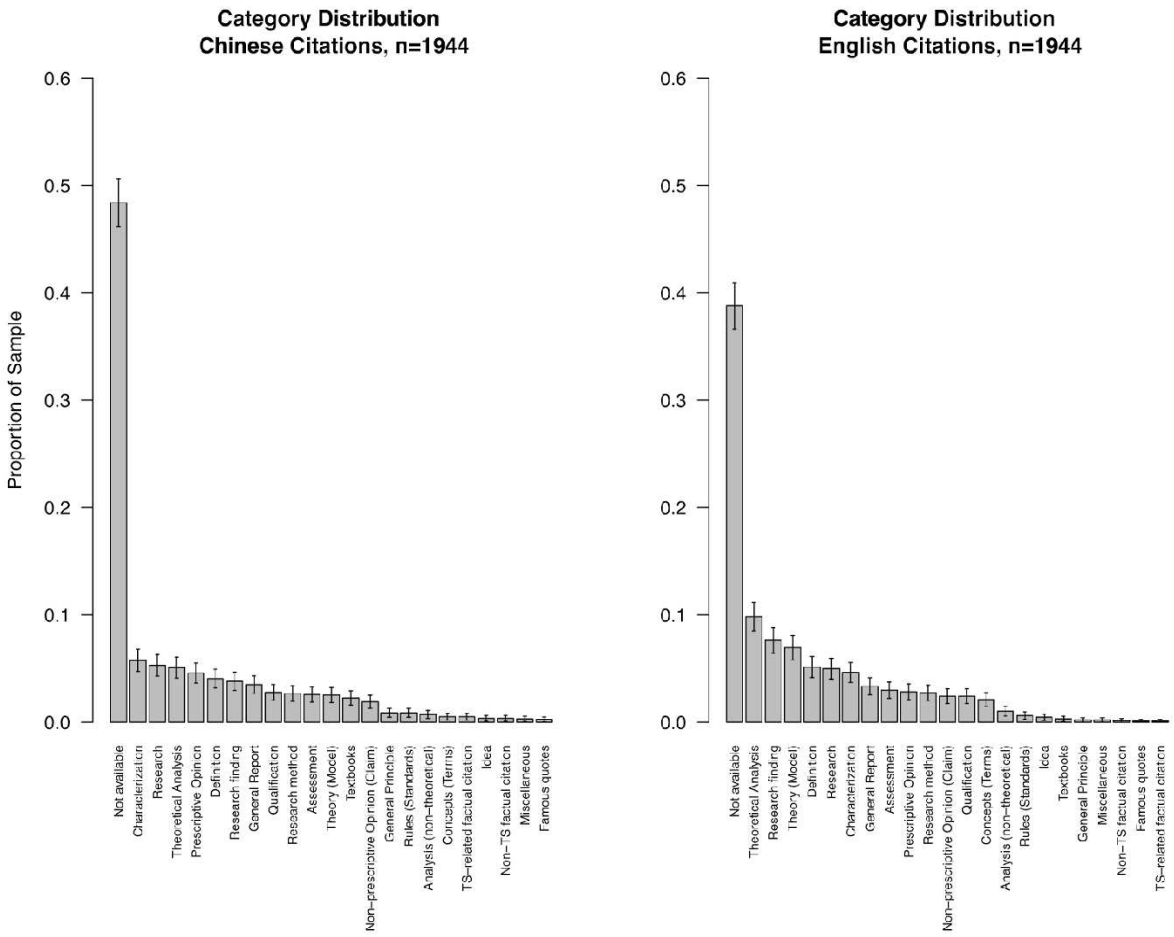


Figure 33: Proportions of all English and Chinese citation type distributions with confidence bars.

Confidence intervals of 95% were constructed to compare the differences in the proportions of each category between Chinese and English references. If for a given category the means of the two populations have non-overlapping confidence intervals, this indicates a statistically significant difference between the Chinese and English citations – non-overlapping 95% confidence intervals guarantees a test for differences at alpha level of 0.05 (Knezevic 2008). For example, the proportion of N/A citations was [46.2% ; 50.6%] for Chinese against [36.6% ; 41.0%] for English. The differences in proportions suggest that English references listed in bibliographies are more likely to have corresponding in-text citations than their Chinese counterparts, which would indicate that CIS authors are more indirectly influenced by their Chinese than their Western colleagues.

PeerJ PrePrints

The analysis also revealed that there was a significant difference in the Theoretical Analysis citation type between Chinese (95% CI: 4.1-6.0%) and English (95% CI: 8.5-11.2%) (Figure 34). This would appear to indicate that, in cases of theoretical discussions, CIS authors are roughly twice as likely to cite English authors as their Chinese colleagues. In addition, it was found that Prescriptive Opinions (95% CI: Chinese 3.7-5.6% vs. English 2.1-3.6%) and Textbook citations (95% CI: Chinese 1.6-2.9% vs. English 0.1-0.5%) were less common in English than in Chinese citations, while Research Findings were more common (95% CI: Chinese 3.1-4.8% vs. English 6.5-8.8%). While advice from all practicing professionals is highly valued in the global interpreting community, these findings appear to suggest that the opinions of Chinese interpreters carry more weight with CIS authors than those of their Western counterparts. Textbooks are usually seen as repositories of established fact rather than sources of cutting-edge ideas. The fact that Chinese textbooks are cited more often than Western ones highlights that CIS researchers frequently turn to them for well-established facts.

The most frequently occurring category of Chinese citation besides N/A was Characterization (95% CI: [4.6 ; 6.6%]). Given that experience-based, intuitive writing was dominant in CIS' early developmental stage, and that practicing interpreters often resorted to summarizing the features of a phenomenon or idea rather than theorizing or providing empirical support, its popularity is understandable. Conversely, the most frequent category in English citations after N/A was Theoretical Analysis (95% CI: 8.5-11.2%), which suggests that CIS authors were influenced by the theoretical work of Western authors.

