Effects of Subject Specificity: Part II: Relationship of LC Subject Headings Specificity and Class Notation Length

A. Khosh-khui

ABSTRACT. Subject headings and class notations in 7,834 LC bibliographic records were analyzed to determine if there was a correlation between the degree of subject specificity and class notation length, and if there was a relationship between the number of subject headings per bibliographic record and class notation length. The findings of the study indicated that as the degree of specificity of subject increased, the length of both LCC and DDC notations increased. However, the difference between class notation lengths corresponding to subject headings with a different level of specificity was not statistically significant. There was not a statistically meaningful association between the number of subject headings per bibliographic record and class notation length.

INTRODUCTION

Subject analysis of library materials involves both classification and subject headings which complement each other in subject retrieval. Classification is unidimensional in nature, while the subject-heading approach is multidimensional. Although multiple subject headings can be assigned to a document, there is usually only one class notation corresponding to the several subject headings assigned to that document. While a book can be listed under several subject headings in the public catalog and can be retrieved by any one of them, there is normally only one classification notation.

A. Khosh-khui is serials cataloger at the Southwest Texas State University Library in San Marcos. He received his Ph.D. from Indiana University. Address all correspondence to the author at: 705 West River Road, Apt. 701, San Marcos, Texas 78666.

This notation is an artificial language and a means by which a topic can be expressed in terms of signs or symbols. The degree of detail is related to the notation length. Maltby mentions that specificity with short notation is unrealistic.[1] Not only is length dependent on the degree of detail, but on the type of symbols used, the hierarchical nature of the classification system, and the location of notations for given subject areas. Class notations in conventional systems have structural limitations because they are restricted by the rigid framework by which they are represented. In fact, most classification systems fit into a structure of notational symbols and consequently do not follow the structure of the classification. "The notation factor limits these classifications unnecessarily. There is nothing particularly limited about classification, in theory, but the best theory can be ruined in the process of codification."[2]

Classification notations should be used ideally for shelf arrangement as well as symbolizing subject matter. Daily identifies the differing purposes of these two systems of organizing knowledge. "Headings are meant to bring together what the classification separates."[3] He points out that the purpose of subject headings is the organizing of knowledge and the purpose of classification is the arrangement of books on the shelf.

The Library of Congress uses the initial subject heading to determine the class notation.[4] Consequently, notations are used economically, i.e., the document initially is classed to strictly adhere to the class closely, subject and then, secondarily, to uniquely identify the particular work. Of course, this process is not always easy and for some subject areas is quite difficult. Both DDC and LCC are enumerative with limited synthesis capability and adhere generally to taxonomic and hierarchical principles.[5] LCC is mixed, using both letters and numbers, while DDC is numerical with a few exceptions. The notational base of LCC, consequently, is broader than that of DDC. DDC is hierarchical but not expressive while LCC is neither hierarchical more expressive.[6] Therefore, both are linear and unidimensional in nature.

Hyman states that shelf classification in American libraries is not based on objectively valid theory or practice.[7] Classification notations in this country are mostly used, as Fairthorne states, to inscribe and order materials.[8] Their potential as subject analysis tools has been disregarded.

Hill discusses the practical considerations of using the classifica-

tion number in online subject access.[9] She describes as problems difficulty of use, changing from one edition to another, and inconsistency of use within a library. She concluded that

> Online subject searching by classification number would offer an approach to information well beyond the capabilities of card catalogs, and it is undoubtedly worth aiming for.[10]

The ability to search by classification notation in an online environment encourages research into the relationship between class notation length and subject specificity. This study attempts to measure the degree of that relationship.

LITERATURE REVIEW

Gorman suggests using different levels of complexity in classifying library materials to solve the dichotomy between class notation as a shelving device and class notation as a retrieval device:

> The solution to the classification problem is to use two different schemes (or two levels of the same scheme): One to arrange the materials, and the other to provide subject-searching capability on machine systems.[11]

Richmond states that both classification and subject heading assignments may be inadequate if a monograph is dealing with several subjects, unless a deeper subject analysis is provided by the librarian:

> . . . Classification is made a tool for achieving [a] consistent, fixed location of like material. On the other hand, realization that a classified shelf list leaves something to be desired in the way of subject analysis has forced catalogers to add subject entries to the card catalog, each corresponding to one of the major topics covered by the book and chosen from a standardized subject heading list containing a mixture of terms gathered through literary warrant . . . If the book is a monograph on a homogeneous subject, the system of shelf classification, supplemented by subject headings is usually adequate. On the

other hand, if the book represents a crossfertilization among subject fields or if it is a collection of papers, possibly on one central theme, but from many points of view representing many different fields, and if the information in it is to be sought from all these subject angles, then it is impossible to present its true subject coverage without either a multiple-classification approach or a much deeper subject analysis than can be made with standard methods.[12]

METHODOLOGY

Ten percent of the bibliographic records on LC MARC tapes were selected as a controlled sample. This was 7,834 records with 15,072 subject headings. The selected records were analyzed by the degree of specificity of subject headings, the length of DDC and LCC notations, and the number of subject headings per bibliographic record.

RESULTS

The results of the analysis showed that there was not a significant correlation between class notation length and the number of subject headings per bibliographic record. While there was a negative correlation of $-.016$ between number of subject headings and the Library of Congress Classification, there was a .02 correlation for the Dewey Decimal Classification. The correlation between the notation length in both classification systems and the number of subdivisions per heading was statistically significant. However, the magnitude of this association was not more than 0.25. The two classification systems had a correlation of .20 which was significant at the 0.001 level. The measure of association for both DDC and LCC and their corresponding subject headings is shown in Table 1.

