J.C. Sager, H.L. Somers, J. McNaught
University of Manchester
Institute of Science and Technology, England

Thesaurus Integration in the Social Sciences Part I: Comparison of Thesauri

Sager, J.C., Somers, H.L., McNaught, J.: Thesaurus integration in the social sciences. Pt. I: Comparison of thesauri
In: Int. Classif. 8 (1981) No. 2, p. 133–138, 45 refs.

This part I of a series of three articles directed towards the creation of an integrated multilingual thesaurus for the social sciences establishes criteria relevant for the comparison of thesauri from the point of view of their integration and examines a number of thesauri in the social sciences with this goal in mind. The comparison considers (1) the knowledge structure (coverage, general descriptors, classification, hierarchies), (2) the linguistic structure (terms and descriptors, word categories, choice of descriptors and their form, relationships), and (3) formal features (typography, address codes, symbols, ordering, spelling variants).
(ace. Authors)

1. Introduction

This series of three articles presents the results of a study entitled "Guidelines for the Establishment of Comparison and Compatibility Matrices between Thesauri in the Social Sciences", carried out by the authors at the Centre for Computational Linguistics of the University of Manchester Institute of Science and Technology for the Division for the International Development of Social Sciences at Unesco.[1]

The study is based on the assumption that an integrated thesaurus for the social sciences is desirable and feasible and that the first step towards the creation of such a thesaurus should be the conflation of selected data from existing thesauri into a descriptor bank, as proposed by the Consultative Meeting on the Establishment of an Integrated Thesaurus in the Social Sciences (39). The guidelines evolved in this study are, therefore, in the first instance intended for the successful implementation of such a descriptor bank, but also, by implication, for the data to be entered and so for the various thesauri to be integrated. Since there are at present certain important linguistic, formal and technical differences between existing thesauri, their parallel use with an integrated thesaurus or the process of subsuming them into an integrated thesaurus can only be effectively achieved if, in the process and over a period of time, these thesauri converge in their linguistic, formal and technical characteristics.

To this end the study presented a synthesis of existing guidelines and standards relevant to the creation and management of multilingual thesauri (34).

A comparison of a small number of thesauri was carried out to determine typical differences between them. These are exemplified in Part I and have given rise to a checklist for the comparison of thesauri which is recommended for use in determining the suitability from a formal point of view of existing thesauri for incorporation in the integrated thesaurus (Part 2).

Subsequently the problems associated with creating an integrated thesaurus were examined from a linguistic, formal and technical point of view and the functions of a descriptor bank analysed in order to present practical and realistic guidelines for the comparison (matrices) to be carried out by the descriptor bank. These guidelines are discussed in Part 3.

This study is one of a series of investigations, some still to be carried out, for the improvement of information and documentation services in the social sciences. It lays no claim to originality but attempts to provide a realistic analysis of the obstacles which have to be overcome and to offer some suggestions towards the aim of achieving greater understanding and easier reference in the social sciences.

2. Selection Criteria[2]

The following selection criteria were established for the thesauri to be analysed:

- The thesauri should be fully implemented and in use;
- The thesauri should be multilingual and collectively cover Romance and Germanic languages in view of the different word-formation patterns;
- The thesauri should collectively present both detailed and more general aspects of the social sciences;
- The thesauri should be available in machine-readable form;
- The thesauri should generally conform to international standards and recommendations on the compatibility of content and form.

The choice in each individual case was further guided by the following considerations:

Unesco Thesaurus: its scope, its detailed analysis of relationships, its link to other subject fields, its exemplary compilation and documentation.

Macrothesaurus: its successful transformation from a monolingual to a multilingual thesaurus, its wide use, its use as a reference tool in terminology.

Eudised Multilingual Thesaurus: its structure, its duplication and overlap of languages and content with the *Macrothesaurus,* and the *Unesco Thesaurus.*

Existing guidelines for thesaurus compilation have been evolved in considerable detail for monolingual thesauri, but less work has been done on establishing guidelines for multilingual thesauri. Most examples in the guidelines are taken from science and technology which might indicate that all the implications of culturally conditioned descriptors may not have been considered. Deviations from existing guidelines are partly accounted for by divergent national standards and partly by the particular requirements of the social sciences. All the documents studied, however, made recommendations relating only to formal and linguistic aspects of thesaurus construction. It was thus possible to establish compatibility on two levels only and the resulting measure of conformity to international standards or accepted guidelines does not provide any indication about the compatibility of the contents of the thesauri nor of the technical, computational conditions for successfully merging thesauri.

