Jean Aitchison

Integration of Thesauri in the Social Sciences*

Aitchison, J.: Integration of thesauri in the social sciences.
In: Intern. Classif. 8 (1981) No. 2, p. 75–85, 18 refs.

The paper discusses the objectives, characteristics and compilation of the proposed Integrated Thesaurus of the Social Sciences. The Integrated Thesaurus is not intended to supplant well-established thesauri, but to act as a master reference tool, which could serve as an aid in searching across data bases or as a switching mechanism. The first step in compilation should be the merging of terms to form a Descriptor Bank. This can reveal both matches and inconsistencies, but is limited by its verbal and alphabetical approach. To compare the merged terms at the concept level analysis in sub-fields is proposed within a classification framework, preferably one with a faceted structure, such as the second edition of the Bliss Bibliographic Classification. The classification would serve as the master against which the terms from the merged information languages, arranged in a compatibility matrix, would be compared. The master classification would be modified during interaction with the matrix terms, and emerge as the Integrated Thesaurus in a format similar to that of the BSI ROOT Thesaurus, having a detailed classified display and full alphabetical section. The compatibility data could be displayed either in the classified or alphabetical section. An example is worked out in the sub-field of Unemployment using terms from 5 thesauri and the UDC. Finally consideration is given to multilingual implications, testing and updating. (Author)

0. Introduction

Universal classification systems, such as the Dewey Decimal Classification (DC) or the Library of Congress Classification (LC) have long been an essential feature of bibliographic classification, whereas the construction of an in-depth Thesaurus extended to all fields of knowledge has not yet been attempted. The thesaurus has always been predominantly oriented towards special subject fields or specific organisations and not amenable to containment or integration within a possibly restricting global system. However, the time may now be ripe for a bid to create not a universal, but at least a pan-disciplinary thesaurus covering the subject areas which make up the social sciences. The move towards integration is being given impetus by the increasing diversity of terminology in existing social science thesauri, which hinders effective communication and efficient information processing and retrieval, not least when searching across several online databases. Unesco has already taken the first steps towards the establishment of an Integrated Thesaurus of the Social Sciences and some initial investigations are underway following a Consultative Meeting in June 1980 organised in Paris by the Division for the International Development of the Social Sciences.

1. Objectives

From a cursory study of the Final report (1) of the Consultative Meeting, it might seem that the objective is a monolithic compilation which will eventually devour all existing information languages (ILs) apart from the highly specialised. On closer examination, however, it emerges that the Integrated Thesaurus is not intended to supplant well-established thesauri in the social sciences but will be a master reference document providing:

(a) A standard against which existing thesauri and classification systems may be compared and subsequently amended and improved.
(b) A means of indicating compatibility between existing ILs and to facilitate movement from one IL to another in database searching.
(c) An aid in translation of descriptors from one natural language to another.
(d) A switching tool or intermediate lexicon for the transfer of indexed data between systems using incompatible ILs.
(e) The tagging of discrete subject fields and sub-fields in the social sciences for use in characterising sources, centres and services providing information in these fields.
(f) A ready-made system to be adopted by any organisation not wishing to develop its own thesaurus.
(g) A source of terms and their relationships in the fringe areas of a specialised thesaurus covering other disciplines.

2. Characteristics of the Integrated Thesaurus

An IL which can perform the tasks required of the Integrated Thesaurus must encompass all social science fields, operate at a high level of specificity, comply with standards on thesaurus construction, provide a classified as well as an alphabetical display and be available in multilingual versions.

2.1. Subject scope

Although the Integrated Thesaurus is intended to cover all social sciences fields there has not yet been a ruling on the inclusion or exclusion of certain disciplines such as Education or Psychology or Religion. A decision will have to be made soon on what constitutes the social sciences for the purposes of the Integrated Thesaurus.

2.2. Specificity

The Integrated Thesaurus must be as specific as the most specific IL included in the system if it is to operate effectively as a switching mechanism for the transfer of indexed data and as a master in compatibility studies. Research carried out in the United Kingdom by Verina Horsnell (2, 3) on an Intermediate Lexicon established that there is loss of information when the switching language is broader than the ILs to be translated, since specificity in the source language translated into the broad intermediate language cannot be restored in the target language. Specificity is equally important to the Integrated Thesaurus in its role as the master in comparison studies between ILs. If the Integrated Thesaurus has a broad level cut-off point, the compatibility between terms from ILs below this level will not be so easy to discern. A ruling still has to be made on the problem of whether to exclude from integration highly

* Invited Paper presented at CONTA Conference, Bielefeld, May 24–27, 1981.

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

specialised thesauri and classification systems in narrow subject areas and those which have a purely national, regional or local bias. If these are excluded the overall specificity level of the Integrated Thesaurus will be reduced.

2.3. Conformity with standards

The Integrated Thesaurus should be compiled following closely standards on thesaurus construction. For monolingual thesauri International Standard 2788 (4) reviews the basic relationships and vocabulary control. It is now undergoing revision and a new edition is expected. British Standard 5723 (5) deals with relationships and vocabulary control, but also provides rules on retention and factoring of compound terms and recommends forms of layout and display. For multilingual thesauri a comprehensive proposal has now been issued (6) with a UNISIST imprint. The requirements of these standards are summarised in "Observations on standards and guidelines concerning thesaurus construction" (7). He recommends the use of international reference codes, given in the standards, as an alternative to monolingual codes. This implies that the codes BT/NT/RT would be replaced by ‹/› – for 'Broader term', 'Narrower term' and 'Related term', in the Integrated Thesaurus.

