

I.F. THESAURUS OF BUILDING SCIENCE AND TECHNOLOGY  
(Thésaurus I.F. - Sciences et Technologie du Bâtiment)

Colin H. Davidson  
Faculté de l'Aménagement  
Université de Montréal

ABSTRACT

The I.F. Team have been struggling with the problems of information storage and retrieval in building, to the point that the basic tools had to be developed. These tools had to be explicit in their structure and construction, to compensate for the diversity of the building industry audience. The first of these tools is a hierarchical thesaurus, distinguished by its structure (a semi-lattice with nine levels) and its construction (based on a set of logical propositions). Candidate terms are treated in pairs, using a set of questions to ascertain systematically what are the relationships between them. Data processing helped with the construction of the thesaurus, particularly in terms of methodology. There is a scope for further work, particularly in preparing the bilingual (English-French) version of the thesaurus, now in draft.

(L'équipe I.F. a lutté avec les problèmes de l'enregistrement et de la recherche de l'information dans le domaine de la construction; fort de cette expérience, les membres de l'équipe se sont rendu compte qu'il fallait développer les outils de base. Ces outils devraient avoir une structure et une construction claires, afin de compenser la diversité de l'audience; les membres de l'industrie du bâtiment. Le premier de ces outils est un thésaurus hiérarchique, sa structure (un semi-réseau à 9 niveaux) et sa construction (basée sur une série de propositions logiques) représentent des innovations. Des termes candidats sont traités par paires, en se servant d'une série de questions afin d'identifier systématiquement la nature des liens entre les termes. L'informatique a aidé avec la construction du thésaurus, surtout en ce qui concerne la méthodologie. Il y a de grandes possibilités de continuer le travail, surtout pour la préparation d'un thésaurus bilingue (anglais-français) actuellement à l'étude).

ACKNOWLEDGEMENTS

The following people were involved with the project described in this paper: Michel Jullien (chief researcher), Pierre Garneau (responsible for data processing), Roger Camous, Claude Diacon, John Roberts, Leonard Wert.

## I.F. THESAURUS

### INTRODUCTION

This paper describes work that has been carried out by the "I.F. Team" at the Faculté de l'Aménagement, Université de Montréal. For reasons that are described below, we decided to develop a Thesaurus in the domain of building science and technology; for reasons that are also explained below, this lead to a number of special problems, the solutions to which - we believe - are of general significance for Thesaurus construction.

### THE CONTEXT

Members of our team had been struggling with problems of information storage and retrieval in building for a while; the frustrations of trying to use the existing classification systems (devised to deal with bricks and mortar and quite unsuited for environmental phenomena or building processes) need not be described here. In addition, we were publishing a quarterly magazine (called Industrialization Forum (Ref.1) - whence our name "I.F. Team") which had a built-in coordinate indexing system - complete, but for a controlled vocabulary.

The nearest thesaurus was E.J.C. (Ref.2), but attempts to use it showed that (i) it did not cover our domain adequately, and (ii) it seemed that its structure was not appropriate - particularly in its liberal use of the related term (RT) cross references. A preliminary enquiry showed that one other English language thesaurus existed in the domain of building, - a loosely structured working document prepared by the Building Research Station in the U.K. (Ref.3); two were about to be prepared - one concerning building products in Canada (Ref.4) and a faceted thesaurus being considered in the U.K. (Ref.5). A French thesaurus, dealing with construction technology, had also been drafted (Ref.6).

### THE PROBLEM

We were concerned with a domain of activity (building - science and technology) and not a "discipline"; an immediate consequence of this was that the people in the domain of activity have varied backgrounds, functions, vocabularies... to the point that communications were already difficult because of varied usages of terms. As a result, we were going to have to use every conceivable device - in the construction and presentation of our Thesaurus - to help standardize the use of terms.

### THE I.F. THESAURUS

A thesaurus is a controlled and dynamic vocabulary, which covers a particular domain of knowledge and activities adequately. A thesaurus can be characterized by:

- (i) its domain:
  - particular domain,
  - neighboring domains,
  - degree of depth of coverage of these domains,
  - extent of coverage of these domains;
- (ii) its structure:
  - internal structure,
  - compatibility with other thesauri;
- (iii) its different functional parts, (figs. 1 and 2);
- (iv) its physical form;
- (v) its method of use;
- (vi) its process of evolution.

The I.F. Thesaurus can be characterized against each of these criteria. However, this paper emphasizes aspects of: (i) its structure, and: (ii) its process of evolution.

#### The Structure.

Because of the nature of the building industry "public" (varied, with no natural standardization of terms), we felt the need to adopt a "scientific" approach to structuring and constructing the thesaurus. By "scientific" we meant that there should be clear rules against which any and all decisions could be made; these same rules could be referred to in any subsequent argument about the outcome of these decisions. The more rigorous the scientific approach, the better we would be compensating for the "undisciplined" (pluri-disciplined?) nature of our public.

The structure we adopted is rigorously hierarchical. Each descriptor has a defined level in the hierarchy (except for some general terms (Ref.7)) and each descriptor has relationships to descriptors one level "up", one level "down" or on the same level in the hierarchy. The structure as a whole (a semi-lattice, with SCIENCES and ENVIRONMENT at the top level - level 1) is built up by systematic accretion of these individual relationships.