3. Differences and divergences among existing thesauri

For the purpose of their inclusion in an integrated thesaurus detailed features have to be compared at various levels:

- the knowledge structure
- the linguistic structure
- formal features
- the computational representation

In the comparison of the knowledge structure we are concerned with the subject classification, whether expressed in numerical notation, class marks, non-descriptors, etc. which provides the broad system of ordering of a thesaurus. The linguistic structure is concerned with the expression form of descriptors, their choice, identification and the number and types of relationships declared in the thesaurus, as well as with the selection and declaration of non-descriptors.

Details of the representation, such as classified thesaurus, hierarchical display, alphabetical display and various indices together with the information they contain are relevant if they are the only available output formats. They are irrelevant if the thesaurus is available in machine-readable form and other combinations of data can be generated automatically. The computational representation is of the utmost importance as only those thesauri which have compatible computational specifications can be integrated economically.

Abbreviations

The thesauri examined and used for exemplification are abbreviated as follows:

C	Cultural Development Thesaurus (43)
E	Eudised Multilingual Thesaurus (42)
M	Macrothesaurus for Information Processing in the Fields of Economic and Social Development (44)
P	Population Multilingual Thesaurus (45)
S	Thesaurus for Information Processing in Sociology (41)
U	Unesco Thesaurus (2)

Other abbreviations

USE	to refer from a non-descriptor to a descriptor
UF	to refer from a descriptor to a non-descriptor
SN	scope note
BT	broader term (descriptor)
NT	narrower term (descriptor)
RT	related term (descriptor)
TT	top term of a hierarchical array

Conventions

- Examples of thesaurus descriptors are given in capital letters
- References throughout are to numbered sections of this article and to the bibliography.

3.1 Knowledge structure

3.1.1 Coverage

Differences in coverage have been described in detail in the study carried out by Meyriat (32).

General descriptors

There are differences in the perception of the need for a general section:

U has a large general section;
S has separate sections on research and methodology;
M has a single chapter on research methodology.

Other thesauri include general terms in other parts of the lists and under the headings Documentation/Information.

3.1.3 Classification

Differences in broad classification have a fundamental effect on the lower levels of associations among descriptors, e.g. SOCIAL SYSTEM as represented in a number of thesauri:

E	25000	Sociology of Education
	25100	Social Systems
	25120	SOCIAL STRUCTURE
M	05.	Culture. Society
	05.03	Society
	05.03.01	SOCIETY
		SOCIAL SYSTEM
	05.03.04	SOCIAL STRUCTURE
P	05.	Social Organisation
	05.01	Society. Social System
	05.01.02	SOCIAL SYSTEM
	05.03	Social Differentiation
	05.03.01	SOCIAL DIFFERENTIATION
		SOCIAL STRATIFICATION
		SOCIAL STRUCTURE
C	13000	Social Structure
	13400	Social System. Society
	13410	SOCIAL STRUCTURE
S	14	Social Structure
	14100	SOCIAL STRUCTURE
		SOCIAL SYSTEM
	14200	Social Stratification
	14210	SOCIAL STRATIFICATION
U	R50/69	SOCIETY
	R52/64	SOCIAL STRUCTURE
		NT ROLE
		SOCIAL MOBILITY
		SOCIAL STRATIFICATION
	R53/57	SOCIAL STRATIFICATION

In other cases there is considerable coincidence of class marks at the intermediate level despite different starting points of classification, e.g.

Labour and Employment U	–	Labour relations M, U
Labour M, S	–	Labour management relations S
Living conditions C	–	Labour C
Teaching profession E	–	Working conditions E
Society P	–	Group P

TRADE UNION C, E, M, P, S, U

This example clearly demonstrates the hybrid nature of thesauri mixing subject classification and terminological relationships.

The analysis of classification systems of thesauri lies outside the scope of this study, but the ordering system affects the relationships among terms and has therefore to be adjusted in an integrated thesaurus. This, however, can only be done after agreement has been reached about the relationships among descriptors.