PeerJ PrePrints

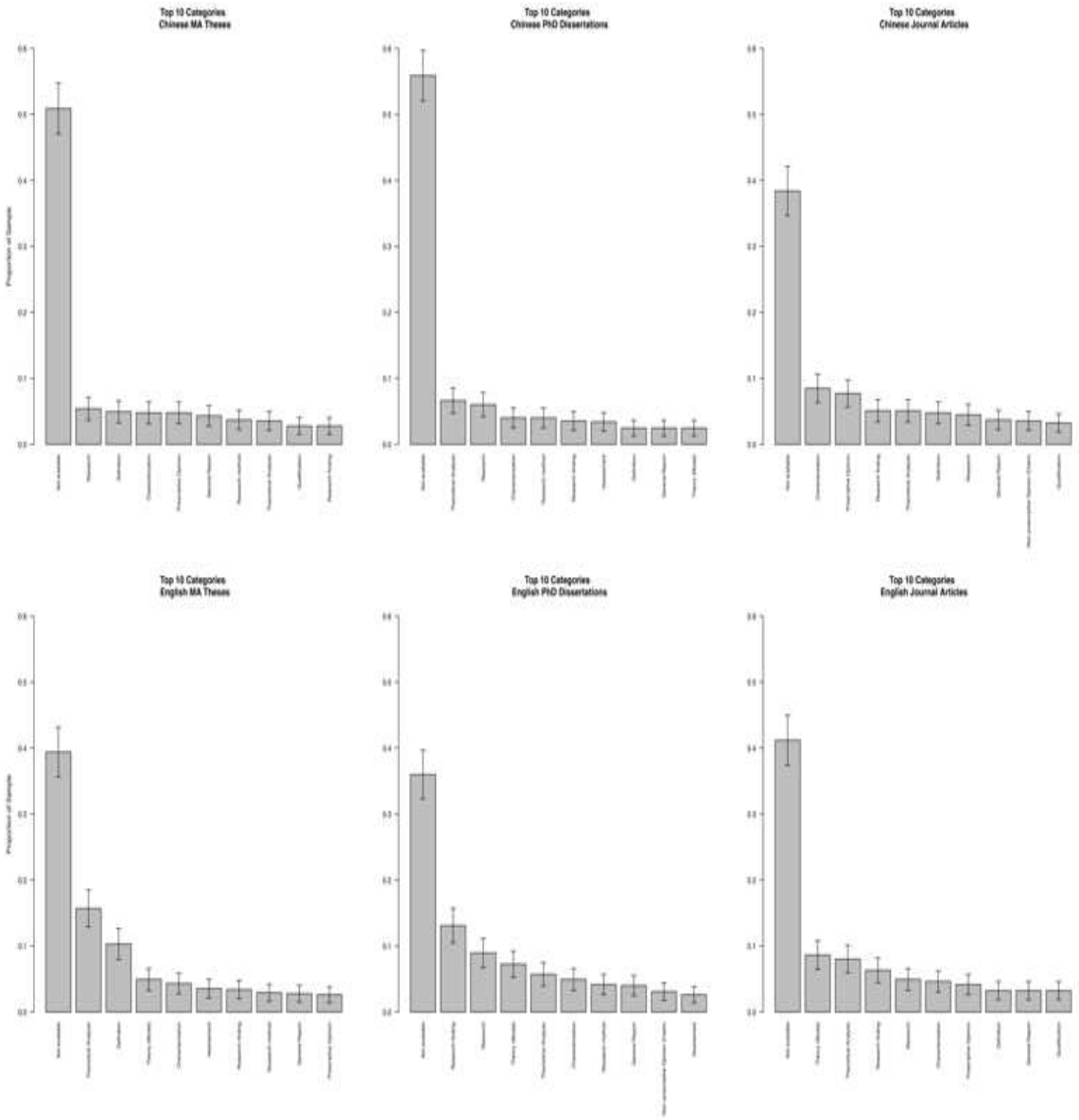


Figure 34: Proportions of the top ten citation types for all Chinese and English citations in theses, dissertations and papers

