



REVIEW ARTICLE



<https://doi.org/10.1057/s41599-024-03080-8>

OPEN

A systematic synthesis and analysis of English-language *Shuōwén* scholarship

Melissa Xiaohui Qin^{1,2✉}

This study documents the past and present of English scholarship of *Shuōwén jiězì*, the most important masterpiece of classical Chinese philology and lexicography. In an attempt to investigate the current status (synchronically) and developmental trajectory (diachronically) of *Shuōwén* scholarship, this study surveyed the bibliographical, theoretical, methodological and thematic state of *Shuōwén* literature manifested in an analytical framework of six parameters (geographic distribution, publication outlets, citations, theoretical approaches, research methods and research strands) by adopting a thorough bibliometric analysis subsequent to meticulous coding in terms of these six parameters in *Shuōwén* scholarship. The results showed that *Shuōwén* scholarship has displayed distinctive distribution patterns and chronological trends along bibliographical, theoretical, methodological and thematic dimensions. Synchronically, the research output is regionally based, thematically restricted and qualitatively examined; diachronically, the research quality is fluctuating, though research quantity multiplies steadily. To look ahead, the study sketched a few directions for advancement of future research agendas: a more global profile of researchers, a larger degree of internationalization of the native circle, a wider disciplinary range, and increased specialization streamlined into data-driven approach to core issues of liùshū, bùshǒu and GPS (graphic-phonetic-semantic) of head graphs. The study is useful in portraying an overarching disciplinary landscape for *Shuōwén* research and providing the basic reference statistics for possible comparisons across classical Chinese studies and for more general conclusions about the state of scholarship on sinology.

Introduction

S *huōwén jiězì* 说文解字, “the most important monument of ancient Chinese linguistic and lexicographical scholarship” (Miller, 1953, p.vi), or the “first kind of book under the sky” (Tang, 2018, p.1), has been the object of commentaries and textual criticisms in traditional Chinese philology for two thousand years, culminating in the Qing dynasty with the appearance of Four Great *Shuōwén* Scholars (Duàn Yùcái, Guì Fù, Wáng Yún, Zhū Jùnshēng). A survey of native *Shuōwén* Scholarship (*Shuōwén xué* 说文学) (Dong and Zhang, 1988) has indexed over 1200 works dating from Tang dynasty (618-907AD) through 1985. An initial search in CNKI (the most inclusive platform for Chinese research resources), returns 3798 results from 1935 to 2023 (by 15 Jan 2023). While a gigantic body of Chinese *Shuōwén* scholarship has accumulated

¹University of Science and Technology Beijing, Beijing, China. ²University of California, Berkeley, US. ✉email: melissaxiaohuiqin@berkeley.edu

covering an extraordinary spectrum of facets and features surrounding linguistic and extralinguistic themes, the state of field of overseas scholarship is, on the contrary, rather vague. Other than one study (Boltz, 1993) which sketched some fundamental information concerning its author, context, structure and textual history and presented an essential bibliographical guide comprising mainly Chinese scholarship plus a few western works (7 studies) and Japanese works (1 study), there has not been a systematic synthesis depicting the panoramic landscape of English-language *Shuōwén* scholarship.

The study aims to address the urgency of filling the critical void in the realm of English *Shuōwén* scholarship by systematically presenting and analyzing the bibliographical, theoretical, methodological and thematic parameters of English *Shuōwén* literature. As a pioneering systematic synthesis, its significance is twofold. In terms of findings, the study charted the state of field, identified knowledge gaps and determined fruitful research agendas for scholars interested in *Shuōwén*-related topics in particular and Chinese script and language in general. In terms of methods, the study established an analytical framework incorporating six parameters to render a comprehensive as well as nuanced picture of topic-specific scholarly literature. To be precise, English-language *Shuōwén* scholarship is defined in the present study as scholarly works published in the medium language of English devoted to either *Shuōwén* or its pertinent commentaries (Table 2 presents the specific inclusion and exclusion criteria). To limit the publication language to English is a major methodological flaw, as it excludes a significant portion of works written in languages other than English (such as French, German and Japanese) especially before the mid-20th century when English had yet acquired the global dominance as the lingua franca of academic publication (Canagarajah, 2002). However, the demarcation line is artificially drawn to form a definitive scope for the study to be manageable.

Rationale of reviewing parameters

The reviewing taxonomy of the present study is adapted from the framework of Veletsianos and Shepherdson (2016) in their efforts to synthesize the empirical MOOC (Massive Open Online Courses, an ecosystem of online learning environments with a spectrum of course designs) literature in 2013–2015. The original framework consists of five parameters which form the bases of scientometrics: geographic distribution, publication outlets, citations, data collection and analysis methods and research strands. For the purposes of the present study, three adaptations were made. First, a new parameter “theoretical approach” was added to investigate the range of theoretical stances adopted (consciously and explicitly) or embedded (unconsciously and implicitly) in the extant scholarship. Second, the parameter “publication outlets” was adjusted into “publication types” as the range of literature in the original framework was limited to journal articles and conference proceedings while the present study aims at a totality survey of publication types. Third, the parameter “data collection and analysis methods” was replaced by “research methods” for the reason that the original framework was utilized to map only empirical literature (to which data collection and analysis are particularly pertinent) while the present review captures a preliminary overall glimpse of *Shuōwén* publications, in which circumstance, “research methods” is a suitable parameter for assessing its methodological status.

Therefore, the reviewing taxonomy in the present study consists of six parameters: geographic distribution, publication types, citations, theoretical approaches, research methods and research strands (Table 1), by the combination of which, it is hoped to produce a comprehensive and systematic synthesis of the English *Shuōwén* scholarship from its genesis till present.

To address the research gaps identified in Section 1, the following research questions are formulated in light of the six aforementioned reviewing parameters:

RQ1: What is the current status of English-language *Shuōwén* literature?

RQ1a: How is the *Shuōwén* literature geographically distributed?

RQ1b: What are the publication types of *Shuōwén* literature?

RQ1c: Which *Shuōwén* studies are cited the most?

RQ1d: What theoretical approaches are adopted in the *Shuōwén* studies?

RQ1e: What research methods are used in the *Shuōwén* scholarship?

RQ1f: What research strands have appeared in the *Shuōwén* literature?

RQ2: What are the diachronic trends of English-language *Shuōwén* literature?

Research methods

This study adopts the systematic literature review as the research design. The systematic literature review uses systematic, explicit and reproducible methodology to identify, evaluate and synthesize the body of completed works produced by researchers (Moher et al. 2009; Okoli and Schabram, 2010; Petticrew and Roberts, 2006). Compared with other review designs such as narrative review (mainly theoretical and qualitative), the rigorous methodology of a systematic review is able to address specific quantitative research questions, prevent possible procedural bias, allow replicability of data and enhance reliability and validity of the review (Xiao and Watson, 2019).

Data collection

Literature search. While the search scope is limited to English-language publications, a publication type restriction was not set in order to yield the entire range of the literature. Likewise, a time frame was not set either. The literature discovery process consisted of three sequential stages.

Stage (1) Keywords syntax—In order to identify the keywords that would maximize the number of relevant hits, multiple searches in major databases were performed, taking note of the various forms in which *Shuōwén* appears. This process lasted 3–4 days and generated the final list of eight keywords: *Shuo-wen chieh-tzu*, *Shuo wen chieh tzu*, *Shuowen jiezi*, *Shuo wen jie zi*, *Shuowenjiezi*, *Shuowen*, *Shuo wen*, *Shuo-Wen*. While these keywords will not cover everything related to *Shuōwén* research in English, a search utilizing these eight keywords is likely to identify the works that take *Shuōwén* as their main object of study, which is the focus of this study.

Stage (2) Database queries—Databases of Web of Science, Scopus, ProQuest, WorldCat were chosen to encompass the maximum range of publication types. These four databases were retrieved in June 2022 by applying the eight search terms respectively in each database ($k=199$). As the search operator setting of each database is different, the author did not combine the keywords but applied each one in each database, all together eight queries for each database, and 32 queries for the four databases.

Table 1 The six-parameter taxonomy of the present study.

Parameter	Value	Exemplars
parameter 1	geographical distribution	authors' affiliation Harvard University
parameter 2	publication type	the form in which the study was published book, doctoral dissertation
parameter 3	citation	citation counts in Google Scholar See Fig. 1
parameter 4	theoretical approach	the theoretical stance/ disciplinary framework/ conceptual perspective taken, or theoretical lens through which <i>Shuōwén</i> was approached grammatology, lexicography, intellectual history
parameter 5	research method	procedures or techniques used in collecting and analyzing data qualitative, quantitative, mixed-methods
parameter 6	research strand	the focus of the study, i.e., which area, theme or topic is concentrated on liùshū 六书, bùshǒu 部首

Table 2 Inclusion and exclusion criteria.

