Preparation Terminological Definitions for Indexing and Retrieval Thesauri: A Model

Abstract: A model for standardizing existing definitions and/or writing new definitions for thesaurus descriptors has been developed, within the framework of a research project concerned with the usefulness of terminological definitions for indexers working with a thesaurus. The proposed model is an expansion of a model presented by Sager and L'Honome in 1994. Examples of its application in a thesaurus describing the field of Adult literacy programming and training are introduced.

1. Introduction

A critical function of the indexing and retrieval thesaurus is that of clarifying meaning, of revealing the relationship that exists between a concept and its expression in the controlled language of an information system. When the meaning of all descriptors is clear and unambiguous, indexers are expected to more consistently and accurately select the most precise and appropriate descriptor to express an indexable concept.

Contemporary thesauri, however, have become large and complex vocabulary control tools, and it is increasingly difficult for the indexer to assess the intended meaning of each one of the terms that constitute the entry vocabulary. Many descriptors appear to have very close if not overlapping meanings, and it is often not possible to distinguish among them and to identify correctly the term needed in a given indexing situation. The difficulty for indexers to elucidate meaning is felt even more in social science thesauri.

Traditional thesaural networks of semantic relationships and scope notes are considered as the main sources of descriptor definition. They are not always efficient and sufficient however, especially when the entry vocabulary in a thesaurus is large, and when semantic and terminological ambiguities exist in the field described.

The lack of proper defining information creates difficulties in using a thesaurus as indexing aid, and, more importantly, leads to inconsistencies in descriptor assignment, with assumed consequences on representational predictability, indexing correctness, and overall indexing quality.

The thesaurus is, and is likely to remain for years to come, an essential indexing tool, and it is important to consider at this time alternate ways of providing indexers with much needed semantic information about individual subject descriptors. Within the larger context of a study focusing on the potential usefulness of terminological definitions in the thesaurus used as indexing tool, this paper presents a model for the development of such semantic addition.

2. Definitions in Thesauri

The idea of providing definitions for descriptors has surfaced regularly for as long as thesauri have been in existence and use. Mooers (1963) believed that definitions were necessary to transform index terms into descriptors, giving them a very distinct meaning determined from context of use. Mooers excluded the possibility of using regular dictionary definitions.

Over the years, some real definitions, to be distinguished from scope notes, were provided for thesaurus descriptors. Definitions were either grouped in a separate section, or presented in
the main section of the thesaurus, as part of the defined term record. These definitions, however, have been of questionable quality and usefulness. There were never any precise guidelines available to help in the process of writing and standardizing them.

Recognizing obvious links between terminology work and thesaurus design, interesting proposals for developing terminological thesauri came from researchers in the field of Terminology. Among others, Sager (1982 and 1990) envisioned a semantically rich tool which would be useful to a wide audience of specialists interested in information production and communication. In a terminological thesaurus, each descriptor record would include a standardized definition, as well as semantic relationships with other terms.

3. Terminological Definitions: Nature and Functions

Within the field of Terminology, the objective of describing accurately the meaning of a concept is normally pursued through the development of one or more conceptual or true definitions for each term in the language of a specialty (LSP). The terminological definition is domain dependent; it describes the intension of a concept and it determines its position in the specific system of concepts to which it belongs.

Although many types of acceptable terminological definitions have been identified and used, analytical definitions, which define a concept by identifying its genus proximus and its differentia specifica, are seen as the necessary foundation of most terminological systems, becoming an essential component of terminologies and termbanks.

4. Defining Rules and Models

The primary defining rules are well-known. The rules of adequacy, completeness, and substitution, as well as comments on tautology and circularity, refer to the semantic contents of the definiens. The rules of brevity, simplicity, clarity, and comments on the use of negative forms and synonyms, refer to the style of the defining text.

It appears that, even in termbanks, definitions often leave much to be desired in clarity and adequacy. The rules are not sufficient to ensure that the semantic content of definitions is sound, complete and appropriate.

Although the need for defining templates, which would call for and organize in a standard order of appearance the various elements of a well-formed analytical definition has long been recognized, few of the defining patterns that were presented were formal enough to be of much help as content and style regulators.