The relationships (which are registered by cross-references) were selected to avoid some of the uncertainties that prevail in many existing thesauri (Ref.8). These are: Broader Term (BT) and its reciprocal Narrower Term (NT) - restricted to the meaning: "type of"; Whole Term (WT) and its reciprocal Part Term (PT) - restricted to the meaning: "part of"; Related Term (RT) - restricted to the meaning: "associated with". The BT/NT and WT/PT relationships exist - and only

EXPERIMENTAL PSYCHOLOGY	***	FACING BRICKS	*7*
...USE PSYCHOLOGY		FACINGS	***
EXPERIMENTS	*0*	...USE CLADDINGS	
EXPERTISE	***	UTILIZATION FACTOR	*6*
...USE SKILL		ZONAL FACTOR METHOD	*6*
EXPLOSIONS	*6*	FACTORIES	***
EXPLOSIVES	*6*	...USE INDUSTRIAL FACILITIES	
EXPONENTS	*5*	FACTORS	*0*
EXPOSED CEILINGS	*7*	DANGER FACTORS	***
EXPOSED CONCRETE FINISHE	*7*	...USE HAZARDS	
EXPOSURE	*0*	WEIGHTING FACTORS	*0*
EXPRESSION	*0*	FAADING	***
EXTERIOR LIGHTING	*6*	...USE DECREASE	
EXTERIOR SPACES	*5*	FAIENCE	*7*
EXTERNAL ENVELOPES	*6*	FAILURES	*0*
EXTERNAL FORCES	***	FAIR GROUNDS	***
...USE LOADS(FORCES)		...USE LUNA PARKS	
FIRE EXTINGUISHERS	*8*	FAIRS	*6*
FIRE EXTINGUISHING EQUIP	***	FALL-OUT SHELTERS	***
...USE FIRE FIGHTING EQUIPMENT		...USE ATOMIC SHELTERS	
EXTRACT VENTILATION	*7*	FALL(AUTUMN)	***
EXTRACTION(CHEMISTRY)	*6*	...USE AUTUMN	
EXTRACTS(DOCUMENTATION)	*6*	WATER FALLS	*6*
EXTRAPOLATION	*5*	FALSE DROPS	*7*
EXTRUDING	*6*	FAMILIES	*4*
EXTRUSIONS	***	FANLIGHTS	*7*
...USE SECTIONS(MATERIALS)		FANS	*8*
EYE WINDOWS	*7*	FAR EAST	*7*
FABRICATION	***	FARM HOUSES	*6*
...USE MANUFACTURING PROCESSES		FARMS	*6*
FABRICATORS	***	FASHION	*4*
...USE MANUFACTURERS		MODE(FASHION)	***
FABRICS	*6*	...USE FASHION	
FACADES	*6*	MECHANICAL FASTENERS	*6*
FACADES	*7*	BOLTS(FASTENERS)	*7*
FACES	*7*	SPACEPS(FASTENERS)	*7*
FACES	*7*	FASTENING	***
FACETED CLASSIFICATION I	*7*	...USE JOINING(PROCESS)	
FACIL	*5*	FAT CONCRETE	*7*
ADMINISTRATION FACILITIE	*5*	FATIGUE LIFE	***
PUBLIC SERVICE FACILITIE	*5*	...USE FATIGUE RESISTANCE	
FACILITIES	*4*	FATIGUE LIMIT	*7*
AGRICULTURAL FACILITIES	*5*	FATIGUE RESISTANCE	*6*
CATERING FACILITIES	*5*	FATIGUE STRENGTH AT N CY	***
CIVIL DEFENSE FACILITIES	*5*	...USE FATIGUE RESISTANCE	
COMMERCIAL FACILITIES	*5*	FATIGUE(MATERIALS)	***
COMMUNITY FACILITIES	*5*	...USE FATIGUE RESISTANCE	
CULTURAL FACILITIES	*5*	FAUCETS	*8*
JWELLING FACILITIES	***	FEASIBILITY	*0*
...USE RESIDENTIAL FACILITIES		FEASIBILITY STUDIES	*5*
EDUCATIONAL FACILITIES	*5*	GLACIAL FEATURES	***
ENTERTAINMENT FACILITIES	***	...USE HYDROGRAPHIC FEATURES	
...USE RECREATIONAL FACILITIES		HYDROGRAPHIC FEATURES	*5*
EXHIBITION FACILITIES	*6*	TOPOGRAPHIC FEATURES	*4*
HEALTH FACILITIES	*5*	FEDERAL SCOPE	***
INDUSTRIAL FACILITIES	*5*	...USE NATIONAL SCOPE	
MAINTENANCE FACILITIES	***	FEEDBACK	*5*
...USE STORAGE FACILITIES		NEGATIVE FEEDBACK	*6*
MILITARY AIR FACILITIES	***	POSITIVE FEEDBACK	*6*
...USE MILITARY FACILITIES		THRESHOLD OF FEELING(SOU	*7*
MILITARY FACILITIES	*5*	FEELINGS	***
MOBILE FACILITIES	*5*	...SEE SCOPE NOTE	
MOTOR RACING FACILITIES	*6*	FELTS	*6*
MULTI-PURPOSE FACILITIES	*5*	FENCES	*7*
NAVAL FACILITIES	***	FENESTRATION	*6*
...USE MILITARY FACILITIES		FERMENTATION	*6*
OLYMPIC FACILITIES	*6*	FERROUS METALS	***
RECREATIONAL FACILITIES	*5*	...USE METALS	
RELIGIOUS FACILITIES	*5*	FERTILIZERS	*7*
RESEARCH FACILITIES	***	FIBER	*6*
...USE SCIENCE FACILITIES		FIBER GLASS	***
RESIDENTIAL FACILITIES	*5*	...USE GLASS FIBER	
SCIENCE FACILITIES	*5*	FIBERS	*6*
SPORTS FACILITIES	*5*	FIBROUS AGGREGATES	*7*
STORAGE FACILITIES	*5*	FIELD TESTS	*0*
TEMPORARY FACILITIES	*5*	FIELD THEORY(MATHS)	*4*
TERMINAL FACILITIES	*5*	ABELIAN FIELDS	*5*
TOILET FACILITIES	***	ELECTRIC FIELDS	*6*
...USE TOILET SPACES		ELECTROMAGNETIC FIELDS	*5*
WELFARE FACILITIES	***	ELECTROSTATIC FIELDS	***
...USE HEALTH FACILITIES		...USE ELECTRIC FIELDS	
BANKS(FACILITIES)	*6*	MAGNETIC FIELDS	*5*