Criterion	Inclusion	Exclusion
language	English	non-English studies
publication type	studies that fall into the range of formally published items such as journal articles, monographs, book chapters, edited chapters, conference proceedings, doctoral dissertations, MA theses, etc.	studies that were not published and therefore not publicly accessible, e.g., oral presentations in academic conferences.
research scope	studies that took <i>Shuōwén</i> as its exclusive or primary object of academic inquiry, or addressed it extensively.	studies that scrutinized more general topics such as Chinese writing systems which made a mere passing or token reference to <i>Shuōwén</i> or which utilized <i>Shuōwén</i> as an inevitable reference work instead of object of inquiry.

Xu Shen's Scholarly Agenda: A New Interpretation of the Postface of the *Shuowen jiezi*

T O'neill - *Journal of American Oriental Society*, 2013 - JSTOR

... As Xu Xuan states, “分部相自許始也; dividing words into mutually distinct bushou first began with Xu Shen” (Xu ... the standard—which is less than half of Xu Shen's original 540 bushou. ...

☆ Save  Cite Cited by 25 Related articles All 6 versions Web of Science: 7

Fig. 1 An example of how Google Scholar was used in the forward referencing process. This figure shows the number of works that cited the original. The bibliographical information of the 25 works could be obtained by clicking “Cited by 25”.

Stage (3) Google Scholar retrieval—Google Scholar was chosen for its comprehensive coverage of publication types (“any type” for the literature type) and unlimited time period (“any time” for the time range). It was manually screened with the same search terms to locate relevant hits (k = 74).

Literature selection. The literature search has yielded an original dataset of 273 items, based on which, automatic and manual selections were conducted. The literature selection comprised four stages.

Stage (1) Duplication removal—94 duplicates were removed from the original dataset, the summed search results of database queries and Google Scholar retrieval (For instance, as both Scopus and Google Scholar returned the item “Xu Shen's scholarly agenda: A new interpretation of the postface of the *Shuowen jiezi*”, only one item was kept in the dataset.).

Stage (2) Title and abstract screening—Titles and abstracts of the 179 records were examined and 112 articles were excluded either because their topics are apparently irrelevant (such as “mining and metallurgical technology”) or because they were written in non-English languages (although the search language is limited to English, WorldCat would still return multiple-language results). To establish reliability of the screening process, every record was independently screened by two researchers (the author and a research assistant). Disagreements were resolved by repeated discussion.

Stage (3) Full text evaluation—The full texts of the remaining 67 publications were assessed for eligibility by applying the inclusion and exclusion criteria presented in Table 2 (For instance, the item “Self as the intersection of traditions: The autobiographical writings of Ssu-ma Ch'ien” has only one occurrence of *Shuōwén* in the main body as a reference source to validate the meaning of a character. In this case, it meets the exclusion criterion of research scope in Table 2). In terms of screening consistence and reliability, the author independently analyzed the whole texts of the 67 items and invited an experienced researcher to evaluate four borderline cases. A consensus was reached after thorough discussion. By the end of the screening process, 33 studies were included in the dataset.

Stage (4) Citation checking—Simultaneous with the full text reading, references of the identified 67 items were manually checked, and additional records not returned by the previous searches were manually included (k = 3). The dataset obtained so far comprised 36 items.

Stage (5) Forward referencing—Google Scholar provides information on how many times a study is cited and allows researchers to view all publications citing the original (Fig. 1). To utilize this information, the author entered each of the 36 items in Google Scholar, and examined all the studies that cited each original. This process returned one new additional document.

Table 3 Coding framework for the six reviewing parameters of *Shuōwén* literature.

Variable	Coding scheme	Value
geographical distribution	open	authors' affiliation as recorded in the study
publication types	open	the form in which the study was published
citations	open	citation counts in Google Scholar
theoretical approaches	open	theoretical stances adopted
research methods	pre-constructed	non-empirical, qualitative, quantitative and mixed methods
research strands	open	study focus

Additional literature search and selection after data analysis. The literature search and selection were completed in two months (June and July 2022). Five months later (when the data coding process ended in November 2022), in order to capture any item published in this lapse of time, the identical literature search and selection were performed once more with the time frame of January 2022 to November 2022, as a result of which, one item was located, creating a 38-item dataset.

Independent literature search and selection in the manuscript revision. The independent literature search and selection arises from the additional bibliography (14 bibliographical items and two bibliographical lists) provided by the anonymous reviewers in the first round of manuscript review (May–Nov 2023). This process consists of three sequential stages.

Stage (1) Full text evaluation—The full texts of the 14 bibliographical items were read. As a result, two items were excluded for the reason that the texts made sporadic mention of *Shuōwén*. The dataset amounted to 50 items.

Stage (2) Citation checking—The references of the 14 bibliographical items and the references included in the two bibliographical lists were scrutinized. As a result, four new items were located, creating a 54-item dataset.

Stage (3) Database queries and Google Scholar retrieval—A new round of literature search and selection identical with the previous phase was conducted to capture any item published after the manuscript submission. The time frame was set between December 2022 and November 2023. Finally, two items were selected, resulting in a 56-item dataset.

Limitations. The literature search methods are most effective in identifying stand-alone pieces of work with an exclusive or primary focus of *Shuōwén* but less effective for non-independent items (embedded pieces of work such as monograph sections) which encompass *Shuōwén* as one of a plethora of sub-themes. That obscure items might have escaped the search procedures limited the size and diversity of the sample.

Data analysis

Qualitative analysis. The literature dataset consists of 56 publications, each of which forms the basic unit of analysis. The author independently coded all the items for parameters 1–6 and invited the experienced researcher (the one I consulted in the full-text screening) to settle uncertain cases for parameters 4–6. The coding of parameters 4–6 involved whole-text analysis, which turned out to be the most time-consuming yet most rewarding stage of the research. For the present analysis, the author has intensively read and meticulously coded all 56 publications and consulted the external researcher for unsettling items. The coding framework is summarized in Table 3. Coding procedures for each parameter are reported as follows.

To determine the geographical distribution of the *Shuōwén* literature, affiliations of authors as recorded in the publications were coded in two ways: 1) by the country in which their

institution or organization was located (or, if unaffiliated, by the country in which the author was located), and 2) by the associated region. For example, the same author Rickard Gustavsson was coded as locating in “Netherlands” (Leiden University) for his MA thesis (Gustavsson, 2016) and “Hong Kong” (City University of Hong Kong) for his PhD dissertation (Gustavsson, 2022). Notably, to record the affiliation as shown in the publication has a limitation as “in the world of academia, geographic determinism and national boundaries are breaking down” (Mair, 2013, p.392), and locations prior to (including academic training) and after the publication are not reflected.

The publication types of the collection were coded by the type in which the study was published and the results were classified into journal articles, monographs, monograph chapters, monograph sections, edited book chapters, edited book sections, article sections, doctoral dissertations, MA theses and translation.

The citations of each article were determined by identifying each item in Google Scholar and noting its citation counts accordingly (by 30 Nov 2023).

The coding of theoretical approaches is of an open nature, for which the theoretical stance or disciplinary framework was identified. As it is not uncommon to find one study defines and highlights different dimensions of its topic from two or more approaches, only the explicitly stated, or in many cases the implicitly stated but predominating approach was designated as the theoretical approach for each article. Notably, establishing a one-to-one corresponding relationship for classification purposes is methodologically flawed as it nevertheless simplifies the multifaceted disciplinary landscape.

The identification of research methods adopted a four-item coding scheme: non-empirical, qualitative, quantitative and mixed-methods. Strictly speaking, none of the 56 studies has adopted the IMRD structure (Introduction, Methods, Results, and Discussion, the common organizational structure for most scientific articles) and reported rigorous data collection and analysis procedures. By consulting methodology resources (Christensen et al. 2021; Patten and Newhart, 2018) and considering the identified dataset, empirical studies were regarded in a broader and looser sense in the present circumstance as studies which have consciously sought to establish the correspondence between findings and evidence, and derived their conclusions from a certain set of data, either primary or secondary, without explicitly elucidating an identifiable research design (from problem formulation to conclusion generation) and especially without explicitly specifying a systematic set of data gathered and analyzed under controlled conditions with rigorous strategies. On the other hand, non-empirical studies are those predominantly conceptual and speculative with impressionistic resort to sporadic examples (e.g., position paper, discussion, introductory overview, book review).

Though the diversity of techniques and philosophical stances that constitute qualitative and quantitative research eludes simple definition, it would not invite much controversy to say their key distinction lies in the types of data collected by the researchers (Friedman, 2012; Phakiti and Paltridge, 2015). In the present

Table 4 Frequency (%) of each region among author affiliations.

Region	Frequency (%)
North America	30 (48.4%)
Europe	18 (29.0%)
Asia	13 (21.0%)
Oceania	1 (1.6%)

Table 5 Frequency (%) of each location among author affiliations.

