Information Processing & Management Vol. 25, No. 6, pp. 677–691, 1989
Printed in Great Britain.

0306-4573/89 $3.00 + .00

DYNAMIC THESAURAL SYSTEMS: A BIBLIOMETRIC STUDY OF TERMINOLOGICAL AND CONCEPTUAL CHANGE IN SOCIOLOGY AND ECONOMICS WITH APPLICATION TO THE DESIGN OF DYNAMIC THESAURAL SYSTEMS*

LORNA K. REES-POTTER
Graduate School of Library and Information Studies, McGill University
Montreal, Quebec, H3A 1Y1

(*Received* 5 *July* 1988; *accepted in final form* 3 *February* 1989)

Abstract—Thesauri have been used in the library and information science field to provide a standard descriptor language for indexers or searchers to use in an information storage and retrieval system. One difficulty has been the maintenance and updating of thesauri since terms used to describe concepts change over time and vary between users. This study investigates a mechanism by which thesauri can be updated and maintained using citation, co-citation analysis and citation context analysis.

Data in twenty-six specialty areas in economics and sociology from the *Social Sciences Citation Indexes*, 1966–67, 1973–74, and 1980–81 was used. Candidate thesaurus terms were developed semi-automatically from the citation contexts of papers citing in these specialty areas. Experts in these disciplines concluded that these terms were appropriate terms to describe these specialty areas. These candidate thesaurus terms compared favourably to terminology found in a standard subject heading list (L.C. subject headings). The terms generated provided a rich source of alternate terminology.

1. INTRODUCTION

A thesaurus provides a level of conceptual structure and terminological control to the storage and retrieval of documents. One difficulty with thesauri is their maintenance and updating. Thesauri tend to be developed at a particular point in time and for a particular user group. Terms used to describe concepts in documents may change over time and vary between users. New concepts for which there is no accepted term arise and new terms then appear. Terms fall into disuse and disappear as the concepts they represent are no longer used. Both indexers and searchers will have problems translating new relevant subject descriptions into the terms of the thesaurus. New documents will be added to the collection which describe new or varying concepts. New terms and new cross references must be added to the thesaurus. Thus, from time to time, the thesaurus must be updated and changed.

Thesauri are difficult and time-consuming to update, and they do not always solve the problem of language incompatibility between the indexer, the searcher, and the documents. There is a need for thesauri that better reflect both the language of the documents in the collection and the current users. Thus there is a need for dynamic thesauri which can reflect these changes.

2. BACKGROUND

Documents contain connections to other documents through the practice of bibliographic reference or citation. References in a document demonstrate to some extent the documents' conceptual antecedents. This practice forms the basis of citation indexing. Both direct citation and co-citation have been used in citation research to measure the strength

*This paper is based on a doctoral dissertation, University of Western Ontario, London, 1987. The research was supported by the Research Grants Program Project 410-82-0282-R1 of the Social Science and Humanities Research Council of Canada.

of these conceptual antecedents. Some investigators have used these methods to describe the nature of knowledge and concepts in a specialty [1–4]. Citation and co-citation networks have been used as a means of reflecting the conceptual structure of a field [5,6]. As well, citation and co-citation analysis have been used to study the history of concepts in a specialty area [7–11].

Small has noted that the footnote number in a document points to a portion of the text to which it refers and the context of the footnote number or citation relates conceptually to this context. Small then suggested cited documents could be seen as concept symbols and individual citation contexts as instances of symbol deployment [12]. The citation context was defined as the two or three sentences in which the footnote was embedded. A high percentage of uniformity in the term or phrase used was found in analyzing frequently cited documents in a chemistry database. In other words, there are recurring patterns of term use when referring to frequently cited documents. Small and Greenlee further used this idea to see if the method could be used to trace consensus in a specialty, and the history of consensus in a specialty over time [7]. The change in the percentage uniformity was not dramatic over time, however there was some change. Cozzens using citation context analysis method for historical analysis in the field of economics found striking changes in how a document is cited over time [13]. The present study uses the bibliometric techniques of citation and co-citation context analysis and further extends this research to an application in the area of terminological analysis and control. If citation analysis reflects concepts in a specialty, reflects the history of concepts in a specialty, and reflects term use in a specialty, it seems reasonable to expect that this technique of citation context analysis could be used to trace term change in a specialty over time and variation among researchers.

An important advantage when considering the objective of the study, (i.e. that of thesaurus development), is that the actual terms and concepts found in the documents themselves are used to generate thesaurus terms rather than the terms selected by indexers who may not be aware of terminological variation and change within the field. As term usage changes in a discipline, so the thesaurus used by indexers and searchers can be changed.

Two aspects of term change are considered in this study. First, as time passes, the terminology used to express concepts may change, (i.e. diachronic change). The phrase "term change in a specialty over time" is used to express this concept. Secondly, at any one time different terms will be used to express concepts by different individuals and groups, (i.e. synchronic change). The phrase "variation among individuals and groups" is used to express this concept. The ability to trace change and variation form the basic pieces of information needed in dynamic thesaurus development.

3. METHODOLOGY

The general methods of data gathering, citation and co-citation analysis used by previous investigators such as Small and Griffith to identify concept groups in a discipline were followed in this study [14].

Social science disciplines were selected because its terminology may be "softer," (i.e. less precise than terminology in a physical science), and present greater problems of control to indexers and provide a more strident testing of the statistical techniques used in this study [15]. The particular disciplines selected were sociology and economics. Using two different social science disciplines which may have different levels of ambiguity in their terminology may add some additional insights.

3.1 *Dataset development*

Social Sciences Citation Index data tapes were purchased from the Institute for Scientific Information. Six data bases were constructed using all the records from the sociology and the economics source journals from three 2-year periods (1966–67, 1973–74 and 1980–81), providing a time element of fifteen years. The number of journals included in these

categories changed through the time periods used; generally around eighty economics journal titles and 120 sociology journal titles were included.

