

## Structuring of Compound and Complex Subjects in Social Sciences: A Users' Survey\*

Raghavan, K. S., Hemalata Iyer: Structuring of compound and complex subjects in social sciences: a users' survey.

In: Intern. Classificat. 5 (1978) No. 1, p.8–14

Presents an analysis of the detailed survey on facet-analysed subject strings in the field of Social Sciences. Subject-strings of four types were administered as a stratified sample of responses to identify the gaps in semantic connotation of terms. This survey has helped to identify the possible aberrations in the semantic connotations due to the presence or absence of terms. It was found that use of prepositions, conjunctions and other role indicating terms give semantic clarity. The syntax of speciator terms may have to be varied according to the respondent groups. It was found that if the number of component terms in a subject string increases beyond a certain limit it gives rise to problems in interpretation of ideas. This indicates a breakeven point for formulating subject headings.

Authors

### 1. Introduction

Information systems collect, process, store and retrieve information. Most information systems handle subjects of documents or readers' queries. Among other things, the success of such information systems depends to a very large extent on the ability of the system to analyse, represent and communicate subjects of documents and readers' queries. In other words, the success of an information system depends to some extent on how a system user interprets a subject heading used in the system. Research in this area of identifying a method of structuring a subject, which will facilitate easy and precise identification of the subject represented by the heading, has led to the conjecture that there may be an 'absolute syntax' of components of a subject.

### 2. Scope of present study

The work reported in this paper aims at:

a) Applying the principles enunciated in the general theory of classification developed in India to a set of

\* Based on a paper presented to the DRTC 'Seminar on Classification and Indexing in Social Sciences', Bangalore, Dec.5–9, 1977.

subjects in the social sciences area to arrive at structured subject headings to represent such subjects; (1) b) Ascertaining whether such structuring helps in obtaining a facet sequence parallel to the absolute syntax (2); and c) Identifying and suggesting modifications if any to the existing method of subject structuring.

### 3. Methodology

The methodology adopted in this study could be discussed under the following heads:

#### 3.1 Selection of documents/subjects for indexing

This was done on a random basis. The only two criteria adhered to were:

- Subjects for indexing should be in the social sciences area; and
- Each subject should have at least four component ideas (facets and/or speciators).

#### 3.2 Indexing procedure & preparation of subject strings

- In order to derive a subject heading to represent each of the subjects, the Postulate-based Permuted Subject Indexing (POPSI) procedure with the general theory of classification developed in India was used. (3)
- In order to identify the difficulties in comprehending the semantic connotation of the subject strings, four sets of subject strings were prepared as below:

*Survey 1:* Strings with subject-field terms and superordinate links of facet terms: subjects of such strings were general in nature and were drawn from different disciplines in the social sciences area (Appendix C).

*Survey 2:* Strings without subject-field terms and superordinate links: In some of the strings apparatus terms (such as of, for, with, in) were incorporated to find out if these helped in the correct interpretation of the strings. The subjects were general in nature and were drawn from different disciplines in the social sciences area (Appendix D).

*Survey 3:* Strings representing subjects going with the subject-field 'Education' were structured in the following way (Appendix E):

Environment, system, attribute, action, method ... etc. The subjects selected were of specialized nature (micro-documents) (e.g. journal articles) involving a number of facet and speciator<sup>1</sup> type ideas.

*Survey 4:* Strings representing the same subjects as mentioned under Survey 3 were structured in the following manner (Appendix F):

System, environment, attribute, action, method ... etc.

#### 3.3 Questionnaire

Questionnaires along with a covering letter of instructions, were circulated among different types of information users.

Users were requested to indicate in the form of a title-like phrase the subject represented by each of the structured subject headings. They were instructed not to use any extraneous words other than connectors such as prepositions, conjunctions, etc.

##### 3.3.1 Population of Respondents

The respondents to whom the questionnaires were circulated were as below:

a) For surveys 1 and 2, the respondents were students and teachers with different subject backgrounds.  
 b) For surveys 3 and 4 the respondents were research scholars/faculty members in education.