	Location	Frequency (%)
1	United States	30 (48.4%)
2	China mainland	8 (12.9%)
3	France	5 (8.1%)
4	United Kingdom	3 (4.8%)
5	Germany	3 (4.8%)
6	Israel	2 (3.2%)
7	Denmark	2 (3.2%)
8	Netherlands	2 (3.2%)
9	Australia	1 (1.6%)
10	Czech Republic	1 (1.6%)
11	Slovakia	1 (1.6%)
12	Thailand	1 (1.6%)
13	China Taiwan	1 (1.6%)
14	Sweden	1 (1.6%)
15	China Hong Kong	1 (1.6%)

context, studies which collected numerical data (numbers) to determine relationships between variables were coded as quantitative; studies which collected non-numerical data (words) to explore the context of phenomena were coded as qualitative; studies which included “a mix of qualitative and quantitative methods” were coded as mixed-methods (Brannen, 2005, p.4). It is noteworthy to mention that most of the empirical studies identified in the corpus were qualitative. While some of them employed occasional frequency counting or percentage calculation, they were still counted as “qualitative” instead of “mixed-methods”. Only qualitative studies with systematic statistics were coded as “mixed-methods”. In terms of coding reliability, six out of the 56 studies invited additional assessment for disagreement reconciliation.

To determine the research strands in the dataset, each study was designated with a code describing its study focus. No pre-determined limits were set on the number of emerging codes. While most studies focus on one primary issue, studies which address multiple themes were coded as “comprehensive”. In terms of coding reliability, five out of the 56 items were further discussed with the external researcher to resolve disputes.

Quantitative analysis. The frequencies and percentages for the values of the six parameters in the coding framework were calculated and compared both synchronically and diachronically, providing a descriptive numerical analysis of the state of *Shuōwén* scholarship.

Results

RQ1: What is the current status of English-language *Shuōwén* literature?

*RQ1a: How is the *Shuōwén* literature geographically distributed?.* Table 4 shows that the 56 *Shuōwén* studies were authored by 49 researchers dispersed in four regions but heavily concentrated in North America (48.4%) and Europe (29.0%). The rest of the

Table 6 Frequency (%) of each publication type.

	Publication types	Frequency (%)
1	peer-reviewed journal article	16 (28.6%)
2	book review	7 (12.5%)
3	monograph section	7 (12.5%)
4	doctoral dissertation	6 (10.7%)
5	monograph chapter	5 (8.9%)
6	edited book chapter	4 (7.1%)
7	MA thesis	3 (5.4%)
8	conference proceeding	2 (3.6%)
9	monograph	1 (1.8%)
10	edited book section	1 (1.8%)
11	journal article section	1 (1.8%)
12	doctoral dissertation chapter	1 (1.8%)
13	doctoral dissertation section	1 (1.8%)
14	translation	1 (1.8%)

Table 7 Journals publishing the *Shuōwén* literature.

	Journal name	Frequency (%)
1	<i>Journal of the American Oriental Society</i>	3 (16.7%)
2	<i>Journal of Chinese linguistics</i>	2 (11.1%)
3	<i>Journal of Chinese Writing Systems.</i>	2 (11.1%)
4	<i>Bulletin of the Institute of History and Philology</i>	2 (11.1%)
5	<i>China Review International</i>	1 (5.5%)
6	<i>International Journal of Humanities and Arts Computing</i>	1 (5.5%)
7	<i>Bulletin of the Museum of Far Eastern Antiquities</i>	1 (5.5%)
8	<i>Monumenta Serica: Journal of Oriental Studies</i>	1 (5.5%)
9	<i>Oriens</i>	1 (5.5%)
10	<i>Journal of the International Folk Music Council</i>	1 (5.5%)
11	<i>Archív Orientální</i>	1 (5.5%)
12	<i>Asia Major</i>	1 (5.5%)
13	<i>International Journal of Lexicography</i>	1 (5.5%)
14	<i>Journal of Pragmatics</i>	1 (5.5%)
15	<i>Studia Orientalia Slovaca</i>	1 (5.5%)
16	<i>Bulletin of the School of Oriental and African Studies</i>	1 (5.5%)
17	<i>International Journal of the Sociology of Language</i>	1 (5.5%)
18	<i>Mínsú diǎnjí wénzì yánjiū 民俗典籍文字研究</i>	1 (5.5%)

research was conducted in Asia (21.0%) and Oceania (1.6%). According to Table 5, these authors were distributed in 15 locations, the majority of whom were affiliated with institutions from United States (48.4%). One third came from four locations: China mainland (12.9%), France (8.1%), United Kingdom (4.8%), Germany (4.8%). The rest 21.0% represented 10 locations that had one or two authors each.

*RQ1b: What are the publication types of *Shuōwén* literature?.*

Table 6 displays that 14 publication types have accommodated the 56 items: peer-reviewed journal articles (28.6%), book reviews (12.5%), monograph sections (12.5%), doctoral dissertations (10.7%), monograph chapters (8.9%), edited book chapters (7.1%), MA thesis (5.4%), conference proceedings (3.6%), monograph (1.8%), edited book section (1.8%), journal article section (1.8%), doctoral dissertation chapter (1.8%), doctoral dissertation section (1.8%) and translation (1%). The 16 peer-reviewed journal articles and the seven book reviews were published in 18 journals dispersed in a wide range of linguistic, history and arts topic areas such as sinology, philology, pragmatics, lexicography, sociology, philosophy, archaeology and music. Except for one journal which has published three studies and three journals which have published two studies, the other 14 journals have published one study each (Table 7). Notably, *Mínsú diǎnjí wénzì yánjiū* 民俗典籍文字研究

Table 8 Studies cited most frequently (over 10 citations by 30 Nov 2023 in Google Scholar).

Publication	Citation count
1 Bottéro, F. & Harbsmeier, C. (2008). The "Shuowen Jiezi" dictionary and the human sciences in China. <i>Asia Major</i> , 21 (1), 249-271.	63
2 Boltz, W. G. (1993). Shuo wen chieh tzu 說文解字. In L. Michael (Ed.), <i>Early Chinese texts: A bibliographical guide</i> (pp. 429-442). Society for the Study of Early China.	40
3 Bottéro, F. (2002). Revisiting the wén 文 and the zì 字: The great Chinese characters hoax. <i>Bulletin of the Museum of Far Eastern Antiquities</i> , 74, 14-33.	40
4 Miller, R. A. (1953). Problems in the study of the Shuo-wen chieh-tzu. [Doctoral dissertation.] Columbia University.	28
5 Thern, K. L. (1966). <i>Postface of the Shuo-wen chieh-tzu: The first comprehensive Chinese dictionary</i> . Wisconsin: University of Wisconsin.	24
6 O'Neill, T. M. (2013). Xu Shen's scholarly agenda: A new interpretation of the postface of the <i>Shuowen jiezi</i> . <i>Journal of the American Oriental Society</i> , 133(3), 413-440.	24
7 Coblin, W. S. (1978). The initials of Xu Shen's language as reflected in the <i>Shuowen duruo</i> glosses. <i>Journal of Chinese linguistics</i> , 6 (1), 27-75.	17
8 Creamer, T. B. (1989). Shuowen jiezi and textual criticism in China. <i>International Journal of Lexicography</i> , 2(3),176-187.	14
9 Hulsewé, A. F. P. (1959). The Shuo-Wen Dictionary as a source for ancient Chinese law. In S. Egerod, & E. Glahn (Eds.), <i>Studia Serica Bergard Karlgren Dedicata</i> (pp.239-258). Copenhagen: Ejnor Munksgaard.	11
10 Serruys, P. (1984). On the system of pu shou in the Shuo-wen chieh-tzu. <i>Bulletin of the Institute of History and Philology, Academia Sinica</i> , 55(4), 651-754.	10

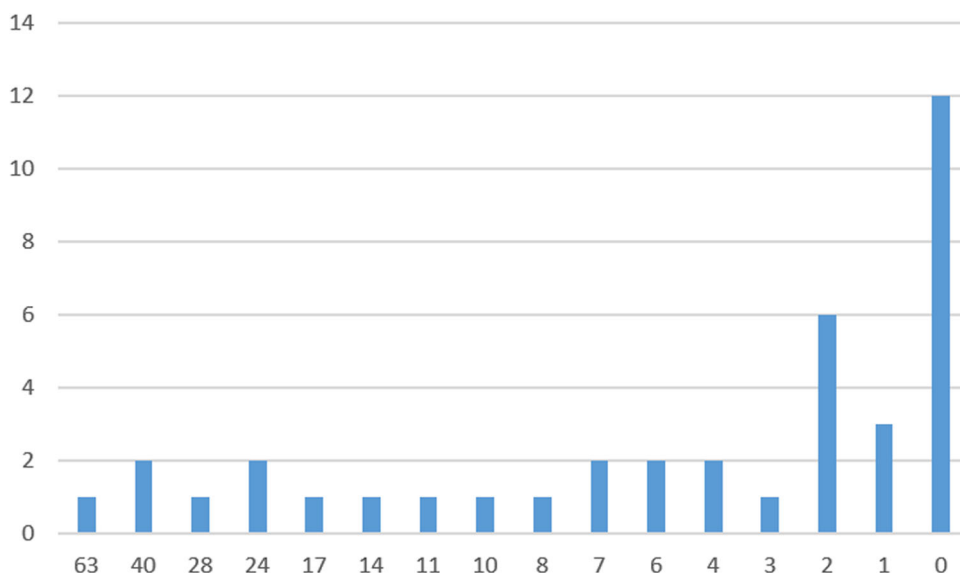


Fig. 2 Distribution of items for each citation count. The horizontal axis represents the citation counts; the vertical axis represents the number of publications.

is predominantly a Chinese journal hosted by Beijing Normal University in mainland China and only occasionally publishes English-language scholarship.