2.4. Method of term display

In accordance with standard requirements the Integrated Thesaurus should have both alphabetical and systematic sections. In order to 'tag subject fields and sub-fields' it must include a disciplinary classification structure as well as displays of hierarchies. The usefulness of the classificatory approach in the compilation of the Integrated Thesaurus is discussed fully in paragraph 5.

2.5. Multilingual characteristics

The final document should be a multilingual tool in French, German, Spanish and other languages, but it is likely that the first version will be in English. Multilingual problems are discussed more fully in paragraph 10.

2.6. Size

An estimate of the size of the Integrated Thesaurus is difficult to make. The number of terms (preferred and non-preferred ones but excluding identifiers) might total anything from 15–30,000. The size of the Integrated Thesaurus will depend in part on the overlap of terms in existing thesauri, which could be substantial. A glance at the contents of thesauri in Education reveals numerous sociological, cultural and psychological terms, whilst a look through a Political Science thesaurus shows many terms from Economics and Sociology. The size will also depend on whether highly specialised ILs are included.

3. Bibliography and checklist of Information Languages

An essential preliminary task in the establishment of the Integrated Thesaurus is the listing of existing ILs in the social sciences. A comprehensive bibliography on dictionaries, thesauri and special classification schemes is in course of compilation by Mrs. M. Krommer-Benz. This bibliography will be kept up-to-date and should include also advanced notice of ILs in preparation. Sager (8) in his *Guidelines* recommends, in addition, that a checklist be compiled for all thesauri in the social sciences which would not only supply comprehensive information on coverage, relationships and format but also give computational details of thesauri in machine-readable form.

4. Descriptor Bank

In the Final report (1) of the Consultative Committee it is proposed that the feasibility should be examined of the setting up of a computerized Descriptor Bank as an interim step in the construction of an Integrated Thesaurus. The Descriptor Bank would incorporate and merge the contents of existing social science thesauri already in machine-readable form and include scope notes and definitions, information on foreign language equivalents and term interrelationships. As well as providing a source of terms for the Integrated Thesaurus the Descriptor Bank would give, in its own right, a useful service, not least in indicating how great the overlap of terminology is in the social sciences. Appendix 1 contains a set of merged terms within the sub-field of Unemployment, giving some indication of the value of a Descriptor Bank. The five ILs, coded A-E represent A: Macrothesaurus (9), B: Political Science Thesaurus (10), C: SPINES Thesaurus (10), D: UNESCO Thesaurus (12), E: Thesaurus on Youth (13) (to be published late 1981). Some narrower and related terms have been omitted in order to reduce the size of the entries to manageable proportions for the purpose of this paper. The entry for the term Unemployment is reproduced here.

UNEMPLOYMENT A:13.01.03 B:S06.0655 C:7335 D:Q55.40 E:HG

< Deprivation	–	–	–	–	E
< Employment	A	–	C	D	E
< Employment characteristics	–	B	–	–	–
< Social disadvantage	–	–	–	D	–
< Social problems	A	–	–	–	–
> Cyclical unemployment	A	–	–	–	E
> Frictional unemployment	–	–	–	–	E
> Seasonal unemployment	A	–	–	–	E
> Structural unemployment	A	B	C	–	E
> Technical unemployment	A	–	C	–	–
> Technological unemployment	–	B	–	–	–
> Youth unemployment	A	–	–	D	E
– Business cycle	A	–	C	–	–
– Deflation	–	–	C	–	–
– Economic recession	A	–	–	–	–
– Employment opportunities	–	–	C	–	–
– Employment policy	–	–	C	–	–
– Employment services	–	–	C	–	–
– Full employment	A	B	C	–	–
– Job training	–	B	–	–	–
– Labour market	–	B	–	–	–
– Labour supply	A	B	C	–	–
– Right to work	–	B	–	–	E
– Unemployed	A	–	–	D	–
– Unemployed people	–	–	–	–	E
– Unemployed persons	–	B	–	–	–
– Unemployment benefits	–	–	–	–	E
– Unemployment insurance	A	–	C	–	–
– Unemployment insurance policy	–	B	–	–	–
– Unemployment policy	–	B	–	–	–
– Welfare policy	–	B	–	–	–

This entry and the ones in the Appendix 1 show that

(a) The Descriptor Bank reveals exact matches and non-matches between terms in different ILs. For exam-

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

ple, Unemployment matches exactly in all five ILs, but Structural Unemployment coincides in no more than three and Frictional Unemployment occurs in one only.

(b) It includes definition and scope notes and draws attention to conflicting interpretations of the meaning of a term.

(c) It shows inexact and partial equivalences between terms, provided this information is built into the equivalence relationship structure of the constituent ILs. For example:

Unemployment benefits	IL:E
→ Social security	IL:D
→ Unemploment insurance	IL:A

(d) It acts as a visual guide to assist the user in moving from one IL to another. The searcher wishing to transfer from a database indexed by IL A to one indexed by IL C can deduce from the entries under Unemployment that although he can continue to use the term Structural Unemployment he must broaden the search to Unemployment if he wishes to search for Cyclical Unemployment, as that term is absent from IL C.

(e) The Descriptor Bank also highlights variations in depth of coverage and also differences in the treatment of narrower and related terms.

(f) The Descriptor Bank also gives the natural language equivalents of merged terms from multilingual thesauri and discloses any variations in translation which might occur.

