

THE COMPILATION OF THE THESAURUS ON WOMEN IN DEVELOPMENT

by

Melling Simanjuntak

Chief editor of the Thesaurus on Women in Development;
Technical Service Librarian at the PDII-LIPI

The coordinator of the Indonesian Clearinghouse on Women in Development, Luwarsih Pringgoadisurjo, established a thesaurus working group in February 1985. This was in realization of the agreement of the ASEAN Women's Programme (AWP) made in Penang during its Second Expert Group Meeting on the Clearinghouse on Women in Development, November 1984. A group of five persons was assigned with the task of preparing a thesaurus to facilitate retrieval systems in clearinghouse of AWP's member countries.

The working group began its first session in mid-February 1985 by identifying the various preliminary steps to be taken. The first was to formulate the nature of the thesaurus needed. It was decided that the thesaurus in question should meet the specifications set forth in ISO 2788 (1) or follow the Guidelines for the Establishment and Development of Monolingual Thesauri (2). The subject covered should be women's studies, which was broken down into 12 categories--this was later expanded to 14 (Appendix 1).

Paper presented at the 2nd Workshops on Thesaurus Construction for the Clearinghouse on Women in Development, Jakarta 4 - 5 December 1987.

Realizing the tediousness and expensiveness of constructing a thesaurus, the working group agreed to seek for better alternatives by: 1) searching for a thesaurus that would suit AWP need; 2) searching for a thesaurus that seemed to be appropriate to AWP needs, but written in a language other than English; and 3) searching for a thesaurus that would be appropriate to AWP needs, with some minor adaptations.

The first action taken was to search for such kind of thesauri through secondary literature, particularly the *Library and Information Science Abstracts and Library Literature*. A number of bibliographies on thesauri were also consulted including those compiled by Gilbert (3), Walkley and Hay (4), and Hafner (5). The first two bibliographies list the thesauri held by the Aslib Library, while the third one lists the 256 titles of all published thesauri. Unfortunately, there was not one appropriate thesaurus could be found.

Furthermore, the working group contacted the Bibliographic System Center - Case Western Reserve University, which is Unesco Clearinghouse for Thesauri (4). It was informed of the existence of two thesauri on women, namely:

1. A special collection thesaurus : women studies/ G.K. Burgess. Seattle : Women's Studies Program, University of Washington, 1979. 152 p.
2. On equal terms: a thesaurus for non-sexist indexing and cataloging/J.K. Marshal. New York: Neil Sherman, 1977. 152 p.

The thesaurus by Burgess, however, was out-of-print. Even worse, we were informed that the publisher itself had not held a single copy of the material.

Inquiries were also sent to a number of relevant information centers, such as the library of the International Center on Women in Washington, D.C., hoping that at least one of them possessed a copy of the thesaurus and would kindly send us one. However we did not receive any information from them.

We managed to obtain a copy of the second thesaurus, on Equal Terms. It is, as we found out later, more a subject heading list than a thesaurus. Although it has the thesaural relationship among terms, i.e. the BTs, the NTs, and the RTs, it lacks a systematic or structural display of terms. It therefore fails to meet the specification outlined by ISO 2788, that "....a thesaurus should in any case include a systematic and an alphabetic display...."

Another thesaurus on women, compiled by Capek (9, 10), might have offered us much help if it had only been available earlier. Scheduled for publication in August 1987, this thesaurus is expected to provide a vast quantity of terms relevant to our thesaurus. We only became aware of its preparation in September 1987, about the same time the preliminary edition of our thesaurus was in press. We expect to have a copy in the next few weeks, and will look at it carefully in order to compare it with our publication. Capek's description of her thesaurus however, enables us to suspect that it differs from ours in some respects. Its compilation follow ANSI 239.3, it is designed to set national (USA) standards for terms to use in information retrieval activities, and it is a feminist thesaurus (9, 10) while ours follows ISO 2788, set regional (Southeast Asia) standards, and reflects the role of women in development. These differences in approach and background will evidently yield differences between the two theasuri.

Compiling the thesaurus

As we were quite sure of the absence of better alternatives for constructing the thesaurus, we proceeded to formulate the details of the thesaurus and how the compilation would be undertaken. In brief, the thesaurus was constructed in three parts: the hierarchical thesaurus, the alphabetical thesaurus, and the permuted index. The compilation involved four major steps: 1) selecting and collecting terms in domain of women's studies; 2) organizing the selected terms into a hierarchical structure; 3) converting the hierarchi-

cal structure into an alphabetical display, and 4) preparing a permuted index to the terms.