RQ1c: Which Shuōwén studies are cited the most?. As the citation counts of non-independent items (where the *Shuōwén* text was attached to a larger work such as a chapter embedded in a monograph) could only return the citation results of the mother work (instead of the embedded item), 16 items (seven monograph sections, five monograph chapters, one edited book section, one journal article section, one doctoral dissertation chapter and one doctoral dissertation section) were excluded from the citation analysis. In the remaining 40 stand-alone items, one item was not found in Google Scholar. Therefore, 39 items were searched for their citation counts. By 30 November 2023, 10 studies were cited over 10 times (Table 8) and 29 items below 10 times, among which 12 items were cited 0 time (Fig. 2).

RQ1d: What theoretical approaches are adopted in the Shuōwén studies?. The open coding process has identified 20 theoretical

approaches (Table 9). While nine approaches have two to eight studies (grammatology 8; intellectual history 8; content review 7; lexicography 5; computer linguistics 4; historical linguistics 4; archaeology 4; translation 3; hermeneutics 2), the other 11 approaches have only one study each.

RQ1e: What research methods are used in the Shuōwén scholarship?. According to Table 10, nearly half of the studies (25, 44.6%) are non-empirical while the rest (31, 55.4%) are empirical. The 31 empirical studies included 26 qualitative, one mixed-methods and four digitalization. Digitalization is a new code which has emerged in data analysis, which includes four computerization projects of the *Shuōwén* text. None of the empirical studies has used a pure quantitative method. The quantitative portion of the one mixed-methods study is descriptive statistics.

RQ1f: What research strands have emerged in the Shuōwén literature?. The open coding yielded 29 codes designating the focus of each study and 16 categories emerged describing these codes. According to Table 11, apart from the most frequently appeared

strands book review (7, 12.5%) and compiling principles (7, 12.5%), other major research strands are GPS of head graphs (6, 10.7%; GPS is a term borrowed from Boodberg, 1937, meaning the graphic-phonetic-semantic scheme of character analysis), introductory overview (6, 10.7%), liùshū 六书 (six categories of character formation, 5, 8.9%), text computerization (4, 7.1%) and native scholarship (4, 7.1%). The rest nine strands occurred much less frequently from 1.8% to 5.4%. A more refined picture is provided by looking at the specific codes in each major research strand. For instance, in the five liùshū studies, three provided an

overview of the theory, and the other two focused on jiǎjiè 假借 and zhuǎnzhù 转注 respectively. In the six studies exploring the GPS of head graphs, two aimed at reconstructing the phonetic systems of certain head graphs, one at reconstructing the GPS of certain head graphs, one at classifying the feminized entries according to their semantic glosses, one at scrutinizing the social historical wealth of entry semantics and one at uncovering the script styles and sources used by Xǔ Shèn.

Table 9 Theoretical approaches of the *Shuōwén* literature.

	Research approach	Frequency (%)
1	grammatology	8 (14.3%)
2	intellectual history	8 (14.3%)
3	content review	7 (12.5%)
4	lexicography	5 (8.9%)
5	computer linguistics	4 (7.1%)
6	historical linguistics	4 (7.1%)
7	archaeology	4 (7.1%)
8	translation	3 (5.4%)
9	hermeneutics	2 (3.6%)
10	bibliography	1 (1.8%)
11	history of science	1 (1.8%)
12	law	1 (1.8%)
13	lexical pragmatics	1 (1.8%)
14	music	1 (1.8%)
15	social history	1 (1.8%)
16	structural linguistics	1 (1.8%)
17	textual criticism	1 (1.8%)
18	theoretical linguistics	1 (1.8%)
19	metalinguistics	1 (1.8%)
20	sociolinguistics	1 (1.8%)

Table 10 Research methods of *Shuōwén* literature.

	Research methods	Frequency (%)
1	non-empirical	25 (44.6%)
2	qualitative	26 (46.4%)
3	mixed-methods	1 (1.9%)
4	digitalization	4 (7.1%)

Table 11 Research strands present in the *Shuōwén* literature.

	Research strands	Frequency (%)	Codes (frequency)
1	book review	7 (12.5%)	
2	compiling principles	7 (12.5%)	authorial intention (2); compilation purpose (1); compiling methodology (1); entry gloss (1); entry organization (2)
3	GPS of head graphs	6 (10.7%)	phonetic systems of certain head graphs (2); semantics of certain head graphs (1); GPS of certain head graphs (1); feminized entries (1); script styles and sources (1)
4	introductory overview	6 (10.7%)	
5	liùshū	5 (8.9%)	overview of liùshū (3); zhuǎnzhù (1); jiǎjiè (1)
6	text computerization	4 (7.1%)	
7	native scholarship	4 (7.1%)	Qing scholarship (2); 10 th century (1); 18–20 th century (1)
8	bùshǒu	3(5.4%)	GPS of bùshǒu (2); semantic system of bùshǒu (1)
9	comprehensive	3 (5.4%)	
10	encyclopedic content	2 (3.6%)	music terms (1); law (1)
11	character evolution theory	2 (3.6%)	
12	postface translation	2 (3.6%)	
13	origin of writing	2 (3.6%)	
14	Xǔ Shèn's other works	1 (1.8%)	
15	key term reinterpretation	1 (1.8%)	
16	recension comparison	1 (1.8%)	

RQ2: What are the diachronic tendencies of *Shuōwén* literature in the past century? This section reports the quantitative results of the diachronic trajectory of *Shuōwén* literature from 1906 to 2023. To delineate chronological trends, three time periods were designated (period I-III: 1906–1979, 1980–1999, 2000–2023) to clarify analysis. As the correlation between citation and time is self-evident, the parameter of citation was not examined diachronically. Instead, the number of publications in each time period was calculated and compared. Therefore, diachronic trends of numerical values of six parameters are reported below: publication number, geographical distribution, publication types, theoretical approaches, research methods and research strands.

Figures 3–6 illustrate the chronological division of number of publications, author affiliations and publication types respectively. According to Fig. 3, the number of publications in the three periods are 12 (21.4%), 15 (26.8%) and 29 (51.8%) with an apparent rise in the last period. According to Fig. 4, an increasing number of locations have accommodated *Shuōwén* publications, namely four in period I (United States 9, 75.0%; United Kingdom 1, 8.3%; Netherlands 1, 8.3%; Sweden 1, 8.3%), six in period II (United States 10, 66.7%; United Kingdom 1, 6.7%; Australia 1, 6.7%; Germany 1, 6.7%; France 1, 6.7%; China mainland, 1, 6.7%) and 13 in period III (United States 11, 31.4%; China mainland 7, 20.0%; France 4, 11.4%; Denmark 2, 5.7%; Germany 2, 5.7%; Israel 2, 5.7%; United Kingdom 1, 2.9%; Netherlands, 1, 2.9%; Taiwan 1, 2.9%; Thailand 1, 2.9%; Slovakia 1, 2.9%; Czech Republic 1, 2.9%; Hongkong 1, 2.9%). This trend could also be observed in the region distribution (Fig. 5): North America dominates the landscape in period I (9, 75.0%) and period II (10, 66.7%) respectively; the three regions, North America, Europe and Asia divide evenly the share in period III. According to Fig. 6, diverse composition of publication types characterizes each period: journal articles (50.5%) occupies half weight in the five publication types in period I; book reviews (4, 26.7%) and