Sager in his *Guidelines* (8) discusses the merging of thesauri into a Descriptor Bank and appears to suggest that this would include automatic confounding of singulars and plurals and of alternative spellings, for example, 'Labour' and 'Labour'. With some prior editing he implies it might also be possible to merge terms which are alternative wordforms and near synonyms of the same concept, for example, Unemployed, Unemployed Person, Unemployed People. It is also assumed that the Descriptor Bank will be able to show how some terms occur as both preferred and non-preferred terms.
For example:

TECHNICAL UNEMPLOYMENT	IL:A IL:C
< Unemployment	IL:A IL:C
Technical unemployment	
→ Structural unemployment	IL:E

The Descriptor Bank *per se* has its immediate uses, especially in database searching, but it does not provide all the requirements of the Integrated Thesaurus. For example, if it embraces only machine-readable items it excludes the many manually produced thesauri and special classification systems which still exist.

Another defiency is its purely verbal approach; this means that it does not show inexact and partial equivalences between concepts which have not been revealed by the cross-reference structure in the merged ILs. Technical Unemployment for instance in ILs A & C and Technological Unemployment in IL B are separate terms in the Descriptor Bank (see Appendix 1). They are probably synonymous, but this fact is not recorded. Similarly, Unemployment Insurance Policy in IL B seems likely to be the nearest equivalent to Unemployment Insurance in IL A and IL C, but mere merging of terms would not establish this. Verbal merging also excludes concepts from classification systems where the terminology is imprecise or phrase-like, and which can only be handled by a conceptual compatibility matrix as described by Dahlberg in her *Guidelines* (14). The Descriptor Bank, in addition, lacks a subject framework in which to evaluate the constituent entries. Individual alphabetical term entries can show differing levels of specificity and inconsistency between ILs, but the entries need to be brought together in context, within a sub-field so that gaps in coverage, redundancies, and inaccuracies in relationships may be discerned more easily.

5. The role of classification

Since conceptual analysis is more readily accomplished within limited subject areas, such as Marriage, the Family, or Social Class or Unemployment an overall classified structure is required which will locate and reveal these areas within main disciplines and subject fields. The subfield Unemployment, for example, would be placed mainly within Economics but also in Social Welfare, and Marriage found within Sociology but also in Psychology and Social Welfare.

The classified approach has certain advantages over the purely alphabetical. During compilation, terms can be considered in context, and using facet techniques the underlying structure of the sub-field can be discovered showing gaps in coverage and new interrelationships. Translation of terms between natural languages is also facilitated when terms can be considered together in subject areas. The structure provided by a faceted classification assists at the updating stage the insertion of new terms and suggests their relationships. In published thesauri the classified display provides an overview of the extent and quality of coverage in any subject area while the notation (i. e. class codes) may be used to tag and retrieve sub-sets in database searching of to make explosion-type searches on hierarchies using truncated notation, and so reduce the labour of compiling the search program.

Sager in his *Guidelines* says that

> the classification system must be catholic and broad enough to accommodate all the branches of the social sciences to be encompassed by the Integrated Thesaurus

This specification could be satisfied by any of the universal classification systems whether DC, LC, the Universal Decimal Classification (UDC) or the Broad System of Ordering (BSO). However, breadth of coverage is not the only requirement, as depth of treatment is also essential. BSO has been shown by Dahlberg (15) to be insufficiently detailed for the Integrated Thesaurus; but the universal classification systems, especially LC and UDC have the necessary depth of coverage. However, a further characteristic is needed for a classification for use in the Integrated Thesaurus: it must have an internal structure compatible with that of thesauri. This means it must organise concepts in a logical way, at the term rather than at the subject level, so that it can serve as the master in a compatibility matrix for comparing terms from thesauri as well as notation from classification systems in the sub-fields of the Integrated Thesaurus.

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Fig.1.

BLISS BIBLIOGRAPHIC CLASSIFICATION OUTLINE

OUTLINES OF THE COMPLETE CLASSIFICATION

FIRST OUTLINE

2	Generalia : Physical forms and forms of arrangement
4	Phenomena : attributes, activities, entities
7	Knowledge, information, communication
	Disciplines
A	Philosophy
AM	Mathematics
AZ	Science
B	Physical sciences
E	Biological sciences
H	Man, anthropology
	Physical : medicine, psychology
J/Z	Social sciences and humanities

SECOND OUTLINE

2	Generalia : physical forms of documents
3	Generalia : forms of arrangement
	Phenomena
	* For multi-disciplinary treatments of particular phenomena
4	Attributes, activities and processes, entities
7	Universe of knowledge
	Methods of enquiry, research
	Disciplines, forms of knowledge
	Communication and information
	Media : semiotics, aural and visual communication
	Recorded information : information sciences and technologies
	Data processing, computer science
8	Records, documentation
9	Mass communication, publicity, public opinion
	Disciplines
A	Philosophy
AL	Logic
AM	Mathematics
AW	Statistics and probability
AX	Systemology, organisation theory, management
AY	Science and technology
AZ	Science
B	Physics
BC	Technologies based primarily on physics (* Alternative is UF)
C	Chemistry
CT	Materials science and technology (* Alternative is UG)
D	Astronomy
DG	Space science
DH	Earth sciences
DHA	Geophysics, geology, hydrology, meteorology
DT	Geography
DTA	Regional geography
DU	Systematic, analytical geography : physical, human
E	Biology
EV	Microbiology
F	Botany
G	Zoology
H	Man, anthropology
HA	Human biology
HH	Health sciences, medicine
I	Psychology
J	Education
K	Social sciences
KA	Sociology
KC	Social anthropology
KK	Customs
KM	Folklore
KO	Ethnography
KT	Environment, social ecology
KY	Travel and description
	[Travel and description]
L	History
L9	Biography
LA	Ancillary historical studies : archeology ...
LG	History, by subject
LI	History, by period
M	Europe
N	America
OA	Australia
OH	Asia
OS	Africa
P	Religion
PB	Systematic theology
PF	Religious systems, institutions
PG	Ancient religions
PI	Modern religions
PX	Occult
PY	Morals, ethics
Q	Social welfare and administration
QE	Social services
QG	Causes of need, persons in need
R	Political science
RO	Public administration
S	Law
T	Economics
TT	Economic systems
TX	Management of enterprises
U	Technology, useful arts
UA	Agriculture and animal exploitation
UD	Mining technology
UE	Engineering and production technology
UF	Physics-based technologies (* Alternative to BS)
UG	Materials science and technology (* Alternative to CT)
UHC	Construction technology : civil engineering, building
UM	Architecture and physical planning
UN	Public health engineering
UO	Transport technology
US	Military science and technology
UT	Other technologies, by product
UV	Food, clothing, housing
UY	Recreative arts
V	Arts, fine arts
V7	Subjects in art, style in art
VB	Architecture as an art (see also UJ)
VE	Plastic arts : sculpture, glyptics, ceramics
VJ	Graphic arts : painting, drawing...
VP	Reprographic and decorative arts
VV	Music
VY	Performing arts
W	Philology : language and literature
WA	Linguistics
WI	Individual languages and their literatures
YU	Literature : general and comparative