A. Selecting and collecting terms

The selection and collection of terms was performed by the method of literary-warrant, i.e., the terms selected were those that appeared frequently in literature. This approach is deemed to be appropriate for several reasons. Firstly, it promotes a higher rate of conformity between document language and the indexing language to be selected for the thesaurus. Secondly, it requires no excellent capability of recollecting the many terms associated with the subject--the capability that none possesses, but urgently required if one is to chose the theoretical topdown approach for terms selection. Additionally, the control of obtaining relevant terms can directly become a built-in mechanism once the sources of terms have been properly selected.

We used, as sources of terms, publications on women studies. Secondary literature such as abstracts and annotated bibliographies were of particular interest since they contain important terms in the subject field. Others sources of terms were articles from women periodicals, a number of thesauri not specialized on women studies, and publications of encyclopedic kinds. We had all together, 35 titles of sources of terms, of which 18 titles were abstracts, bibliographies, periodicals, and 7 titles of thesauri (Appendix 2).

We collected, out of all sources, about 2200 terms. These figures increased in January 1986 due to the inputs from distinguished social scientists and librarians participating in a workshop held to evaluate the first draft of the thesaurus. The present edition contains about 2400 terms.

B. Constructing the hierarchy

The construction of the hierarchical thesaurus was performed inductively using a bottom-up approach. Every single terms was identified. Terms which belong to a given, more generic term were grouped together in such a way that, for example, the terms "Lactogenic hormones", "Sex hormones", and "Steroid hormones" were grouped under "Hormones". So were "Gays" and "Lesbians" under "Homosexuals".

The objective of hierarchical construction is to organize terms according to their level of specificity, following as closely as possible, these conventions:

1. A term B is considered to be at one level of greater specificity than a term A if it is a kind of A; both terms have a kind of genus-species relationship.

2. A term B is considered to be at one level of greater specificity than a term A, if it represents a part of A; both terms have a kind of partitive relationship. This convention is only applicable for terms which relate to a) parts or organs of the body, b) geographical names, c) subject disciplines/scientific fields. A partitive relationship between two terms, other than the types mentioned here is considered to be an associative relationship.

3. A term B is considered to be at one level of greater specificity than a term A, if B is an example of A; both terms have a kind of Instance Relationship.

The organization of terms were completed in several stages. The first stage was to group the terms according to the conventions given above, producing hundreds of groups of terms. Each group was identified under a given broader term, which was at one level of lesser specificity than its subsumed terms. The second stage was to organize those broader terms in the same way we organized terms during the first stage.

The result was hundreds of groups of broader terms. So, the organization of terms was done in stages until we had the broadest terms, or the top terms, for each group. There are 217 top terms in this thesaurus.

Top Term ----- Movements
- Political movement
-- Liberation movement
--- Womens liberation movement
--- National liberation movement
- Religious movement
etc.

Fig 1. Hierarchy of terms subsumed under "Movement"

Identifying some 2400 terms in order to organize them into hierarchies was the most tedious part of the work. We managed to do this with the help of a number of reference books such as Broader System of Ordering (6), Occupational Classification and Industry Index (7), as well as various kinds of classification systems in addition to the thesauri which we previously used as sources terms.

We took the opportunity to reselect terms during the stages of terms organization. Some terms have been deleted or factored during these stages. If there was more than one term belonging to the same concept or meaning, the most frequently used one has been chosen to be descriptor and include in the hierarchy, while the rest was reserved for SEE reference, i.e. lead-in-terms.

C. Compiling the alphabetical thesaurus

The compilation of the alphabetical thesaurus was performed in two stages, namely 1) conversion of the hierarchical thesaurus into an alphabetical list of descriptors. The list shows the Broader Term (BT), the Narrower Term (NT), and the

Top Term (TT) for each descriptor, and 2) supplementing this alphabetical list with Scope Note (SN), Used For (UF), interfiling the lead-in-term for USE references, and identifying Related Term (RT) to be assigned to each descriptor.

The conversion stage was performed with the aid of MINISIS/HP 3000 computer under the auspices of the computer staff of the PDII. The hierarchical thesaurus was firstly loaded into the computer, using the EDITOR facility to make necessary modifications before it was ready for loading into the PDIITHES data base system.

```
!REC-ID
!T980! Movements
!T300! Liberation movement
!T100! Political movement
!T400! Womens liberation movement
!T400! National liberation movement
```

Fig 2. Modified hierarchy with address for each term

The modified hierarchical thesaurus was next loaded from the EDITOR file into the PDIITHES data base system, using BATCHIN processor. It was in this new system that the modification was once again performed to supply the main term (T100) with the necessary information such as Scope Notes (T940), Used For Term (T920), and Related Term (T500), or to post USE Term (T900) in the case of lead-in-terms.