PeerJ PrePrints

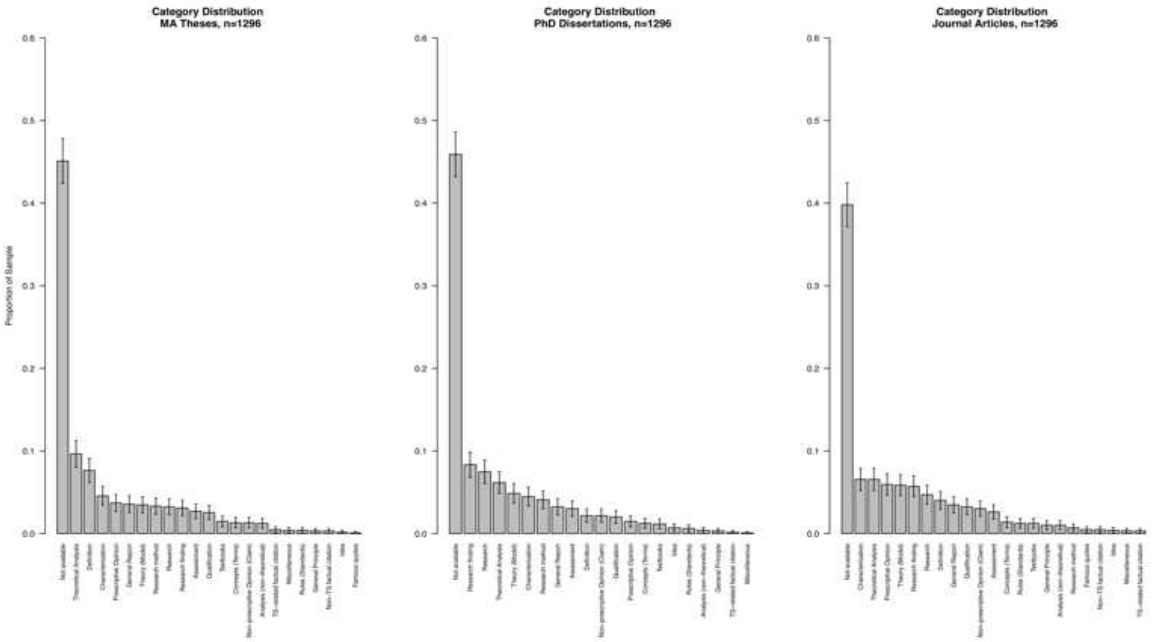


Figure 35: Proportions of all citation types for MA theses, dissertations and papers

From this point forward the citation distribution in each document type was examined, to spotlight the differences in how different sub-groups of CIS authors make bibliographic references. Citation practice differed widely across the document classes. In MA theses statistically significant differences between Chinese and English references were found for the following citation types: N/A (95% CI: 47.1-54.7% for Chinese vs. 35.6-43.2% for English), Tangential Research (95% CI: 3.7-7.1% for Chinese vs. 0.3-1.9% for English), Theoretical Analysis (95% CI: 2.1-5.0% for Chinese vs. 12.9-18.5% for English) and Textbooks (95% CI: 1.5-4.0% for Chinese vs. 0-0.5% for English). The finding for N/A tallies with the earlier result for all citations. As for the Tangential Research type, the result suggests that Chinese authors' research is more frequently mentioned in passing without specific reference to its contents than is the case for Western authors. One might wonder if some thesis authors cite their compatriots as a way of paying tribute rather than because their works inspire or influence them. In addition, this finding is in line with our interpretation of the differences in proportions for the N/A category between Chinese and English references. Of course, there is also the possibility that thesis authors prefer a style of literature review that summarizes more than it analyzes. It is worth noting that no statistically significant differences were observed

PeerJ PrePrints

between Chinese and English references for citation types generally associated with empirical research, such as Research Methodology or Findings.

It was also revealed that Theoretical Analysis was the second most popular type of all thesis citations, though English authors were far more frequently referred to than Chinese for this category. Its popularity across both Chinese and English strongly suggests that theoretical research is influential on Chinese thesis authors. It was somewhat surprising to find that citing someone for their definitions was a moderately popular citation type (English: 7.9-12.5%; Chinese: 3.4-6.8%) — one might have expected researchers mainly to cite others for their research findings or theories, rather than regularly resorting to them for the definitions of certain terms. Close examination of the contexts in which this type of citation occurs revealed that a significant number of MA students reviewed the history of interpreting at the beginning of their theses, citing Western researchers to define various types of interpreting and clarify the differences between it and translation.

In journal articles and conference proceedings there were statistically significant differences between Chinese and English references for the Prescriptive Opinion (English<Chinese) and Theory citation types (English>Chinese). This finding suggests that in developing their work CIS article authors were more likely to turn to Western scholars for theories and models and to their Chinese colleagues for intuitive understanding of interpreting. It is worth remembering here that the preference for citing Chinese colleagues for Prescriptive Opinion was also observed in theses. It is understandable that Western scholars are more often cited for Theory, because the first generation of Chinese SI trainers received their education in Brussels — theories such as the Interpretive Theory of Translation and the Effort Models have served as the foundations for many a Chinese author's research. It should be noted here that unlike the earlier findings, no statistically significant difference was observed in N/A between English and Chinese references in research papers[10]; this might be explained by the fact

[10] The statistic of a two-sample z test is 0.79, which is smaller than the critical value of 1.96. Thus, there is no evidence to suggest that there is a statistically significant difference between the proportions of English and Chinese N/A citations.

PeerJ PrePrints

that the comparison was made on a smaller sample for papers (the average number of citations in papers was 10, compared to 43 for theses and 278 for dissertations).

In doctoral dissertations the following citation types yielded statistically significant differences between Chinese and English references: N/A (Chinese>English), Research Findings (English>English), Concepts (English>Chinese), and Theory (Chinese>English). As in the dissertation category, there were statistically fewer corresponding in-text citations for Chinese references than for English ones, which is consistent with all previously discussed findings in the overall, article and thesis categories. It should be noted that the proportion of Research Findings was higher here than in any other document type examined. This suggests that PhD students particularly rely on predecessors' empirical findings to shape their own work.

In sum, CIS researchers displayed different citation behavior across languages and document classes, and no particular citation type yielded significant differences between Chinese and English references across all three document classes. However, the N/A citation type was more frequently used in the Chinese references of theses and doctoral dissertations than in the English ones, while the opposite was the case for Theory. Other than N/A, no citation type occurred more than 20% of the time across languages and document classes, suggesting that research is cited in diverse ways in the CIS community and no particular citation type is predominantly used.

5.5 What effect does the choice of empirical vs. theoretical research have on the use of citations in journal articles, MA theses and doctoral dissertations?

A two-proportion one-sided z-test was used to evaluate whether the proportion of empirical citations was greater than non-empirical ones. The test yielded the following p-values for the four tests (p-values were rounded to 3 digits): the comparison between the proportion of Research Methods and Findings cited by articles and MA theses (empirical citations) was not significant ($p = 0.532$). However, the comparison between doctoral dissertations and theses was significant ($p < 0.001$). The proportions for citations of

PeerJ PrePrints

research methods and findings in theses, articles and doctoral dissertations were 0.064, 0.065 and 0.124 respectively.