Sager and L'Homme have recently proposed a model for standardizing existing term definitions that "consists of a regularized form of the traditional analytical definition by categorizing and restricting the modes of description and thereby reducing the free-text element in the defining phrase" (1994, 352). Sager and Ndi-Kimbi are also searching for "more sensible rules and proper patterns for definitions that can be used for authoring packages for the writing of definitions" (1995, 72).

The model for rewriting existing definitions is an uncomplicated one. It divides the defining phrase into seven components, as shown in Fig. 1.

The model does not offer a solution to the problem of identifying the characteristics which are necessary and sufficient in a well-formed definition. The choice of characteristics to be included in (6) above will depend on what type of concept is being


**Definiendum** =

1. Subject-field attribution
2. Concept class of *definiendum*
3. Defining concept or *definiens* (i.e., the *genus proximus*)
4. Concept class of *definiendum*
5. Relationship of *definiendum* and *definiens*
6. Essential distinguishing characteristics of the *definiendum* (i.e., the *differentia*)
7. Non-essential characteristics

---

Fig. 1: The structure of a standardized definition (Sager and L'Homme, 1994)

---

Defined, the concept system that it is part of, and the needs of the user. It has been suggested that one way to construct an effective *differentia* would be to answer one or more of the news reporter's question: Who? What? When? Where? Why? and How? (Strehlow, 1983). Dahlberg (1989) has proposed a similar approach.

---

5. Description of a Corpus

Our test thesaurus is made of 361 descriptors extracted from an existing and widely used thesaurus in the field of Adult literacy. Of these descriptors representing core concepts in the field, a large majority (275) refer to entities/objects. 259 descriptors are Level 2 terms (i.e., a precoordination of two concepts which may or may not be part of the thesaurus).

The conceptual structure and terminology of the field of Adult literacy illustrate most of the semantic and linguistic problems inherent to the social sciences. Most terms used by Literacy specialists are borrowed from the neighbouring fields of Education and Adult education; the rest comes from a large number of disciplines (Linguistics, Psychology, Sociology, etc.) As is common practice in the social sciences, the meaning of most borrowed terms has been slightly modified, and often more than once, to reflect new general points of view or even personal opinions. Through a process of unregulated precoordination, terms have been created on demand, and lack of standardization now leads to communication of information difficulties among specialists, which a thesaurus in the area is expected to fix. But since series of terms with obviously overlapping meanings have made their way into the thesaurus (e.g., *literacy consultants, literacy coordinators, literacy facilitators, literacy instructors, literacy practitioners, literacy tutors, literacy workers*), it looks as if nothing short of a verbal definition will be of any help to the indexer of literacy-related information in the process of index term selection.

---

6. Expanding Sager and L'Homme's model

Using Sager and L'Homme's defining model as a tool for transforming existing definitions, or for writing new definitions for all descriptors in our corpus, this project develops section (6) of the model with a view to standardizing the nature and order of conceptual characteristics that will be emphasized and considered sufficient for the completeness of definitions within our context.

The expanded model now being tested is illustrated in Fig. 2. The template shows how
Section (6) has been subdivided further, using seven facets (eC1 to eC7) which we consider as being of primary importance in establishing clear distinctions among concepts. The model allows for great specificity in the description of the various types of concepts (entities, activities, properties) that are represented in our thesaurus.

Table:

"Source definition(s)"

(1) Domain:
(2) Concept class:
(3) Genus:
(4) Genus class:
(5) C/G:

(6) eC1: [nature: being..., having..., doing...]
(6) eC2: [purpose/function]
(6) eC3: [means/instrumentation]
(6) eC4: [origin]
(6) eC5: [destination]
(6) eC6: [place]
(6) eC7: [time]

(7) iC1:

Fig. 2: Proposed defining template for an entity/object (type of concept)

7. Application of the Expanded Model

The proposed model is used to add semantic information in a thesaurus which describes a limited domain and which we consider as a logical conceptual system. In this thesaurus, all definitions will be interdependent. The use of the templates requires prior categorization of all concepts represented, as well as a thorough search for existing definitions of all concepts and/or terms.