Most thesauri have hybrid structures, operating with classification schemes of non-descriptors and class marks at the higher levels and descriptors only at the fourth level of classification. The classification schemes are variously developed in the thesauri examined, e.g.

	No. of digits in classif.	Fields 1	Sub-Fields 2	Descriptor groups approx. 3	Ratios 1:2	2:3
C	5	14	78	350	5.6	4.5
E	5	21	95	220	4.5	2.3
M	6	19	128	500	6.7	3.9
P	6	18	89	220	4.9	2.5
S	5	11	62	194	5.6	3.1
U	7	7	24			

Only U is fully structured hierarchically; because it is a general thesaurus, comparison with specific thesauri is inappropriate in this area.

3.1.4 Hierarchies

Thesauri are developed with different depths of hierarchies. This lack of detailed structure may be considered an advantage for integration since it is likely to be easier to achieve agreement on as yet undeclared relationships than on those already declared and therefore fully incorporated into information systems. A depth of 7 or 8 levels of structure is encountered in U and M; C, E and S go to three or four levels at most. There are therefore more RT than BT or NT; e.g. in E we have

25000 Sociology of Education
25500 Labour Market. Employment
25520 LABOUR MARKET
LABOUR SHORTAGE
MANPOWER
MANPOWER NEED
} All are RT to each other; none have a BT or a NT.

In many cases descriptors have no declared BT or NT relationships, but occur simply next to each other within the same group and at the same level; e.g.

14. Demography. Population
14.02. Age groups
14.02.05 CELIBACY
DIVORCE
HUSBAND
MARITAL STATUS
MARRIAGE
MARRIED PERSONS
NUPTIALITY
POLYANDRY
POLYGAMY
WIFE
RT
– NT to MARRIED PERSONS
– NT to MARRIED PERSONS

3.1.5 Class marks

The class mark may be no more than a convenient label to group descriptors so that terms occur both in the heading of the class mark and as descriptors; e.g. in D

132000 Community. Group. Organisation

BT/NT	No. of other NT outside array	RT	No. of other RT outside array
13210 ASSOCIATION	4		1
CLUB	6		0
COMMUNITY	0		8
GROUP	7		2
LOCAL COMMUNITY	2		3
ORGANISATION	6		0
SOCIAL MOVEMENT	3		0

Some class marks are quite inappropriate as in the example of 'Age groups' in M above, or, also in M, where nationalities are mixed with ethnic groups and adjectives for regions, e.g.

14.03 Ethnic groups
14.03.02 AFGHAN . . . AUSTRIAN . . . BRITISH . . .
CZECHOSLOVAK . . . EUROPEAN . . .
GERMAN . . . SCANDINAVIAN . . . SOVIET . . .

3.2 Linguistic structure

Differences have been established at all levels of linguistic description.

3.2.1 Terms and descriptors

There can be conflict between terms which are descriptors in some but non-descriptors in other thesauri, e.g.

FARMER	C M P S U	
PEASANT	M	
	P	use FARMER
	S	use FARMER
PEASANT CLASS	P	use PEASANTRY

The presence of both descriptors in M is explained by the fact that FARMER is listed under 'occupation' and PEASANT under 'class'; the 'class' term is represented by PEASANTRY in C, P and S.

Special difficulties arise when a non-descriptor in one system is a NT in another; e.g.

AGRARIAN REFORM	M	NT	LAND REFORM
	P	use	LAND REFORM
LAND REFORM	M	BT	AGRARIAN REFORM
	S	use	AGRARIAN REFORM
	U		

Compound terms are variously referred to uniterms or to other compounds; reference from uniterms to compound terms is rare; e.g.

POLITICAL ACTIVITY	P	use POLITICS
POLITICAL ATTITUDE	S U; E	use POLITICAL BEHAVIOUR
POLITICAL BEHAVIOUR	E M S U	
POLITICAL ECONOMY	M	use ECONOMICS
POLITICAL DEVELOPMENT	M U; S	use POLITICAL MODERNISATION
POLITICAL STRUCTURE	P	use POLITICAL SYSTEM
POLITICAL SYSTEM	C M P S U	
POLITICIANS	U; S	use POLITICAL MAN
POLITICS	C E M P S U	

The conversion from an adjectival phrase to a prepositional phrase seems unjustified and not even explicable by adjustment to the foreign language; e.g.