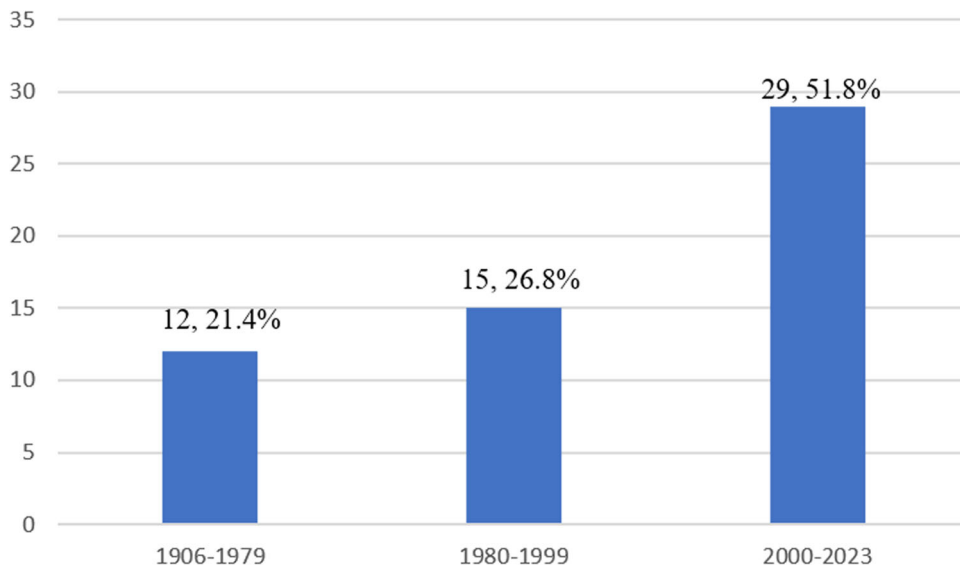


Fig. 3 Number of publications in each time period. The horizontal axis represents the three time periods; the vertical axis represents the number of publications.

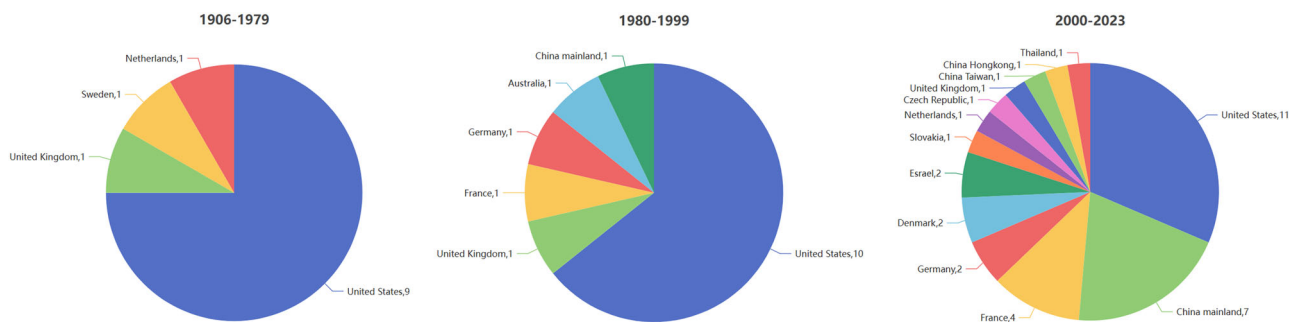


Fig. 4 Geographical distribution of Shuōwén literature in periods I to III. Each panel represents the share of the designated geographical location for the specific time period.

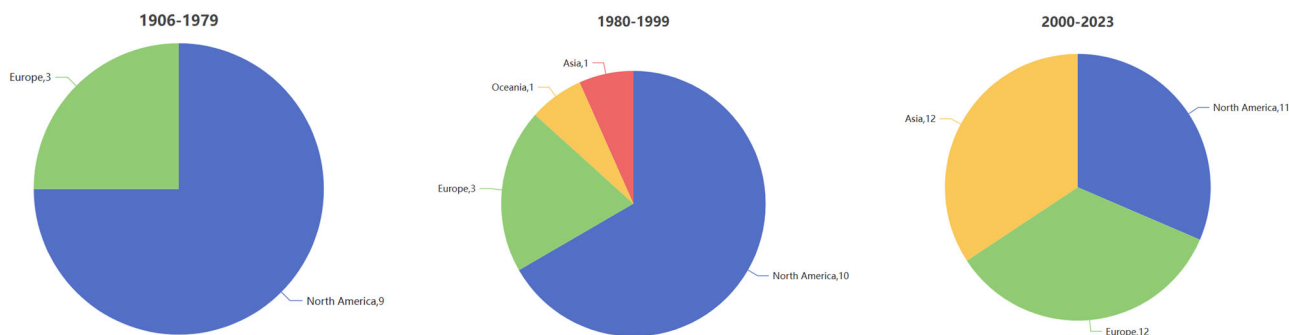


Fig. 5 Region distribution of Shuōwén publications in periods I to III. Each panel represents the share of the designated region for the specific time period.

monograph sections (4, 26.7%) stand out in the six types in period II; period III has seen the largest number of types and a more balanced distribution of each type (journal article 7, 24.1%; monograph chapter 4, 13.8%; book review 3, 10.3%; monograph section 3, 10.3%; MA thesis 3, 10.3%; doctoral dissertation 2, 6.9%; conference proceeding 2, 6.9%; edited book chapter 1, 3.4%; edited book section 1, 3.4%; monograph 1, 3.4%; doctoral dissertation section 1, 3.4%; doctoral dissertation chapter 1, 3.4%).

Figures 7–9 present respectively the graphical illustration of the distribution of theoretical approaches, research methods and

research strands for each time period. According to Fig. 7, the three periods have seen respectively eight, ten and twelve theoretical approaches. While traditional approaches such as archaeology, lexicography and grammatology have appeared in each period, distinctive approaches also feature each period such as encyclopedic approach (law and music) in period I, socio-linguistics in period II and computational linguistics in period III. According to Fig. 8, period I includes two thirds qualitative studies (8, 66.7%) and one third non-empirical studies (4, 33.3%); period II 86.7% non-empirical and 13.3% qualitative studies; period III 58.6% qualitative, 24.1% non-empirical, 13.8%

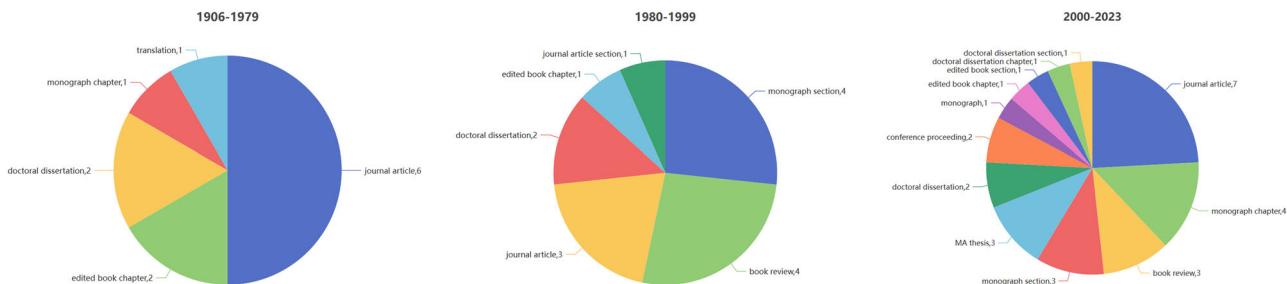


Fig. 6 Publication types of *Shuōwén* literature in the three time periods. Each panel represents the share of the designated publication type for the specific time period.

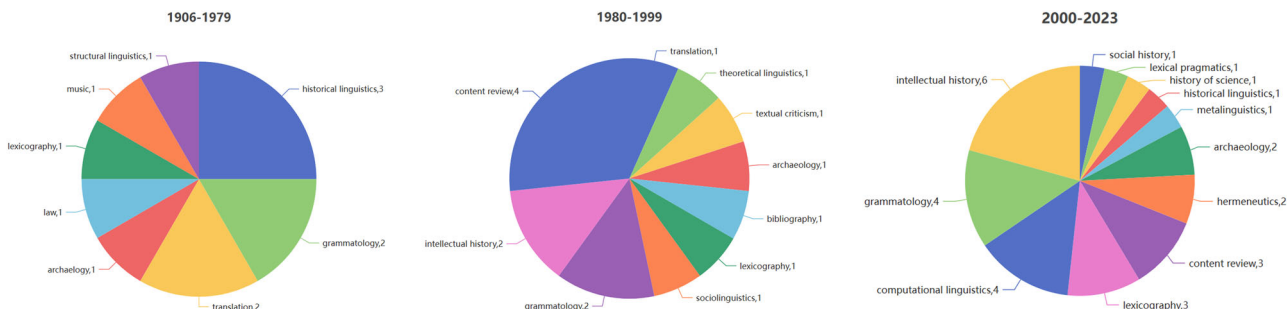


Fig. 7 Theoretical approaches of *Shuōwén* literature in three time periods. Each panel represents the share of the designated theoretical approach for the specific time period.