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

There is one Universal Classification system, the second edition of the Bibliographic Classification of H. E. Bliss (BC2) (16) which meets most of these requirements. The work on the revision of BC2 is in progress under the direction of Jack Mills, assisted by his colleagues, Vanda Broughton, Eric Coates and others at the Polytechnic of North London. The second edition retains the carefully planned order of basic classes for which the first edition has been praised (see Fig. 1). The concepts and terminology are up-to-date and specificity is pitched more towards periodical than monograph literature. Publication began in 1977 and by mid-1981 a number of classes falling within the social sciences have been published including Class I Psychology and Psychiatry, Class J Education, Class P Religion, Class Q Social Welfare. Other classes are near publication and available in penultimate drafts, for example, Class T Economics and Management and Class K Society (includes Culture and Sociology). There are also early drafts of Class R Political Science and Class S Law, which might need updating from other specific faceted classifications, before they are ready to use in a compatibility matrix.

6. Subject field analysis

Within each sub-field, analysis of the terms from the ILs to be integrated may be accomplished using a conceptual compatibility matrix as shown in Dahlberg's *Guidelines* (14). One way to create such a matrix is to provide a baseline document or master classification system. Margaret Block in her research at the United States Department of Health (17) compiled a classification to act as the baseline against which existing ILs were compared. In the case of the Integrated Thesaurus it is suggested that BC2 might be used as the master. It would not, however, be an inflexible standard, but be open to change and amendment under the influence of the ILs in the matrix. At the end of the exercise it would emerge as the Integrated Thesaurus, transformed possibly in content and certainly in format. For the purpose of illustrating the process of analysis and development at the sub-field level the subject of Unemployment is again taken as the example.

6.1. Selection of master classification

The main place for Unemployment in BC2 is in Class T Economics and Management, although it also occurs as a social problem in Class Q Social Welfare. In Class T Unemployment is placed under Employment within Labour Economics. An extract from the relevant section of Class T is shown in Fig. 2.

6.2. Comparison of master and ILs in matrix

Terms and class codes related to the subject of Unemployment in the ILs to be integrated are at this stage brought together. If the Descriptor Bank exists, a printout of relevant terms could be obtained provided the entries have been broadly classified by subject field. The resulting collection of merged terms might include a batch of entries similar to that found in Appendix I. Supposing, however, that there is no Descriptor Bank, a similar collection of merged terms on Unemployment could be produced by checking the terms in the master and against the ILs and then by following the network of cross-references in the various thesauri. An operation of this sort would be necessary, even if a Descriptor Bank is available, for the selection of terms from non-computerised vocabularies and classification systems.

Once all the merged terms and codes are assembled, they may then be compared with the master classification by listing them against the master, in parallel columns, according to ILs, to form a compatibility matrix.

7. Integrated Thesaurus formation from matrix

During the comparison with the ILs, the BC2 master will undergo change and amendment. The revised classification is then set out in a format suitable for input to the computer. The resulting output will form the sub-field for Unemployment in the new Integrated Thesaurus. (See Figs. 3–5)

A computer program which can handle this type of classified display is British Standards Institution's ROOT system ,developed for the ROOT Thesaurus (18), but capable of handling social science as well as technical subjects. A feature of the ROOT systems is that unlike BC2, which enters a concept in as many subject fields and sub-fields as it may legitimately occur, the ROOT system requires that a term should be entered once in its most appropriate location and be shown as a cross-reference elsewhere. From the schedule on Unemployment in Fig. 3 set out in the ROOT format it can be seen that a decision was made to consider Unemployment Insurance and Unemployed People in Class Q Social Welfare. Cross-references to these related terms now placed in other schedules are given immediately below the term Unemployment using international symbols preceded by asterisks.

TLE Unemployment
*— Unemployed people QGX
*-- Unemployment insurance QFG.G

Extracts showing corresponding entires for these terms in Social Welfare Q are given in Figs. 4 und 5.
Changes in the original BC2 master shown in Figs. 3, 4 and 5 are due to its interaction with terms from the ILs to be integrated. These ILs A-E are the same as those represented in the Descriptor Bank illustrated in Appendix 1, with the addition of IL F, Abridged English Edition of the UDC. The terminology has again been reduced in size and modified to make the displays more meaningful. The amendments to the BC2 master include:

7.1. Addition of new terms

The BC2 master is extended by the addition of new terms, for example Youth Unemployment TLE.QY in Fig. 3 taken from ILs A, D and E and Unemployed Young People at QGX.LR in Fig. 5 taken from IL E.