Both methods used in previous studies to look at concepts in specialty areas, direct citation analysis and co-citation clustering, were used for data production. For direct citation analysis the records were sorted and counts were made of the number of occurrences of each reference document, i.e. the number of times each reference was cited by other documents. The method previously used by Small and Griffith was followed for the co-citation clustering [1]. The reference files developed (above) were used for data. Only the items cited four times or more were selected from each dataset for the clustering ensuring a large enough sample in each data set for the clustering. The Jaccard coefficient, often used in co-citation clustering as a normalization procedure, was followed [7]. The clustering was carried out using a complete link clustering algorithm developed at the University of Western Ontario [16].

3.2 *Concept group selection*

This research involves time analysis to identify any change in terminological usage. As a consequence, candidate concept groups were restricted to the cited documents and co-citation clusters that appear in all three time periods used in this study. The requirement for inclusion of co-citation clusters is that at least two documents (a co-citation) in a cluster must appear in the three time periods.

The cited document data produced 235 continuing cited documents in sociology and 130 continuing cited documents in economics. The most frequently cited documents (i.e. the top ten in each discipline) were used since such highly cited documents are uniformly associated with particular concepts by a large number of authors [12]. Table 1 presents the most cited titles and the number of citations to them in each time period.

The co-citation clusters data produced only three co-citation clusters in each of the two disciplines that continued through the three time periods. This small number of continuing clusters is not surprising and fits in with earlier findings of Small and Crane that fewer clusters in sociology and economics emerge in co-citation clustering [6]. Table 2 presents the constituent titles for each cluster, the co-citation items as well as any other titles that appeared in any of the three time period clusters. One of the sociology clusters is in fact a tri-citation, i.e., the three books by Blau, Homans and Thibaut clustered together in each of the three time periods.

3.3 *Citation context selection*

The data sets were again scanned to provide at most, ten citation contexts in each time period for each highly cited document or co-citation document pairs or triples. A set of rules were developed to define a citing article and a citation context:

1. The citing document must be in English.
2. A footnote or reference is required that is actually embedded in the text of the citing document, following Small's idea that a footnote functions as a pointer to a portion of the text.
3. The footnote or reference must give a full bibliographic reference to the specific highly cited or co-cited pairs. This rule was established to ensure that the particular cited document was being referenced, not another document by the same author.
4. The context of co-cited documents must include the co-citations within the same context, i.e. a discussion of all co-cited documents within a specific section of text.
5. Exact quotations from the cited documents are not used. This study is interested in alternate terminology.
6. Only the context in the citing document that is actually describing the cited document's concepts is to be used. At most two to three sentences were found to be appropriate, supporting Small's earlier findings [12].
7. Multiple citation contexts within one document may be used as long as the above rules are followed.

Table 1. Highly cited documents

Document	Number of Citations 1966–67	1973–74	1980–81
Economics			
Arrow, K.J. *Social Choice and Individual Values.* Wiley, 1951.	14	41	56
Becker, G.S. *Human Capital.* National Bureau of Economic Research, 1964.	20	41	49
Debreu, G. *Theory of Value.* New York: Wiley, 1959.	7	64	65
Friedman, Milton. *A Theory of the Consumption Function.* Princeton, N.J.: Princeton University Press, 1957.	24	47	51
Goldberger, A.S. *Economic Theory.* New York: Wiley, 1964.	9	64	47
Johnston, J. *Econometric Methods.* New York: McGraw-Hill, 1963.	23	94	137
Keynes, J.M. *The General Theory of Employment, Interest and Money.* New York: Harcourt, 1936.	14	60	130
Marshall, A. *Principles of Economics.* New York: Macmillan, 1890.	16	46	79
Musgrave, R.A. *The Theory of Public Finance.* New York: McGraw-Hill, 1959.	26	52	32
Samuelson, P.A. *Foundations of Economic Analysis.* Cambridge, Mass.: Harvard University Press, 1947.	10	69	89
Sociology			
Becker, H.S. *Outsiders:* Studies in the Sociology of Deviance. New York: Free Press, 1963.	14	43	40
Berger, P.L. and Luckmann, T. *Social Construction of Reality.* New York: Doubleday, 1966.	4	54	69
Durkheim, E. *Division of Labor in Society.* New York: Free Press, 1947. 1st French ed. 1893.	10	58	61
Goffman, E. *Asylums.* New York: Doubleday, 1961.	19	58	45
Goffman, E. *The Presentation of Self in Everyday Life.* New York: Doubleday, 1956.	16	41	49
Homans, G.C. *Social Behavior.* New York: Harcourt, 1961.	18	57	85
Kuhn, T.S. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press, 1962.	4	47	87
Merton, R.K. *Social Theory and Social Structure.* Glencoe, Ill.: Free Press, 1949.	49	132	118
Parsons, T. *The Social System.* Glencoe, Ill.: Free Press, 1951.	25	75	74
Weber, M. *The Theory of Social and Economic Organization.* Oxford University Press, 1922.	18	61	77

8. Citations that are merely perfunctory, i.e., the document mentioned or acknowledged but not described, were discarded.

All other citation contexts, even if the reference was negative, were used. Any articles or contexts that did not meet the rules were discarded and another article or context substituted if available.

3.4 *Phrase selection*

The next step involved the analysis of the citation contexts and the phrases extracted, and a set of rules governing what phrases and terms were to be extracted from the context was established. Noun phrases as the basic term units were used followed by the international standards for thesaurus development [17]. Complex noun phrases were broken down into their constituent parts as much as possible. Phrases such as "supply of money," "presentation of self," were considered separate units. Appendix A lists the phrases from two concept groups, one in economics and one in sociology. All noun phrases from the contexts were included, none were discarded.