(1) Articles > theses IS NOT significant, $p = 0.532$

(2) Doctoral > theses IS significant, $p < 0.001$

It was observed that doctoral dissertations cited research methods and findings more than theses, therefore the null hypothesis that the reverse would apply can be rejected; but there was little evidence to support the same claim for theses as compared to journal articles, therefore that null hypothesis cannot be rejected. However, given that the p-value in hypothesis (2) is significant, it can be stated with confidence that doctoral dissertations cited research methods and findings more than papers. The data suggest the following relations:

Research methods and findings: dissertations > (theses ? journals)[11]

It was observed that theses used fewer citations of Concepts, Ideas and Opinions (theoretical citations) than papers, but there was scant evidence to suggest that doctoral dissertations used fewer such citations than theses. The overall proportions for these citations in the three types of publication were 0.094, 0.086 and 0.133.

(3) Theses > doctoral IS NOT significant, $p = 0.246$

(4) Articles > theses IS significant, $p < 0.001$

[11] 'Theses? journals' means that the null hypothesis that theses have fewer research methodology citations than articles cannot be rejected, therefore the author cannot comment on any relationship that might exist between theses and journals in this specific case. Unlike a two-sided test, where the null hypothesis is equality, the null hypothesis for our one-sided test is less than or equal to zero.

PeerJ PrePrints

However, the present analysis demonstrates that the differences between the two categories are significant, which is in line with the idea that theses use fewer concepts:

Concepts, ideas, and opinions: (dissertations? theses) < journals

Because questions (1) and (3) compare MA theses and research papers, and questions (2) and (4) compare doctoral dissertations and MA theses, and because there is not a new random sample for each test, the tests are correlated. To control the familywise error rate, the present author implemented the Bonferroni-Holm correction method, a sequential method where the p-values do not have to be adjusted but can, rather, be compared to a different significance level to keep the familywise error rate for all of the tests at 0.05. Without such correction, there would be an increased chance of rejecting correct null hypotheses, thus increasing Type-I errors.

After the correction, even the two smallest p-values remained significant when compared to the new cutoffs:

(2) Research methods and findings: doctoral > theses IS significant at 0.0125, p-value: 7.75e-08
(4) Concepts, ideas, and opinions: articles > theses IS significant at 0.0167, p-value: 0.000639

This strongly suggests that the previous findings — that dissertations cite more research methods and findings than theses, and that dissertations and theses cite fewer concepts, ideas and opinions than papers — still hold true even after adjusting for familywise errors. In other words, empirical research has a greater influence on how doctoral scholars cite papers than it does on thesis and paper authors; and theoretical research has a more significant impact on how paper authors make citations than it does on MA and doctoral authors. This finding is interesting because it indicates that data-driven research is more popular among MA and doctoral students, whereas theoretical research is favored by established academics. If MA and doctoral students represent the future of CIS, one

PeerJ PrePrints

can expect empirical research to continue to expand its sphere of influence in the foreseeable future.

6. Conclusion

It is hoped that this scientometric survey has demonstrated the merits of blending quantitative with qualitative analysis to paint a panorama of Chinese Interpreting Studies. Citation data was used to measure the progress of CIS: the field is a dynamic one with new papers being constantly cited, though a few influential older papers have withstood the process of 'obliteration'. Citation sampling and labelling were employed to describe how scholars exhibit different citation patterns across languages and document types, and how authors' choosing of empirical or theoretical research variously influences journal articles, MA theses and doctoral dissertations.

Thanks to its comprehensive data collection, the study was not faced with the issues of sample size typically associated with quantitative analysis of interpreting studies (IS) in the past. In the West only a few hundred individuals are dedicated to interpreting research; by contrast, no fewer than 3,500 scholars are documented in this study's database: this wealth of available data makes it possible to adopt some of the latest statistical techniques to assess the evolution of CIS. The qualitative elements of the study serve to spotlight unique patterns of behavior exhibited by its researchers when citing their predecessors.

The focus of this paper being IS, no attempt was made to explore translation-related publications. Given that TS has a longer history and many more participants than IS in China, it would be an interesting future line of inquiry to research the themes and patterns of the former, so as to offer a balanced view of how the whole discipline of translation and interpreting studies has developed and is developing. Also, with the ever-increasing level of academic exchange between East and West, a comparative study of the differences and similarities between the two would provide valuable insights to policy-makers charged with shaping the future direction of academic research.

PeerJ PrePrints

References

APA. (2010). *Publication manual of the American Psychological Association*. Washington, DC: American Psychological Association.

Baumgartner, H. & Pieters, R. (2003). The structural influence of marketing journals: A citation analysis of the discipline and its subareas over time. *Journal of Marketing,* 67 (2), 123-139.

Carr, J. E. & Britton, L. N. (2003). Citation trends of applied journals in behavioral psychology 1981-2000. *Journal of Applied Behavior Analysis,* 36 (1), 113-117.

Chen, B. (2009). Interpreting research in China: Past and present (In Chinese). *The Science and Education Article Collects,* 264-265.

Fan, J., Zhang, C., & Zhang, J. (2001). Generalized Likelihood Ratio Statistics and Wilks Phenomenon. *The Annals of Statistics*, 29 (1), 153-193.

Franco Aixelá, J. (2004). The study of technical and scientific translation: An examination of its historical development. *The Journal of Specialised Translation,* 1, 29-49.

Franco Aixelá, J. (2013). Who's who and what's what in Translation Studies: A preliminary approach. In C. Way, S. Vandepitte, R. Meylaerts, & M. Bartłomiejczyk, (Eds). *Tracks and Treks in Translation Studies*. John Benjamins, 7-28.