7.2. Deletion of terms

Regional Unemployment in Fig. 3 is assumed to have been deleted from BC2 after being discarded by subject experts.

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Fig.2.
BLISS BIBLIOGRAPHIC CLASSIFICATION
SUB-FIELD UNEMPLOYMENT
(in Class T Economics and Management penultimate draft)

ECONOMICS BC1

(ECONOMIC RESOURCES)
((Types of resources) Factors of production)
(LABOUR ECONOMICS)
((Supply, demand and exchange))
(Mobility of labour, immobility)
(Skill structure, career structure)
((By various characteristics)

TLD LR — Physical characteristics
LS — Educational characteristics

M — (Demand for labour) Employment, aggregate employment, unemployment and employment together
* Includes works on employment and unemployment levels
(Policy)
MB — Employment policy
(Selection) see TLF L
(Reduction) see TLF XG
MBR — (Fluctuations)

N — Levels of employment
P — Full employment
Q — Employment to capacity (special to Management)
R — Partial employment
S — Short-time working
T — Under-employment

TLE → Unemployment, involuntary unemployment
(Theories)
4FA — Philips curve
9T — (Welfare economics)
Add to TLE 9T letters A/Y following QF in QFA/QFY so far as is necessary.
9TF K — Unemployment benefits, relief, dole
* Economic aspects only
See also under Social Welfare Q
9TG — Unemployment insurance
AN — Rate
AP — Duration of unemployment
(Regulation and control)
See also Employment policy TLD MB
TLE J — Job creation programmes
JL — Public works, useful public works
JN — Direct labour
JP — Compensatory employment
Special grants to employers
(Reflationary measures) see TES H
K — Even distribution of work, work sharing
* By reducing working hours, banning overtime, etc.
L — Restricting entry of certain categories of workers, numerus clausus, quotas
M — Retraining schemes, government retraining schemes
MP — Government retraining centres

MB — (Forms of unemployment)
MC — Redundancies (general)

Fig.2. (Cont'd.)

ECONOMICS BC1

(ECONOMIC RESOURCES)
((Types of resources) Factors of production)
(LABOUR ECONOMICS)
((Demand for labour) Employment, aggregate employment,
(Levels of employment)
→ (Unemployment, involuntary unemployment)
((Forms of unemployment))
(Redundancies (general))
TLE ND — Sectoral unemployment
NE — Regional unemployment
NF — Frictional unemployment
* Resulting from immobility of labour and time gaps when changing jobs.
NH — Temporary, layoffs
NK — Induced retirements
(Unemployables) see Persons TLW WB
P — Mass unemployment
PQ — Cyclical unemployment
PR — Seasonal unemployment
PS — Casual unemployment
PT — Structural unemployment
* Caused by longterm changes in consumer demand and in technology
PV — Technical, technological unemployment TGM
(Specific causes)
PW — Automation
PX — Disguised unemployment
See also Participation rate TLD EF
TGM PY — Invisible unemployment
* Potential labour not registered (e.g. married women, illegal immigrants.)
Q — Lowered intensity of labour
QR — Overmanning
QS — Short time working
QV — Voluntary unemployment
* Persons who choose not to work for some reason.

(Types of employment)
Apprenticeships
(Compensatory) see Unemployment control TLE JN
(By time factor)
R — Permanent employment (usually assumed)
RT — Temporary employment
S — Casual employment
ST — See also Casual unemployment TLE PS
Seasonal employment
SV — See also Seasonal unemployment TLE PR
T — Contract work
TU — Homework, outwork
* Done at home, supplied to factories, etc.
TW — Lump labour
TX — Full-time work (usually assumed)
V — Part-time work
VW — Moonlighting
VX — Specialized employment
VY — Illegal, unadeclared (e.g. of illegal immigrants)
(Distribution)

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

7.3. Addition of synonyms and quasi-synonyms

In Fig. 4 Unemployment Insurance Policy from IL B is added as a synonym to QFG.G.

7.4. Change of hierarchical levels

The hierarchical levels of the BC2 master may also be changed during analysis. In Fig. 3 IL A makes Seasonal Unemployment a narrower term to Cyclical Unemployment. It is assumed that the subject experts have accepted this interpretation and the master is modified accordingly.

7.5. Addition of polyhierarchies

The original BC2 schedule does not show the broader and narrower terms in other subject fields which are present in the ILs to be integrated. These are added as asterisked terms, for example, in Fig. 3 Deprivation QGP is added as an asterisked broader term cross-reference under Unemployment taken from IL E.

7.6. Addition of related terms

The original BC2 shows only a few related terms from other schedules. It was necessary to compare the related terms of the ILs in the matrix and to add these as asterisked related terms to the master to form the Integrated Thesaurus. Reference to ,the terms in the Descriptor Bank in Appendix 1 shows that there are numerous related terms occurring in the ILs, mainly in ILs B and C. It is assumed in Figs. 3, 4 and 5 that these have been checked by subject experts and a selection has been added to the BC2 master. For example, in Fig. 3 Deflation from IL C, and Economic Recession from IL A have been added. An unanswered question is how many related terms should be added to the Integrated Thesaurus: all related terms or only those recognised by experts to be the most closely associated?

7.7. Compound term treatment

In Figs. 3–5 it will be seen that some compound terms have been retained in the BC2 master which would have required factoring, according to BS 5723 and revised IS 2788 rules. It is possible that many unauthorised compound terms may have to be retained because of their familiarity and frequency of occurrence. However, if it is decided that the factoring rule should be strictly adhered to, the split terms should nevertheless be shown in the classified display, using the ROOT system device for synthesised terms, e. g.

