Meta
Journal des traducteurs
Translators' Journal

META

Termium: System Description

Robert Dubuc

Volume 17, numéro 4, décembre 1972

URI : https://id.erudit.org/iderudit/002642ar
DOI : https://doi.org/10.7202/002642ar

Aller au sommaire du numéro

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

Découvrir la revue

Citer cet article

Dubuc, R. (1972). Termium: System Description. *Meta*, *17*(4), 203–219. https://doi.org/10.7202/002642ar

TERMIUM
System Description

0. INTRODUCTION

0.1. *General*

0.1.1. The Banque de terminologie de l'Université de Montréal was founded to provide clients with a remedy for the present dearth of reliable terminological documentation, its dispersion and its inaccessibility. To this end, therefore, the Bank has set itself three basic objectives : to collect terminological documentation which already exists in various forms and is often difficult to obtain, to promote a methodical approach to terminological research in sectors as yet unexplored, and to make all information available to clients at reasonable cost through fast and efficient processing.

0.1.1.1. To define the basic conception of the project, a clear distinction had to be made between a word bank and a terminology bank. A word bank, on the one hand, provides methodical organization of terms in a given language into etymological, morphological and semantic families in order to furnish clients with improved means of expression. The purpose of a terminology bank, on the other hand, is case-method study of vocabulary ; from the study of scientific or technical fields in terms of several languages, it is quite evident that the two are very different in operational outlook. For the terminology bank, the basic terminological problem (concept definition and equivalents in other languages) acts in a sense as the information nucleus to which all the related secondary data (terminological parameters) adhere.

0.1.2. In developing the project, it was decided to create a « basic stock » of from 100 000 to 300 000 bilingual terminological data units or records, stored in a computer's magnetic memory, which could be consulted directly through a keyboard terminal. *Termium* was the name chosen for the computer system governing the project. We shall attempt to describe it and to define its basic conception, content and operation.

0.1.3. The *Termium* file involves a mass storage phase concurrent with the design and developmental phases of the system. With the aid of the mass storage technique, the Bank has produced a preliminary version of the file containing approximately 30 000 data units.

0.2. *The Termium file*

0.2.1. The rationale behind the design of the file was to provide the client with sufficiently comprehensive terminological data to enable him to determine their scope. For this reason, each data unit will give terminological information in context or a definition, including reference to source and date.

0.2.1.1. Moreover, in view of our geographical, economic and demographic position, we felt it essential for the file to be fully bilingual, so that it would be accessible both from French and from English.

0.2.1.2. The *Termium* « store » or « corpus », in short, is a file and not a dictionary. In other words, the information given is perfectly valid from the documentary standpoint but in itself does not set a standard ; any standardized expressions appearing in the file will have been authorized by agencies outside the Bank which have the power to do so. In addition, lexicography is not part of *Termium's* principal operation. Neither the development of definitions nor the publication of conventional-type technical dictionaries is envisaged. *Termium's* role is to collect *evidence of usage.*

0.2.1.3. This concept keeps us constantly alert to usage, enables us to direct research efforts where most needed, make fast changes and corrections, and incorporate the latest information instantaneously. Through the technique of direct access and time-sharing, the necessary corrections and additions to the file can be made daily, so that clients receive the latest and most accurate information possible.

0.2.2. *Access to the information.* — *Termium* information can be looked up through the English or French keywords in the terminology data units or through any component words considered relevant. For example, the record « water-waste treatment plant — *station d'épuration* » could be looked up not only under these two phrases in their entirety, but also under the words « water-waste », « treatment », « plant », *station* and *épuration.* This highly flexible look-up mode permits optimum utilization of the store of terminology, lessens the problem of synonymous usage and can point the client towards a solution when the file contains no direct information on the problem submitted.

0.2.3. An essential factor for establishing the bounds of a query, as well as for amplifying the information given, is indication of the field involved. It is a well-known fact that the English word « range » has several very different meanings according to whether it is used militarily, in acoustics, or electronics. Apart from adding not inconsiderable semantic precision, field indication enables the Bank to establish various terminological directories by subject and these can be of great practical use to editors, lexicologists and language specialists.