EDUCATIONAL REFORM/
M P S U ; E use REFORM OF EDUCATION
BILDUNGSREFORM/
REFORME DE L'ENSEIGNEMENT/
REFORMA DE LA EDUCACION

In some cases descriptors seem to have been selected by the need to adjust to foreign language equivalents – though this is not consistent – and can lead to overlap of descriptors; e.g.

DEATH RATE	P S U; E	use MORTALITY/ MORTALITE/ STERBLICHKEIT
MORTALITY RATE	P	use DEATH RATE/ TAUX DE MORTALITE/ TASA DE MORTALIDAD
MORTALITY/ MORTALITE/ MORTALIDAD/ STERBLICHKEIT	E M P S U	
DEATH/MORT/ MUERTE/TOD	E M P S U	
DEATHS	P	use MORTALITY

3.2.2 Word categories

There is a risk of confusion between word categories; e.g. INTELLECTUAL is identified as a noun in E only through the parallel German descriptor INTELLEKTUELLER.

Adjectives as descriptors are unusual; where they occur they can create unnecessary divergencies; e.g.

MARRIED is BT to HUSBAND and WIFE in E, where other thesauri have MARRIED PERSON as the broader term.

UNMARRIED does not occur in C, E and M but is listed as a descriptor noun in P, S and U, as BT to such descriptors as

UNMARRIED MOTHER	P U
BACHELOR	P
DIVORCED PERSON	P
MARRIAGEABLE	P

P has a great number of adjectives, though none derived from countries; M does have such adjectives classified under ETHNIC GROUPS, but few others.

3.2.3 Choice of descriptors and their form

There is conflict in the choice of foreign language equivalents; e.g. in P

English	SPOUSE	MATE	MATE SELECTION
French	CONJOINT	PARTENAIRE	CHOIX DU CONJOINT
Spanish	CONYUGE	COMPANERO	ELECCION DEL CONYUGE

There is inconsistency between languages in proper names; e.g.

English:	GERMAN DEMOCRATIC REPUBLIC	U
	GERMANY DR	C E M P
French:	ALLEMAGNE RD	C E
	RD ALLEMANDE	M P
German:	DEUTSCHLAND DDR	E
Spanish:	RD ALEMANA	M
	RD ALEMANIA	
	GERMAN FEDERAL REPUBLIC	U
	GERMANY FR	C E M P
	ALLEMAGNE RF	C M P
	ALLEMAGNE FR	E
	DEUTSCHLAND BRD	E
	ALEMANIA RF	M P

Antonyms are unevenly represented; e.g.

Plists	DE FACTO POPULATION	– DE JURE POPULATION
	OVERPOPULATION	– UNDERPOPULATION
	OPEN POPULATION	– CLOSED POPULATION
	MARRIAGEABLE POPULATION	– NON MARRIAGEABLE POPULATION

but no antonyms to
INACTIVE POPULATION
DECREASING POPULATION
POPULATION GROWTH
NON AGRICULTURAL POPULATION

There is wide divergence in the use of singular and plural forms as a result of conflicting national standards and considerations of the use of descriptors for abstracting. This question is discussed in detail in part. 3.

3.2.4 Relationships

There is a great unevenness in the number of declared NT and RT between thesauri;

	No. of	NT	RT			NT	RT
SOCIAL CLASS	C	5	5	POPULATION	C	1	4
	E	6	3		E	2	9
	M	5	3		M	4	15
	P	6	3		P	36	16
	S	7	1		S	4	7
	U	3	3		U	6	21

The overlap of RT and NT for SOCIAL CLASS is as follows:

SOCIAL CLASS	NT	BOURGEOISIE	C E M P S –
		LOWER CLASS	– E – – S –
		MIDDLE CLASS	C E M P S U
		PEASANTRY	C – – P – –
		PROLETARIAT	– – – P S –
		RULING CLASS	C E M P S –
		UPPER CLASS	– E M – S U
		WORKING CLASS	C E M P S U
	RT	CLASS CONFLICT	– E – P – U
		CLASS CONSCIOUSNESS	C – M – – U
		CLASS STRUGGLE	C – M – – –
		CLASS DIFFERENTIATION	– – – – – U
		SOCIAL INEQUALITY	C E – – – –
		SOCIAL ORIGIN	– – – P – –
		SOCIAL STRATIFICATION	C – M P – –
		SOCIAL STRATUM	– – – – S –
		SOCIAL STRUCTURE	C E – – – –