### 3.3.2 Organisation of questionnaire

The questionnaire was in three parts:

a) Part 1 was the covering letter (Appendix A) describing the purpose of the survey and instructing the respondents in answering the questionnaire. The respondents were instructed to indicate in the form of a title-like phrase the subject which, according to them, the string represented. A couple of examples were provided by way of illustration.  
 b) Part 2 consisted of a list of subject strings.  
 c) Part 3 solicited the bio-data of the respondents (Appendix B).

## 4. Analysis

### 4.1 Survey 1

The responses for string sets included in surveys 1 and 2 were analysed and the data tabulated. For each of the strings, the responses were grouped as correct returns/wrong returns/near-correct returns. Table 1 gives the tabulated data of the responses to string set in survey 1.

Table 1. (With Subject-field & Superordinate Links)

String Number	Percentage of Correct Returns	Percentage of Wrong Returns	Percentage of 'Near-Correct Response'
1	10	90	..
2	30	70	..
3	50	50	..
4	78	..	22
5	60	30	10
6	80	20	..
7	70	10	20
8	44	56	..
9	67	33	..
10	33	67	..

The wrong and 'near-correct' returns were checked to identify the reasons for misinterpretation. Nearly 62% of the wrong and 'near-correct' returns were due to the presence of subject-field terms (43%) and superordinate links (19%).

For example:

*String (1)*

History. OPEC. International relations. Economic assistance. Developing countries.

*Responses:*

- 1) History of the economic assistance for developing countries by OPEC in international relations.
- 2) OPEC economic assistance to developing countries and international relations.

*String (6):*

Sociology Tribes – Soligas. Marriage custom.

*Response:*

Marriage customs of tribes – Soligas. is social.

In string (1), the basic facet term 'History' and the matter property facet term 'International relations' are superfluous and caused considerable misinterpretation.

In string (6), the basic facet term 'Sociology' gave rise to misinterpretation.

### 4.2 Survey 2

The response to string set in survey 2 is tabulated in table 2.

Table 2 (Without Subject field and Superordinate Links)

String Number	Percentage of Correct Responses	Percentage of Wrong Responses	Percentage of Near-Correct Responses
1	75	..	25
2	69	15	16
3	91	9	..
4	8	23	69
5	100	..	..
6	69	31	..
7	92	8	..
8	46	54	..
9	8	17	75
10	92	8	..

The significant increase in the percentage of correct returns is obvious. This can be attributed to the non-inclusion of subject-field terms and super-ordinate links in the subject string and the inclusion of apparatus terms in certain cases. In strings (4) and (9), where a very small percentage of the returns was correct, the responses and strings were checked to identify the reasons. It was found that the structuring of string (4) was not correctly done.

*String (4):*

Students-Higher Secondary School. Intelligence level. Effect (on) aspirations – Occupation. Delhi

*Responses:*

- 1) Effect of intelligence of students at higher secondary school level on aspirations and occupations in Delhi.
- 2) The intelligence level of higher secondary school students' effect on aspirations for occupation in Delhi.
- 3) The intelligence level of students in higher secondary school level and the effect of their aspirations and occupation in Delhi.

The string should have been structured as follows:

Students – Higher Secondary School – Delhi. Intelligence level. Effect (on) aspirations – Occupation.

There was a terminology problem in string (9).

*String (9):*

India, Women, Behaviour – Employment seeking *influenced by* Socio-economic change.

*Responses:*

The behaviour of employment-seeking women in India influenced by socio-economic change.

In this string, the term 'behaviour' was used to connote 'pattern'. This was not understood by most of the respondents. It must be recalled in this context that the respondents of this survey were drawn from students/teachers with different subject background. The other misinterpretations related mainly to Facet-Facet-Specialist relations. These were studied in detail in the surveys 3 and 4.

### 4.3 Survey 3

In this survey, strings were prepared to represent subjects of microdocuments in Education. Certain apparatus words, like prepositions, etc. were introduced in certain strings in order to ascertain if these made the strings more meaningful to the information users.

*String (1):*  
East Los Angeles College Students – Physically handicapped; Curriculum – Special guidance needs: Designing.

*String (2)*  
College freshmen – Mexican American; Curriculum – (in) Literacy composition: Designing.

The two strings represent similar subjects. An apparatus word (in) was introduced in the String (2). That the introduction of apparatus words make the string more intelligible is obvious by a comparison of the percentage of correct responses to these strings (27% and 73% respectively). Some responses to these strings are given below.