0.2.3.1. The Bank's field directories have a three-level logical structure : 26 classes with 26 divisions, subdivided into 26 sections, offering a theoretical capacity of about 17 500 fields.

0.2.4. Finally, the words themselves are also rated as to value. The weighting index 0 means unacceptable, 1, doubtful, 2, acceptable and 3, standard. Through appropriate programming therefore, for purposes of standardization, it will be possible to have a list of doubtful words drawn up automatically in, for example, the field of electrical engineering.

That, in essence, is the *Termium* file. We have purposely given it modest objectives and geared it directly to the needs of the community to be served.

0.3. *Computer operation*

0.3.1. In order to meet the immediate and future needs of that community as effectively as possible, *Termium's* directors, after consulting with directors of similar experiments in Europe, opted for a direct access system, that is, for a system operating through keyboard terminals which handles multiple and instantaneous look-ups of the terminology file in conversational mode.

0.3.1.1. This method of operation is based on an explicit language for the declaration of entries, or access words, starting from keywords, the significant components of which are declared accessible. In addition, the problem of graphic disparity is eliminated by reference to a « blocked » word in which all sources of disparity (hyphens, upper and lower-case letters, etc.) are removed.

0.3.1.2. To ensure the practicability of direct access, the system is designed so that look-up and answer time is virtually unaffected by the size of the file or the length of the answer sought.

0.3.2. Although completely bilingual, the system nevertheless does not exclude the eventual incorporation of other languages. By virtue of its mode of language notation, as many as ten languages could be introduced.

0.3.3. *Termium*, in essence, contains three main files : a master file containing primary data units (keywords and terminological parameters) as well as the corresponding secondary data units (contexts or definitions), a file with all entries stored in blocks in alphabetical order, and finally an index file which locates the file block of the word sought.

0.3.3.1. Since it allows room for the incorporation of any future development or adaptation, the system can always keep pace with the latest developments in its market's requirements and in technical progress.

0.3.4. In the *Termium* system, moreover, provision has been made for auxiliary files, the principal ones being lists of sources and authors, a coded fields index, statistical and bibliographical files, and back-up files for preserving the information stored.

Termium also has a potential for sequential processing, especially for purposes of standardization, correction, updating and lexicography. This inexpensive and convenient mode of operation will be of great practical use.

0.3.5. It should be mentioned besides that *Termium* possesses a full stock of typographical characters, complete with capitals, small letters and accents, which, although expensive, was considered necessary in view of the very strict processing essential in the handling of terminological data, as well as the demands of a clientele that is particularly conscious of the importance of such graphic marks.

0.3.6. Direct access will be the preferred mode of look-up and could be used as well for storage and correction. However, due to the high cost and complex control procedures involved in on-line storage, a method for the mass generation of input data has been developed using the Flexowriter. With this device, both a coded and a hard copy of the terminology data unit can be made in a single operation, facilitating verification and correction of errors. This method makes possible an input of 200 data units per normal machine day, a speed which is more than sufficient for the growth rate anticipated for the Bank's store of terminology.

0.3.7. During initial operation, up to 100 000 data units will be stored on a unit 841 « disk-pack » in a CDC 6600 computer at the University's Computer Centre. By late 1972, the system will be using a unit 844 « disk-pack ».

0.3.7.1. Any terminal of any make that has a bilingual keyboard can communicate with *Termium*. The Bank, however, has chosen an Olivetti (model TE-318), since its 94-character keyboard (ASC II) most closely meets the recommendations made by the Canadian Standards Association (CSA) to the International Standards Organization (ISO) for standardizing teletypewriter bilingual keyboards.

1. OPERATION OF THE TERMINOLOGY BANK

1.1. *The problems* in terminology which must be solved in order to provide effective communication in scientific and technical fields are as follows :

1.1.1. *Dispersal of research effort.* — At the present time, a great deal of research is being done in a number of fields, but for want of an effective coordinating agency, this effort often duplicates work already done or under way, or else it does not take sufficient account of the most urgent needs of the community to be served.