The SN, UF, and USE terms were readily integrated into the PDIITHES data base, since they had been prepared during the organization of terms for the hierarchical thesaurus. The associative relationships among descriptors or RT, however, required another identification process before it was ready for integration. The associative relationships between terms was determined by following the guidelines given in Appendix 3.

T100 MAINE : Liberation movement
T980 TP'TERM : Movements
T920 USEFTE : Emancipation
T500 RELATE : Political participation
T300 BROAD : Political movement
T400 NARROW : Womens liberation movement
National liberation movement
T960 SCOPEN : Movement or action of trying to
achieve equal rights or status,
or to attain cultural, social,
political and economic indepen-
dence

Fig 3. The alphabetical thesaurus in PDIITHES data base upon completion of necessary information

Liberation movement
SN Movement or action of trying to achieve
equal rights or status, or to attain
cultural, social, political and economic
independence
UF Emancipation
TT Movements
BT Social movement
NT National liberation movement
Womens liberation movement
RT Political activities

Fig 4. The alphabetical display of the thesaurus

Preparing the permuted index

The permuted index was prepared with the help of a word processor. All descriptors included in the thesaurus are presented in the index, with all the component words of descriptors treated as an entry element (except stop-words) to

allow users to find multiword descriptors by looking for any one of its component words. The juxtaposition of similar component words within the index helps users to find descriptors that contain a given component word.

Evaluation

The performance of a thesaurus can best be tested through its use in actual indexing. There are no specific rules as to how many articles should be indexed before one can justify performance of the thesaurus used. Lancaster (8) given an example of evaluating thesaurus performance by indexing 100 articles. But we have our doubts that such a few articles would be sufficient to test a thesaurus of, say, 2000 descriptors. As a rule of thumb, we would say that the more articles that are indexed the more we know about the weaknesses and strengths of the thesaurus at hand.

In evaluating the Thesaurus on Women in Development, we propose that it should be tested against at least 600 articles picked at random from the universe of literature on women. Assuming that each of these articles requires an average of 4 descriptors to represent its subject content, the whole indexing will then give at least one chance for every descriptors to be consulted.

The test should be in a real indexing manner. The indexer first writes down the results of her conceptual analysis as a series of words or phrases representing the subject matter of the article indexed. Next, the thesaurus is consulted and an attempt is made to translate the conceptual analysis into descriptors (8). The translatability of the concept written down by the indexer into descriptors listed in the thesaurus which may occur in some level of translatability, is then measured.

CONCEPTS =====Translated into===== DESCRIPTORS

Abortifacients	Abortifacients
Women	Women in Islam
Islam	
Emancipation	Liberation movement
Social responsibility of parents	Social responsibility Parents responsibility
Indonesia	Indonesia
ASEAN Women's Programme	????????

Fig 5. Translation of concepts into descriptors

The performance of the thesaurus in terms of descriptor completeness and specificity can now be justified. Descriptor completeness is expressed in the presentage of translatable concepts at the required level of specificity. Eighty five percent of descriptor completeness and eighty percent of descriptor specificity might be considered good enough.

Conclusion

The completeness and specificity of descriptors are surely not the only measures of thesaurus performance, but they indicate the success of the thesaurus in indexing. Nevertheless, through examination of the thesaurus is strongly needed, involving every aspect of the thesaurus. This includes the examination of reciprocity of descriptor relationships, review of the hierarchical structure, and ease of use of the thesaurus.

We are proud of what we have achieve so far though we are not fully satisfied with the achievment. A lot more improvements are required before the thesaurus can satisfy the emerging need for standarized terms for information networking among AWP's Clearinghouse on Women in Development.

We are committed to making continuous improvement and updating of the thesaurus. Any suggestion to further our commitment are greatly welcome.