3.3 Formal features

Thesauri consist of several different listings:

- The classified thesaurus usually consists of two lists,
  a summary or broad structure, also called 'subject category fields' as in C E M P S U;
  a detailed list, ordered alphanumerically as in U or numerically as in C E M P S
- Permuted index or KWOC index as in C E M P S U
- Hierarchical display as in M P U
- Alphabetical display of descriptors as in M U

The amoung of information given in each of the above lists varies considerably; this is irrelevant when the information can easily be rearranged automatically. For everyday parallel use of thesauri it is however convenient if identical patterns exist for similar uses; e.g.

U is the most complex and detailed thesaurus. The alphabetical display lists term number, SN, BT, NT, TT, RT as well as USE/UF. M follows this pattern except for TT, but the classified display only gives term group numbers and USE references, thus making the alphabetical display the main part of the thesaurus. C, E, P and S, which are only accessible via the alphabetical index, do not have individual term numbers nor TT. The alphabetical indexes only provide reference to the group numbers of the descriptors and also USE references. The top term (TT) is only identified in U.

Multilingual thesauri are inconsistent in their language representation in any one language version. E and P give descriptors in all languages in the systematic display. E and P have a monolingual index in each language version. M is multilingual only in the descriptors of the alphabetical display (the main part of this thesaurus). C and S are fully bilingual in all parts and therefore have no other language versions. C, however, gives class marks only in English.

3.3.1 Typography

E uses block capitals only;
C uses block capitals throughout except for scope notes which are in capitals and lower case;
M uses block capitals but differentiates entry terms in bold, scope note and symbols in italics;
P uses block capitals, in bold for entry terms, varying type sizes for class marks, and italics for SN and symbols;
S and U use upper and lower case, bold and italics, but again in different functions.

There are many other variations; e.g. multilingual descriptors in C, M, P and S are separated by a slash, and in E by a dash.

3.3.2 Address codes

These vary from numeric with or without decimal points to alphanumeric. The recommendation that address codes should precede the descriptor in the systematic display is observed by all thesauri. They are variously placed in the other parts. Most thesauri do not have a unique address code for each descriptor but group a great number under one code number; only U provides a unique reference for each descriptor. Contrary to expectation address codes are provided for non-descriptors in C, E, P and S.

3.3.3 Symbols

All thesauri examined use the English symbols SN, USE, UF, BT, NT, RT for all language versions.

3.3.4 Ordering

Alphabetic ordering varies considerably; e.g. in compound words the following methods were observed:

Left before right in
S COLLECTIVE ATTITUDE
ATTITUDE CHANGE

Right before left in
C AIR POLLUTION
OPEN-AIR CENTRE

E ATTITUDE CHANGE
BODY ATTITUDE

P AREA STUDIES
ATTRACTION AREA

Mixed in M
AGRICULTURAL INCOME
CASH INCOME
INCOME
INCOME DISTRIBUTION
INCOME TAX
LOW INCOME
NATIONAL INCOME

3.3.5 Spelling variants

The spelling follows British English usage, and uses relatively few hyphens. A number of variations were encountered; e.g.

ROMANIA – RUMANIA
EROTISM – EROTICISM

Notes:

1 The authors wish to express their appreciation for the assistance and advice offered by Messrs. Derek Austin, J. Litoukhin and Jean Viet.
2 No criticism is implied of the thesauri analysed and used for exemplification, nor indeed can it be intended as these thesauri were compiled for different purposes for a variety of agencies with their own conventions and requirements.