1.1.2. *Disparities in value of research.* — Compounding the lack of coordination, terminological research works which reach publication are often of dubious quality, for want of strict research methodology and of established criteria for validating the information.

1.1.3. *Inaccessibility of information.* — Terminological studies, often of only limited circulation, soon become very difficult to obtain. The multiplicity of sources adds a further complicating dimension to the problem. Besides, with science and technology evolving constantly and rapidly, up-to-date terminology data in key sectors is virtually unobtainable in an immediately usable form.

1.1.4. *The burden of translation research.* — Increasing contact between peoples results in an ever-increasing spate of translation. Translation offices must cope with texts that are not only more numerous from day to day but also increasingly complex. Working under present conditions, a translator spends at least a third

of his time on terminological research. Worse still, this research is most often non-productive, in the sense that the translator does not systematically record his findings ; consequently, more often than not, each translator does research that has already been done.

1.1.5. *Delays in conventional methods of disseminating terminological information.* — Owing to delays inherent in the traditional publishing process, it is safe to say that a dictionary when barely off the presses is already out of date. Moreover, answers to problems in terminology must often be bound in very short order, and rapid access to information is quite impossible using manually assembled files or conventional technical lexicons.

1.2. *Solutions*

1.2.1. These problems in terminology can be solved through a systematic approach to terminological research that will put the validity of the information beyond doubt and ensure the collection of all work done in the various fields.

1.2.2. From the outset, the computer seemed indicated as the only possible vehicle for ensuring fast input of terminological data, its massive organization incorporating the various criteria for its use, and also its quick and continuous updating.

1.2.3. Moreover, because of the evolution of terminological problems and their repercussions, particularly on translation, it was considered necessary to provide automatic access to the store of information, or more specifically, direct, conversational access through a keyboard terminal with all the typographical capabilities for meeting the exacting criteria expected of a terminology bank by its clients (i.e. upper and lower-case letters, accents and other diacritics). Using these techniques, a client could increase his output, particularly in translation, to an extent which would amply justify the cost of look-up.

1.3. *Information validity*

Since *Termium's* information must not only satisfy its clientele but also meet the criteria observed by responsible terminology agencies throughout the world, the Bank has adopted the validation criteria recognized by the Geneva International Conference on Terminology of December 1969, namely :

a) that terminological information always be given with mention of context or definition ;

b) that reference be given to its source and date ;

c) that it be taken from reliable *original* texts.

1.4. *Guiding principles*

Despite the adoption of these criteria, however, the Bank must not neglect the special needs of the potential clientele of a computerized file.

1.4.1. The basic fact which predicates all terminological research in Canada is the presence of two official languages in the country. Problems in terminology arise most often in situations involving translation, particularly from English (or

American) into French. We believe that this is sufficient justification for the essentially bilingual nature of the *Termium* file.

1.4.2. *The entry.* — By entry we mean any term which gives access to the stored information. For example, in the case of the data unit appearing in paragraph 2.2.3., the two terms (which we shall call keywords) each represent an entry, i.e. « waste-water treatment plant » and *station d'épuration ;* if the terminologist considers that the words « treatment » and « plant » should also give access to this data unit, we would then have two additional entries, and also a third if it were felt that the word *épuration* warranted access as well. Thus, the data unit could contain five entries.

1.4.2.1. The declaration language used in the entries, the words which give access to the terminological information, depends essentially on the terminology problem considered in relation to the basic working languages of the Bank, English and French. On storage, the terminologists decide what the access words are to be. Therefore, the data unit « velocity operated type microphone » will be obtainable from English by the expression in its entirety and by « microphone », since it is reasonable to assume that no one would look for the translation of this term under « velocity », « operated », or « type » either separately or together. The data unit will also be accessible through the corresponding French expression *microphone à vitesse* and only that, since the words *microphone* and *vitesse* used separately have no particular terminological significance. The same is true for complex units corresponding to a simple unit. For example, with the data unit « bird's eye view — *plongée* », it would be pointless to obtain the separate components of the complex unit. Even though this solution does incur some risk of subjective judgment on the part of the terminologists, it nevertheless avoids useless retrieval efforts, overloading of the memory and risk of an information avalanche upon look-up.