REFERENCES

- 1) ISO 2788-1974. Guidelines for the establishment and development of monolingual thesauri, in *ISO Standard Handbook 1*, 1977: 396-409.
- 2) Guidelines for the establishment and development of monolingual thesauri. --2nd and rev. ed., Paris: Unesco, 1981.
- 3) Gilbert, Valarie. "A list of thesauri and subject headings held in the Aslib Library", *Aslib Proceedings*, 31 (6) June 1979: 264-267.
- 4) Walkley, Janes and Hay, Barbara. "An annotated list of thesauri held in Aslib Library", *Aslib Proceedings*, 23 (6) June 1971: 292-300.
- 5) Hafner, Cornelia. *Bestandsverzeichnis thesauri*. Frankfurt : Documentationszentrum fuer informationswissenschaften, 1977. 14 p.
- 6) Coates, Eric (et al.). *Broad system of ordering*. The Hague: FID, 1978.
- 7) Occupational classification and industry index (Dictionary of occupational titles vol. 2. Washington, DC: US Department of Labor, 1979.
- 8) Lancaster, F.W. *Thesaurus construction and use: a condensed course*. Paris: Unesco, 1985. 89 p.
- 9) Capek, Mary Ellen S. "Wired words: developing an on-line thesaurus and data base for improving access to women's information resources", in *NWSA Perspectives*, 5 (2):
- 10) Capek, Mary Ellen S. *A Women's thesaurus: an index of language used to describe & locate information by & about women*.

Appendix 1

Women in Development identified in fourteen categories

- 1) Generalitis
- 2) Religion and Philosophy
- 3) Women in Social Development
- 4) Women in Political Development
- 5) Women in Economic Development
- 6) Legal Status, Law, etc.
- 7) International Relations
- 8) Population and Demography
- 9) Education of Women
- 10) Biological Condition of Women
- 11) Psychology
- 12) Literature and Arts
- 13) Science and Technology
- 14) History

Thesauri

- 1) Unesco Thesaurus: a Structured List of Descriptors for Indexing and Retrieving Literature in the field of Education, Social Science, Culture and Communication, compiled by Jean Aitchison. Paris, Unesco, 1977..
- 2) Macrothesaurus for Information Processing in the field of Economic and Social Development. Paris, OECD, 1978.
- 3) APAIS Thesaurus: a List of Subject Terms Used in the Australia Public Affairs Information Service. Canberra, NLA, 1980.
- 4) Population/Family Planning Thesaurus, by Carolina Lucos and Margaret Osburn. Chapel Hill, N.C., University of North Carolina, 1975.
- 5) Population/Fertility Control Thesaurus, by Helen K. Kolbe and Rita F. Bergman. Washington, D.C., Population Information GWU, 1976.
- 6) Thesaurus on Descriptors Used for Information Processing in the ILO Library, 2nd ed., Geneve, ILO, 1978.
- 7) U.S. AID Thesaurus. Washington, D.C., US AID, 1984.

Periodicals and Abstracts

- 1) Asian Women. Pakistan, Asian Women's Institute, 1976.
- 2) Women at Work. Geneve, International Labour Organization.
- 3) Women in Action. Rome, Isis International.

- 4) Women's Studies Abstracts. Rush, N.Y., Rush Publishing, 1972.
- 5) Women's Studies: an Interdisciplinary Journal. London, Gordon and Breach, 1978.
- 6) Women's Studies International Forum. Oxford, Pergamon Press, 1978.
- 7) Women's Studies Quarterly. New York, Feminist Press, 1972.
- 8) Women's World. Geneve, Isis-Wicce, 1976.
- 9) CWAS: Newsletter. Minneapolis, M.N., Committe on Women in Asian Studies, 1983.
- 10) Feminist Studies. College Park, University of Maryland, 1972.
- 11) INSTRAW News. Santo Domingo, INSTRAW.
- 12) Isis International Women's Journal. Rome, Isis International.
- 13) SIGN: Journal of Women in Culture and Society. Chicago, Ill., University of Chicago Press.
- 14) Women in Asian Newsletter. Salisbury East, Salisbury University.
- 15) Women in Southeast Asia: a Bibliography, by Fan Kok-Sim. Boston, G.K. Hall, 1982.
- 16) Bibliographic Guide to Studies on the Status of Women: Development and Population Trends. Paris, Unesco, 1983.

- 17) Women's Studies: a Recommended Core Bibliography, by Esther Stineman. Littleton, Colo., Libraries Unlimited, 1979.
- 18) Women's Studies: a Bibliography of Dissertations 1907-1982, by V.I. Gilbert and D.S. Tatla. New York, Basil Blachwell, 1985.

Appendix 3.

Guidelines for determining associative relationship

- 1) Associative relationship is only given for terms which are not subsumed under a Top Term.
- 2) Associative relationship is reciprocal.
- 3) It is commonly applied to two terms which demonstrate one of the following relationship:
 - a) Thing/application
 - b) Result/cause
 - c) Thing/property
 - d) Raw material/product
 - e) Complementary activities
 - f) Opposites
 - g) Activity/agent
 - h) Activity/product
 - i) Whole/part.