Figure 1: Typical page of Alpha-Permuted Index.

(E6) Figure 2: Typical page of Alpha-Hierarchical Index.

exist - between descriptors on one level in the hierarchy and appropriate descriptors one level "up" or one level "down" the hierarchy; The RT/RT relationships exist - and only exist - between descriptors on the same level of the hierarchy (Ref.9).

This set of relationships is shown graphically in fig.3 and typographically (as used in our alpha-hierarchical index) in fig.4. These "hierarchical groups" as has been stated - are joined by systematic accretion to form the whole structure or "conceptual space".

Before dwelling on the importance of this rigorous structure, we wish to describe the procedures that were used to generate the individual hierarchical groups. The key for this construction work is a set of so-called "logical propositions".

Obviously there are preliminary stages in which candidates are collected, their usage evaluated and their meanings agreed; once this has been done a preliminary sortation is made (which used certain computerized techniques) making clusters of candidates that apparently were likely to be related. At this moment, terms in these clusters were taken - two by two - and "processed" with the logical propositions (fig.5).

Let us consider, for example, some of the descriptors in the hierarchical group TOWNS (figs. 3 and 4).

(i) Let I be NEW TOWNS and J be TOWNS

P 1: NEW TOWNS can be a type of TOWNS - true  
P 3: NEW TOWNS is always a type of TOWNS - true

therefore: NEW TOWNS is a NT of TOWNS

(ii) Let I be BUILT-UP AREAS and J be TOWNS

P 1: BUILT-UP AREAS can be a type of TOWNS - false  
P 2: TOWNS can be a type of BUILT-UP AREAS - true  
P 4: TOWNS is always a type of BUILT-UP AREAS - true

therefore: BUILT-UP AREAS is a BT of TOWNS

(iii) Let I be STREETS and J be TOWNS

P 1: STREETS can be a type of TOWNS - false  
P 2: TOWNS can be a type of STREETS - false  
P 5: STREETS can be an element, a sub-set, a sub-system or an aspect of TOWNS - true  
P 7: if STREETS exist then TOWNS exist - true  
P 8: if TOWNS exist then STREETS exist - true

therefore STREETS is a PT of TOWNS (PT<sub>s</sub> meaning immediately part of it).