References (of the three parts)

(1) Aitchison, J.: Thesaurofacet. Whetstone (1969): English Electric Co.
(2) Aitchison, J.: Unesco Thesaurus: a structured list of descriptors for indexing and retrieving literature in the field of Education, Science, Social Science, Culture and Communication. (2 vols). Paris (1977): Unesco.
(3) Aitchison, J., Gilchrist, A.: Thesaurus construction: a practical manual. London (1972): Aslib.
(4) Austin, D.: Precis: a manual of concept analysis and subject indexing. London (1974): Council of the British National Bibliography.
(5) Austin, D., Waters, J.: UNISIST Guidelines for the Establishment and Development of Multilingual Thesauri, (Revised text), PGI/80/WS/12. Paris (1980): Unesco; also ISO/TC 46/SC5 33E.
(6) Beling, G., Schuck, H.-J., Wersig, G.: Procedural guide for the translation of foreign language thesauri into German. Wachtberg-Werthoven (1973): Forschungsinstitut für Funk und Mathematik / Deutsche Gesellschaft für Dokumentation.
(7) Beling, G., Wersig, G.: The implementation format MATER, in: International Cooperation in Terminology, Infoterm Series Volume 3, Miinchen (1975): Verlag Dokumentation.
(8) British Standards Institution: Transliteration of Cyrillic and Greek characters. BS 2979 (1979).
(9) British Standards Institution: Transliteration of Arabic characters. BS 4280 (1976).
(10) British Standards Institution: The romanization of Japanese. BS 4812 (1976).
(11) British Standards Institution: Guidelines for the Establishment and Development of monolingual thesauri. BS 5723 (1979).
(12) Canisius, P.: Automatic partial translation in a multilingual information system, in: Commission of the European Communities (C.E.C.): Overcoming the language barrier (Third European Congress on Information Systems and Networks, Luxemburg, May 1977), München (1977): Verlag Dokumentation, Vol. I, 259–269.
(13) Commission of the European Communities (C.E.C.): Astute: automated system for thesaurus updating, testing and editing. Luxembourg (1976): C.E.C.
(14) Commission of the European Communities (C.E.C.): Definition of the essential characteristics of thesauri. Bruxelles (1976): Bureau Marcel van Dijk.
(15) Commission of the European Communities (C.E.C.): State of the art on multilingual activities in the field of scientific and technical information, Final report (2 vols). Bruxelles (1976): Bureau Marcel van Dijk.
(16) Commission of the European Communities (C.E.C.): Etude de faisabilité de la représentation graphique de thésaurus gérés par Astute, Rapport final. CETIL/213/80 (1980).
(17) Deutsches Institut für Normung – Richtlinien für die Erstellung und Weiterentwicklung von Thesauri. DIN 1463 (1976).
(18) Engineers Joint Council (E.J.C.): Rules for preparing and updating engineering thesauri. New York (1965): E.J.C.

