Comparing Citation Characteristics Based on Reference Entries, In-text Citations, and Essential Citations

Chi-Shiou Lin
Dept. of Library & Information Science, National Taiwan University
No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan
chishioulin@ntu.edu.tw

ABSTRACT
In this poster, we describe partial findings from a larger study that explores the influences of three different citation bases (reference entries, in-text citations, and essential citations) on the results of citation analysis in social sciences and humanities. Various citation characteristics such as document type, document language, self-citation rate, cited half life, and subject distributions of cited works will systematically compared to understand whether the selection of citation base did have an influence on the results. The completed analyses of the sociology discipline will be presented in this poster.

Keywords
Citation analysis, citation base, citation characteristics

INTRODUCTION
Citation analysis has been a popular method to study the use and exchange of scientific information in the academic world. A citation appearing in a research paper is viewed as a concrete evidence of information usage. However, exiting citation studies are almost based exclusively on the references list attached to the end of a paper rather than based on the actual in-text citations. Although theoretically the reference list does contain each unique bibliographic entry cited in the article, an entry may be cited more than once in the text, thus using the reference list as the basis for citation counting inevitably undermines the influence of certain cited references. However, counting the actual in-text citations is tedious and difficult, so it is an understandable compromise to use reference lists in most citation analyses. The use of reference lists is further reinforced by the development of large-scaled citation indexes in which a mega quantity of reference data have been prepared for researchers to download for immediate analyses.

The use of reference lists, however, aggravates a long existent problem of citation analyses, i.e., the questionable assumption that each citation is viewed as having equal contribution to the citing paper. As most citation analysts are aware of this problematic assumption, few research have examined if the citation analysis results will be different if the actual in-text citations that are more reflective of the times of usage (or intensity of use) are taken for analysis.

The problem is further complicated by the highly diverse and sophisticated citation behaviors of social scientists and humanities researchers. Citations may be used for rhetorical purposes or provide only minor supports to the citing paper. Hence, even within the in-text citations, not all citations are of equal importance to the paper.

There has been a long and lasting interest in the citation usages of social scientists and humanities researchers (see Huang & Chang, 2008 for an extensive review). In recent years, the validity of citation-based research evaluations applied to social sciences and humanities have also been questioned and debated (Huang & Chang, 2006). As such, this study compares the citation analysis results from three different bases of use evidence: reference entries, in-text citations, and essential citations (in-text citations that are directly related to the construction and argumentation of the core thesis of a paper). The purpose is to observe whether citation characteristics (aspects of information used) in social sciences and humanities will be different when the evidence base is altered. This poster will describe the results of the sociology literatures.

METHODOLOGY
The findings reported in this poster is from a larger project aiming to compare the citation disciplines in four subject disciplines in Taiwan, i.e., Chinese literatures, Chinese history, sociology, and economics. A cross-sectional design was used to draw paper samples from academic journals published during 1991-1995 and 2006-2010. In each subject, three representative research journals were chosen as the paper sources. For the two time periods, the selection of journals was based on the journal list of Taiwan Social Sciences Citation Index (TSSCI), the journal evaluation reports by the National Science Council (now the Ministry of Science and Technology), and recommendations of subject experts. Systematic sampling was used to draw 20 papers written in Mandarin Chinese from each journal, that

ASIST 2016, October 14-18, 2016, Copenhagen, Denmark.
is, 60 papers for each subject discipline, excluding book review, commentaries, and other non-research articles.

For each paper, the reference entries and the actual in-text citations are manually recorded, excluding those to primary sources (e.g., classic works, antique and rare books, newspapers, archives and manuscripts). And the in-text citations were carefully mapped to the reference list. At this stage, we found that some reference entries had not been cited in the texts, the percentages of the uncited reference entries were 16.3% (433/2642) for 1991-1995 and 5.7% (243/4216) for 2006-2010. This further shows the problems of using reference lists as the foundation of citation analyses.

Each paper was then read by 2-3 human coders to determine whether an in-text citation was an essential citation. This content coding follows our previous study (Lin, Chen, Chang, 2013) in which we used a citation functions scheme based on Murugesan & Moravcsik’s (1975) seminal conception of essential versus perfunctory citation. Operationally, an essential citation is defined as one that provides direct and major support to the construction of the major theses or the argumentation of the major concepts/theses in the paper. Citations that are used for background description, to acknowledge previous related works, indirect to the core thesis, or purely for rhetorical purposes, are considered non-essential. The content coding was conducted by trained graduate and undergraduate LIS students. The inter-coder reliability rates were .966 (1991-1995) and .997 (2006-2010) (Kassarjian, 1977).

### FINDINGS

We observed the document types, document languages, the authors’ self-citation, and the cited half life of the cited works in the sociology journals. Table 1 shows the results from the three citation data bases.