1.4.3. One of the criteria for selecting contexts is that they be meaningful : the context chosen must enable the client to form at least some idea of the concept involved in the data unit. A second criterion is that contexts be similar in both languages. This is often difficult, but, if accomplished, the data unit will provide maximum guarantees of validity. Besides, overlapping contexts can often be used as material for other data units, affording appreciable savings in time and research effort. Thus, if in the definitions of « frequency response » and of *réponse de fréquence,* the expression « reference frequency — *fréquence de référence* » occurs with the same two contexts, a new data unit can easily be made under this new heading without any additional research.

1.4.4. All information must be drawn from reliable original texts, « reliable » meaning recognized to be competently written. Translations, except for rare exceptions, are not considered a sound source of terminological information ; the danger of contamination is too great and often the translator has been unwilling or unable to verify his vocabulary strictly enough. Using original texts can also bring to light the existence of idiomatic expressions peculiar to each language.

1.4.4.1. Each data unit must carry a reference to source and date so that anyone wishing to check or amplify the reference may refer to the work quoted. Conse-

quently, bibliographical and quotation references (page, volume, etc.) must be complete and the notation of the date is considered necessary in view of the rapid evolution of technology and science. In electronics, for example, information dated 1940 is obviously less valid than material dated 1971.

1.5. *Terminological parameters*

A terminology bank must also take into account certain semantic and grammatical parameters which have a distinct influence on terminology ; it is therefore necessary on occasion to record the language level of the term used (a colloquialism or slang, for example, when used generically or specifically). As an instance of grammatical influence, gender and number are often of special semantic relevance and must therefore be noted.

1.5.1. Owing to the Bank's policy of deliberately steering clear of strict lexicography for a more properly terminological approach (which is closer to actual usage and more flexible), the client, to be sure of his ground, must be able to evaluate the validity of the information he receives. A weighting index is provided for this purpose.

1.6. *Fields classification*

As the Bank's « working capital » has been set at about 250 000 data units, the field or fields in which a particular term is used must be specified for easier look-up and classification of the information in orderly groups.

1.6.1. For this purpose, the Bank has developed a classified fields code with a three-level logical structure. 26 classes form the skeleton of the classification and contain the basic fields of activity which lend themselves to terminological development. These 26 classes each comprise 26 divisions representing the main areas of activity within each class, and these divisions are in turn subdivided into 26 sections each representing a specific nucleus of activity within the division. Thus, in the code AFF, the first letter represents the class *Administration,* the second, the division *Gestion du personnel* and the third, the section *Perfectionnement et formation du personnel.* This mode of classification affords a theoretical possibility of 17 500 vocabulary sections. Its advantage over the standard codes already in existence (Universal Decimal Classification, Dewey, Library of Congress) is that it is designed exclusively for terminological utilization, and has all the flexibility necessary for that purpose. The code will be published in both French and English.

2. THE TERMINOLOGY DATA UNIT

2.1. *Limitations*

The terminology data unit must be designed to perform two types of function. First, to act as a vehicle for reliable terminological information, and then to enable its communication to clients in concise form and in the manner most useful to them.

2.1.1. For the first of these functions, the terminology data units bear a reference to source, year of publication and page for each keyword. Coupled with this information are terminological parameters indicating level and other particularities

of language, language notation, weighting index and grammatical notes on the nature of the expression, its gender and number, whenever such indications are terminologically relevant.

2.1.1.1. The contexts or definitions are recorded in modules with 225 characters. Experience has shown that this size, in the large majority of cases, allows for inclusion of the essence of a context or definition in one single module.

2.1.2. Finally, there are three indications covering the data unit as a whole : fields of application, in code (as many as three) ; a data unit author key certifying the author's qualifications ; the overall weighting index which evaluates not only each keyword but also the equivalence of the terms.