Freedman, D. (1997). Some issues in the foundation of statistics. *Topics in the Foundation of Statistics,* 1, 19-39.

Gao, B. (2008). *Conjecture and refutations: An inquiry into researches on cognitive process in simultaneous interpreting* (In Chinese). *PhD dissertation, Shanghai International Studies University*.

Garfield, E. (1972). Citation Analysis as a Tool in Journal Evaluation. *Science,* 178(4060), 471-479. http://www.jstor.org/stable/10.2307/1735096?ref=no-x-route:8f4816e110f43a76725a95940101096a (accessed 17th February 2015)

PeerJ PrePrints

Garfield, E. (1977). The 'obliteration phenomenon' in science -- and the advantage of being obliterated. In E. Garfield (Ed.), *Essays of an information scientist*. 2. Philadelphia, PA: Institute for Scientific Information. 396-398.

Garfield, E. (1979). *Citation indexing: Its theory and application in science, technology, and humanities*. New York: Wiley.

Gile, D. (1994). Opening up in interpretation studies. In M. Snell-Hornby, F. Pöchhacker, & K. Kaindl (Eds.), *Translation studies: An interdiscipline*. John Benjamins. 149-158.

Gile, D. (1999). Use and misuse of the literature in interpreting research. *The Interpreters' Newsletter,* 9, 29-43.

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

Gile, D. (2005). Citation patterns in the T&I didactics literature. *Forum,* 3 (2), 85-103.

Gile, D. (2006a). L'interdisciplinarité en traductologie : Une optique scientométrique (In French). In Ö Kasar, (Ed.) Interdisciplinarité en traduction (*11e Colloque International sur la Traduction*). Istanbul: Isis. 23-37.

Glänzel, W. (2003). *Bibliometrics as a research field: A course on theory and application of bibliometric indicators*. Lecture.

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

Grbić, N. & Pöllabauer, S. (2009). An author-centred scientometric analysis of Daniel Gile's œuvre. *Effort and models in interpreting and translation research*. John Benjamins. 3-24.

Hajjem, C., Harnad, S. & Gingras, Y. (2005). Ten-year cross-disciplinary comparison of the growth of open cccess and how it increases research citation impact. *ArXiv*.

Hargens, L. L. (2000). Using the literature: Reference networks, reference contexts, and the social structure of scholarship. *American Sociological Review,* 65 (6), 846.

Harnad, S. & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles in the same journals. *D-lib Magazine,* 10 (6).

PeerJ PrePrints

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York: Springer.

Hu, G. (1990). An overview of interpreting research (In Chinese). *Foreign Language Teaching and Research,* 4 (84), 1-6.

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). *Doctoral dissertation, Shanghai International Studies University.*

Ivancheva, L. (2008). Scientometrics today: A methodological overview. *Collnet Journal of Scientometrics and Information Management,* 2 (2), 47-56.

Kalaitzidakis, P., Mamuneas, T. P. & Stengos, T. (2003). Rankings of academic journals and institutions in economics. *Journal of the European Economic Association,* 1 (6), 1346-1366.

Kempthorne, O. (1952). *The design and analysis of experiments*. New York: Wiley.

Knezevic, A. (2008). Overlapping confidence intervals and statistical significance. *StatNews*, 73, 1.

Kuhn, T. S. (Ed.) (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lawrence, S. (2001). Online or invisible. *Nature,* 411 (6837), 521.

Lederberg, J. (1972). Reply to H.V. Wyatt. *Nature,* 239 (234).

Lowry, P. B., Karuga, G. G. & Richardson, V. J. (2007). Assessing leading institutions, faculty, and articles in premier information systems research journals. *Communications of the Association for Information Systems (CAIS),* 20 (16), 142-203.

Meho, L. I. (2006). The rise and rise of citation analysis. *Physics World,* 1-15.

Merton, R. K. (1967). *On theoretical sociology*. New York: Free Press.

Moed, H. F. (2005). *Citation analysis in research evaluation*. Dordrecht: Springer.

Nasr, M. (2010). *Pour l'obtention du titre de Docteur en Traductologie* (In Chinese). *PhD dissertation, UNIVERSITÉ PARIS III.*

Pöchhacker, F. (1995). Writing and research in interpreting: A bibliographic analysis. *The Interpreters' Newsletter,* 6, 17-31.

PeerJ PrePrints

Rao, J. N. (1963). On Three Procedures of Unequal Probability Sampling without Replacement. *Journal of the American Statistical Association,* 58 (301), 202-215.

Ren, S. & Rousseau, R. (2002). International visibility of Chinese scientific journals. *Scientometrics,* 53 (3), 389-405.

Rice, J. A. (1995). *Mathematical statistics and data analysis*. Belmont, CA: Duxbury Press.

Ru, M. (1996). *Interpreting quality and the role of the interpreter from the perspective of users* (In Chinese). *PhD Dissertation, Fu Jen Catholic Universit.y*

Schulz, K. F. & Grimes, D. A. (2002). Generation of allocation sequences in randomised trials: Chance, not choice. *The Lancet,* 359 (9305), 515-519.

Tang, S. & Zhou, Y. (1958). The nature of interpreting (In Chinese). *Western Languages,* 2 (3), 321-327.

Vallmitjana, N. & Sabaté, L. G. (2008). Citation analysis of PhD dissertation references as a tool for collection management in an academic chemistry library. *College & Research Libraries,* 69 (1), 72-82.

White, H. D. & Mccain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science,* 49 (4), 327-355.

White, H. D. (2004). Citation Analysis and Discourse Analysis Revisited. *Applied Linguistics,* 25 (1), 89-116.

