

## TOWARD A THESAURUS OF RADIOCARBON DATING AND RELATED TERMS

DILETTE POLACH

Radiocarbon Dating Research, PO Box 43, Garran, ACT, 2605, Australia

**ABSTRACT.** The development of a *Thesaurus of Radiocarbon Dating and Related Terms* was undertaken to provide a basic common vocabulary of terms for the purpose of indexing an *Annotated Bibliography of Radiocarbon Dating (1948-68)* and form the basis of a language for information storage and retrieval in  $^{14}\text{C}$  dating and related term areas. The interdisciplinary nature of  $^{14}\text{C}$  dating dictates that selected terms have meanings in the context of archaeology, geology, earth sciences, oceanography, and environmental sciences, to name but a few, as well as chemistry and nuclear instrumentation on which the technique of  $^{14}\text{C}$  determinations is based. This paper describes the method used for selection of terms and explains the structure of the proposed thesaurus. A limited edition *Draft Thesaurus* was distributed to a selected panel of colleagues for critical evaluation.

### INTRODUCTION

An Information Storage and Retrieval system (ISAR) is designed to store information in a form suitable for a specific purpose (documents, bibliographic data, research projects, data elements, etc, hereafter called documents) and retrieve or select the information stored in the form required by the user of the system. Two main requirements must be met to obtain satisfactory retrieval.

1) The documents in store must be organized according to a properly conceptual structure. This enables, at the indexing stage, to characterize the document adequately at the appropriate level. In the same way it enables, at the retrieval stage, to search the store from a broad spectrum of aspects so that the resulting retrieval will be comprehensive, yet free of irrelevant material.

2) There needs to be terminologic control, both at the indexing and searching stage. The control will also ensure that the indexer or searcher is led to the appropriate term, no matter which variant for the concept first came to mind. If the conceptual structure and the terminologic control are inadequate, the end result will be either under-retrieval (relevant documents missed) or over-retrieval (non-relevant documents retrieved) or both. Further, as the size of the store as well as the number of queries increase, losses due to poor performance become even more significant.

These two principles apply to all kinds of ISAR systems, manual, mechanical or electronic, such as library catalogues, punched card files, or computer data bases.

The list of terms that describe the information contained in a document or a file of documents forms the index to that document or file. For the choice of analytical and abstract terms (hereafter called descriptors) a Thesaurus is used, which is central to an ISAR system.

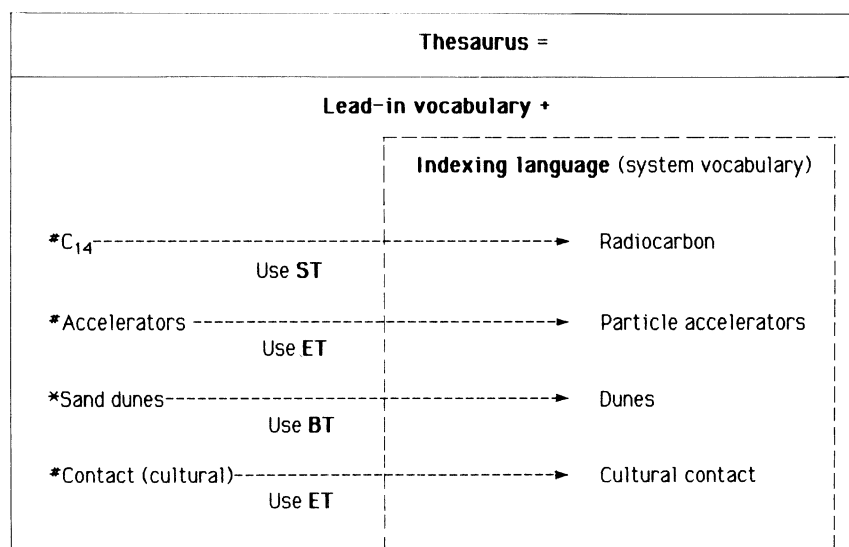
### THE ROLE OF A THESAURUS OF $^{14}\text{C}$ DATING

A thesaurus consists of an indexing language (list of descriptors) presented as a structured vocabulary. Included in this structure are all the relationships between descriptors, and a lead-in vocabulary which leads from

terms not used as descriptors to the appropriate descriptors in the indexing language. The role of this structure is four-fold (Fig 1):

- 1) It gathers under one chosen (preferred) term all the synonyms for the same concept—this may also include spelling variants.
- 2) It removes the ambiguity of homographs by adding qualifiers.

1. <i>Synonym-homonym structure</i>	Relates <b>terms</b> to <b>concepts</b>
2. <i>Equivalence structure</i> (quasi-synonyms or equivalent terms)	Groups together widely overlapping concepts resulting in " <b>ISAR-concepts</b> " expressed by <b>preferred terms</b> (1-1 correspondence)
3. <i>Classificatory structure</i>	Made up of <b>ISAR concepts</b> as expressed by preferred terms, and their interrelationship. May refer from an ISAR concept not used for indexing and searching to the appropriate descriptor(s).
4. <i>Indexing language</i>	Made up of descriptors, i. e. preferred terms designating ISAR concepts that actually are used in document representations and search request formulations.



Legend: **ST** Synonymous term                      \*) Not a preferred term  
**ET** Equivalent term                            \*) Preferred term but not a descriptor  
**BT** Broader term

Fig 1. Summary of definitions (based on Soergel, 1974, Fig 6A)

3) It puts each descriptor within a hierarchical class structure, enabling the user to choose the degree of specificity desired.

4) It leads to other descriptors not related hierarchically but related by association or by close meaning without being synonymous.