2.1.3. Each *Termium* data unit contains virtually all of the information necessary to certify the validity of the data appearing on it. The size of the original file card (3 x 5 in/ 75 x 125 mm) compels recording of the strict essentials and obviates the pitfall of extraneous information. This size, moreover, fits in with most existing card files in Canada and is compatible with the international standard size.

2.1.4. The information must also be arranged clearly and in an orderly manner for easy understanding by the client. By virtue of the divisions in the data unit, all information can be related to the pertinent keywords without any possible ambiguity. Similarly, the information pertaining to the data unit as a whole stands out clearly from all the rest. With its concise, synthetic structure, this data unit should be an ideal vehicle for disseminating terminological information.

2.2. *Data unit content*

2.2.1. This is what the basic original data unit file card looks like :

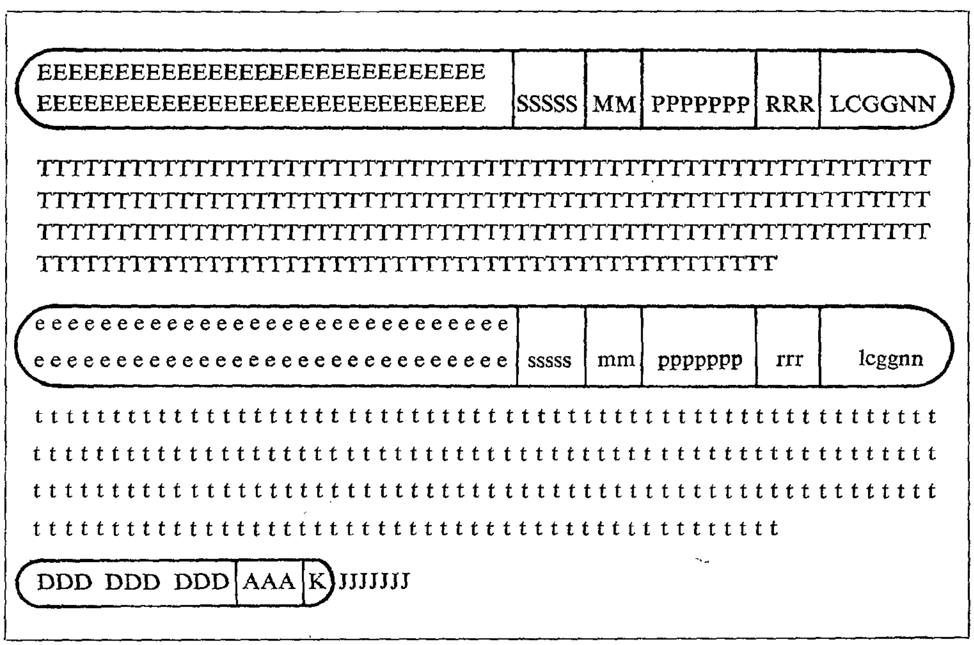


FIGURE 1

2.2.1.1. It is divided into three sections :
1) English section (upper part in fig. 1) ;
2) French section (middle part) ;
3) Identification section (lower part).

2.2.1.2. In this format and in this size (3 x 5 in/ 75 x 125 mm), the data unit card contains a pair of corresponding terms or expressions in English and French for one or more fields of application.

2.2.1.3. All information is recorded on it in this form, with a choice of characters including upper and lower-case letters and accents.

2.2.2. *Contents of sections 1 and 2*

E,e : Keyword (potential for 58 characters including blanks).

S,s : Source (5-character code developed by the Bank).

M,m : Source publication date (last two figures of the year).

P,p : Source page indication (7 characters) includes volume number, if any, and then part number or issue number of a periodical, separated from page number by a hyphen.

‡ R,r : Particularities (3 characters).
Identification of the linguistic particularities and notation of the language levels.

‡‡ L,l : Keyword language.

‡‡ C,c : Keyword weighting.

‡ G,g : Grammatical category : verb, adjective, adverb, noun, etc.

‡ N,n : Gender and number.

T,t : Keyword definition or context (225 character maximum).

2.2.3. *Contents of section 3*

D : Fields code (maximum of 3).