White, M. J. & White, K. G. (1977). Citation analysis of psychology journals. *American Psychologist,* 32 (5), 301-305.

Xu, Z. (2014). Chinese Interpreting Studies: Genesis of a discipline. *Forum: International Journal of Interpretation and Translation,* 12 (2), 159-190.

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

Zuckerman, H. (1987). Citation analysis and the complex problem of intellectual influence. *Scientometrics,* 12 (5-6), 329-338.

PeerJ PrePrints

Appendix:

Justification for the use of Varying Coefficient Models (VCM)

Varying Coefficient Models (VCMs) are more generalized versions of regression. Regression expresses the value of an output as a combination of different type of input (or predictors). In varying coefficient models, the coefficients associated to each predictor vary with other variables (that may or may not be connected to the predictors). For example, in the context of a chemistry experiment which can be modelled using linear regression, we may get very different coefficients in the regression depending on outside parameters like the temperature. It is then better to use a varying coefficient model: the coefficients in the linear regression will in this case be functions of the temperature (this is not the same as including the temperature as a predictor, since the dependence of the output to the temperature is not necessarily linear). An overview of the theory behind VCM models can be found in Hastie & Tibshirani (1993).

In this appendix I explain VCMs in the context of CIS citation. To avoid confusing notation, I refer to cited papers written in a given year as 'target papers', and those doing the citing as 'source papers'.

My goal in this analysis was to model the number of target citations in the source year i referring to articles written in the target year j. To understand how VCMs can help in that goal, it is first useful to consider the range of possible outcomes. To do this, I constructed three models: two models examine the extremes of the relationship between target and source citations, and the third captures the middle ground between these two points. These three models described in detail below were applied to analyze my three data sets (journal articles and conference proceedings, MA theses, and doctoral dissertations).

Model 1: Research stagnation

PeerJ PrePrints

In this model, target citations depend only on the target year and not the source year. One example of such a model is "only papers written in 1996 were cited in years following 1996". I will refer to this model as "research stagnation", as newly written papers are never cited.

Model 2: Perfect research flow

At the other extreme, we can imagine that target citations depend only on the target year relative to source year. An example of this model is "only papers written 3 years before the source paper were cited.'' This model is an example of what I will call research flow, as newly written papers consumes information written in the past, and all of it is transferred through current work. In this model, a given paper cannot be cited in more than one year. There is therefore no durability in the papers cited. I refer to this as model 2.

Necessity to generalize these models

A third option also exists: it is possible that citations evolve on a path somewhere in-between the two previous models. One way of modeling this is to assume that target citations depend on the number of years between the target year and the source year (i.e. the time between the redaction of the article and its citation), but also on the source year itself.

The extent to which this relationship changes with the source year determines where between the two extremes we are:

- The first extreme is a case where the relationship does not change at all with time. This is the 'perfect flow' model (model 2)
- The second extreme corresponds to a relationship that can change unhindered with time. Model 1 is a specific case of this extreme, where we only allow articles from a given year to be cited.

PeerJ PrePrints

For example: in 1999 citations may refer primarily to articles published the two previous years (the target years 1997 and 1998), and in 2000 citations could correspond only to articles published in 1998 and not 1999 nor 1997. In this case, the coefficient associated with the target years (relatively to the source year) depend on the source year (so model 2 is not sufficient enough to describe the data)), and do not depend only on the target year (so model 1 is not sufficient either). An intermediate model as described previously is therefore necessary.

Model 3: Varying coefficient model

Model 3 generalizes both model 1 and model 2, as there can be both year-to-year dependence and relative age dependence. A parameter, alpha, controls the amount of year-to-year change:

- alpha = 0 corresponds to the first extreme mentioned, where there is no change at all with the source year (model 2)
- alpha = 1 corresponds to the second extreme (unhindered change from year to year, which is only a slight generalization of model 1 since model 1 would have alpha very very close to 1).

This model fits precisely in the frame of Varying Coefficient Models as the coefficients in the linear regression are allowed to depend on the considered source year.

To summarize, model 3 is a varying coefficient model where the predictor is the number of years between the source year and the target years (i.e. the age of the article at the moment of its citation; if we call i the source year, and j the target year, the predictor for the target year j is $(i-j)$). The varying part of the model comes from the fact that the coefficients of the regression are allowed to vary with the source year, and the variable on which the coefficients from the linear regression depend is the source year.

Mathematical expressions of the different models

PeerJ PrePrints

To speak in mathematical terms, if we call N_ij the number of target citations sourced from year i and targeting year j, we have:

- $N_ij = B_(i-j)(i) * (i-j)$ for model 1, with $B_(i-j)(i)$ only being different of 0 for a given target year j. In our example only papers from 1996 were cited, which means that N_ij is different from 0 only if $j=1996$, (but does not depend on i, the source year). So $B_(i-j)(i) = 0$ for every j but $j = 1996$, and its value (when it is non 0) is allowed to vary with the source year i.

- $N_ij = B_(i-j) * (i-j)$ for model 2, where $B_(i-j)$ is a coefficient that depends only on the age of the article at the moment of its citation ($i-j$), but does not change with the source year i. Indeed, for any source year, only articles published in certain target years (relatively to the source year) were cited. In our previous example for model 2 (in which only articles published 3 years before were cited), the coefficients corresponding to the predictors ($i-j$) would be 0 for ($i - j$) =4 years or more, 2 years or less, and different from 0 only for ($i-j$) =3 years.

- $N_ij = B(i) * (i - j)$ for model 3. This notation describes a varying coefficient model, which blends the last two models together: the number of citations in year i targeting year j depend on the difference $(i-j)$, and the coefficients of the regression itself depends on the considered source year i. We write B(i) to describe a relationship between age of target paper, $(i-j)$, and N_ij. This relationship can then change with i.