*Responses to string (1):*

- 1) Curriculum designing and special guidance needs of physically handicapped students of East Los Angeles College.
- 2) Designing a curriculum and special guidance to physically handicapped students in East Los Angeles College.

The responses to Survey 3 are tabulated in the table 3 given below.

*Table 3*

String Number	Percentage of Correct Returns	Percentage of Wrong Returns	Percentage of 'Near-correct' Returns
1	27	63	10
2	73	27	..
3	45	27	28
4	55	45	..
5	72	18	10
6	81	9	10
7	81	9	10
8	46	27	27
9	73	27	..
10	28	45	27
11	73	..	27
12	70	20	10
13	80	10	10
14	70	10	20
15	90	10	..

A total picture of the overall correct returns, wrong returns and near-correct returns is provided in Table 4 given below.

*Table 4*

Total Percentage of correct Responses	Total percentage of wrong Responses	Total percentage of 'Near-correct' Responses
64	21	15

It is evident that a fairly large percentage of responses was accurate. In order to ascertain the nature of misinterpretations and to seek suitable solutions, the wrong and 'near-correct' responses were further analysed.

#### *4.3.1 Analysis of "near-correct" and wrong responses*

Table 5 presents the tabulated data regarding the nature of misinterpretations of subject strings.

It is evident that a vast majority of 72% of the misinterpretations were of the nature of either omission of facets/speciator type idea or misinterpretation of facet-speciator relation.

*Table 5*

Nature of misinterpretation	Percentage
Omission of facet/speciator type idea	46
Misinterpretation of facet speciator relation	26
Misinterpretation of facet relation	0.5
Misinterpretation due to complexity of strings	14
Unpredictable responses <sup>3</sup>	13.5

The omission included facet as well as speciator type idea. Terms denoting speciator type idea accounted for nearly 80% of such omissions. Here are some sample responses indicating the nature of omission.

#### *String (1) (of Survey 3)*

East Los Angeles College, Students – Physically handicapped; Curriculum – Special guidance needs: Designing

##### *Response:*

Designing curriculum for physically handicapped students of East Los Angeles College.

#### *String (12):*

Pre-schol, Children – Female – Japanese; learning ability – (for) English language Morphemes – (as) Second language.

##### *Response:*

Learning ability of Japanese female preschool children.

#### *String (13):*

Secondary School, Students – Non-science; Curriculum – (in) Science and Technology: Model

##### *Response:*

Curriculum in science and technology for secondary school students.

Omission of a component idea and misinterpretation due to complexity in the subject string occurred mostly in strings with more than seven component terms. This must be viewed in the context of the findings of Surveys 1 and 2 where the average number of component terms in subject strings was four. In the response to these strings (in surveys 1 and 2), there were hardly any omissions of component ideas.

#### *4.3.2 Analysis of misinterpretation relating to facet-speciator relation*

*Table 6*

Nature of misinterpretation	Percentage
Involving [P] and speciator	13.3
Involving [M] and speciator	20
Involving [E] and speciator	13.3
Speciator interpreted as 'and'	40
Speciator read as [S]	13.4

In a large number of cases, it could be observed that the speciator sign was read as 'and'. In most of the interpretations involving facets and speciators, the term denoting a speciator idea was attached to the facet term following it, rather than to the preceding facet term as a qualifier, e.g. of strings where the speciator sign was interpreted as 'and'.

#### *String (1):*

East Los Angeles College, Students – Physically handicapped; Curriculum – Special guidance needs: Designing

##### *Response:*

- 1) Designing the curriculum and special guidance needs of physically handicapped.

2) Special guidance needs and design of curriculum for physically handicapped students of East Los Angeles College.

Misinterpretation of facet speciator relation:

*String (3):*

High School, Teacher – (of) science; Beliefs – (regarding) classroom behaviour: Analytical study

*Response:*

Analytical study of high school teachers regarding science beliefs in classroom behaviour.

*String (10):*

Teacher Training College, Student teachers – 'High Risk' type; academic performance: prediction – (using) standardised tests

*Response:*

Prediction of high risk type academic performance of student teachers of teacher training colleges.

*String (13):*

Secondary school, Students – Non-science: Curriculum – (in) science and technology: Model

*Response:*

Model of non-science curriculum in science and technology to students in secondary schools.