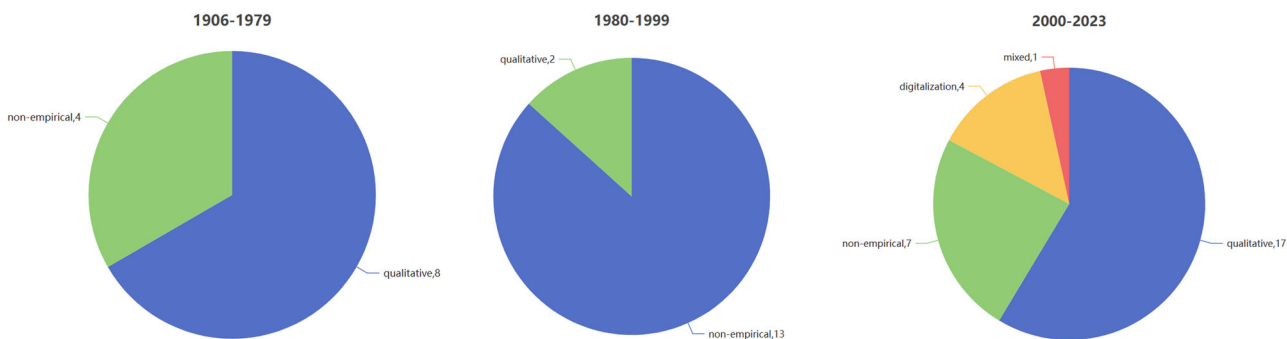


Fig. 8 Research methods of *Shuōwén* literature in three time periods. Each panel represents the share of the designated research method for the specific time period.

digitalization and 3.4% mixed-methods studies. According to Fig. 9, the three periods have respectively seen nine, eight and 20 research strands. While the strands of liùshū, bùshǒu and GPS of head graphs are prevalent across all three periods, distinctive strands exist for each period such as encyclopedic content, postface translation and Xǔ Shèn’s other works in period I, origin of writing in period II and text computerization and compiling principles in period III.

Discussion

The current status of English-language *Shuōwén* scholarship. Given the statistics along the six parameters (geographical distribution, publication types, citation counts, research approaches, methods and themes) reported in the results section, it is not hard to find that *Shuōwén jiězi* is, by no means, a “hot” topic. It is neither thoroughly examined nor highly advanced judging from external parameters such as output quantity, impact, geography and internal parameters such as research methods, approaches and themes. In terms of output quantity, the total number of 56

publications (including 7 book reviews and 20 embedded works such as chapters and sections of larger works) in over one century could not be said impressive if compared with the thousands of returned results by searching Chinese languages and linguistics in the four above-consulted databases (Scopus, Web of Science, ProQuest and WorldCat), where the overwhelming majority is (understandably) devoted to modern standard Mandarin. This observation could be further enhanced by the citation counts in Google Scholar where over half of the studies were cited below 10 times.

In terms of geographical distribution, the past-century’s *Shuōwén* scholarship is a highly regional enterprise, represented in a very small number of locations in over 100-year time span (to illustrate, a glance at Veletianos and Shepherdson 2016 shows that three years’ MOOC literature were contributed by 38 countries). This trait is not surprising as “sinology is not a universally cultivated discipline” (Schafer, 1990, p. 35). English-language *Shuōwén* research is heavily concentrated in Europe (29.0%) and United States (48.4%), the former of which has

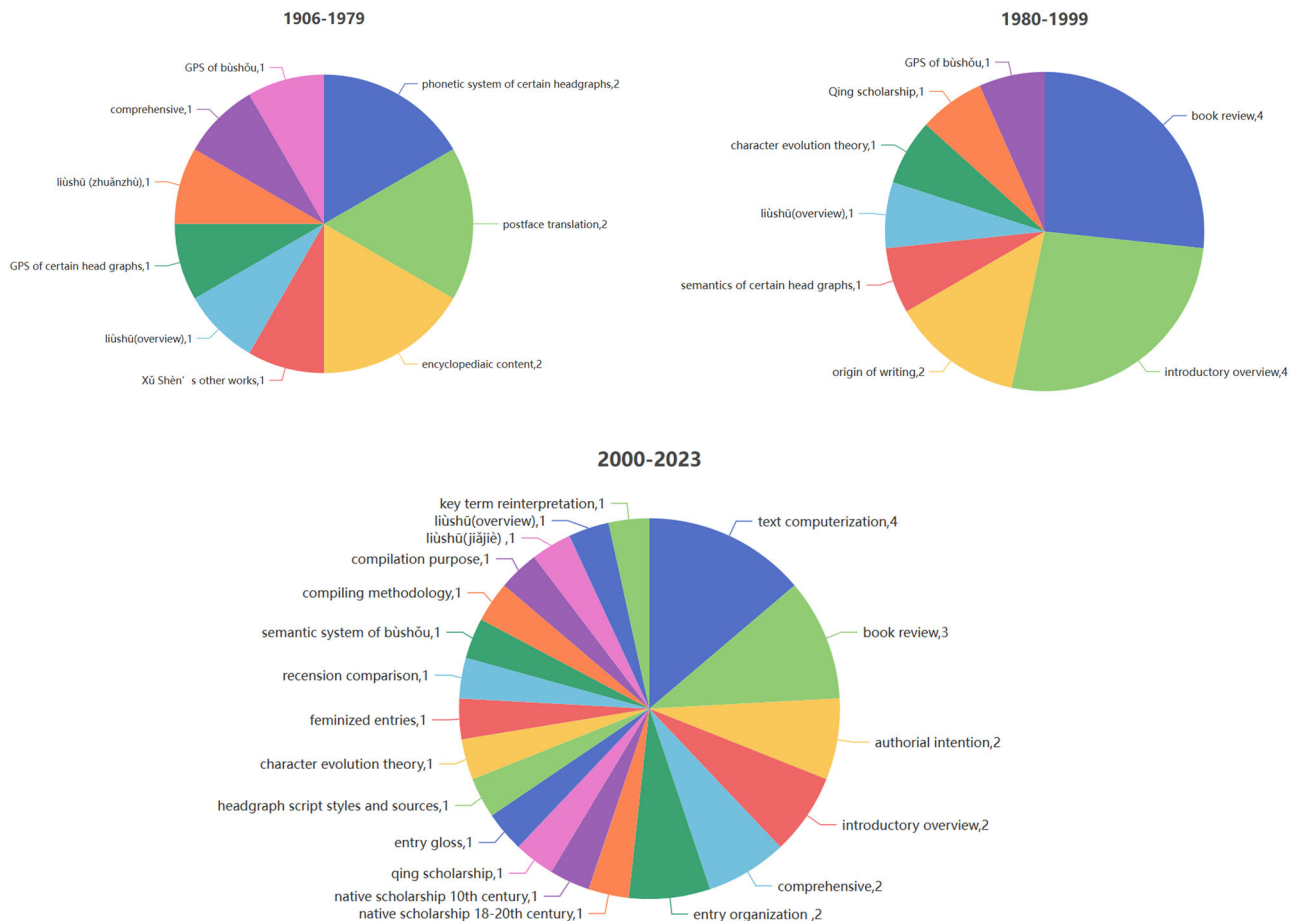


Fig. 9 Research strands of *Shuōwén* literature in three time periods. Each panel represents the share of the designated research strand for the specific time period.

occupied a unique eminence in sinology from the 18th to the mid-20th century and the latter from the mid-20th century till now (Godin, 2002; Schafer, 1990). Among the 15 locations, France, Germany, Netherlands, Sweden, United Kingdom, United States are traditional centers of sinological research, Australia is a major force that has emerged in last century (Schafer, 1990), while Czech Republic, Denmark, Slovakia, Thailand, Israel, Denmark entered the scene in the new millennium. Notably, the handful Chinese mainland authors without international collaboration (8, 12.9%) demonstrate that native scholarship has produced extremely limited English-medium output and is largely isolated from the international academic community, which forms a sharp contrast with the substantial degree of internationalization in other subjects of Chinese languages and linguistics (Mair, 2013).

In terms of publication types, the *Shuōwén* literature has emerged in a wide range of forms with journal articles occupying less than one third (28.6%), which is characteristic of humanistic scholarship with its heavy reliance on various media of knowledge dissemination and communication (Hicks, 1999; Larivière et al. 2006; Pedersen et al. 2020). While journals may be the predominant outlet for research output for most NS [natural sciences] disciplines, SS&H [social sciences and humanities] researchers have a wider range of research outlets aside from journal publishing (Archambault et al. 2006; Glänzel and Schoepflin, 1999). Books are the major publishing outlet, or the “dominated format of cited sources” in the humanities literature (Kniewel and Kellsey, 2005, p.142; Sivertsen, 2016). That there is only one monograph which appeared as late as 2016 might suggest a lack of accumulative efforts and in-depth output,

though the situation was somewhat remedied by the six doctoral dissertations identified in the dataset.