This is not a completely general model, which would have a separate, unconstrained coefficient for each source year-target year pair. Instead we ask that the coefficients $B(i)$ vary smoothly with i, meaning that the relationship between number of citations and difference is similar from (source) year to year. The smoothness of the function $B(i)$ *is directly linked to the parameter alpha: alpha=0 corresponds to an infinite smoothness (constant function), and alpha=1 to no particular smoothness (with possible jumps in the function).*

PeerJ PrePrints

The coefficients of *B(i) * (i-j)* are also constrained in the index *(i-j)* for each fixed source year i to be a natural cubic spline. This schematically takes the form of a cubic polynomial expansion

B(i)(i-j) = B_0(i) + B_1(i) * (i-j) + B_2(i) * (i-j)^2 + B_3(i) * (i-j)^3,*

with further linear constraints at the boundaries *(i-j)=0* and the maximum age of a citation. The added constraints make *B(i)* a length 3 vector for each source year *i*. Note that the vector is only length 3 and the expansion above is only schematic because of the additional boundary constraints. Since the *N_ij* are counts restricted to the non-negative integers, a Poisson regression VCM was a natural fit for all three models.

Continuing with our hypothetical example, an implementation of the VCM (model 3) may show that the *B(i) * (i-j)* changed only a little with source year *i* (e.g. 50 percent of papers were cited from *i-j=3* years before in 1997, and 40 percent of papers were cited from *i-j=3* years before in 1998).

Implementation

All three models were fit using a generalized linear model with Poisson link function since the response is count data – integers in the non-negative half space. Additionally, the VCM was fit using locally weighted least squares and a Gaussian kernel. This meant that the VCM fit a separate, weighted Poisson regression for each source year, $^{i}0$, with weights proportional to exponential squared fall of

$$w_i \propto e^{\frac{(i-i_0)^2}{\log \alpha}}$$

with α defined as above. For more information on fitting VCM models, see Chapter 6 of Hastie, Tibshirani and Friedman (2009).

Model Assessment

PeerJ PrePrints

Supplemental table 1 shows deviance, and p-values for a test of deviance for models 1, 2, and 3 applied to all three data sets (journals, MA theses, and PhD dissertations), as well as the value of alpha chosen for each data set, and the degrees of freedom (Df) for the first two models. Deviance is related to the amount of variation not explained by the regression, so smaller deviance values indicate a better model fit.

Supplemental Table 1: Evaluation of VCMs.

	Deviance 1	Df 1	Dev. 2	Df 2	Dev. 3	Alpha	P(D3<D1)	P(D3<D2)
MA	341.9	17.00	845.40	6.00	247.22	0.37	<0.001	<0.001
Journal	3871.26	39.00	10124.87	21.00	1909.24	0.37	<0.001	<0.001
PhD	67.0	6.00	80.45	4.00	29.25	0.14	<0.001	<0.001

The p-values reported in the table are for null hypotheses: "model 1 does as well at predicting as model 3" and "model 2 does as well at predicting as model 3." Since the p-values are all less than .001, it is safe to say that the null hypotheses were rejected in favor of the alternative: model 3 is better at predicting the number of citations than either model 1 or model 2. As described in the following section, a standard test for deviance does not apply for VCMs. Hence p-values were calculated using a bootstrap procedure.

For all datasets, the VCM did significantly better than either stagnation or perfect flow, which indicates that while papers were being re-cited year to year, research was progressing in CIS.

Note that the value of alpha was not optimized, except to lower it by half on the exponential scale (smaller change of coefficients from year to year) for PhD dissertations due to the smaller sample size of the dataset. A supplementary analysis showed an optimal deviance for alpha strictly in between 0 and 1 for all three datasets. This showed that model 3 fitted the data better than model 1 and 2 (it is worth reiterating that alpha = 1 does not correspond to model 1, but to a generalization of model 1 where the variation of the coefficients from year to year is the most important), even if the change of the

PeerJ PrePrints

coefficients from one source year to another was not that important (not that far from model 2): the pattern of papers cited in the year n+1 is highly correlated with the pattern of paper cited in the year n.

Bootstrap procedures for test of deviance

Deviance calculations are the standard for assessing goodness of fit for generalized linear models - the Poisson regression was used to estimate expected number of citations in this paper. A test for the better fit of a model with more variables compared to one with fewer variables is achieved by comparing the deviance difference of the larger model from the smaller model to a null Chi-squared distribution with degrees of freedom depending on the number of parameters in each model.

Unfortunately, the same test does not carry over to VCM models. This is because the VCM model is non-parametric, meaning the parameter space is infinite dimensional, hence standard degrees of freedom calculations do not apply. Instead I apply a result in Fan, Zhang and Zhang (2001), which states that the deviance difference, call it $\tilde{D}(H_1,H_0)$, still has a limiting normal distribution as the number of observations, n, gets large, under the null hypothesis, H_0, and this holds even for infinite dimensional models.

Furthermore, the mean and variance of the Normal distribution is free of nuisance parameters. For details see Theorem 5, 10, and the remarks that follow in (Fan et al., 2001)

This motivates the following bootstrap procedure for calculating p-values to test the null hypotheses considered in Supplemental Table 1:

(1) Fit regression coefficients $\hat{\beta}_0$ under the null model.

(2) Sample the citation counts, Y_n, by a parametric bootstrap. In particular, for b = 1,…, B, Y_n^b is sampled from a Poisson distribution with parameter $\hat{\beta}_0$.

(3) Re-estimate VCM model coefficients on sampled data, $\hat{\beta}_1^b$, for each sample Y_n^b. Do the same thing for null model coefficients, $\hat{\beta}_0^b$.