In preparing definitions for all 361 descriptors in our corpus, three situations have been encountered:

a) one or more appropriate definitions already exist for a concept, and all that needs to be done is to standardize their content and style. A standardized definition for library literacy programs is provided as an example of this situation in Fig. 3. Note that the genus proximus, literacy programs, is itself defined in the same corpus and will be marked as an entailed term in the definitions of more specific concepts.

b) one or more inadequate definitions exist, and they can be used as a basis for the creation of an adequate standardized definition. A standardized definition for literacy specialists is provided as an example of this case in Fig. 4. Note common defining mistakes in the source definitions: the first definition shows an inappropriately restrictive definiens, while the second definition defines the term by itself.

c) no definition has been obtained for a specific concept; a standardized definition is
Term: library literacy programs

Source definitions: = "programs designed, administered, and staffed by a library"

(1) Domain: Literacy
(2) Concept class: AE
(3) Genus: literacy programs
(4) Genus class: AE
(5) C/G: a type of
(6) eC4: designed, administered, and staffed by libraries
(7) iC1: usually conducted in a library setting

library literacy programs = literacy programs designed, administered, and staffed by libraries, and usually conducted in a library setting.

Fig. 3: Standardization of an existing adequate definition

developed using elements of adequate existing definitions of the more general constituting concepts, or using elements of adequate definitions of specific concepts with which the definiendum shares essential characteristics. A standardized definition for the descriptor literacy centres is provided in Fig. 5 as an example of such a situation. In this example, the specification of the genus proximus is made necessary by the fact that the natural genus, Centre, has no meaning attached to it.

Term: literacy specialists

Source definitions: = "individuals involved in literacy research"
"person who is specialized in literacy"

(1) Domain: Literacy
(2) Concept class: MEA
(3) Genus: [Individuals]
(4) Genus class: MEA
(5) C/G: a subset of
(6) eC1: who have extensive theoretical and practical knowledge of issues pertaining to literacy assessment, literacy instruction, and literacy programming

literacy specialists = individuals who have extensive theoretical and practical knowledge of issues pertaining to literacy assessment, literacy instruction, and literacy programming.

Fig. 4: Modification and standardization of existing inadequate definitions
Concept: literacy centres

Source definitions:
"centres = facilities serving as a focal point for activities or services"
"community centres = facilities at which social, educational, recreational and other activities are held for the benefit of the community"

(1) Domain: Literacy
(2) Concept class: MEI
(3) Genus: [Centres]
(4) Genus class: MEI
(5) C/G: a

(6) eC1: physical facilities
(6) eC2: at which a variety of literacy services are provided
(6) eC2: at which literacy related activities such as literacy classes and literacy workshops are held

literacy centres = physical facilities at which a variety of literacy services are provided, and at which literacy related activities such as literacy classes and literacy workshops are held.

Fig. 5: Creation of a standardized definition

Terminological definitions are not complete definitions. Their only function is to make clear what difference there is between two or more concepts in a particular domain or LSP, and only the required essential characteristics needed to do so will be used in the description of the concept. The standardized definitions given here as examples show that prior knowledge is required from their reader. The meaning of terms like "physical facilities", "individuals", or "libraries" is considered common knowledge; these terms, which do not belong to the same LSP as the term which they are used to define, are not themselves defined in the thesaurus.

8. Conclusion

Our application of the proposed expanded defining model has shown that, even in a field such as Adult literacy where the terminology is far from being standardized and has a tendency to vary significantly from one geographical area to another, and from one social environment to another, it remains possible to describe concisely through a verbal definition the semantic coverage of a descriptor, adding in this manner to the semantic information already available to the indexer to assist in term selection.

The same defining model could also be used to modify definitions when the meaning of a concept changes over the normal course of evolution of a field of knowledge, or under the influence of new points of view, new connotations, etc.
Notes

1. In the *NASA Thesaurus*, for example, the grouped definitions constitute a *thesaurus dictionary*, designed as a complement to the thesaurus rather than as one of its parts.

2. This has been, and still is, the most common practice in thesauri. Definitions appear as scope notes. Scope notes then play the dual role of describing the intension of concepts and providing rules for using the terms that represent them.

3. Analytical definitions are also known as true, generic, Aristotelian, and referent-oriented definitions.


5. The *Canadian Literacy Thesaurus / Thesaurus canadien d'alphabétisation*. Toronto, ON: Canadian Literacy Thesaurus Coalition, 1992.

References