In terms of research methods, the substantial quantity of non-empirical studies (25, 44.6%), the predominance of qualitative studies in empirical studies (26/31) and the absence of quantitative studies characterize the methodological status of English *Shuōwén* scholarship. Although “humanities departments do not have a tradition of empirical research” (Peer et al. 2012, p.xxi), the identified methodological status clearly contradicts the dominant position of number-based statistical research in linguistics since the middle of 20th century (Dörnyei, 2007; Duff, 2010) and the shift “from a more rationalist mode of inquiry to a more empirical mode of investigation in various domains of Chinese linguistics” which has been clearly visible since the 21st century (Jing-Schmidt, 2013, p.1). This seemingly outlier trend might indicate that the *Shuōwén* scholarship falls outside the mainstream trajectory of linguistic studies in general and Chinese linguistics in particular.

In terms of theoretical approaches, the 56 studies were approached from 20 theoretical stances including conventional angles of grammatology, historical linguistics, archaeology and lexicography as well as contemporary angles of computer linguistics, intellectual history and lexical pragmatics. While this wide array of approaches showcases the diverse backgrounds of scholars and the massive wealth of the *Shuōwén* text for interdisciplinary inquiries, a prominent gap is the lack of prevailing data-based and data-driven approaches to linguistic matters such as corpus linguistics, which is partly accountable for the absence of quantitative studies.

Table 12 Research strands of native scholarship till 1985 (statistics based on the bibliographical list in Dong and Zhang, 1988).

	Research strands	Frequency (%)	Corresponding strands in the present study
1	liùshū	350 (29%)	liùshū (8.9%)
2	commentaries on the prefaces and postfaces of <i>Shuōwén</i> works	154 (13%)	
3	comprehensive	118 (10%)	comprehensive (5.4%)
4	collation and annotation	116 (10%)	
5	miscellaneous notes (zhá jì 札记)	81 (7%)	
6	phonology	67(8%)	phonetic systems of certain head graphs (3.6%)
7	Duàn Yùcái's <i>Shuōwén jièzì zhù</i>	61 (6%)	
8	bùshǒu	55 (5%)	bùshǒu (5.4%)
9	citation	45 (4%)	
10	others	33 (3%)	
11	graph shape	32 (3%)	GPS of certain head graphs (1.8%)
12	macrostructure	21 (2%)	compiling principles (12.5%)
13	graph variants	19 (1%)	
14	excluded graphs	16 (1%)	
15	recensions	12 (0.9%)	recension comparison (1.8%)
16	dictionary retrieval	10 (0.8%)	
17	additional graphs	9 (0.7%)	
18	Xǔ Shèn's biography	8 (0.6%)	

In terms of research strands, the 29 themes represent the linguistic and extralinguistic facets and features concerning and surrounding the *Shuōwén* text. For comparative purposes, the research theme composition of native scholarship is introduced here. Dong and Zhang (1988) included a bibliographical list of native *Shuōwén* works (monographs and journal articles) from the Tang dynasty up until 1985 subsumed under various themes, totaling 1207 studies (counted by the present author). The present author calculated the number of works under each theme and summarized the numerical distribution of all works in Table 12 (for a convenient comparison, corresponding strands identified in the present study were added in the fourth column).

Common themes between domestic and international literature are liùshū, comprehensive, bùshǒu, graph shape, phonology, macrostructure and recension, though each theme occupies slightly different weight. For instance, comprehensive is a major theme in the native circle but a minor one in the overseas circle; the minor theme macrostructure in the former becomes a major one in the latter. The shift of weight does not surprise much as “comprehensive” works involve solid expertise and long-term devotion while “compiling principles” (sub-theme of macrostructure) do not require sophisticated knowledge of the ancient script, which means they form circumjacent areas surrounding central issues of character formation, shape, semantics and phonetics. The other notable observation is the substantial attention from both circles to liùshū (ranking first in the native circle and fifth in the overseas circle), accentuating the fact that liùshū is the first and foremost theory that governs Chinese character formation, and the most significant theoretical breakthrough achieved by Xǔ Shèn.

However, other important themes attracting native attention are not seen in the international circle: commentaries on the prefaces and postfaces of *Shuōwén* works, collation and annotation, miscellaneous notes, citations, graph variants, excluded and additional graphs, dictionary retrieval and Xǔ Shèn's biography. Though some are characteristic of traditional Chinese scholarship such as commentaries on the prefaces and postfaces of *Shuōwén* works and miscellaneous notes, all other themes are fundamental and pertinent issues comprising the macro and micro structure of the dictionary. Therefore, they pinpoint significant directions for future research agenda in the international circle. On the other hand, notable international themes are missing in the native scene: postface translation, introductory overview, book review,

and text computerization. While the absence of postface translation and introductory overview are easy to understand (both are characteristic for cross-language studies), the absence of the latter two might be attributed to the time frame and selection criteria of Dong and Zhang (1988). An initial search in CNKI returned results of book reviews and text computerization (Jan 2023), though not in considerable quantities. While text computerization has appeared in both the native and overseas circles (a minor in the former but a major in the latter), all publications are presenting the process and outcome of text digitalization, that is, infrastructure building. There has so far not emerged a study which conducts graphical inquiries by sourcing from the electronic data. This, apparently, pinpoints a very promising direction for future research endeavors.

The diachronic trends of English-language *Shuōwén* scholarship. Judging from the research quantity in the three periods, *Shuōwén* scholarship is “coldest” in period I and increasing slightly and steadily in the subsequent two periods. The 73-year span of period I has seen 12 studies, which consist of nine United States studies and three European studies. Notably, this distribution might be skewed as publishing in non-English languages before the mid-20th century was still common (see the introduction section for this methodological flaw which particularly affects period I). However, despite this limitation, focusing on English language scholarship, even in the early period, helps us understand the overall state of *Shuōwén* research and see some of the diachronic developments of this field. The first sizable modern work (and the first comprehensive study), is a doctoral dissertation of the Columbia University in the year of 1953 (Miller, 1953). The time and location are not accidental. The initial intent of developing Chinese studies was explicitly declared in the United States in the 1920s and the infrastructure took shape over the next thirty years (Yu, 2021). Therefore, 1953 was ripe for the appearance of the first substantial study. Colombia is one of the three pioneering university centers of sinology (the other two being located at California Berkeley and Harvard, see Yu, 2021) where a professorship for Chinese studies was established as early as 1902. With the pace and scale of support accelerated from the 1950s on, Chinese studies flourished and developed rapidly (Mair, 2013). With this background, the *Shuōwén* scholarship has grown steadily in terms of output quantities, countries and regions,

citations, theoretical approaches and themes (as shown by the statistics in the results section).

Judging from the research *quality* in the three periods, *Shuōwén* scholarship is most innovative and insightful in period I and is declining in originality, robustness and depth in the successive two periods, reaching its bottom in period II. To illustrate, period I has yielded only 12 studies in 73 years, but it claims the highest proportions of journal articles (50.0%) and empirical qualitative studies (66.7%). Apart from the one introductory overview (of the six *liùshū* categories) and two postface translations, and the rest are all primary studies dealing directly with the *Shuōwén* text on central graphical, textual and linguistic issues contributed by revered sinologists and real pioneers (e.g., Roy Andrew Miller, Weldon South Coblin, Paul L.-M. Serruys). The scale, rigor and depth of their studies remain unrivaled to this day. Period II represents a very different picture. It has the highest percentage of non-empirical studies (86.7%), book reviews (26.7%), non-independent items (40%, items embedded in larger works such as monograph sections) and introductory overviews (33.3%, four overviews of *Shuōwén* and one overview of *liùshū*) and the lowest number of research strands.

Period III presents a more complicated picture. On the one hand, the output quality of period I has not been surpassed. The majority of period III studies either remain at the level of initial introduction or focus on circumjacent areas such as compiling principles and secondary scholarship (Of course, the author does not mean these circumjacent issues are not important) instead of central graphical issues. Few can compete with the seminal works in period I in yielding significant and original insights for the core issues of *liùshū*, *bùshǒu* and GPS of head graphs and for the nature of Chinese writing. On the other hand, the scholarly scenes are prospering. Quantitatively, period III has the largest number in terms of all substantial parameters (largest number of publications, publication types, theoretical approaches and research strands). Qualitatively, it witnessed the appearance of the first monograph in over one century's course, saw the blossom of extralinguistic approaches such as intellectual history which have utilized the *Shuōwén* text beyond the traditional philological domain, and witnessed the emergence of text computerization, a technological breakthrough in the history of *Shuōwén* studies (though the technological innovation remains at the infrastructure level with no study utilizing the digital text as the primary data in the textual and graphical inquiries). At this point, it might be fair (and sounds not too controversial) to designate period III as "reviving" (renaissance is too big a word): while unmatching the premium quality of linguistic output of period I, it is extending the *Shuōwén* scholarship into an increasingly multi-disciplinary landscape.