The citation context term data was selected document by document, context by context, and was organized into a database and then sorted in term lists by cited document and data. The disappearance or appearance of terms in a particular time frame may reflect change in concepts.

A comparison of the three phrase lists within each concept group representing the different time periods was carried out. The best method to try to note any change in the lists

Table 2. Co-citation clusters

Co-Citation Documents and Others in Cluster	Number of Co-citations 1966/67	1973/74	1980/81
Economics			
1. Cagan, P. *Determinants and Effects of Changes in the Stock of Money, 1875-1960.* National Bureau of Economic Research, 1965.	3	4	3
Friedman, M. and Schwartz, A.J. *A Monetary History of the United States, 1867-1960.* National Bureau of Economic Research, 1963.			
Bruner, K. and Meltzer, A. *An Alternative Approach to the Monetary Mechanism.* Staff Analysis... House Committee on Banking and Currency, 88th Congress, 2dSess. Aug. 1964.	x		
Kareken, J. and Solow, P.M. "Lags in Monetary Policy," In Commission on Money and Credit, *Stabilization Policies.* Englewood Cliffs, NJ: Prentice-Hall, 1963.	x		
Laidler, D. *The Demand for Money.* 2ded. New York: Dun-Donnelly, 1977.			x
2. Kaldor, N. "Alternative Theories of Distribution." *Review of Economic Studies*, 22(1956):83-100.			
Pasinetti, L. "The Rate of Profit and Income Distribution in Relation to the Rate of Economic Growth." *Review of Economic Studies*, 29(1962):267-79.	1	9	5
Hahn, F.H. "A Review of Pasinetti More Often Covered in the Literature." *Economic Journal*, 74(1964):779-902.	x		
Harrod, R.E. *Towards a Dynamic Economics.* London: Macmillan, 1948.		x	
Robinson, J. *Essays in the Theory of Economic Growth.* London: Macmillan, 1962.		x	
3. Lipsey, R.G. "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom 1862-1957, A Further Analysis." *Economica*, 105(Feb. 1960):1.			
Phillips, A.W. "The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861-1957." *Economica*, 100(Nov. 1958):285.	6	13	6
Friedman, M. "The Role of Monetary Policy." *American Economic Review*, 58(1968):1-17.		x	
Phelps, F.S. "Phillips Curves, Expectations of Inflation and Optimal Unemployment in Our Time." *Economica*, 34(1967):254-81.		x	
Sociology			
4. Heider, F. "Attitudes and Cognitive Organization." *Journal of Psychology*, 21(1946):107-112.			
Heider, F. *Psychology of Interpersonal Relations.* New York: Wiley, 1958.	7	5	3
Festinger, L.A. *Theory of Cognitive Dissonance.* Evanston, Ill.: Row, Petersen, 1957.	x		
Jordan, N. "Behavioral Forces that are a Function of Attitudes and Cognitive Organization." *Human Relations*, 6(1958):273-82.	x		
Osgood, C.E. and Tannenbaum, P.H. "The Principle of Congruity in the Prediction of Attitude Changes." *Psychological Review*, 62(1955):42-55.	x	x	
Newcomb, T.M. "An Approach to the Study of Communicative Acts." *Psychological Review*, 60(1953):393-404.		x	
Cartwright, D. and Harary, D. "Structural Balance." *Psychological Review*, 63(1956):227-93.			x
5. Blau, P.M. *Exchange and Power in Social Life.* New York: Wiley, 1964.			
Homans, G.C. *Social Behavior.* New York: Harcourt, 1961.	2	2	2
Thibaut, J.W. *The Social Psychology of Groups.* New York: Wiley, 1959.			
6. Cloward, R.A. and Ohlin, L.E. *Delinquency and Opportunity.* New York: Free Press, 1960.			
Cohen, A.K. *Delinquent Boys.* New York; Free Press, 1955.	15	13	11
Lander, B. *Towards an Understanding of Juvenile Delinquency.* New York: Columbia University Press, 1954.	x		
Miller, W.B. "Lower Class Culture as a Generating Milieu of Gang Delinquency." *Journal of Social Issues*, 14(1958):5-19.	x		

Note: x indicates the year in which the document appears in the cluster.

over time was to note what sub-concept areas continued through the three lists and what sub-concept areas appeared in one or two time periods only, i.e., were non-continuing. (See Table 3 for example). This process involved careful reading of the phrase lists noting phrases that were visibly the same or similar and those that were not visibly the same but were conceptually similar, i.e., "money supply" and "stock of money." These phrases were then organized into these sub-concept groups (See Table 3). Specialty area phrases noted under Original Documents in the table refer to phrases from a published review or abstract of the original document(s).

3.5 *Evaluation of the candidate thesaurus terms*

To test the resulting data from this methodology, experts in the disciplines of economics and sociology were used to verify and comment on the resulting candidate thesaurus terms. The phrase lists were divided into four sets, two in sociology and two in economics. Two tests were carried out by the different experts on each data set. The test sets were divided to put similar subject documents in the same test set (see Tables 4 and 5). With three phrase lists for each concept group, each economic or sociology test set con-

Table 3. Quantity theory of money

Original Documents[1]	1966–67 Citations	1973–74 Citations	1980–81 Citations
Continuing Concepts			
Stock of money			*Growth in the stock of money
Money stock			
Supply of money	*Money supply		*Quantity theory of money
			Quantity theorists
		*Monetarist view	*Monetarist approach
	Decline in the money supply		
Behavior of Money			
Money supply	*Variations in the rate of change of money	Monetary contractions up to 1896	Monetary changes
Changes in money			
Changes in money		*Monetary variables	*Monetary variables
Non-continuing Concepts			
	Monetary history	Economic history	
Business cycles	*Theory of the cycle		
Banks( )			
Banking system	Central bankers		
Banking and financial structure			
Federal Reserve system	Federal Reserve( )		
	*Departures from equilibrium( )		
		*Full employment model	
			*Expectations on interest
			*Interest rates
	*Stable relationships		

_____ Library of Congress subject heading
_ _ _ _ Library of Congress "see from" reference
( ) Library of Congress subject heading similar to but not the same as phrase extracted
* Considered an essential phrase by expert

[1]Phrase from a published review or abstract of the original document.