TLE.QY **Youth unemployment
→ Unemployment
+ Youth

8. The compatibility matrix

While the BC2 master is emerging as the Integrated Thesaurus, the parallel listing of the corresponding terms of the ILs to be integrated forms a compatibility matrix for use in the comparison and switching between ILs and between the ILs and the master Integrated Thesaurus. The layout of the matrix can be seen in Figs. 3, 4 and 5. An example see bottom of this page.

For concepts from constituent ILs which match exactly the terms in the master, only the identification number or notation is given, but where the term is different, whether it is a synonym as in IL B or a broader term as in IL D, the wordform also is given. For classification systems both the class number and the wordform is shown. The symbol < before the code 331.063 in IL F indicates that the class number is for a broader concept than the one in the master. There is no match for the term in IL D and it can be seen from the display in Fig. 3, that the broader term to use must be Unemployment.

The terms from the ILs in the matrix are matched not only against the main terms in the BC2 master but also against the asterisked cross-references. This information could be omitted from the matrix here and appear only against the term in its main place in the classification system. For example, the matrix display would be omitted against Unemployed People where it appears as a cross-reference under TLE, Fig. 3, but appear in Fig. 5 at its main location at QGX. To do this would reduce the value of the sub-field display for Unemployment when searching speedily for compatible terms in different ILs.

From Figs. 3, 4 and 5 it can be seen that the compatibility matrix can act as a switching mechanism for transferring indexed data between thesauri and classification systems, suggest improvements to the constituent ILs whether in vocabulary control, interrelationships or specificity, and act as a guide to searchers moving between databases. In this latter operation it is possible that it may not serve as well as the Descriptor Bank, in suggesting related terms to search if the matrix excludes a proportion of the related terms in existing ILs. The Descriptor Bank might also be better at indicating the relationships in the constituent ILs than the matrix, as for example in Fig. 3 the matrix does not show that Technical Unemployment in IL A is not NT to Structural Unemployment as it is in the master. Some code might be devised to indicate this type of discrepancy.

9. Presentation and layout of published thesaurus

9.1. The classified display

As has already been discussed, the classified section of the Integrated Thesaurus could be set out as for the ROOT system as illustrated in Figs. 3, 4 and 5. This

TLE.PV Technical unemployment = Technological unemployment	A:13.01.03	B:S06.0657 Technological unemployment	C:6949	D	E Structural unemployment HGGS	F:<331.063 Unemployment types

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

layout is the same as that of the classified section in the UNESCO Thesaurus, except that asterisked entries are given under the term in the display for all, rather than a selection of broader, narrower and related terms appearing elsewhere in the schedules. From this style of 'embedded' classified display a full alphabetical section may be generated by computer. The ROOT program includes scope notes in the classified display, but it might be preferable if the lengthy definitions and detailed scope notes likely to occur in the social sciences were printed out in the alphabetical section only of the Integrated Thesaurus. The notation used for the Integrated Thesaurus could be similar to that of BC2 although some adjustments might be required to make it more amenable to computer searching.

9.2. Printed display of the matrix

A suitable form for the published matrix display will not be easy to devise. The matrix presentation in Figs. 3, 4 and 5 is already overcrowded with six ILs and it is likely that the average matrix will include more than this number. A larger size page or photoreduction might be required. An alternative to displaying the matrix in the classified section could be to transfer the information to the alphabetical section. (See para. 9.3.)

9.3. The alphabetical section

The alphabetical section is derived automatically from the classified display in the ROOT system and contains all the information (to one hierarchical level) in a conventional thesaurus format, using the same international symbols as found in the classified display.

Descriptor. Notation
= Foreign language equivalent(s)
[Definition or scope note]
≈ Non-preferred synonym or quasi-synonym
< Broader term in the same part of the display
> Narrower term in the same part of the display
– Related term in the same part of the display
*< Broader term in another part of the display. Notation
*> Narrower term in another part of the display. Notation
*– Related term in another part of the display. Notation

Examples:
Unemployment TLE
< Employment
> Cyclical unemployment
> Frictional unemployment
> Structural unemployment
> Temporary unemployment
> Youth unemployment
*< Deprivation QGP
*– Deflation TEJ.F
*– Economic recession TDO
*– Redundancy TUF.XK
*– Unemployed People QGX
*– Unemployment insurance QFG.G

Technical unemployment TLE.PV
[Unemployment due to displacement of certain occupations by technology]
= Technological unemployment
< Structural unemployment
*– Technological change UAH

The non-descriptors may be synonyms or synthesised terms, having the following

Synonym
—→ Descriptor. Notation

**Synthesised term. Notation
—→ Descriptor A.
+ Descriptor B.

Descriptor A
+ Descriptor B
= **Synthesised term. Notation

and the same for Descriptor B.

For example:

Technological unemployment
—→ Technical unemployment TLE.PV

**Youth Unemployment TLE.QY
—→ Unemployment
+ Youth

Unemployment
+ Youth
= **Youth unemployment TLE.QY

Youth
+ Unemployment
= **Youth unemployment TLE.QY

The alphabetical section serves mainly as an entry to the classified display, where the compatibility matrix will be found. Alternatively, the alphabetical section might be used instead of the classified for the display of compatibility data, with the information tabulated immediately below each descriptor entry (see figure 6), e.g.:

TECHNICAL UNEMPLOYMENT TLE.PV
A 13.01.03
B S06.0657 Technological Unemployment
C 6949
etc.

This type of tabulated display can be handled more readily by the computer, but it lacks the comprehensive overview of subject field compatibility provided by the matrix in the classified display.

9.4. Other display

A permuted index should be included to supplement the alphabetical section by ensuring the retrieval of second and third words within compound terms. In addition it might be helpful to have, as in the UNESCO Thesaurus, a Top Term (TT) display to gather together under top-terms hierarchies of broader and narrower terms which may cut across the classified arrangement.