(19) European Brewery Convention (E.B.C.): European Brewery Convention Thesaurus. (no place) (1976): E.B.C. Information and Documentation Group.
(20) Field, B.J.: INSPEC Thesaurus, 2nd. ed. London (1975): Institute of Electrical Engineers.
(21) Foskett, D.J.: A study of the role of categories in a thesaurus for educational documentation. Strasbourg (1972): Council of Europe.
(22) Iljon, H.: Le logiciel Astute: un outil informatique pour la création, la mise à jour, l'édition et l'impression de thésauri mono et multilingues. Documentaliste 14 (1977), 5/6, 25–33.
(23) International Organisation for Standardization (I.S.O): Lexicographical symbols particularly for use in classified defining vocabularies. ISO 1951 (1973).
(24) International Organisation for Standardization (I.S.O): Documentation: Formats for bibliographic information interchange on magnetic tape. ISO 2709 (1973).
(25) International Organisation for Standardization (I.S.O): Documentation: Guidelines for the establishment and development of monolingual thesauri. ISO 2788 (1974).
(26) International Organisation for Standardization (I.S.O): Magnetic Tape exchange format for terminological / lexicographical records (MATER). ISO/DP 6156.
(27) International Organisation for Standardization (I.S.O): Supplement to the ISONET Thesaurus. ISO/INFCO/WG2–87 (1980).
(28) International Organisation for Standardization (I.S.O): International system for the transliteration of Slavic Cyrillic characters. 2nd ed. ISO/R 9 (1969).
(29) International Organisation for Standardization (I.S.O): International system for the transliteration of Arabic characters. ISO/R 233 (1961).
(30) International Organisation for Standardization (I.S.O): Transliteration of Hebrew. ISO/R 259 (1962).
(31) International Organisation for Standardization (I.S.O): International system for the transliteration of Greek characters into Latin characters. ISO/R 843 (1968).
(32) Meyriat, J.: Social Science Documentary Languages: a comparative analysis. Paris (1980): International Committee for Social Science Information and Documentation, Unesco.
(33) Sager, J.C., McNaught, J.: Feasibility study of the Establishment of a Terminological Data Bank in the U.K. (Report prepared for the British Library Research and Development Department), CCL/UMIST Report No. 81/1. Manchester (1981): Centre for Computational Linguistics, University of Manchester Institute of Science and Technology.
(34) Somers, H.L.: Observations on standards and guidelines concerning thesaurus construction. In: Int. Classif. 8 (1981) No. 2, p. 69–74.
(35) Tell, B.V., Wersgren, K., Heinberg, W.: The use of ERIC tapes in Scandinavia: searching with thesaurus terms in natural language. DECS/Doc (72) 15. Strasbourg (1972): Council of Europe.
(36) Thompson, G.K.: Abstracting services in Education and the Social Sciences, in: EUDISED Technical Studies, Strasbourg (1971): Council of Europe.
(37) UNESCO: Meeting of experts on 'Interconcept', principles and strategies. SS/77/CONF. 601/9. Paris (1977): Division for the International Development of Social Sciences, Unesco.
(38) UNESCO: First session of the Ad Hoc Committee on Social Science Information. SS/77/CONF. 607/8. Paris (1977): Division for the International Development of Social Sciences, Unesco.
(39) UNESCO: Consultative meeting on the establishment of an integrated thesaurus of the Social Sciences. SS/CS/26/80/8. Paris (1980): Division for the International Development of Social Sciences, Unesco.
(40) Viet, J.: EUDISED: Report of the steering group. Strasbourg (1971): Council of Europe.
(41) Viet, J.: Thesaurus for Information Processing in Sociology, (Maison des Sciences de l'Homme Service d'Echange d'Informations Scientifiques Publications Série D, Méthodes et Techniques III). Paris (1971): Mouton.
(42) Viet, J.: EUDISED Multilingual Thesaurus for Information Processing in the field of Education of the Council of Europe. Paris (1974): Mouton.
(43) Viet, J.: Cultural development thesaurus. Strasbourg (1975): Council for Cultural Cooperation, Council of Europe.
(44) Viet, J.: Macrothesaurus for Information Processing in the field of Economic and Social Development. (New English Edition). Paris (1978): Organisation for Economic Cooperation and Development.
(45) Viet, J.: Population Multilingual Thesaurus, (English Edition). Paris (1979): CICRED/UNFPA.

Research and Development Reports from UMIST

Feasibility Study of the Establishment of a Terminological Databank in the U.K.

Linguistic databanks containing the vocabulary and phraseology of special languages are established services in leading industrial countries providing tools for technical writers, translators, information scientists, standardization agencies and publishers. This study summarizes these developments and reports on a market survey on the basis of which user types and their needs were determined. Certain desirable characteristics of a linguistic databank are identified and the conditions are established under which such a new information tool could be realized in the U.K. An analysis of the responses of the market survey is appended as well as an extensive bibliography. (CCL/UMIST 81/8)

Brit.Libr.R and D Report No.5642. £ 9.50 +

Selective Survey of Terminological Databanks in Western Europe, November 1980

The most recent comprehensive survey of linguistic/terminological databanks in Western Europe, with full descriptions of EURODICAUTOM, LEXIS, NORMATERM, TEAM, TERMDOK under the following headings: Users and uses, origin and size, organization, financial support, costing and costs, data! with tables and a full bibliography. (CCL/UMIST 81/9)

Brit.Libr.R and D Report No.5643. £ 7.50 +

Model Specification of a Linguistic Databank for the U.K.

A model of a multifunctional linguistic/terminological databank is developed based on substantial market research. Detailed specifications of user types and uses, organization, computational requirements, costs and data. Hardware and software requirements and user facilities such as search and output formats are proposed. (CCL/UMIST 81/10).

Brit.Libr.R and D Report No. 5644. £ 9.50 +

All three reports are in single spaced typescript (A4), ringbound and obtainable from: Centre for Computational Linguistics, University of Manchester Institute of Science and Technology (UMIST), P.O.Box 88, Manchester M60 1QD, England.