A STUDY OF INDEXING CONSISTENCY: CONSISTENCY BETWEEN THE LIBRARY OF CONGRESS AND THE BRITISH LIBRARY CATALOGERS

Yasar Tonta

School of Library and Information Studies,University of California, Berkeley, CA 94720

Abstract

The article aims to compare the indexing consistency between the Library of Congress (LC) and the British Library (BL) catalogers with regards to their using the Library of Congress Subject Headings (LCSH). Eighty-two titles, published in 1987 in the field of Library and Information Science (LIS), were identified for comparison, and for each title its LC subject headings, assigned by both LC and BL catalogers, were compared. By applying Hooper's "consistency of a pair" equation, the average indexing consistency value was found for 82 titles. The average indexing consistency value between LC and BL catalogers is 16 percent for "exact• matches, and 36 percent for "partial• matches. The major findings of the study are discussed, and, in the Appendix, the examples of LCSH that assigned by both LC and BL catalogers for the same document are provided along with its consistency value.

Introduction

It has for long been observed that different indexers tend to assign different index terms to the same document. In other words, "the indexers differ considerably in their judgment as to which terms reflect the contents of the document most adequately"(1). Essentially, indexing consistency is seen as "a measure of the similarity of reaction of different human beings processing the same information"(2).

Indexing consistency in a group of indexers is defined as "the degree of agreement in the representation of the essential information content of the document by certain sets of indexing terms selected individually and independently by each of the indexers in the group"(3).

Indexing consistency studies reported in the literature have shown that the consistency values vary a great deal

between the indexers. Hooper(4), Leonard(5), and Markey(6) reported the results of some 25 published and unpublished indexing consistency experiments in which the indexing consistency values ranged from 4 percent to 82 percent. However, the indexing consistency scores of various studies, as researchers rightly caution us, should be considered separately and not compared. For, it appears that consistency values depend on a number of factors under which the indexing was performed. Zunde and Dexter(7) listed 25 factors affecting the indexing performance. (See also, Tarr and Borko(8).) For instance, factors such as the use of classification schedules and other indexing aids, the employment of subject specialists as indexers, and indexer training, among others, have greatly improved the consistency values (9,10). Markey (11) offers a more detailed discussion of these factors, relating some of the factors with the findings of previous studies.

Indexing consistency also depends on the consistency measure used in the evaluation. Studies reported in the literature employed a variety of methods and different formulae to calculate indexing consistency values. In fact, as Cooper (12) puts it, "this circumstance makes generalization about their findings difficult". (For more information about various indexing consistency formulae and statistical techniques involved in consistency studies, see Zunde and Dexter (13), Hooper (14), Leonard (15), Markey (16), and Rolling (17).)

It is assumed that there is a relationship between indexing consistency and the "indexing quality". That is to say, "an increase in consistency can be expected to cause an improvement in 'indexing quality'"(18).

For some authors what is more important, and needs to be thoroughly scrutinized, is the relationship between indexing consistency and the effectiveness of information retrieval. Cooper (19) further suggests that "until this relationship [i.e., the relationship between indexing consistency and retrieval performance] has been investigated, there is little point in measuring interindexer consistency at all". Leonard attempted to investigate this relationship in his doctoral dissertation and found that "inter-indexer consistency and retrieval effectiveness exhibit a tendency toward a direct, positive relationship, i.e. high inter-indexer consistency in assignment of terms appears to be associated with a high

retrieval effectiveness of the documents indexed" (20). However, he feels that "[c]onsiderably more research is needed before the relationship between inter-indexer consistency and retrieval effectiveness can satisfactorily be defined" (21).

Methodology

This study attempts to compare the indexing consistency between the Library of Congress (LC) and the British Library (BL) catalogers. For the comparison, the books published in 1987 in the field of Library and Information Science (LIS) (020 in Dewey) were chosen. First, all the titles published in 1987 in the United Kingdom were identified using the "BNB Subject Catalogue• (Volume 1). There were 237 titles altogether. Secondly, using the ISBN numbers provided, all 237 titles were searched on OCLC database. Of 237, 217 titles were found on OCLC. (The rest were either serials, microform copies or local publications.)

Thirdly, of 217, titles that were cataloged and given the Library of Congress Subject Headings (LCSH) by both LC and BL catalogers were identified. (In this study, the terms "indexing" and "cataloging" are used interchangeably.) The 040 field in the MARC format was used to identify the origin of cataloging information. For instance, UKM stands for UK MARC, i.e., cataloged by BL, and DLC stands for LC, i.e., cataloged by LC. (Items that were cataloged according to LC practices by libraries other than LC (by the National Library of Medicine, for example) are not included in the sample.) By checking the 040 field for each record found on OCLC, it was possible to download all the records that were cataloged by both BL and LC. It turned out that there were 82 items. That is to say, all the subject headings (LCSH) that were assigned to 82 items by BL and LC were compared with regards to consistency.