Table 4. Concept groups – economics

Document(s)	Experts' grouping Test A	Experts' progression[1] Test A	Experts' progression[1] Test B	Experts' label Test A
Expert 1				
Cluster Cagan/Friedman	Correct	3,2,1	1,2,3	Quantity theory of money.
Cluster Lipsey/Phillips	Correct	1,2,3	2,1,3	The Phillips curve.
Freidman	Correct	1,3,2	3,2,1	Permanent income hypothesis.
Keynes	Correct	2,3,1	2,1,3	Keynesian macro-economics.
Samuelson	Correct	1,2,3	1,2,3	Consumer theory.
Goldberger ↔	With Johnston	No progression	2,1,3 2,3,1	Econometrics.
Johnston ↔	With Goldberger	No progression 3,2,1	1,3,2	Econometrics.
Expert 2				
Cluster Kaldor/Pasinetti	Correct	1,3,2	No progression	Macro-economic treatment of aggregate saving behavior.
Arrow	Correct	1,3,2	3,1,2	Public finance. Social choice. Public choice. Welfare economics aspects.
Becker →	With Marshall	2,3,1	No progression	Labor economics with human capital, education.
Debreu	Not correct	1(alone)	3,2,1	Micro-economics.
→	with Marshall/ Musgrave	2,3		General equilibrium aspect of public finance.
Marshall	Not correct	3(alone)	3,2,1	Principles of economics.
→	with Becker	1		Labor economics.
→	Debreu/Musgrave	2		Public finance.
Musgrave ↔	With Debreu/ Marshall	2,3,1	No progression	Public finance.

[1]Progression order as given by expert, earliest to latest. For actual order, 1 = 1966–67, 2 = 1973–74, 3 = 1980–81.
↔ indicates interconnections of phrase lists made by expert in Test A.

sisted of eighteen or twenty-one phrase lists. The experts were given the alternate test set in their field of expertise for each Test A and Test B. Test A gave the experts the phrase lists in random order. Test A asked the experts to organize the phrase lists into concept groups, to label the concept groups and arrange the phrase lists within these concept groups in order from earlier to later. Test B gave the experts the phrase lists already correctly grouped but in random order and asked the experts to judge the appropriateness of the selected terms (i.e. note essential terms and inappropriate terms), to arrange the phrase lists from earlier to later, to describe the concept area and to describe any change found.

A comparison of the central concepts extracted from the contexts with those found in the *Library of Congress Subject Headings* was carried out [18]. If the resulting terms are similar to that found in a standard subject heading list, these terms may very well be appropriate candidate terms for a thesaurus in the specialty area. The level of phrase agreement between the specialty area, continuing and non-continuing phrases and the Library of Congress subject headings, was examined. The level of cross reference between these phrases was noted as well.

4. RESULTS

Citation context analysis does a substantially correct job in developing phrase descriptions of concept groups. Overall, seven of the thirteen economics groups and eight of the

Table 5. Concept groups—sociology

Document(s)	Experts' grouping	Experts' progression[1] Test A	Experts' progression[1] Test B	Experts' label Test A
Expert 1				
Cluster Heider/Heider	Correct	3,1,2	Not done	Not done
Cluster Cloward/Cohen	Correct	3,2,1	3,1,2	Theory of delinquency as subcultural process.
Becker	Correct	1,2,3	1,2,3	Development of the theory of delinquent subculture.
Goffman-Asylums	Correct	1,3,2	Not done	Total institutions/ social control for deviant behavior.
Goffman-Presentation of Self	Correct	3,1,2	Not done	Symbolic interaction-ism/ role-playing/ self-concept and place in social system.
Cluster Blau/ Homans/Thibaut	With Homans	3,1,...2	3,1,2	Development of exchange theory.
Homans	With Blau/ Homans/Thibaut	3,1,2	3,1,2	Development of exchange theory.
Expert 2				
Berger	Correct	1,3,2	1,3,2	Philosophical issues.
Kuhn	Correct	3,1,2	Not done	Research methodology.
Weber	Correct	1,2,3	2,3,1	Bureaucracies and administrative control.
Durkheim	With Parsons/ Merton	2,3	2,1,3	Deviance and social control.
	With Merton/ Parsons	1		Social change.
Merton	With Durkheim/ Parsons	3	Not done	Deviance and social control.
	With Durkheim/ Parsons	1,2		Social change.
Parsons	With Durkheim/ Merton	1	No progression	Social change.
	With Durkheim/ Merton	3		Deviance and social control.
	Merton alone	2		Not labelled

[1]Progression order as given by expert, earliest to latest. For actual order, 1 = 1966–67, 2 = 1973–74, 3 = 1980–81.
↔ Indicates interconnections of phrase lists made by expert in Test A.

thirteen sociology groups were correctly organized together by the experts, even though there were differences between the experts' ability to organize the classes correctly (see Tables 4 and 5). The column titles Experts' Grouping refer to the three applicable phrase lists which make up one concept group and whether they were grouped correctly together, or were grouped with a phrase list from another concept group. Other phrase lists grouped by the expert with the concept group are indicated. It should be remembered that the test set phrase lists were given to the experts in random order in Test A, but in Test B the three phrase lists were grouped together but were in random order in the group. The phrase lists were to be placed in a progression in each Test A and Test B. These placements are indicated in the column titles Experts' Progression. In these columns the time period of the phrase list (i.e. 1, 2, or 3) are indicated in the order as given by the expert. In Table 4, Cagan/Friedman, in Test A when the expert was asked to order the lists earlier to later, the expert placed the 1980–81 list first, the 1973–74 list second and the 1966–67 list third. The lines in Tables 4 and 5 indicate how the experts grouped the phrase lists. For full detail of the experts' progressions the full text of the dissertation must be consulted. In several groups the expert indicated no progression found. The last column (Experts' Label) is the label given by the expert in Test A.