10. Multilingual processes

The first draft of the Integrated Thesaurus is likely to be in English with versions in other languages following in due course. The French version would be compared with monolingual French thesauri, and the German version with monolingual German thesauri, and so on for other languages. This operation could result in alterations to the structure of the French and German and other language versions of the Integrated Thesaurus, which would have to be transmitted back to the original English version.

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

11. Testing and updating

As the draft for each sub-field is completed it would be input to the computer and printed out as a classified display, matrix and alphabetical index. This printout would be discussed, tested, modified and corrections input. Subject fields completed earlier would be enlarged and amended automatically as reciprocal asterisked entries are input from broader, narrower and related terms in other sub-fields completed later. The full alphabetical thesaurus would not be printed out until the whole of the classified section is completed and corrected. When both classified and alphabetical parts are output, checked and corrected, the thesaurus would be ready for photocomposition and printing. Continuous updating would be necessary after publication so that new published editions might be produced at regular intervals, and an up-to-date version be available for online display.

12. Administrative aspects

It is not within the scope of this paper to suggest the administrative machinery necessary to organise the construction of the Integrated Thesaurus. The work would probably be undertaken by working groups, supervised by an Editorial Committee, with access to subject experts. The elapsed time for the completion of the English language version might be three years and the multilingual edition a further two years.

13. Conclusions

An Integrated Thesaurus could have a valuable role to play as a reference standard, switching tool and comparison and compatibility matrix in the improvement of communication and information retrieval in the social sciences. It would encourage the enhancement of existing ILs and increase the overall understanding of social science concepts. The compilation of such an Integrated Thesaurus is likely to be a time-consuming and intellectually demanding task. However, if fuller consideration is given to compilation techniques, solutions may emerge which will simplify the process, making the task of compilation well within the capability and resources of information and social science experts.

References:

[1] UNESCO: Division for the International Development of the Social Sciences. Consultative Meeting on the Establishment of an Integrated Thesaurus of the Social Sciences; Paris, 9–11 June, 1980. Final Report. Paris: Unesco 1980.

[2] Horsnell, V.: Intermediate Lexicon for Information science: a feasibility study. London: Polytechnic of North London, School of Librarianship 1974.

[3] Horsnell, V. Merrett, Alicia: Intermediate Lexicon research project, phase 2. London: Polytechnic of North London, School of Librarianship 1978.

[4] International Standards Organisation: International Standard 2788: Guidelines for the establishment and development of monolingual thesauri. Geneva: ISO 1974.

[5] British Standards Institution: British Standard 5723: 1979. Guidelines for the establishment and development of monolingual thesauri, London: BSI 1979.

[6] UNISIST: Guidelines for the establishment and development of multilingual thesauri. Revised text. (PGI/80/WS/12). Paris: Unesco 1980.

[7] Somers, H. L.: Observations on standards and Guidelines concerning thesaurus construction. In: Intern. Classif. 8 (1981) No. 2. p. 69–74

[8] Sager, J. C. et al.: Guidelines for the establishment of comparison and compatibility matrices between thesauri in the social sciences, submitted to the Unesco Division for the International Development of Social Sciences. Manchester: University of Manchester, Institute of Science and Technology 1981. 3 vols.

[9] Viet, Jean: Macrothresaurus for information processing in the field of economic and social development. New English ed. prepared by Jean Viet. Paris: OECD 1978.

[10] Beck, Carol et al.: Political science thesaurus 2; revised and expanded second ed. Pittsburg: University Center for International Studies 1979.

[11] SPINES Thesaurus: a controlled and structured vocabulary of science and technology for policy-making, management and development. Paris: Unesco Press 1976. 4 vols.

[12] Aitchison, Jean: UNESCO thesaurus: a structured list of descriptors for indexing and retrieving literature in the fields of education, science, social science, culture and communication. Paris: Unesco 1977. 2 vols.

[13] Aitchison Jean et al.: Thesaurus on youth: an integrated classification and thesaurus for youth affairs and related topics, compiled in association with Inese A. Smith and Susan Thompson. Leicester: National Youth Bureau 1981.

[14] Dahlberg, Ingetraut: Toward establishment of compatibility between indexing languages. In: Intern. Classif. 8 (1981) No. 2, p. 86–91

[15] Dahlberg, Ingetraut: The Broad System of Ordering (BSO) as a basis for an Integrated Social Sciences Thesaurus: In: Intern. Classificat. 7 (1980) No. 2, p. 66–72.

[16] Mills, J., Broughton, V.: Bliss Bibliographic Classification, vol. 1. Introduction and auxiliary schedules; Class H Human biology and health sciences Class I Psychology and psychiatry; Class J Education; Class P Religion; Class Q Social welfare. London: Butterworth 1977.

[17] Block, Margaret L.: Analysis of existing controlled vocabularies: Women's Educational Equity Communications Network. Princeton, New Jersey. United States Department of Health Education and Welfare. Educational Testing Service 1978.

[18] British Standards Institution: BSI ROOT Thesaurus. 2 vols. Hemel Hempstead: BSI 1981.

Technical unemployment TLE.PV

A	13.01.03	
B	S06.067 Technological unemployment	Compatibility Data
C	6949	
D	–	
E	→ Structural unemployment HGCS	
F	< 331.6.063 unemployment types	
[	Unemployment due to displacement of certain occupations by technology]	
=	Technological unemployment < Structural unemployment	
*	– Technological change UAH	

Fig. 6 Compatibility Data Displayed in Alphabetical Section

Jean Aitchison
12 Sollershott West
Letchworth
Herts. SG6 3PX,
UK

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access -

Fig.3.