For each of the 82 items in my sample, subject headings assigned by LC and BL catalogers were identified from the downloaded data, and then compared. LC subject headings were readily available as the 600 (personal name), 610 (corporate name), 611 (conference, congress, meeting, etc. name), 630 (uniform title), 650 (topical LCSH) and 651 (geographical LCSH) fields in the MARC format are exclusively used for all kinds of LC subject headings.

Finally, the "consistency of a pair of indexers" formula, defined by Rodgers and developed by Hooper, was applied to find out the indexing consistency value for each title cataloged by LC and BL catalogers. According to Hooper's equation, "the consistency of one indexer with respect to a second is based on the number of times the two indexers agree on the use of a term, divided by the total number of terms used by either indexer (based on the

specific document)22.

Hooper's "consistency of a pair" formula is as follows:

$$CP(\%) = A / (A + M + N$$

where:

CP: is the consistency of term assignment between two indexers (consistency expressed as a percentage);

A: is the number of term agreements between 'M' and 'N' for a specific document;

B: is the number of terms used by 'M' but not used by 'N'; and,

C: is the number of terms used by 'N' but not used by 'M'.

Having obtained the indexing consistency value for each title, I calculated the average indexing consistency value between BL and LC catalogers for 82 titles.

Further explanation is due here. It was assumed that each individual cataloger at LC approaches the same document in the same way and assigns the same subject heading(s), which in fact may not be true. This assumption was made for BL catalogers, too. I am aware of the fact that what I found is "not• the individual interindexer consistency value between the two indexers but, rather, the indexing consistency value between LC and BL catalogers as two different groups.

Findings

The major findings of the study are as follows:

1. LC catalogers assigned 282 subject headings for 82 items while BL catalogers assigned 127. In other words, on the average, LC assigned 3.44 subject headings per title. The same average isÜf _ Ü_1.55 for BL catalogers. The marked difference with regards to the average number of subject headings between LC and BL is understandable. It is obvious that BL relies on PRECIS (Preserved Context Index System) for subject access rather than LCSH, whereas LC completely depends on LCSH for subject retrieval.

It appears that BL catalogers tend to keep the number of LCSH assigned for each title to the minimum. Only for two titles (2.4 percent) did BL catalogers assign more subject headings than LC catalogers. For 17 titles (20.7 percent) both BL and LC catalogers assigned the same number of subject headings. For the rest of 63 titles (73.9 percent) LC catalogers assigned more LCSH than BL catalogers.

2. Each and every subject heading for the same title that assigned by LC and BL catalogers was compared. It turned out that 49 out of 127 BL-assigned subject headings ″exactly• matched the LC-assigned subject headings. "Exact matches" were allowed for variants in spelling (i.e., catalog - catalogue) and punctuation (i.e., on-line - online), but not for synonyms (i.e., non-book - audio-visual). In the second run, I softened my rule and tried to identify ″partial• matches, too. Forty-four BL-assigned headings partially matched further. A synonym in a multiple-word-subjectªheading was treated as a "match" as long as it was not the first word in that subject heading. The lack of a subdivision in a subject heading was also allowed for partial matches if the main part of the subject heading matched exactly. Seventeen BLªassigned subject headings for 12 titles were completely different from LC-assigned ones. (Examples of subject headings assigned by LC and BL catalogers for the same titles are given in Appendix.)

3. Average indexing consistency value for exact matches between BL and LC catalogers is 16 percent.

For both exact and partial matches, average indexing consistency value between BL and LC catalogers is 36 percent. (Several examples of consistency values are given in Appendix.)

It is tempting to think what would the indexing consistency value have been if the number of subject headings assigned by BL catalogers were equal to that of LC catalogers. For 17 titles that have the same number of LCSH assigned by both LC and BL indexers, the following indexing consistency values were obtained:

For exact matches, average consistency value is 14 percent.

For both exact and partial matches, average consistency value is 41 percent.

Although there is a slight difference between the two averages, there seems to be no strong relationship between the indexing consistency value and the assumption that if the equal number of subject headings were assigned by both LC and BL for all titles the consistency value would have been different.

Conclusion

It should be borne in mind that the sample used in this study is small. Moreover, since two different subject indexing tools (i.e., LCSH and PRECIS) are used in LC and BL, it may not be very meaningful, if at all, to compare the two groups of catalogers. Yet the indexing consistency value found in this study is similar to those of reported in other consistency studies.

In conclusion, the indexing consistency value between LC and BL catalogers for the books in the field of LIS is 16 percent for exact matches and 36 percent for both exact and partial matches, which is pretty low.

References

1. P. Zunde and M.E. Dexter, "Indexing Consistency and Quality," ”American Documentation• 20, no. 3 (1969): 259–267.

2. Zunde and Dexter, "Indexing Consistency," p.259.

3. Zunde and Dexter, "Indexing Consistency," p.259.

4. R.S. Hooper, "Indexer Consistency Tests - Origin,•
"Measurements, Results and Utilization•. (Bethesda, MD:
IBM Corp., 1965).