The mean of class notation length was computed for subject headings with different numbers of words per heading. At the same time, the mean of subject length and the mean of subdivision per subject were computed for each category to find the variation of each variable as the number of words per subject increases. As shown in Table 2, the average length for LCC for uniterm headings was 6.57 characters and it slightly increased as the number of words per title increased. The length of DDC notations increased from

TABLE 1

Comparison of Association Between DDC/LCC Notation Lengths With the
Number of Subject Headings Per Record and Their Specificity Level

Classification Systems	Number of LCSH per Bibliographic Record	Number of Subdivisions Per LCSH	Number of Characters Per LCSH	Number of Characters Per Class Notation
LCC	r = -.016 p = .025	r = .12 p = .001	r = .08 p = .001	r = .20 (DDC) p = .001 (DDC)
DDC	r = -.02 p = .013	r = .25 p = .001	r = .19 p = .001	r = .20 (LCC) p = .001 (LCC)
			N = 15072	

r = The Pearson product moment correlation coefficient, i.e.
the degree of association between each pair of variables.

p = Probability of event "p", i.e. possibility of correlation
"r" to be significant.

N = Number of cases, i.e. subject headings in the study.

TABLE 2

Comparison of DDC and LCC Class Notation Variations Due to
Variations in Subject Specificity

Level of Subject Specificity	Number of LCSH Per Record	Number of Characters Per LCSH	Number of Subdivisions Per LCSH	Number of Characters Per DDC Notation	Number of Characters Per LCC Notation
	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$	$\bar{x}$
Uniterm	2.46	14.56	.82	7.11	6.57
2 Words	2.47	22.83	.90	7.79	6.77
3 Words	2.40	30.80	1.21	8.18	6.93
4 Words	2.37	39.62	1.45	8.23	6.90
5 Words	2.38	47.35	1.62	8.64	6.98
6 Words or More	2.35	47.35	1.65	8.64	6.95
Chi Square	0.439	20.206	0.301	0.053	0.105
Significance	.35	.35	.65	.65	.95

$\bar{x}$ = Mean

7.11 characters for uniterm to 8.65 for subject headings with more
than five words. However, the chi-square test showed that in both
classification systems the differences between mean length are not
statistically significant, as Table 2 indicates.

The correlation of notation length in DDC and LCC to the varia-
tion of subject specificity was calculated to determine any identifi-
able pattern. There was a .67 correlation between DDC notation

length uniterm subject headings. For other categories there was not a significant correlation between the two variables at the .001 level. None of the computed correlations for LCC notation length and subject headings were found to be significant. Table 3 shows the results of this part of the analysis.

CONCLUSION

For both classification systems, there was not a meaningful association between notation length and the degree of specificity. In spite of the fact that Dewey Decimal Classification is a hierarchical system, the notation length did not significantly vary when notations are applied to classify more specific topics. The findings of the study indicated that as the degree of subject specificity increased, the length of both LCC and DDC notations increased. However, the difference between class notation lengths corresponding to subject headings with a different level of specificity was not statistically significant. There was not a statistically meaningful association between the number of subject headings per bibliographic record and class notation length.

TABLE 3

Correlation Between Number of Words Per Subject Heading
and Length of DDC and LCC Notations

Level of Subject Specificity	Number of LCSH Per Group	DDC Notation Length		LCC Notation Length	
	f	r	p	r	p
Uniterm	2849	.67	.001	.005	.398
2 Words	4999	.02	.111	-.02	.059
3 Words	3558	.003	.439	-.02	.092
4 Words	2081	.04	.037	.005	.412
5 Words	1011	-.03	.189	-.02	.223
6 Words or More	573	-.001	.488	-.02	.282
All Subject Headings	15072	.02	.013	-.02	.025

f = frequency count

r = The Pearson product moment correlation coefficient, i.e.
    the degree of association between each pair of variables.

p = Probability of event "p", i.e. possibility of correlation
    "r" to be significant.

NOTES

1. Maltby, A. (1972). "Classification — Logic, Limits, Leves." In *Classification in the 1970's: A Second Look*, rev. ed., edited by Arthur Maltby. Hamden, CT: Linnet Books, p. 14.

2. Richmond, Phyllis A. (1961). "Recommendations for Further Study in Classification," *American Documentation* 12:35.

3. Daily, Jay E. (1957, October). "Subject Headings and the Theory of Classification," *American Documentation* 8:272.

4. Library of Congress, Subject Cataloging Division (Summer, 1978). "Order of Subject Tracings," *Cataloging Service Bulletin* 1:15.

5. Chan, Lois Mai. (1981). *Cataloging and Classification: An Introduction*. New York: McGraw-Hill, p. 210.

6. Ibid, p. 223.

7. Hyman, Richard J. (1980). "Shelf Classification Research: Past, Present — Future?" Occasional Papers, University of Illinois, Graduate School of Library Science; 146, p. 3.

8. Fairthorne, R.A. (1961). *Towards Information Retrieval*. Index compiled by Calvin N. Mooers. London: Butterworths, p. 95.

9. Hill, Janet Swan. (1984, March). "Online Classification Number Access: Some Practical Considerations." *The Journal of Academic Librarianship*, 10:17-22.

10. Gorman, Michael. (1981, September). "The Longer the Number, the Smaller the Spine: Or, Up and Down with Melvil and Elsie," *American Libraries*, 12:498.

11. Richmond, Phyllis Allen. (Spring, 1960). "Some Aspects of Basic Research in Classification." *Library Resources & Technical Services*, 4:140.

12. Ibid, p. 145.