Figure 3: GRAPHIC DISPLAY OF THE HIERARCHY

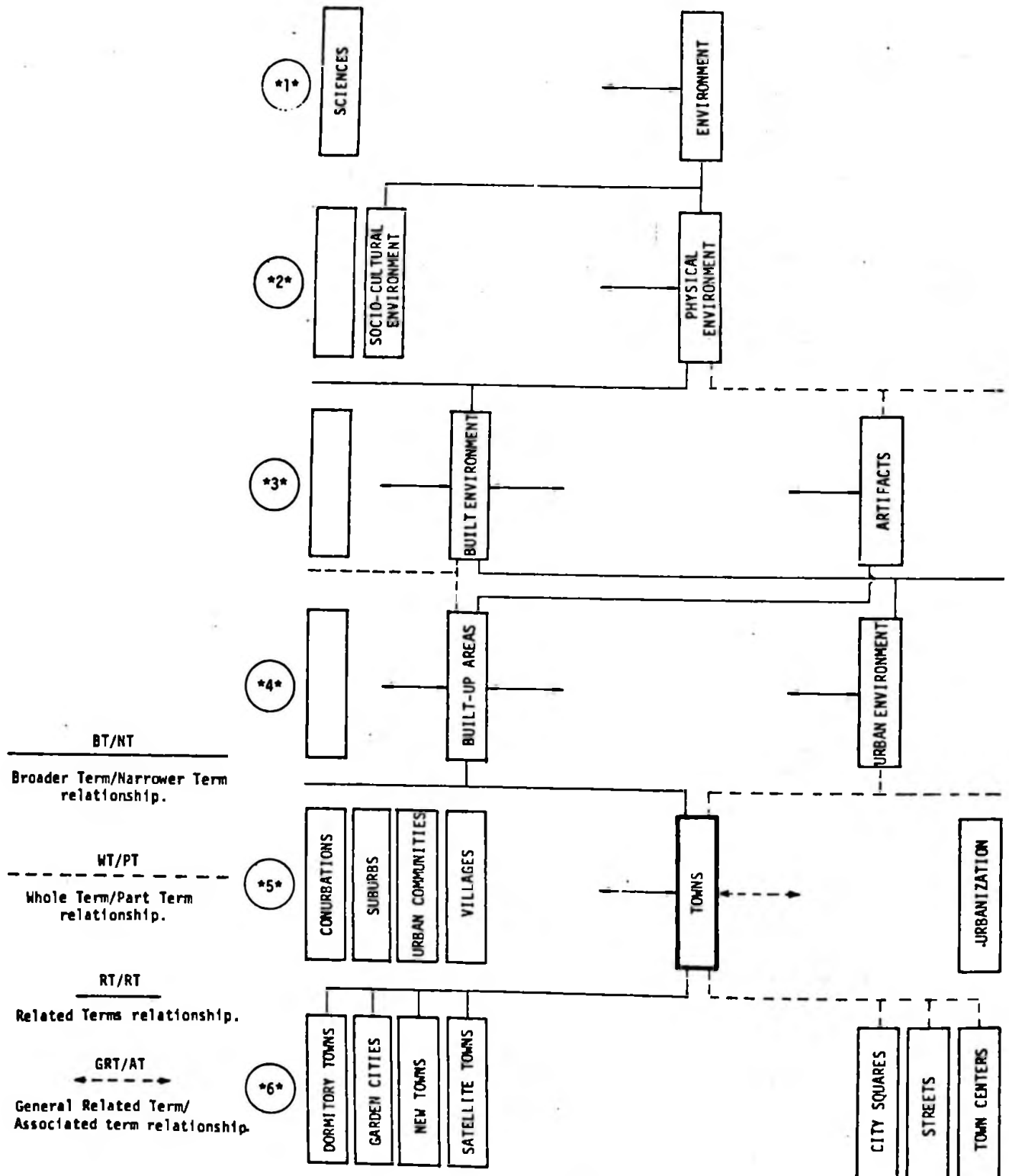


Figure 4: EXAMPLES OF MAIN ENTRIES IN THE ALPHA-HIERARCHICAL LIST

H i e r a r c h i c a l   D e s c r i p t o r

(This example is based on an existing group in the thesaurus, which has been completed, for the purposes of this section.)

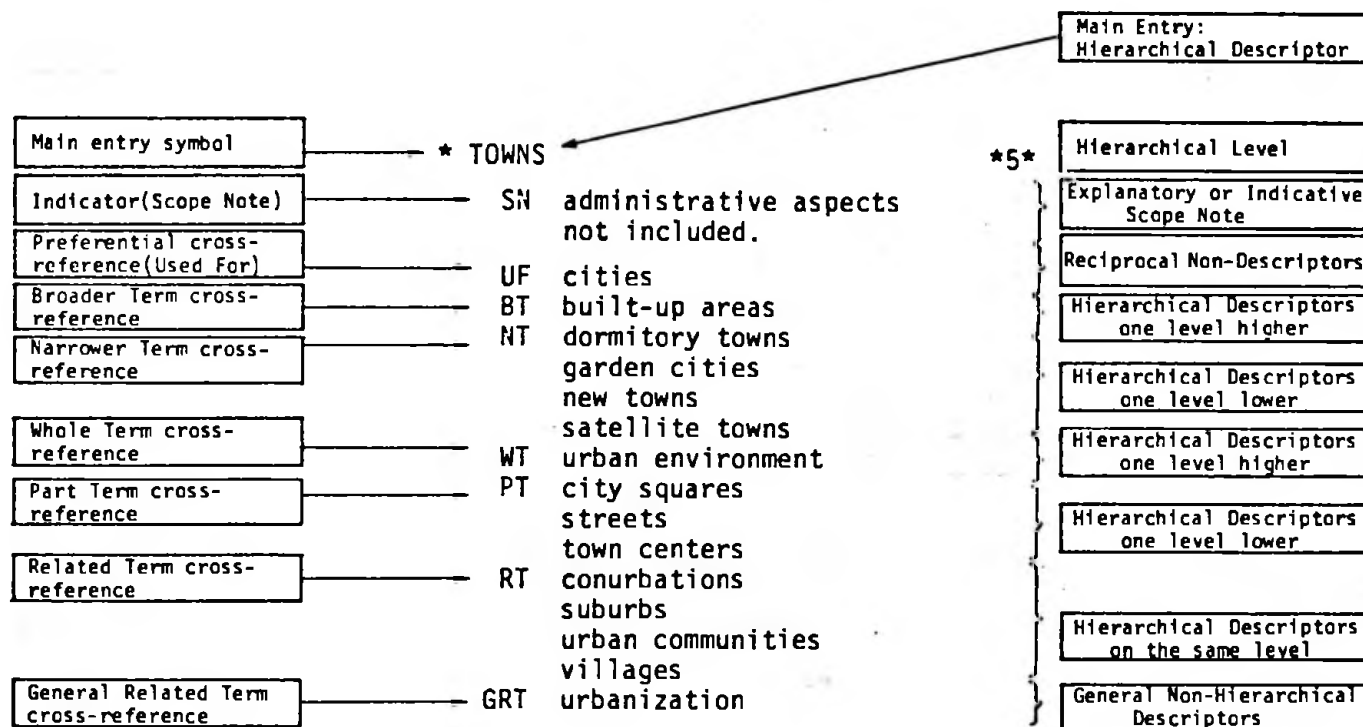
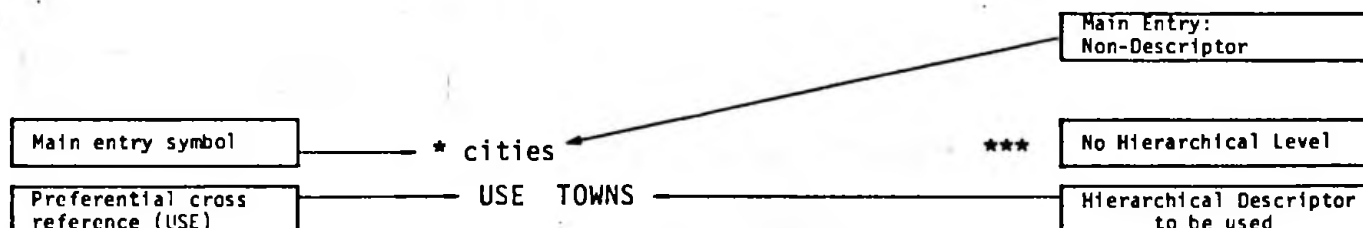
R e c i p r o c a l   N o n - D e s c r i p t o r