5. L.E. Leonard, "Inter-indexer Consistency Studies, 1954–
1975: a Review of the Literature and Summary of Study
Results," "University of Illinois Graduate School of
Library Science• "Occasional Papers• No. 131, December
1977. 51 p.

6. K. Markey, "Interindexer Consistency Tests: a Literature
Review and Report of a Test of Consistency in Indexing
Visual Materials," "Library and Information Science
Research• 6, no. 2 (1984): 155-177.

7. P. Zunde and M.E. Dexter, "Factors Affecting Indexing
Performance," In: "Proceedings of the 32nd Annual Meeting
of the American Society for Information Science•.
(Washington, DC: ASIS, 1969): 313-322.

8. D. Tarr and H. Borko, "Factors Influencing Inter-indexer
Consistency," In: "Proceedings of the 37th Annual Meeting
of the American Society for Information Science•.
(Washington, DC: ASIS, 1974): 50-55.

9. Hooper, "Indexer Consistency Tests," p.7.

10. Zunde and Dexter, "Indexing Consistency," p.260.

11. Markey, "Interindexer Consistency Tests," pp.155-177.

12. W.S. Cooper, "Is Interindexer Consistency a
Hobgoblin?," "American Documentation• œ 20, no. 3 (1969):
268-278. p.269.

13. Zunde and Dexter, "Indexing Consistency," pp.259-267.
14. Hooper, "Indexer Consistency Tests,".

15. Leonard, "Inter-indexer Consistency Studies".

16. Markey, "Interindexer Consistency Tests," pp.155-177.

17. L. Rolling, "Indexing Consistency, Quality and
Efficiency," "Information Processing & Management• 17, no.
2 (1981): 69-76.

18. Cooper, "Is Interindexer Consistency," p.269.

19. Cooper, "Is Interindexer Consistency," p.269.

20. As cited in: Leonard, "Inter-indexer Consistency Studies,"p.33.

21. Leonard, "Inter-indexer Consistency Studies," p.33.

22. Hooper, "Indexer Consistency Tests," p.33.

APPENDIX

Examples of Indexing Consistency Values Between the Library of Congress and the British Library Catalogers

1. Vickery, B.C. and Vickery, A. "Information science in theory and practice"

LC: Information science

BL: Information Science

"Consistency value: 100%

2. Veit, Fritz. "Presidential libraries and collections

LC: Presidents - United States - Archives

BL: Presidents - United States - Archives

"Consistency value: 100%•

3. ur Rahman, Sajjad. "Management theory and library education •

LC: Library education

Library administration - Study and teaching

Library administrators - Training of

BL: Library administration - Study and teaching

Library education

"Consistency value: 67%•

4. "Personnel issues in reference services

LC: Reference librarians

Library personnel management

Reference services (Libraries)

Library administration

Library services - Organization and administration

BL: Library personnel management

Reference services (Libraries)

Reference Librarians

"Consistency value: 60%•

5. Crawford, Walt. "Technical standards: an introduction for "librarians

LC: Library science - Technological innovations - Standards
Library science - Standards

Technology - Standards

Information science - Standards BL: Library science - Standards

Information science - Standards

"Consistency value: 50%•

6. Strickland-Hodge, Barry. "How to use Index Medicus and Excerpta Medica•

LC: Medicine - Abstracting and indexing

Index Medicus

Excerpta Medica

Medicine - Bibliography Methodology

MEDLARS-MEDLINE information system - United States

Abstracting and Indexing

BL: Index Medicus

Excerpta Medica

Medicine - Abstracting and indexing

"Consistency value: 50%•

7. Burton, Paul F. "The librarian's guide to microcomputers for "information management

LC: Libraries - Automation

Library science - Data processing

Microcomputers - Library applications BL: Microcomputers - Library applications "Consistency value: 33%

8. Harrod, Leonard Montague. "Harrod's librarians' glossary of• "terms used in librarianship, documentation, and the book• "crafts and reference book

LC: Library science - Dictionaries

Information science - Dictionaries

Bibliography - Dictionaries

Book industries and trade - Dictionaries

BL: Library science - Dictionaries

"Consistency value: 25%

9. "Conservation of library and archive materials and the graphic• "arts•

LC: Library materials - Conservation and restoration Archival materials - Conservation and restoration Graphic arts - Conservation and Restoration

Books - Conservation and Restoration Art - Conservation and Restoration Paper - Preservation

BL: Library materials - Conservation and restoration

"Consistency value: 17%

10. Tracy, Joan I. "Library automation for library technicians•

LC: Libraries - Automation

Library science - Data processing

Library technicians

BL: Processing (Libraries) - United States

"Consistency value: 0%•

11. Hartley, J.; Noonan, A. and Metcalfe, S. "New electronic• "information services• "•

LC: Database industry - Great Britain

Database industry

BL: Information storage and retrieval systems

On-line data processing

"Consistency value: 0%•

12. "International Conference on Application of Microcomputers in• "Information, Documentation and Libraries

LC: Libraries - Automation - Congresses

BL: Microcomputers - Library applications

"Consistency value: 0%•