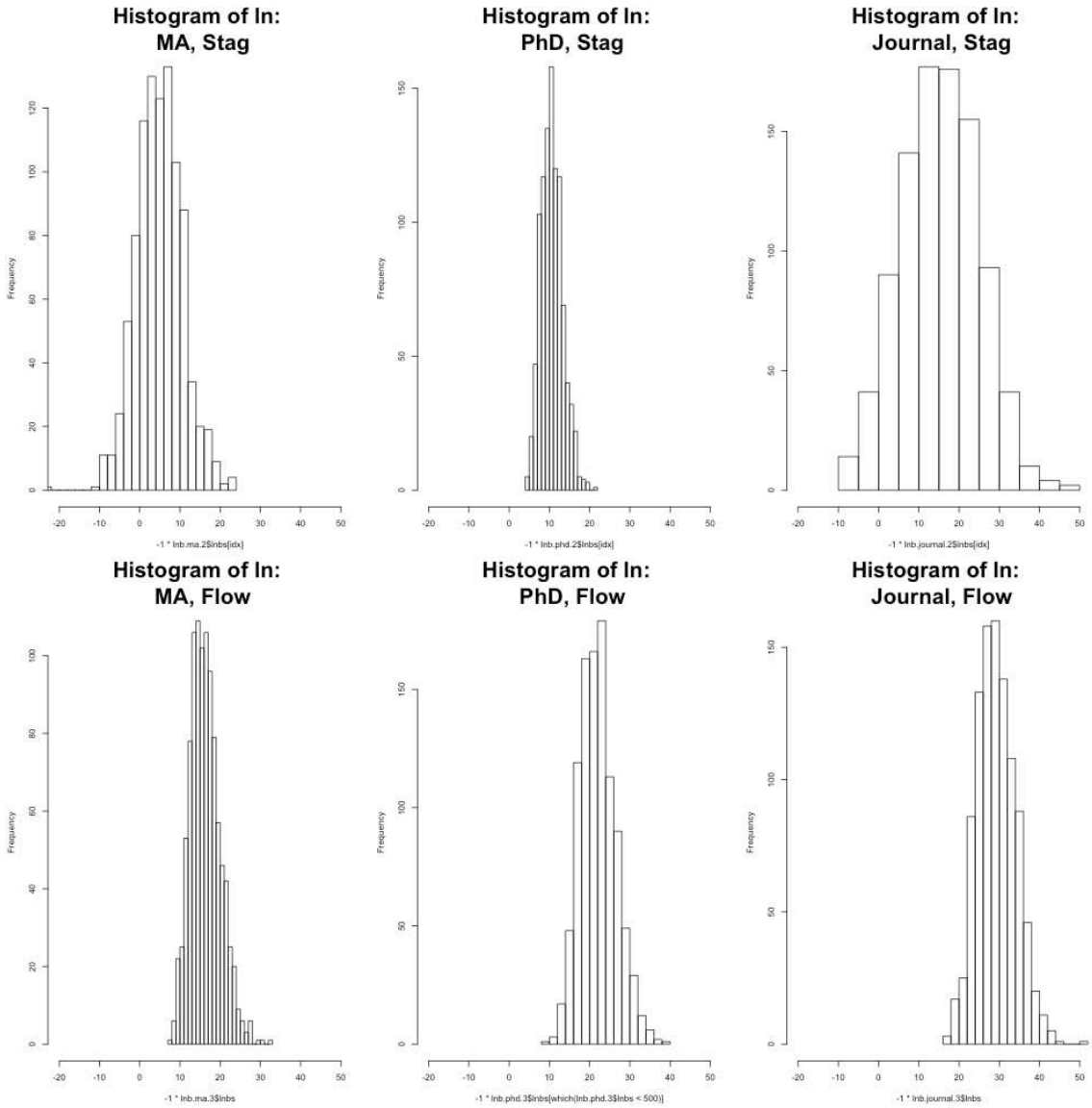
PeerJ PrePrints

(4) Compute bootstrapped deviance differences, $\tilde{D}(H_1, H_0)^b$, now distributed under the null.

(5) The p-value is then the number of $\tilde{D}(H_1, H_0)^b$ which are greater than the original, non-null deviance difference, $D^* = \tilde{D}(H_1, H_0)$, divided by B.

Supplemental Figure 1 shows the results of bootstrapping 1000 samples from each of the 6 null hypothesis examined, research stagnation and research flow for each dataset - MA theses, doctoral dissertations and journal articles. The finding revealed that the original, non-null test statistics, D^*, are all greater than their respective largest bootstrapped test statistic. Thus, by the bootstrap procedure outlined above, p-values were estimated for rejecting the null hypotheses of research stagnation and perfect research flow for each of the three datasets to all be < .001. This gives evidence that the VCM models (model 3) are a more accurate representation of how research citations evolve over time in the field of CIS than their non-varying counterparts (models 1 and 2).

PeerJ PrePrints



Supplemental Figure 1: Empirical null distributions of $\tilde{D}(H_1, H_0)$ using the bootstrap procedure and taking B = 1000 samples.

Each histogram corresponds to either the null hypothesis of research stagnation (Stag) or perfect research flow (Flow) and one of the three citation datasets – MA Theses (MA), Doctoral dissertations (PhD), and Journal articles (Journal). The non-null deviance differences of D^* from left to right, top to bottom are: 117.42, 36.45, 1126.24, 307.36, 52.18, 4295.79. Note these values do not exactly match the differences of the values in Supplemental Table 1 as those values are all normalized to have the saturated model

PeerJ PrePrints

deviance 0. Since all non-null test statistics are greater than their respective largest bootstrapped null test statistics, it is concluded that the p-value for each null hypothesis is < .001.

PeerJ PrePrints