#### 4.4 Survey 4 and comparison of its broad findings with those of Survey 3

Table 7 below gives the general comparative idea of the responses to the subject strings of surveys 3 and 4.

Table 7

String number	Percentage of correct returns in	
	Syntactic Structure 1 (Survey 3)	Syntactic Structure 2 (Survey 4)
1	27	100
2	73	100
3	45	90
4	55	60
5	72	100
6	81	100
7	81	90
8*	46	..
9	73	100
10	28	60
11	73	100
12	70	100
13	80	75
14	70	75
15	90	90

\* Wrong structuring and thus data immaterial.

It can be noted that there is a statistically significant difference in the responses to the two syntactic structures. There is a high percentage of correct responses to sequence 2, although the two variables in the survey viz. subjects and respondents, were controlled. The exact reasons for such a difference need to be further investigated.

## 5. Component ideas in Education

A perusal of the subjects indexed for this experiment in particular and the literature in Education in general suggests that the following types of ideas are associated with the basic subject Education.

1. 'Persons concerned with education' such as students, teachers, parents, administrators, etc.
2. 'Attributes' of 'Persons concerned with education' such as behaviour (in students' behaviour), ability in (ability of negro students), etc.
3. Ideas such as curriculum, syllabus, etc.
4. 'Action' type ideas, such as teaching, designing, assessment, etc.
5. Speciators to the above types of ideas such as 'negro' in 'negro students', 'mathematics' in 'teaching of mathematics', 'science' in 'science teacher', 'audio-visual method' in 'audio-visual method of teaching'.

### 5.1 Categorisation of component ideas

Working within the framework of the general theory of classification developed in India, one can easily categorise some of the types of ideas listed above as manifestations of the five fundamental categories. 'Action' type ideas and 'Attributes' can, for example, be easily deemed as manifestations of the fundamental categories 'Energy' and 'Matter Property' respectively. 'Persons concerned with education' can be deemed to be a manifestation of the fundamental category 'Personality'. It is, therefore, clear that the personality isolates in Education do not consist merely of 'educands', but also of other 'persons concerned with education'. When we come to the idea of curriculum it is difficult to deem it as a manifestation of either of the fundamental categories 'Energy' and 'Matter'. It involves no 'action' to be deemed as 'Energy' and is certainly not a 'Property' of 'persons concerned with education'. Therefore, applying the principle of residues, it can be deemed only as a manifestation of the fundamental category 'personality'.

### 5.2 Speciators to 'Personality'

The subjects selected for structuring indicated the occurrence of the following types of speciators to 'persons concerned with education'.

1. The characteristics of the 'Persons concerned with education'. These are mainly of three types:

- a) Sociological characteristics, such as race, caste, nationality, etc.  
Eg: Negro students, Mexican American teachers, etc.
- b) Physiological characteristics, such as, Age, Sex, etc.  
Eg: Male students, Physically handicapped students etc.
- c) Psychological characteristics.  
Eg: Emotionally disturbed students, Mentally handicapped students, etc.

2. Subject of learning/teaching of the 'Persons concerned with education'.  
Eg: Science teacher, Students of library science, etc.

3. Level of education  
Eg: College student, Secondary school teacher, etc.

#### 5.2.1 Sequence of Speciators to Personality; Sociological, psychological and physiological characteristics

It might be useful to derive a principle helpful in sequencing these speciators to obtain a parallel helpful sequence of subjects in a surrogate file/library shelves. The principle will also be useful in sequencing these speciators when all of them or more than one of them are present with respect to a personality isolate.

Eg: Female Negro students.

It has been found helpful to have the most intrinsic characteristic of the Personality facet as the first speciator and the second most intrinsic characteristic as the second speciator, and so on. Hence the order of the speciators will be,

Physiological characteristics,

Psychological characteristics

Sociological characteristics.

Eg: The concept 'Physically handicapped male Negro students' will be structured as: Students — Male — Physically handicapped — Negro.