A redundancy measure was used to compare the experts' concept groupings to the groupings provided by the high citation and co-citation cluster phrase sets. This measure was introduced by Young [19], and was suggested as a measure of similarity between classifications by Guazzo [20], and used by Tague and Shepperd [21] to assess the relationship between pairs of classifications. This measure describes the proportional reduction in uncertainty regarding concept group (X) provided by the experts' phrase list grouping (Y):

$$I(X/Y) = (H(X) - H(X/Y))/H(X).$$

$H(X)$ is the uncertainty regarding the concept group to which a phrase list belongs:

$$H(X) = \sum_x \frac{f(x)}{n} \log_2 \frac{f(x)}{n}$$

where $f(x)$ is the number of phrase lists for concept group X and n is the total number of phrase lists. $H(X/Y)$ is the uncertainty regarding the concept group to which a phrase list belongs given its classification by the expert.

$$H(X/Y) = \sum_x \sum_y \frac{f(x,y)}{n} \log_s \frac{f(x,y)}{f(y)}$$

where $f(x,y)$ is the number of lists common to concept group X and expert Y. The values for each expert were as follows:

Discipline	Expert	$I(X/Y)$
Sociology	1	.90
	2	.74
Economics	1	.90
	2	.64

High values (1.0) indicate high agreement between the lists, low values (0) indicate no agreement. The above figures show the differences between the two experts in each discipline area and show overall a fairly high level of agreement between the experts organization of the lists and the citation context analysis organization of the lists. The result shows that citation context analysis does produce fairly correct descriptions of concept groups and agrees with the work of Garfield [22], Small [14,23], McCain [24], Cozzens [25]. This analysis is made strictly from the point of the citation-based groupings.

Tables 4 and 5 display the experts' grouping of the concept groups and the specialty areas in order to present clearly the relations selected by these experts. Overall, the experts were able to label the concept groups accurately and appropriately. Specialty area labels from a published review or abstract of the original document(s) were used to identify these areas (data not presented here). The Experts' Label are those given by the expert in Test A to the specialty area. When the subject of the concept group is considered, the experts groupings of some of the phrase lists can be understood. For example, the Goldberger and Johnston concept groups are both in the specialty area of econometrics. The Debreu, Marshall and Musgrave concept groups each concern aspects of the specialty area of public finance. Similar grouping decisions can be seen in the sociology concept groups. In economics, for all of the thirteen concept groups, appropriate labels were applied to the phrase lists. In sociology, ten of the concept groups are clearly labelled correctly. An analysis of the detail of the errors shows a subject historical similarity in the specialty areas (i.e. Merton, Parsons). In other words, the document groupings represented concept groups fairly well and thus can be accepted as concept symbols.

With respect to terminological change, the results of this research were disappointing. Tables 4 and 5 show the results. In economics, out of the thirteen concept groups, two of the experts' progressions in Test A were correct and two in Test B. The consumer theory concept group was correct in both tests. In sociology, out of the thirteen concept groups,

two progressions in Test A and one in Test B were correct. The Becker concept group in both tests were correctly ordered. There is a lack of evidence for the experts that the phrase lists show change over time. The experts' ordering of the phrase lists was compared with the citation year ordering using Kendall's Tau as a correlation coefficient. The resulting value of Tau was .06 indicating virtually no discernable time progression in the phrase lists as seen by the experts.

The third aspect of the concept groups that is important to consider, is the extent to which the phrase lists contain appropriate phrases, considering the specialty area. The experts were asked to indicate the phrases they consider essential and those they consider inappropriate to the specialty area. Appendix A gives an example of this data for the Quantity Theory of Money specialty area and the Presentation of Self specialty area. Table 6 aggregates this data for all concept groups. There is a higher proportion of essential phrases (i.e., 19.8 percent in sociology and 37.8 percent in economics) than of inappropriate phrases, (i.e., 1.3 percent in sociology and 6.3 percent in economics) suggesting that the experts considered the phrases as appropriate in application to the various specialty areas. If all the essential phrases and those not deleted as inappropriate are considered as a group, the judges considered a high proportion of these phrases as appropriate to the specialty areas. On the average, 98.7 percent of the phrases in sociology, and 93.7 percent of the phrases in economics were judged appropriate to the specialty areas. The lowest level of appropriate phrases (i.e., the lowest level of appropriate phrases in any phrase list) is 93.2 percent in sociology and 74.3 percent in economics, which is a larger portion of the phrases.

This project attempts to establish a mechanism to maintain and update a thesaurus. It would be important that the phrases extracted would be appropriate candidates for a thesaurus. Consequently, it is important that these extracted phrases match, (to some extent), subheadings found in a standard subject heading scheme.

Generally L.C. subject heading matching was better in economics than sociology. A brief summary is presented in Table 6. There was no specialty area in either discipline that had no representation by L.C. headings. There was one specialty area in each discipline that was poorly covered by L.C. headings. These were "social choice" in economics and "balance theory" in sociology. The methodology seems to produce a rich supply of alternate terminology for the specialty areas. There was on the average 13.6% of the appropriate terms in the economics groups and 12.7% of the appropriate terms in sociology groups that matched Library of Congress subject terms.