The developmental trajectory of English-language *Shuōwén* scholarship may seem somewhat contradictory with the "tremendous amount of progress" of Chinese linguistics in the past forty years (Mair, 2013, p.390). However, taking into account the classical nature of the target text, it would not come as a surprise as the ability (and interest) to handle classical and literary Chinese is declining globally (Wilkinson, 2000). The same also applies to the Chinese context where the first-rate *Shuōwén* scholarship was produced in the Qing period, the standard of which remained unmatched up till today (though from the Republican to the 1980s, works of considerable value have been cultivated, their writers all belong to the older generation) (Zhang, 1998).

Conclusion

This study depicts the extant status and chronological trends of English-language *Shuōwén* scholarship which might be indicative of the state and development of East Asian classical studies in

general. Synchronically, the totality of works is regionally based, qualitatively intensive, and thematically restricted, which is to say, the *Shuōwén* scholarship has not been among the livelier areas within the field of Chinese language studies. Diachronically, the research quantity multiplies steadily while the research quality is fluctuating: declining for the last two decades of 20th century and reviving since the turn of the new millennium (though from rather different perspectives). While it is reasonable to say the heritage of first-generation scholars has not taken root in succeeding scholars, it is also fair to say the new-generation scholars have worked out new avenues.

To sum up, past and present English *Shuōwén* scholarship, though in general differing from the state and trend of Chinese linguistics in general, is not "abnormal". Classical, ancient and literary studies do not flourish anywhere (Schafer, 1990). Commanding a foreign language is never easy; commanding its classical forms is even more demanding. Classical Chinese is not even easier for native Chinese as their modern language knowledge would hamper, distract and taint the decipher of meaning and sound of the ancient script which has undergone radical changes over the course of the past three millennia (just as imposing the dialect of modern Peking on old texts is unworthy and dangerous).

To look ahead, we would hope for a more global profile of researchers' entry into the community and a larger degree of international collaboration on the part of Chinese native scholars. Increased specialization is a fruitful direction, especially streamlined into important traditional themes such as citations, variants and into new themes such as computer linguistics while core issues of *liùshū*, *bùshǒu* and GPS of head graphs continue to be investigated. The already emerging trend of extralinguistic approaches which situates the text of *Shuōwén* in a broader context of intellectual and ideological concerns is another fruitful direction. Methodologically, one of most promising research agenda is exploring the graphical content by sourcing from digital data, which could cause the shift of approach from qualitative and introspective methods to corpus linguistics and quantitative methods. The data-driven approach in the information era will shed enlightening insight on the traditional *Shuōwén* scholarship just as it has enlightened many other century-old linguistic topics. We believe that the *Shuōwén jiězi* has hardly been mined of its full riches, the digging of which awaits interested and dedicated scholars with diverse conceptual backgrounds and multiple disciplinary approaches.

Data availability

The datasets are accessible in Harvard Dataverse by <https://doi.org/10.7910/DVN/NCK41P>.

Received: 3 May 2023; Accepted: 23 April 2024;

Published online: 17 May 2024

References

- Archambault E, Vignola-Gagné É, Côté G, Larivière V, Gingras Y (2006) Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics* 68(3):329–342
- Boltz, WG (1993) *Shuo wen chieh tzu 說文解字*. In L Michael (Ed.), *Early Chinese texts: A bibliographical guide* (pp. 429–442). Society for the Study of Early China
- Boodberg PA (1937) Some proleptical remarks on the evolution of archaic Chinese. *Harvard Journal of Asiatic Studies* 2(3/4):329–372
- Brannen, J (2005) Mixed methods research: A discussion paper. *NCRM Methods Review Papers*, ESRC National Centre for Research Methods, Southampton. 1–30
- Canagarajah, S (2002) *A geopolitics of academic writing*. Pittsburgh, PA: University of Pittsburgh Press

- Christensen, LB, Johnson, RB, & Turner, LA (2021) *Research methods, design and analysis* (13th ed.). Pearson
- Dong X, Zhang Q (1988) 许慎与<说文解字>研究[Xǔ Shèn and the studies on Shuōwén jiězì]. Henan University Press, Zhengzhou
- Dörnyei Z (2007) *Research methods in applied linguistics*. Oxford University Press, Oxford
- Duff, P (2010) Research approaches in applied linguistics. In R Kaplan (Ed.), *Handbook of applied linguistics* (2nd ed., pp. 45-59). Oxford: Oxford University Press
- Friedman, DA (2012) How to collect and analyze qualitative data. In A Mackey & SM Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp.180-200). Blackwell Publishing Ltd
- Glänzel W, Schoepflin U (1999) A bibliometric study of reference literature in the sciences and social sciences. *Information Processing and Management* 35:31-44
- Godin B (2002) *The social sciences in Canada: What can we learn from bibliometrics? Project on the measurement of the social sciences*. Working Paper No. 1
- Gustavsson, R (2016) *Conforming to heaven: Organizational principles of the Shuōwén jiě zì*. [MA thesis.] Leiden University
- Gustavsson, R (2022) *Script, cosmos, and society: Philology in 10-11th century China*. [Doctoral dissertation.] City University of Hong Kong
- Hicks D (1999) The difficulty of achieving full coverage of international social science literature and the bibliometric consequences. *Scientometrics* 44(2):193-215
- Jing-Schmidt, Z (Ed.). (2013) *Increased empiricism: Recent advances in Chinese linguistics*. Amsterdam/Philadelphia: John Benjamins
- Knievel JE, Kellsey C (2005) Citation analysis for collection development: A comparative study of eight humanities fields. *The Library Quarterly: Information, Community, Policy* 75(2):142-168
- Larivière V, Archambault É, Gingras Y, Vignola-Gagné É (2006) The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology* 57(8):997-1004
- Mair, VH (2013) Developments in the study of Chinese linguistics during the last three decades. In H Zhang, Z Xue, S Jiang, & GL Lugar (Eds.). *A scholarly review of Chinese studies in North America* (pp.390-407). Association for Asian Studies
- Miller, RA (1953) *Problems in the study of the Shuo-wen chieh-tzu*. [Doctoral dissertation.] Columbia University
- Moher D, Liberati A, Tetzlaff J, Altman D, The PRISMA Group (2009) Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Physical Therapy* 89(9):873-880
- Okoli C, Schabram K (2010) A guide to conducting a systematic literature review of information systems research. *Sprouts: Working Papers on Information Systems* 10(26):1-49
- Patten, ML, & Newhart, M (2018) *Understanding research methods* (10th ed.). Taylor & Francis
- Pedersen DB, Grønvad JF, Hvidtfeldt R (2020) Methods for mapping the impact of social sciences and humanities—A literature review. *Research Evaluation* 29(1):4-21
- Peer WV, Hakemulder F, Zyngier S (2012) *Scientific methods for the humanities*. John Benjamins, Amsterdam/Philadelphia
- Petticrew, M, & Roberts, H (2006) *Systematic reviews in the social sciences: A practical guide*. Blackwell Pub
- Phakiti, A, & Paltridge, B (2015) Approaches and methods in applied linguistics. In B Paltridge & A Phakiti (Eds.), *Research methods in applied linguistics: A practical resource* (pp.1-25). Bloomsbury Academic
- Schafer, EH (1990) What and how is Sinology? *Tang Studies*, (8-9), 23-44
- Sivertsen G (2016) Patterns of internationalization and criteria for research assessment in the social sciences and humanities. *Scientometrics* 107:357-368
- Tang K (2018) 说文解字 [Shuōwén jiězì]. Zhonghua Book Company, Beijing
- Veletsianos G, Shepherdson P (2016) A systematic analysis and synthesis of the empirical MOOC literature published in 2013-2015. *International Review of Research in Open and Distributed Learning* 17(2):198-221
- Wilkinson, E (2000) *Chinese history: A manual, revised and enlarged*. Harvard University Asia Center
- Xiao Y, Watson M (2019) Guidance on conducting a systematic literature review. *Journal of Planning Education and Research* 39(1):93-112
- Yu P (2021) ACLS and the promotion of Chinese studies in the United States, 1928-1958. *Journal of the American Oriental Society* 141(4):895-910
- Zhang Q (1998) “说文学”源流考略. [Discerning the development and trends of Shuōwén scholarship]. Guizhou People's Publishing House, Guiyang

Author contributions

The sole author of this article assumes full responsibility for its content.

Funding

This work was supported by China National Office for Philosophy and Social Sciences (Grant number: 23BY008).

Ethical approval

Ethical approval was not required as the study did not involve human participants.

Informed consent

Informed consent was not required as the study did not involve human participants.

Competing interests

The author declares no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1057/s41599-024-03080-8>.

Correspondence and requests for materials should be addressed to Melissa Xiaohui Qin.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024