A thesaurus is most useful to those directly responsible for indexing and retrieving literature: authors, editors, indexers, abstractors, documentalists, librarians, and users. For the author/editor team, the thesaurus facilitates the selection of more meaningful terms for titles, as keywords for indexing, for abstracts and for concepts discussed in the author's publication. For the indexer/user team, the thesaurus enables the selection of the most appropriate term at the desired level of specificity to be used as descriptor for each concept in the index and/or to phrase an inquiry.

When a new discipline emerges, articles are usually published at first in generalist journals. As the discipline grows, so do the number of specialized journals. Monographs, reviews, reports, treatises, etc are written, conferences are held and their proceedings are published. Finally, it may be decided that a data base or bibliography are needed to make the growing literature widely available. In this case vocabulary control becomes necessary.

In the case of a very specialized field, existing thesauri will lack the degree of specificity demanded. If the field is multidisciplinary, the vocabulary may be scattered in various thesauri or dictionaries. The use of several sources for selecting index terms will lack the consistency necessary for good indexing and retrieval.

#### A THESAURUS FOR RADIOCARBON DATING

A study of the  $^{14}\text{C}$  dating literature made on two sets of data, covering the years 1948–1968 and 1978–1980, shows that  $^{14}\text{C}$  dating fits the criteria for special bibliographic treatment. Its rapid growth has been documented (Polach, 1980, Fig 1). Table 1 illustrates multidisciplinaryity.

The data file 1978/80 was compiled from computer retrieval of several DIALOG<sup>TM</sup> data bases. A wide variety of indexing practices were noted, ranging from a well-structured vocabulary (*Geoarchives*) to keywords from titles (*Chemical Abstracts*, *Conference Abstracts*, *Dissertation Abstracts*), to wide ranging descriptors (*INSPEC*, *INIS*).

The index of the bibliography *Twenty years of Radiocarbon Dating, an*

TABLE 1  
Radiocarbon dating literature: articles published in journals, by disciplines

Disciplines	% of total	
	1948/68	1978/80
Science/technology	26.5	27.9
Geosciences	36.0	45.0
Archaeology/Anthropology	24.9	17.8
Others	12.6	9.1
Total (numbers)	100.0 (2629)	100.0 (698)

annotated bibliography, 1948–68 (pilot study) (Polach, 1979) was compiled using *ad hoc* terms from the natural language of the literature indexed. There was no attempt to introduce a structure to the vocabulary or even to achieve consistency. The desire to improve on this effort, to index a further 1600 odd items, and to provide a starting point for terminologic control of an eventual data base, led to the compilation of a draft thesaurus.

In compiling the thesaurus draft several shortcuts had to be made as this was a one-person effort with no financial support. The documentation was kept to a minimum and most of the vocabulary was culled from existing thesauri. The criteria for inclusion were as follows:

- 1) Terms related to the literature
- 2) Expanded terms within the following bounds
  - a) for geologic and archaeological terms, the Pleistocene and Holocene epochs
  - b) for other scientific and technical terms, current terminology of  $^{14}\text{C}$  dating laboratories used in the literature.

Two thesauri which were in part relevant to  $^{14}\text{C}$  dating were examined: *Thesaurus of Engineering and Scientific Terms (TEST)*, and *Australian Thesaurus of Earth Sciences and Related Terms*.

As no thesaurus relevant to archaeology was found, the terms in the literature were supplemented by terms selected from the *Penguin Dictionary of Archaeology* (Bray & Trump, 1982) and a structure devised.

A list of terms used in the Bibliography of Radiocarbon Dating (Polach, 1979) and in the data file 1978–80 (see above) was made. This came to 1875 unique terms. Each term was then checked against the relevant thesaurus. The structure in which it, or an equivalent term, appeared was then used in totality or modified to fit in with the thesaurus criteria. Reciprocal and related terms in their turn were submitted to the same process. This resulted in a list of 3354 terms (descriptors and non-descriptors) meshed together by a class and association relationship and representing all disciplines. The result was input on a word processor. The thesaurus as such is capable of modifications and additions.

#### SUMMARY AND CONCLUSIONS

The role of a thesaurus in an Information Storage and Retrieval system is to enable efficient retrieval by providing terminological control in, 1) bringing synonyms together, 2) clarifying homographs, 3) offering a wider choice of terms through class and association relationships. A strong case for the compilation of a thesaurus for the terminologic control of the  $^{14}\text{C}$  dating vocabulary is substantiated. A decision needs to be made to incorporate a thesaurus into a data base of  $^{14}\text{C}$  dating or to publish it separately.

#### ACKNOWLEDGMENTS

Special thanks are due the panel of colleagues who accepted to critically evaluate the *Draft Thesaurus* and to Henry Polach, Australian National University for encouragement and advice. The figure was prepared by Winifred Mumford.

## REFERENCES

- Australian thesaurus of earth sciences and related terms, 1976: Adelaide, Australian Mineral Foundation.
- Bray, W and Trump, D, 1982, Penguin dictionary of archaeology, 2nd ed: London, Penguin Books.
- Polach, D, 1979, The first twenty years of radiocarbon dating. An annotated bibliography, 1948–68 (pilot study): Canberra, Radiocarbon Dating Research Lab, Australian Natl Univ.
- 1980, The first twenty years of radiocarbon dating. An annotated bibliography, 1948–68; a pilot study *in* Stuiver, M and Kra, RS, eds, Internatl <sup>14</sup>C conf, 10th, Proc: Radiocarbon, v 22, no. 3, 997–1004.
- 1985, Draft, Thesaurus of radiocarbon dating and related terms: Garran, Australia, Radiocarbon Dating Research.
- Soergel, D, 1974, Indexing language and thesauri construction and maintenance: Los Angeles, Melville Pub.
- Thesaurus of engineering and scientific terms (TEST), 1967: New York, Engineers Joint Council.