There is also a level of cross reference between some of these similar L.C. headings. It is expected that there is some level of interrelation through cross reference within these specialty areas. This did prove to be the case. There is an average of 4.3 (see references in the economics groups), and an average of 4.9 (see references in the sociology groups). Considering the "see also references," there is an average of 12.1 "see also references" in the economics groups and 5.8 "see also references" in the sociology groups. It must be remem-

Table 6. Candidate thesaurus terms

	Economics percent	Sociology percent
Average essential phrases	37.8	19.8
Average inappropriate phrases	6.3	1.3
Average appropriate phrases	93.7	98.7
Lowest level appropriate phrases per concept group	74.3	93.2
Average appropriate phrases matching L.C. subject headings	13.6	12.7
	Number	Number
Average number of L.C. subject terms per concept group	13.7	13.8
Average number of see references per concept group	4.3	4.9
Average number of see also references per concept group	12.1	5.8

bered that these counts include all the references among the terms in a specialty area. If the number of L.C. terms in each specialty area is considered, there is generally a level of interconnection between these terms.

As a consequence, it can be said that the methodology established in this study seems to produce a good number of concept phrases that are compatible at a broader level of concept with subject headings in a standard vocabulary listing system such as the Library of Congress subject headings. The methodology produces many more terms that are appropriate and interrelated to concepts in a specialty area than those used in the Library of Congress system and thus many more alternate terms that may be useful to indexers and searchers. This is the core of information needed to produce a dynamic thesaurus.

5. DISCUSSION

This study investigated a mechanism by which thesauri can be updated and maintained using citation and co-citation analysis and citation context analysis. From this research, several conclusions were reached concerning the methodology, the terms generated, and change and variation in the terms. Generally, terminology used in citing paper contexts reflects concepts of a specialty area. Thus, highly cited and co-cited documents do act as concept symbols, as originally suggested by Small. This methodology provides a rich source of alternate terminology to describe specialties and the concept areas within specialties in economics and sociology. This is true for every specialty area studied.

Specialty areas are as easily identified in sociology as in economics. In all, thirteen specialty areas in economics and thirteen areas in sociology were analyzed and two experts in each field were used to judge the results. The difference between the disciplines and the ability to correctly label the specialty is very slight. Overall, the method was able to develop similar high levels of appropriate phrases for the specialties in both disciplines (i.e., 93.7 percent appropriate in economics, 98.7 percent appropriate in sociology).

There is little difference between the two concept-forming methodologies of highly cited documents and co-citation cluster documents with respect to ability to group, label and select appropriate phrases for specialties. The only difference was that co-citation-based clusters provided more essential phrases in sociology than did highly cited documents. However, the number of inappropriate phrases was similarly low for both methodologies in sociology.

Little indication of terminological change over time was found in this study. However, the experts were able to correctly identify the different specialty areas, and were able to identify variation between individuals and groups. One expert was able to differentiate the specialty areas of delinquency as a subcultural process and the theory of delinquent subcultures, in spite of similarity of concept areas covered by the two specialties.

It should be noted that a high proportion of the highly cited and co-cited documents used in this study were monographs. This study used these citations as concept symbols and a focal point for context and terminological analysis. It may be that there is a difference between citation contexts and citation terminological analysis of monographs and periodical articles in the terminology these different concept symbols would produce. This research did not explore this point. As well, highly cited documents appear to represent established concepts in a field. It may be a point for further research to investigate the citation contexts of less frequently cited documents. These may produce concept areas that are less stable and thus have greater tendency to change.

In a few specialty areas there was some indication of a shift in emphasis in the specialty area over time [26]. In the quantity theory of money, there was a shift in methodological emphasis, from "monetary history," to "monetary approach," "monetary variables," "monetary aggregates." In the specialty human capital, there was a shift from emphasis on education to education as skills, development and training, i.e., education as a metaphor. In the specialty area of social and economic organization, there was a shift from a political-social perspective to a psycho-social perspective.

A few specialty areas show an increasing tendency to specialization. In the quantity theory of money specialty, there was a shift from the detail of the approach to a labelling

of the specialty. Finally, there was a greater specificity and sophistication in the concepts, as in the Weber concept group, for example, with the concepts concerning organization, i.e., "organizational rationality," "organizational stability."

No particular type of change and variation in terminology was found in this study other than meaning shift. The time period of fifteen years does not seem to be extensive enough to allow for detection of terminological change.

The central and important concepts in specialty areas may be outlined by organizing the continuing and non-continuing phrases in specialty areas. This is similar to a finding by Small and his concept of percentage uniformity.

It was interesting to note the wide variety of connotations in everyday language and a large number of the technical terms in both economics and sociology. Yet these terms are used with great frequency in the scholarly language of economists and sociologists. This finding certainly agrees with many previous comments on the lack of precision of concepts in these disciplines [27]. It has also led to problems in the original extraction of the noun phrases used in this study. Such phrases as "propensity to save", "outputs", "demand", "utility", "conformity", "human social reality", "label", "audience", "performance", "differentiation", "solidarity", were all considered essential phrases by the experts. On the other hand, in reading the contexts of these phrases, the degree of repetition of these technical terms in many articles points to their standard use and stability in the specialty areas.

The objective of this research is to develop a mechanism to update and maintain a thesaurus with *relatively* efficient procedures. Throughout the study, no truly efficient procedure for extracting the terminology became apparent. However, a computer based system could provide assistance with many other aspects of the procedure carried out. Some of this work is similar to that of O'Connor [28] with the computer use of citing statements for retrieval, and Small [29] with the synthesis of specialty narratives.