This method of structuring provides not only consistency in indexing, but can also be used in mechanical generation of subject index headings for use in an alphabetical index. The experiment regarding the interpretation of subject strings (reported in this paper) has pointed out the difficulties faced by information users in interpreting subject strings. One of the major problems faced by information users is in interpreting inverted headings incorporating speciators. While inverted headings have several advantages in providing a systematic arrangement, it might be useful to avoid inverted headings in the short display of subject index entries. A mechanical method of deriving such direct headings for the short display might, therefore, be useful in index generation. The above structuring also provides for such mechanical derivation of a meaningful direct heading. For example, the above structure can simply be read backwards as 'Negro physically handicapped male students' to provide a meaningful direct heading.

#### 5.2.2 *Subject of learning/teaching as a Speciator*

Subject of learning/teaching is an important characteristic of 'Persons concerned with education'. It determines, more than physiological, psychological and sociological characteristics, the curriculum, the method of teaching, etc. In other words, in the context of education, the subject of learning/teaching is more intrinsic to 'persons concerned with education' than the physiological, psychological and sociological characteristics. Hence it might be useful to consider the structuring of a heading with 'subject as a speciator' occurring before the speciators mentioned above..

For example:

The concept 'Physically handicapped male Negro science students' will be structured as:

Students — Science — Male — Physically handicapped — Negro

and can be read backwards as: 'Negro physically handicapped male science students' for the purpose of short display.

#### 6.2.3 *Level (Environment)*

The 'level' of the 'persons concerned with education' is the most important characteristic. In the subjects structured for the survey reported in this paper, there was hardly a paper in which the level did not appear as a component idea. The 'level' provides the environment in which a student community will learn and an educational programme is implemented. It determines more than anything else the nature of students/teacher suitable for the environment, the curriculum, teaching methods, facilities required, etc. Thus the environment determines even the kind of 'persons concerned with educa-

tion'. It is therefore useful to deem 'level of education as a manifestation of the fundamental category 'Personality' of a level occurring earlier to 'persons concerned with education' and 'curriculum'.

#### 5.3 *'Speciators to Energy'*

Speciators to energy are mainly of three types

1. 'Subject' (physics, library science, etc.)
2. Method of Action (audio-visual method, computer-aided method, etc.)
3. Agent of Action (male teachers, negro teachers, etc.)

Subject determines, to a certain extent, the 'Method of action' 'Agent of action' is the least important in the presence of the other two types of speciators. Hence the suggested sequence is —

[E] — 'Subject' — 'Method' — 'Agent'

#### 6. Conclusions and suggestions for consideration

These suggestions are to be viewed in the context of an alphabetical index to a classified catalogue/bibliography or an index to the classified part of an indexing/abstracting periodical or for the purpose of an alphabetic-specific subject catalogue.

1. Introduction of subject field terms and superordinate links in a subject string where they are superfluous leads to confusion among information system users and therefore possible misinterpretation of the subject strings. It is, therefore, suggested that the subject field terms and superordinate links be omitted in structuring a subject heading except when these are necessary for clarity (e.g. as a context indicator).

2. The use of appropriate terminology selected on the basis of the nature of the information system users will improve the performance of the system.

3. The responses to strings 1 and 2 in survey 3 (table 3) clearly indicate that the use of apparatus terms in subject strings will make the strings more meaningful and intelligible to system users. It is suggested that apparatus terms be used in subject strings.

4. The survey 3 clearly indicates that a large number of system users find it difficult to understand the relation between a facet term and its speciator term. It is therefore suggested that, as far as possible, direct headings be used in the cases involving speciators (in deriving subject headings for an index to a classified catalogue). However, to derive subject headings for use in a dictionary catalogue or for deriving feature headings to present ideas in a text, use of inverted headings with speciator sign may be continued. Direct headings may be used in the short display to avoid confusion to the users. However, the use of speciator sign and the inverted heading may be continued in the full representation to facilitate generation of entries under the facet term as well as speciator term.

5. Subject strings with eight or more component terms present difficulties in interpretation. It is suggested that a study of the level of specificity expected by information system users be made. It may not be necessary to coextensively represent the subject of every document (as users approach is likely to be broader). The suggestion of A. Neelameghan and G. Bhattacharyya to split micro subjects for the purpose of subject heading structuring may also be considered in this context.