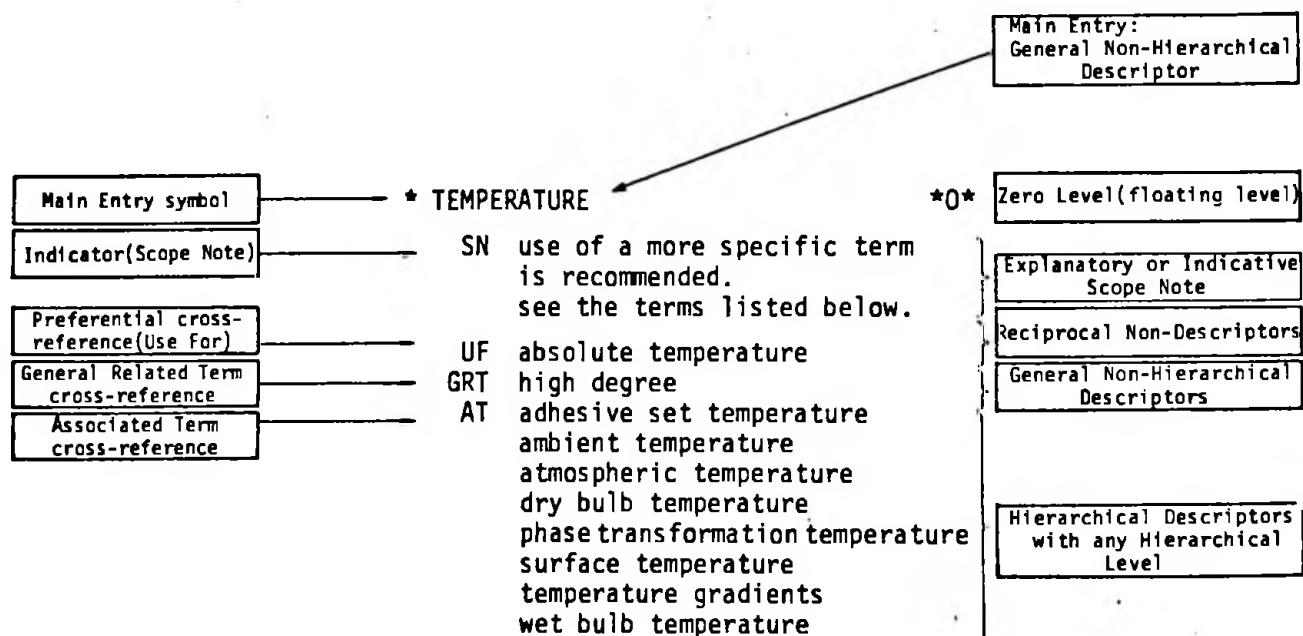
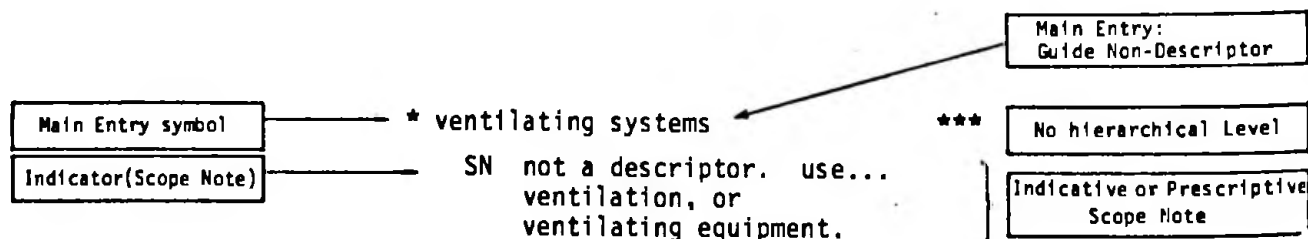
Figure 7: General Non-Hierarchical DescriptorGuide Non-Descriptor

Figure 5: LOGICAL PROPOSITIONS

P1 (I) can be a type of (J).

P3 (I) is always of type of (J).