The basic assumption is that a full text database is available with papers that have references, whether listed at the end of the paper or as footnotes. It is assumed, as well, that a complete and current list of highly cited papers in a discipline is available.

The procedures necessary involve the identification of citations to concept markers (highly cited papers or co-citation clusters), the displaying of the full text of these markers in new papers, and the selection of new terms or new relations (interconnections) by the compiler. New terms or relations could then be introduced to the thesaurus. Overall, the procedures involve the development of three modules:

1. A citation identification module which, given the identification of the highly cited document or co-cited documents, searches the full text database for occurrences of these among the references.
2. A context identification module which finds the actual citation within the text and displays the context.
3. A thesaurus maintenance module which will accept the new phrases into the thesaurus and specify the linkage of the new phrases to a concept and their relation to existing ones for the same concept.

These modules are currently being developed and adjusted to operate in a micro-based environment. The author is currently working on a fourth module, the concept record which will substantially organize the concepts at the term level with the necessary relations to form the syntactic structure of the system.

The next stage of this research involves the testing of the programs on another database containing full text of documents and papers. It would be most appropriate to use documents in a small specialty area to see if an adequate and appropriate thesaurus of specialty terms could be generated. A full text database in the legal field of judicial decisions is being considered to meet the general requirements of the system. Judicial decisions are available in full text form in electronic database format. The full text of legal decisions have references embedded in the text of decisions, i.e., as a reference to precedence. Also, from several published lists of "decision judicially considered," the frequently cited deci-

sions could be organized and used as concept markers. With the increasing development of computer-readable full text bases, high quality thesauri are needed. This methodology has a potential role for thesaurus maintenance.

REFERENCES

1. Small, Henry G.; Griffith, B.C. The structure of scientific literature, I. identification and graphing specialties. Science Studies, 4: 18-40; 1974.
2. Griffith, B.C., et al. The structure of scientific literature, II. Towards a macro and micro structure for science. Science Studies, 4: 339-365; 1974.
3. Mullins, N.C., et al. The group structure of co-citation clusters: A comparative study. American Sociological Review, 42: 552-562; 1977.
4. Griffith, B.C. The social and behavioral sciences literature. Paper presented at the American Society for Information Science Conference, Minneapolis, Minn.; 1979.
5. Small, Henry G. A co-citation model of a scientific specialty: A longitudinal study of collagen research. Social Studies of Science, 7: 139-166; 1977.
6. Small, Henry G.; Crane, D. Specialties and disciplines in science and social science: An examination of their structure using citation indexes. Scientometrics, 1: 445-461; 1979.
7. Small, Henry G.; Greenlee, E. Citation context analysis of a co-citation cluster: Recombinant DNA. Scientometrics, 2: 277-301; 1980.
8. Sullivan, D., et al. The state of a science: Indicators in the weak interactions. Social Studies of Science, 7: 167-200; 1977.
9. Cole, S.; Cole, J.R.; Dietrich, L. Measuring the cognitive state of scientific disciplines. In: Edited by Elkana, Y., et al. Toward a metric of science. New York: Wiley and Sons; 1978: 209-251.
10. Nadel, E. Citation and co-citation indicators of a phased impact of the BCS theory in the physics of superconductivity. Scientometrics, 3: 203-1; 1981.
11. Nadel, E. Multi-variate citation analysis and the changing cognitive organization in a speciality of physics. Social Studies of Science, 10: 449-473; 1980.
12. Small, Henry G. Cited documents as concept symbols. Social Studies of Science, 8: 327-340; 1978.
13. Cozzens, Susan E. Split citation identity: A case study from economics. Journal of the American Society for Information Science, 33: 233-236; 1982.
14. Small, Henry G. Co-citation context analysis and the structure of paradigms. Journal of Documentation, 36: 183-196; 1980.
15. Griffith, Belver C. Introduction. In: Key papers in information science. White Plains, N.Y.: Knowledge Industry Publications; 1980: 1-2.
16. The co-citation clustering work was done by Charles McClelland at the School of Library and Information Science, University of Western Ontario.
17. ISO, Guidelines for the establishment and development of monolingual thesauri, section 3.2.5. Geneva: I.S.O.; 1974.
18. Library of Congress subject headings, 9th ed. Washington, D.C.: Library of Congress; 1980, was mainly used. Updates to Dec. 1985 were checked as well as previous editions of this title, the 7th ed. 1966 and the 8th ed. 1975.
19. Young, J.F. Information theory. New York: Wiley; 1971.
20. Guazzo, M. Retrieval performance and information theory. Information Processing and Management, 13: 155-165; 1977.
21. Tague, Jean; Sheppard, Michael J. Redundancy as a measure of classification similarity. Journal of Informatics, 2: 123; 1978.
22. Garfield, Eugene. Can citation indexing be automated? in: M.E. Stevens, V.E. Giuliano and L.B. Heilprin editors. Statistical association methods for mechanized documentation. Washington, D.C.: NBS; 1965; 189-192.
23. Small, Henry G. The synthesis of specialty narratives from co-citation clusters. Journal of the American Society for Information Science, 37: 97-110; 1986.
24. McCain, Katherine. Co-cited author mapping as a valid representation of intellectural structure. Journal of the American Society for Information Science, 37: 111-122; 1986.
25. Cozzens, Susan. Comparing the sciences. Social Studies of Science, 15: 127-153; 1985.
26. For detail examination of the individual concept group analysis, the original work must be consulted.
27. Riggs, Fred. W. The importance of concepts: Some considerations on how they might be designated less ambiguously. American Sociologist, 14: 172-185; 1979.
28. O'Connor, John. Citing statements: Computer recognition and use to improve retrieval. Information Processing and Management, 18: 125-131; 1982.
29. Small, Henry J. The synthesis of specialty narratives from co-citation clusters. Journal of American Society for Information Science, 37,3: 97-110; 1986.