In regards to document types, for both periods of times, the analysis based on the three different citation sources all showed that books had been cited more than journal papers. Moreover, the percentages of books had consistently increased when the citation base shifted to in-text citations and essential citations. This indicates the importance of books as an information source for sociologists. Further, the importance of journal articles had risen over time. In 2006-01, we can still see that the percentages of this document type have risen slightly when the citation base was altered, but they appeared to drop in 1991-1995. Chi-square test showed that citation base was significantly correlated to document type in the earlier period, but not in the recent one. This means that the use of reference entries to analyze older literatures may cause more distortion.

In document languages, for both periods of time, the Taiwan sociologists had cited more information in English, all over 50%. But the percentages did not vary greatly when the citation base was altered. In contrast, the use of Chinese sources had risen from 1991-95 to 2006-10. Moreover, when the citation base was altered, we saw that the percentages of Chinese sources had dropped in the earlier period; but it had risen at least from the use of reference entries to in-text citations in 2006-10. This suggests that English has consistently been the most influential language to Taiwan’s sociology world, but there had been more use of Chinese sources in recent years. This should be an encouraging sign for the Taiwan sociologists.

![Table 1. Document Type, Language, and Self-Citation by Three Different Citation Bases](image-url)
The changes in the author self citation rates were less consistent with the altering citation bases. But chi-square test suggested that the use of citation base had a significant correlation to the 2006-10 self-citation rates, meaning that the use of reference entries may undermine the importance of author self-citation in this period.

As to the cited half life in sociology, the result for the 1991-1995 period using the three citation bases were 8, 9, and 10 years respectively, while those of the 2006-2010 were 10, 10, and 10 years. Previously, the use of reference entries as citation base may cause the undermining of the value of the older literatures in sociology. But this seemed to be of less threat now.

### Table 2. Distributions of the Cited Subjects Based on Three Different Citation Bases

<table>
<thead>
<tr>
<th>Subject Classes</th>
<th>References</th>
<th>In-text Citation</th>
<th>Essential Citation</th>
<th>References</th>
<th>In-text Citation</th>
<th>Essential Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>000 Generality</td>
<td>108</td>
<td>4.09</td>
<td>106</td>
<td>2.73</td>
<td>37</td>
<td>1.87</td>
</tr>
<tr>
<td>100 Philosophy</td>
<td>71</td>
<td>2.69</td>
<td>106</td>
<td>2.73</td>
<td>53</td>
<td>2.68</td>
</tr>
<tr>
<td>200 Religion</td>
<td>7</td>
<td>0.26</td>
<td>23</td>
<td>0.59</td>
<td>19</td>
<td>0.96</td>
</tr>
<tr>
<td>300 Sciences</td>
<td>12</td>
<td>0.45</td>
<td>15</td>
<td>0.39</td>
<td>10</td>
<td>0.50</td>
</tr>
<tr>
<td>400 Applied Sciences</td>
<td>297</td>
<td>11.24</td>
<td>508</td>
<td>13.10</td>
<td>326</td>
<td>16.46</td>
</tr>
<tr>
<td>500 Social Sciences</td>
<td>2099</td>
<td>76.04</td>
<td>2922</td>
<td>75.33</td>
<td>1414</td>
<td>71.38</td>
</tr>
<tr>
<td>600 Chinese History/Geography</td>
<td>12</td>
<td>0.45</td>
<td>13</td>
<td>0.34</td>
<td>6</td>
<td>0.30</td>
</tr>
<tr>
<td>700 World History/Geography</td>
<td>85</td>
<td>3.22</td>
<td>143</td>
<td>3.69</td>
<td>95</td>
<td>4.80</td>
</tr>
<tr>
<td>800 Linguistic &amp; Literatures</td>
<td>22</td>
<td>0.83</td>
<td>27</td>
<td>0.70</td>
<td>14</td>
<td>0.71</td>
</tr>
<tr>
<td>900 Arts</td>
<td>19</td>
<td>0.72</td>
<td>16</td>
<td>0.41</td>
<td>7</td>
<td>0.35</td>
</tr>
<tr>
<td>Total</td>
<td>2642</td>
<td>100.00</td>
<td>3879</td>
<td>100.00</td>
<td>1981</td>
<td>100.00</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 70.181, df = 18, p < .001; \text{Cramer’s } V = .064 \]

\[ \chi^2 = 27.738, df = 18, p > .05 \]

Note: The subject classification was conducted by the content coders based on the New Classification Scheme for Chinese Libraries (2007 version).

ACKNOWLEDGMENTS

This study was sponsored by the Ministry of Science and Technology, Taiwan (Grant Number: MOST104-2410-H002-197).

REFERENCES


Lin, C.-S., Chen, Y.-F., & Chang, J.-Y. (2013). Citation functions in social sciences and humanities: Preliminary results from a citation context analysis of Taiwan’s history research journals. In *Proceedings of the ASIS&T 2013 Annual Meeting*. 