4. Students – Higher secondary school. Intelligence level. Effect (on) aspirations – occupation: Delhi.
5. Bihar Tribes. Mothers – (with) new-born babies attitudes (to) family planning.
6. Earth-quake. Control. Experiments. Colorado (USA).
7. Education: Higher. Physics. Teaching techniques – audio-visual.
8. Industry. Labour disputes. Tribunal Functioning (influenced by) Supreme Court.
9. India. Women. Behaviour – employment-seeking (influenced by) socio-economic change.
10. Students – Elementary school. Perception – (of) movement influenced by Teachers. Sex.

#### Appendix E (Survey 3)

1. East Los Angeles College, Students – Physically handicapped; curriculum – Special guidance needs: Designing.
2. College freshmen – Mexican American; Curriculum – (in) Literary composition: Designing.
3. High School, Teacher – (of) Science; Beliefs – (regarding) Classroom behaviour: Analytical study.
4. High School – (with) Predominantly white students, students – Negro: Teaching – (of) Skill building, Effects (of) Educational survival.
5. Pre-school, Children; Auditory discrimination skills (influenced by) Training – (in) Structured classical Music.
6. Elementary school, Students – Mentally handicapped: Training – (in) Special schools; Effectiveness (compared with) Training – (in) Mainstream schools.
7. Secondary School, Students; Attitudes – (to) Science: Improvement – (using) student-based science curriculum.
8. College, Students, Comprehension ability – English prose (influenced by) Redundancy (in) Text: Reduction.
9. Elementary School, Students; Academic achievement: improvement – (using) WIST Reading Readiness program: Administered – (at) Preschool stage.
10. Teacher Training College, Student Teachers – 'High risk' Type; Academic performance: Prediction – (using) Standardized Tests: Results: Evaluation.
11. Elementary School, Students: Teaching – Economics – (using) Mini-society model – Effectiveness (compared with) Teaching – (using) Token economy model.
12. Pre-school, Children – Female – Japanese; Learning ability – (for English language Morphemes – (as) second language.
13. Secondary School, Students – Non-science; Curriculum – (in) Science and Technology; Model.
14. University of Maryland Students; Behaviour – (with regard to) selecting courses – offered elsewhere.
15. College, Students – Undergraduate – Male – Negro; Academic performance; Prediction – (using) 'Motivation expectancy theory'.

#### Appendix F (Survey 4)

1. Students – Physically handicapped – East Los Angeles College; Curriculum – Special guidance needs: Designing.
2. Freshmen – Mexican American – College; Curriculum – Literary composition: Designing.
3. Teacher – High School – Science; Beliefs – (regarding) Classroom behaviour: Analytical study.
4. Students – Black (Negros) – (in) High School – (with) predominantly white students: Teaching – Skill building; Effects – (on) Educational survival.
5. Children – Preschool stage: Auditory discrimination skills (influenced by) Training – (in) structured classical music.
6. Students – Mentally handicapped – Elementary school: Training – (in) Special schools; Effectiveness (compared with) Training – (in) Mainstream schools.
7. Students – Secondary school; Attitudes – (towards) science. Improvement – (using) Student based science curriculum.
8. Students – College; Comprehension ability – (of) English Prose (influenced by) Redundancy – (in) Text: Reduction.
9. Students – Elementary school; Academic achievement: Improvement – (using) WIST Reading Readiness program: Administered – (at) Preschool stage.
10. Student Teachers – High risk type, Academic performance; Prediction – (using) Standardized tests; Results: Evaluation.
11. Students – Elementary School: Teaching – Economics – (using) 'Mini-society model'; Effectiveness (compared with) Teaching – (using) 'Token Economy Model'.
12. Children – Female – Japanese – Preschool; learning ability – (for) English language morphemes – (as) Second language: Analysis.
13. Students – Secondary School – Non-science; Curriculum – Science and Technology Model.
14. Students – University of Maryland; Behaviour – (with regard to) Selecting courses – offered elsewhere.
15. Students – Male – Negro – Undergraduate; Academic performance: Prediction (using) Motivation Expectancy Theory.

#### Notes:

- 1 A speciator is an idea used or intended to be used to qualify another idea by which the conceptual wholeness of the idea qualified is not disturbed, e.g. 'steel' in 'steel chair'.
- 2 By 'Near-correct responses' are meant those responses which omitted certain ideas but were otherwise correct in the interpretation of the semantic connotation of the string.
- 3 Responses that could not be analysed.