INTEGRATED THESAURUS – CLASSIFIED DISPLAY AND MATRIX

		Bliss (B32) master (amended)	Matrix of ILs A–F					
			A	B	C	D	E	F
T		ECONOMICS AND MANAGEMENT (Cont'd.)						
TK		ECONOMIC RESOURCES (Cont'd.)						
TL		LABOUR ECONOMICS (Cont'd.)						
TLD.M		Employment						
		.....						
TLE		Unemployment	13.01.03	306.0655	7335	955.40	HG	331.6
	Add BT from E	+< Deprivation QGP	02.04.01 Social problems	-	-	R71 Social disadvantage	LFN	-
		+> Redundancy TUF.XK	-	-	-	-	-	-
	Add RT from C	+– Deflation TES.F	-	-	1638	-	-	-
	Add RT from A	+– Economic recession TDC	03.02.04	-	-	-	-	-
	Add RT from A,B,C	+– Unemployment insurance QFG.G	02.33.02	326.0309 Unemp. ins. policy	7336	-	-	368.44 Unemp. insurance
	Add RT from A,B,C and E	+– Unemployed people QCX	13.01.02 Unemployed	332.0127 Unemployed person	-	R84.15 Unemployed	HGS	-
		.....						
		.....						
		(By form of unemployment)						
	Delete. Treat at TUF.XK	Redundancy						
	Delete	Regional unemployment						
TLE.NF		Frictional unemployment	-	-	-	-	HGOF	<331.6.063 Unemployment types
		Retain BC2 scope note and print out in alphabetical section						
TLE.NR	Retain for future use	Temporary unemployment	-	-	-	-	-	-
TLE.PQ		Cyclical unemployment	13.01.03	-	-	-	HGGP	-
TLE.PR	Make NT as A	Seasonal unemployment	13.01.03	-	6243	-	HGGM	<331.6.063 Unemployment types
	Add RT from E	+ – Seasonal employment TLE.SF	-	-	-	-	HFT	-
	Add RT from A	+ – Seasonal workers TLV.SM	13.03.02	-	-	-	-	-
TLE.PT		Structural unemployment	13.01.03	306.0656	6744	-	HGGS	<331.6.063 Unemployment types
	Consider RTs from B and C here							
TLE.PV	Retain indenting although not NT in A, B and C	Technical unemployment	13.01.03	306.0657 Technological unemployment	6949	-	→ Structural unemployment HGGS	<331.6.063 Unemployment types
		= Technological unemployment						
	Add RT from A	+– Technological change UAH	12.06.00	-	-	-	-	-
		(By age group)						
TLE.QY	Add new term from A, D and E	Youth unemployment	13.01.03	-	-	955.60.20	HI	<331.6.063 Unemployment types
	Add RT from E	+– Unemployed young people QSX.LR	-	-	-	-	HIB	-
	Add RT from A	+– Youth QLR	14.02.02	-	-	-	-	-
	Add RT from D, E	+– Youth employment TLE.RL	-	-	-	955.60	HH	-

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY

Fig.4 / Fig.5.

INTEGRATED THESAURUS – CLASSIFIED DISPLAY AND MATRIX

Bliss (BC2) master (amended)

Matrix of ILs A–F

			A	B	C	D	E	F
Q		SOCIAL WELFARE (Contd.)						
QF		SOCIAL SECURITY						
		.....						
QFG		..... SOCIAL INSURANCE						
		= National insurance						
QFG.G		Unemployment insurance	02.03.02	S26.039 Unemp. ins. policy	7336	→ Social security R95	MWD Unemployment benefits	368.44
	Quasi-synonyms from A and B	= Unemployment benefits						
		= Unemployment insurance policy						
	RT from C	+– Labour administration TLB.E	–	–	3750	–	–	–
	RT from C	+– Labour legislation TLB.F	–	–	3756	–	–	–
	RT from A and B	+– Unemployed people QGX	13.01.02	S32.0127 Unemployed person	–	–	–	–
	RT from A–C	+– Unemployment TLE	13.01.03	S06.0655	7355	–	–	–
		.....						
		.....						
Q		SOCIAL WELFARE						
QG		SOCIAL PROBLEMS						
		.....						
		.....						
QGG		DISADVANTAGED PEOPLE						
		= People in need						
		= Persons in need						
		.....						
QGP		Deprivation						
		= Social disadvantage						
		= Underprivilege						
		> Unemployment TLE						
QGP.G		Deprived people						
		= Socially disadvantaged						
		= Underprivileged						
		.....						
		.....						
		.....						
QGX		Unemployed people	13.01.02 Unemployed	S32.0127 Unemployed person	–	R84.15 Unemployed	HGB	–
		= Unemployed						
		= Unemployed person						
	BT from B	+< Occupational role TLD.LM	–	S32.0061	–	–	–	–
	RT from A	+– Job seekers TLF.AOV	13.02.01	–	–	–	–	–
		+– Redundancy TUF.XK	–	–	–	–	–	–
	RT from A,B,D and E	+– Unemployment TLE	13.01.03	S06.0655	–	Q55.40	HG	–
	RT from A and B	+– Unemployment insurance QFG.G	02.03.02	S26.0309 Unemp. ins. policy	–	–	–	–
	RT from D	+– Workers TLV.B	–	–	–	Q88	–	–
QGX.LR	New term from E	Unemployed young people	–	–	–	–	HIB LEO	–
		+< Deprived young people QGP.LR						
		+– Youth unemployment TLE.QY	–	–	–	–	HH	–

https://doi.org/10.5771/0943-7444-1981-2-75 https://www.inlibra.com/de/page/agb - Open Access - (cc) BY