P2 (J) can be a type of (I).

P4 (J) is always of type of (I).

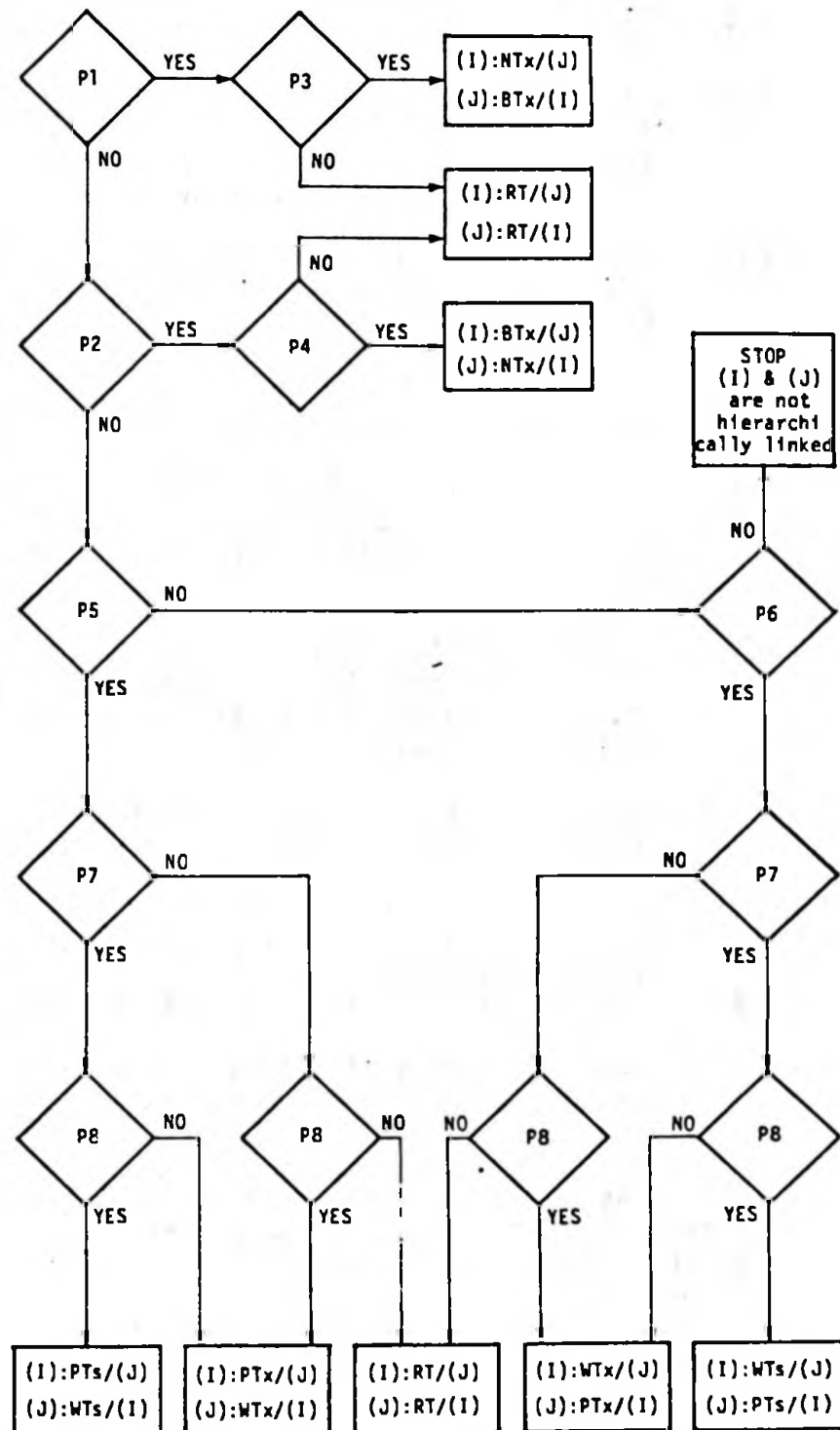
P5 (I) is an element, a sub-set  
a sub-system or an aspect of (J).

P6 (J) is an element, a sub-set,  
a sub-system or an aspect of (I).

P7 if (I) exist, then (J) exist.

P8 if (J) exist, then (I) exist.

(s : specific)  
(x : less specific)



(iv) Let I be CITY SQUARES and J be TOWNS

P 1: CITY SQUARES can be a type of TOWNS - false

P 2: TOWNS can be a type of CITY SQUARES - false

P 5: CITY SQUARES can be an element, a sub-set, a sub-system or an aspect of TOWNS - true

P 7: if CITY SQUARES exist, then TOWNS exist - true (since TOWNS UF CITIES)

P 8: if TOWNS exist, then CITY SQUARES exist - false

therefore CITY SQUARES is a PT<sub>x</sub> of TOWNS (though it is just possible there may be another <sup>x</sup>intermediate term yet to be found)

(v) Let I be URBAN ENVIRONMENT and J be TOWNS

a similar procedure leads from P 1 to P 2 to P 5 to P 6 to P 7 to P 8; therefore URBAN ENVIRONMENT is a WT of TOWNS

(vi) Supposing, by chance, one had stated:

Let I be NEW TOWNS and J be SATELLITE TOWNS

P 1: NEW TOWNS can be a type of SATELLITE TOWNS - false

P 2: SATELLITE TOWNS can be a type of NEW TOWNS - false

P 5: NEW TOWNS can be an element (etc) of SATELLITE TOWNS - false

P 6: SATELLITE TOWNS can be an element (etc) of NEW TOWNS - false

therefore STOP! NEW TOWNS and SATELLITE TOWNS are not hierarchically related. In the event, they are both NT of TOWNS, and could be RT's of each other (though in practice it may not be worthwhile showing the RT relationship)

(vii) Supposing, by chance, one had stated:

Let I be DORMITORY TOWNS and J be GARDEN CITIES

P 1: DORMITORY TOWNS can be a type of GARDEN CITIES - false

P 2: GARDEN CITIES can be a type of DORMITORY TOWNS - after some hesitation: - true

P 4: GARDEN CITIES is always a type of DORMITORY TOWNS - false

therefore GARDEN CITIES is a RT of DORMITORY TOWNS

(In this case they are both NTs of TOWNS, relationships which emerge sooner or later).

This is actually not as laborious as it seems, because one becomes adept at the logical propositions and only has to linger over them in controversial cases.

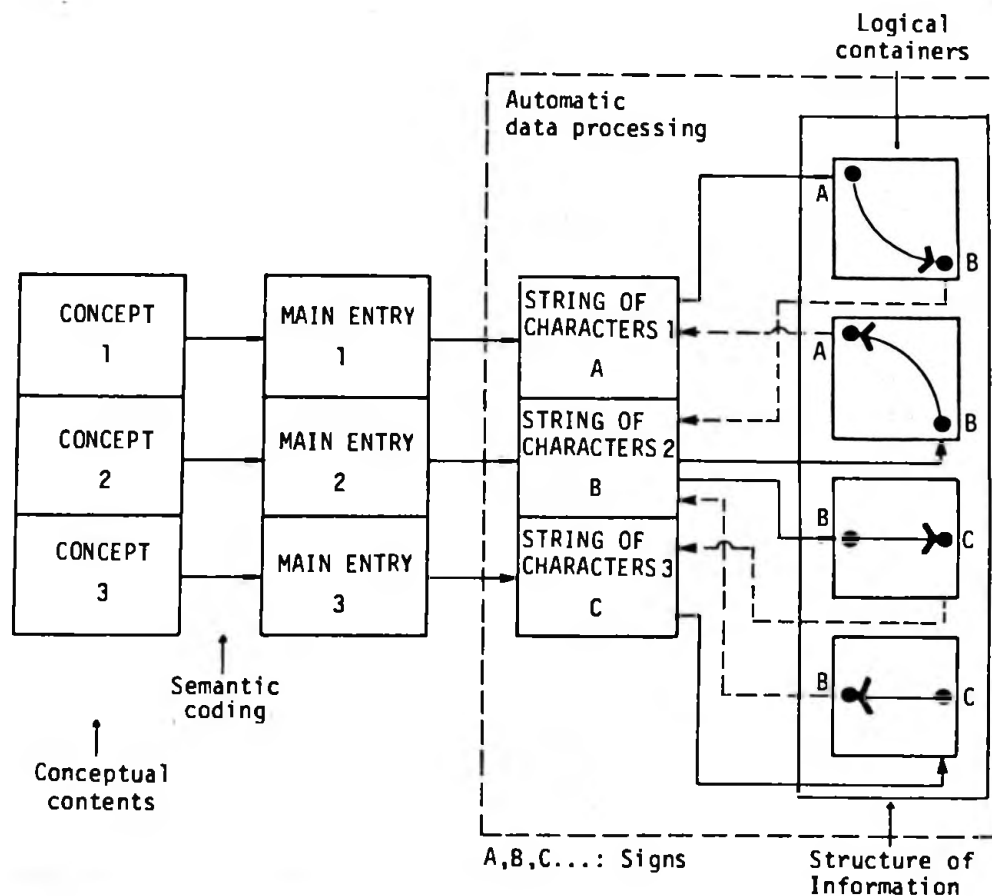
Data processing played an important part in the structure of the thesaurus, particularly in terms of methodology (fig. 6). The relationships between concepts which make up the hierarchy of the thesaurus are reflected at the level of the logical containers within the automatic data processing; the logical containers must be able to be processed

Figure 6: The role of data processing.

Suppose concept 1 is at level  $(n - 1)$ , concept 2 is at level  $(n + 2)$ , concept 3 is also at level  $(n + 2)$ ;

(i) Suppose that the logical propositions show that A is a BT of B - the data processing will (a) check that B is on one level below A and (b) then enter the cross reference and its reciprocal (B is an NT of A; A is a BT of B).

(ii) Suppose that the logical propositions show that C is an RT of B - the data processing will (a) check that C is on the same level as B and (b) then enter the cross reference and its reciprocal (B is an RT of C; C is an RT of B).



without knowing about the conceptual content to which they correspond.

Computer science does not have access to the semantic content of the data that is being processed; it deals only with the processing of symbols and numbers. Because of this characteristic, computer science provides - at all stages - the best possible logical check of the chosen structure and its rules. Identifying and dealing with special cases and "exceptions" have a detrimental influence on the clarity, concision and efficacy of the computer programs; any rule that has no logical relationship to the set of rules will be spotted quickly. In this way, computer processing of data is an effective way of appreciating the elegance and consistency of the structure as a whole.