APPENDIX A

QUANTITY THEORY OF MONEY PHRASES

1966–67	1973–74	1980–81
×1950	ECONOMIC HISTORY	DEFLATION
BANK FAILURES	*FULL EMPLOYMENT MODEL	*EXPECTATIONS ON INTEREST RATES
CENTRAL BANKERS	GREAT DEPRESSION	*GROWTH OF REAL OUTPUT
DECLINE IN THE MONEY SUPPLY	*MONETARIST VIEW	GROWTH OF THE STOCK OF MONEY
*DEPARTURES FROM EQUILIBRIUM	MONETARY CONTRACTION UP TO 1896	INCOME EFFECT
FEDERAL RESERVE	*MONETARY VARIABLES	*INTEREST RATES
INSURANCE PROVISIONS	×ROLE OF DEMAND	LIQUIDITY EFFECT
MONETARY HISTORY		*MONETARIST APPROACH
*MONEY SUPPLY		*MONETARY AGGREGATES
SAVINGS AND LOAN ASSOCIATIONS		MONETARY CHANGES
×SHARES		MONETARY VARIABLES
*STABLE RELATIONSHIPS		PERIODS OF INFLATION
*THEORY OF THE CYCLE		PRICE LEVEL MOVEMENTS
*VARIATIONS IN THE RATE OF CHANGE OF MONEY		QUANTITY THEORISTS
		*QUANTITY THEORY OF MONEY

*Expert's indication of essential term
×Expert's indication of inappropriate term

PRESENTATION OF SELF PHRASES

1966–67	1973–74	1980–81
*ACTORS SELF-PRESENTATION	COGNATE PHENOMENA	ASSOCIATION
ANALYSIS OF REALISM	*CONCEPT OF IDENTITY	*AUDIENCE
DESIRED RESPONSE	*FAMILY	BACK REGIONS
DIFFERENTIAL RELATIONSHIP	*FEEDBACK	BACKSTAGE
FORMAL ORGANIZATIONAL STRUCTURES AND PROCESSES	FRONT REGION	DEFERENCE RITUAL
	GROUP ASSIGNED LABELS	*DRAMATURGIC TECHNIQUES
HUMAN INTERACTION	IMPACT OF A DISCLOSURE	DRAMATURGIC MODEL
IDENTITY DEVELOPMENT AND MAINTENANCE	IMPRESSION	FRONT REGIONS
	*IMPRESSION MANAGEMENT	HIGHER RANKED
*IMPRESSION MANAGEMENT	INFORMAL RULES	*IMPRESSION MANAGEMENT
INGRATIATION	INSTITUTIONAL ROLES	INFORMATION CONTROL
INTERACTION IN THE REAL WORLD	INTERACTIONIST APPROACH	INTERACTION
	INTERNALIZATION AND INTEGRATION	LOWER OBSERVABILITY
INTERPERSONAL CONTROL	ORGANIZATIONAL SOLIDARITY	LOWER-RANK PEOPLE
MAXIMUM BENEFIT	REFERENT	*PERFORMANCE
NORMATIVE PATTERN	REGION BEHAVIOR	*PRESENTATION OF SELF
OBJECTIVE PUBLIC EVALUATION	REGIONS	RITUAL ASPECT
	ROLE OCCUPANCY	SOCIAL AFFECTIVE COMMUNICATION
ORGANIZATIONS		

continued on facing page

PRESENTATION OF SELF PHRASES *continued*

1966-67	1973-74	1980-81
PEER CULTURE	ROLE PARTNERS	SOCIAL CONTROL
PERCEIVED LONGER TERM CONSEQUENCES	ROLE-TYPE BEHAVIOR	SOCIAL INTERACTION
PERCEPTION	*ROLES	TEAM PERFORMANCE
*PERSONAL IDENTIFY	SELF-ASSIGNED LABEL	VEHICLE FOR THEIR INTERACTION
PROCESSES OF SELF-IDENTITY	*SELF-CONCEPT	
ROLE DISTANCE	SELF-CONCEPTION	
ROLE EMERGENCE	SELF-CONCEPTS	
ROLE THEORY	SELF-LABELING	
*ROLE-PLAYING	SELF-PRESENTATION	
RULEFULNESS	*SITUATIONAL IDENTITY	
SELECTION OF BEHAVIOR	SITUATIONALLY SPECIFIC	
*SELF IDENTITY	SOCIAL ACTIVITY	
SELF-EVALUATION	SOCIAL ENCOUNTERS	
SITUATIONAL DEFINITIONS	SOCIAL SITUATION	
SITUATIONAL STIMULLI	SOCIAL SPACE	
SOCIAL ENCOUNTERS	SOCIAL SYSTEMS	
SOCIAL MECHANISMS	SOCIALIZATION	
SOCIAL STRUCTURE	SOCIETAL RESPONSES	
SUB-CULTURE	SOCIOLOGICAL SITUATIONS	
SUBJECTIVE PUBLIC EVALUATION	SOLIDARITY SCALE	
SUBJECTS' KNOWLEDGE	*STATUSES AND ROLES	
SYMBOL VERIFICATION	SUITABLE IDENTITY	
SYMBOLIC INTERACTIONISTS	TEAM	
TARGET PERSON'S COMPLIANCE	TEAM MEMBER	
TOTAL SYSTEM	TEAM PERFORMANCE	
VARIATION IN BEHAVIOR	TEAMMATES	
WORK SITUATIONS	TEAMWORK	
WORKING CLASS BEHAVIOR		

*Expert's indication of essential term

ˣExpert's indication of inappropriate term