What does this structure mean for the user of the Thesaurus? We have used the term "conceptual space" to describe the structure as a whole; each descriptor (which is a "label" for a concept) has a unique position within the conceptual space. Experience shows us that this is a great asset in guiding towards the best descriptor with which to describe a concept - whether in indexing comments, indexing queries or whether for other purposes for which concepts and their relationships are useful. The "conceptual space" is too complicated to represent graphically in its entirety (see, however, fig. 3 for an extract relating to our example TOWNS); the regular user of the Thesaurus - by constantly referring to the hierarchical groups - quickly forms a mental image of the main features of the "conceptual space".

#### The Process of Evolution.

A thesaurus - we stated - is a controlled and dynamic vocabulary... In our case, we developed the Thesaurus up to a certain point prior to using it for indexing; during that time, we drew upon the work of the Construction Industry Thesaurus team (Ref.5) whenever appropriate, to increase the chances of compatibility. Then the draft thesaurus was used for indexing. During this time, some candidate terms were being generated; they were "fitted in" to the structure, using the same logical propositions. Sometimes obvious "gaps" were thus filled in; sometimes term errors were corrected and sometimes (though not often) whole sections of the Thesaurus had to be "slipped" down or up one level (this had major repercussions on RT/RT relationships; in practice we found that any such major upheaval actually often lead to the elimination of some questionable RT/RT relationships).

The English edition of the I.F. Thesaurus was published as soon as we felt confident that the rate of changes was being reduced; this was in May 1972. Since then we have published one set of addenda and corrigenda.

A more significant aspect of the process of evolution is particular to the bilingual status of Canada. We have translated the Thesaurus into French (a "rough" translation, which reflects the structure of the English Thesaurus but contains some term forms we are not

satisfied with). This involved preparing sets of computer programs, enabling monolingual or bilingual alpha-hierarchical and non-hierarchical listings to be produced - with English or with French entry. Possibly thanks to the discipline in the structure, we did not encounter too much difficulty with the descriptors, though many of the non-descriptors did not need to be translated.

More significantly, however, is the fact that we are now collaborating with the Groupe Latin ( a sub-group of Commission W.52 of the International Council for Building Research, Studies and Documentation) in preparing a basic latin-language thesaurus for building. Thanks to the collaboration of the researchers in the Groupe Latin, the coverage of the French Thesaurus is being improved; this in turn will be fed back into the English edition.

Obviously, the massive addition of new terms poses major problems - both in the hierarchical structuring of these terms and in data processing. An ideal we would like to attain would be to have the Thesaurus on-line on a CRT, so that one could "walk-through" the conceptual space, making improvements on the way.

### CONCLUSIONS

This paper has scarcely mentioned the problems of indexing and the related problems of indexing rules. It is our experience, however, that for all aspects of working with, or on, a thesaurus the "scientific" approach is advisable. The rigorous structure enables the user to identify the position of the concept within the conceptual space; the logical propositions enable the hierarchical groups to be fully understood, and the significance of the relationships between terms to be grasped. Once this has been accomplished, indexing rules and search strategies can be built up appropriately.

## Footnotes.

- (1) Industrialization Forum - Building: Systems, Construction, Analysis, Research; published in English and in French jointly at Massachusetts Institute of Technology, Washington University and Université de Montréal; see: Wert, Leonard, "Information Retrieval and Industrialization Forum". Industrialization Forum, Vol.1, No.1 (October 1969), pp.11-17.
- (2) Engineers Joint Council. Thesaurus of Engineering and Scientific Terms. New York, Engineers Joint Council, 1967, 690pp.
- (3) Building Research Thesaurus; compiled in the Library of the Building Research Establishment, revised edition, Garston, Building Research Establishment, Department of the Environment, 1972, 1 vol.
- (4) Canada. Department of Industry, Trade and Commerce. Materials Branch. Thesaurus of Canadian Construction Terminology, preliminary edition. Ottawa, Department of Industry, Trade and Commerce, 1971, 1 vol.
- (5) Roberts, Michael, Chris. Eve, Peter Linn and Ellen MacHale: Construction Industry Thesaurus; second preliminary draft. North Western Polytechnic School of Librarianship and the Polytechnic of the South Bank, 1971, 1 vol.
- (6) Thésaurus du Bâtiment et des Travaux Publics. Paris, Institut Technique du Bâtiment et des Travaux Publics, 1970, 149pp. This document is now being merged into a basic latin-language building thesaurus in preparation by the Groupe Latin of the International Council for Building Research, Studies and Documentation.
- (7) General Terms exist ex-hierarchy (e.g. TEMPERATURE, see fig.7 ) they have associative relationships with other General Terms (designated General Related Term - GRT) or with hierarchical terms (designated Associated Term - AT). Since the General Terms are ex-hierarchy, there can be no question of hierarchical levels in their cross-references.
- (8) For example, E.J.C. puts part terms in the RT groups; whereas the Thesaurus of Canadian Construction Terminology puts them in the NT relationships. For an attempt to sort out RT's, see the introduction to: Barhydt, Gordon C. Charles T. Schmidt and Kee T. Chang. Information Retrieval Thesaurus of Education Terms. Cleveland, Press of Case Western Reserve University, 1968, 133pp.
- (9) This restriction means that the associative relationship is only shown between descriptors that are "comparable" - in the sense that they represent concepts of analogous generality or specificity.