‡‡ A : Data unit author (key provided by the Bank) (3 characters).

‡‡ K : Weighting of data unit as a whole (1 character).

‡‡‡ J : Data unit serial number (7 characters).

2.3. *Codes*

2.3.1. The capsule size of the data unit cards makes it necessary to use codes for indicating sources, particularities, weighting, grammatical notes, fields and data unit authors.

2.3.2. These codes are generally simple but give the most explicit information possible.

\# : Optional reference.

\## : Failing explicit instructions to the contrary, the computer system will adopt a predetermined uniform notation, called default option.

\### : This number is assigned automatically by the system when the data unit is stored in the memory.

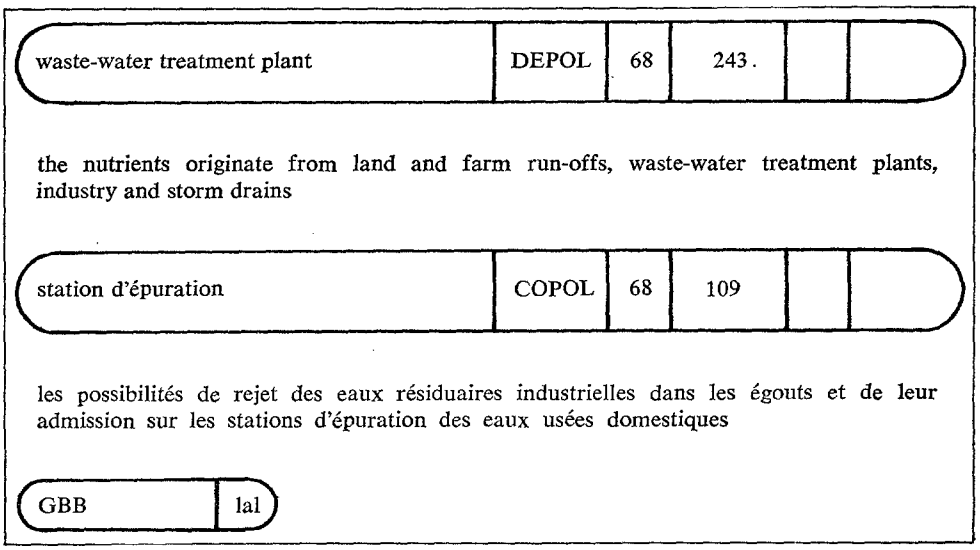


FIGURE 2 : Example of a completed data unit card.

2.3.2.1. The code for source consists of a pronounceable acronym of five characters, a combination of the author's name and the title. The periodical name codes are usually composed of an abbreviation of the title into a pronounceable word to make it easier to remember. In both cases, the code refers back to a complete bibliographical data unit.

2.3.2.2. The particularities code identifies the geographical extension of a term. The main symbols are :

CE : Canadian English
FA : African French
FB : Belgian French
FC : Canadian French
FF : France French
FR : Regional French
FS : Swiss French
FU : Standard French
GB : British English
US : American English

Coupled with this code are semantic specifications :

A : Archaic
C : Current (opposed to scholarly)
D : Derogatory or pejorative
E : Seldom used, exceptional
F : Familiar or colloquial level
G : Generic
I : Limiting adjective
J : Technical jargon
L : Literary level
N : Neologism
P : Popular or slang level
R : Descriptive adjective
S : Specific
T : Scholarly (opposed to current)
V : Old

2.3.2.3. *Languages code* : 1) English, 2) French, 3) Latin, 4) German.

In the case of French and English, this appears only in the pseudo data units, since the system automatically identifies these languages in the basic data unit. This notation allows incorporation of other languages into the system.

2.3.2.4. *Weighting code* for each keyword and for the data unit as a whole : 0) unacceptable, 1) doubtful, 2) acceptable (noted automatically by the system), 3) standard.

2.3.2.5. *Grammatical category code* :

N (plus blank) :	common noun	V (plus blank) :	verb
NP :	proper noun	VI :	intransitive verb
AJ :	adjective	VT :	transitive verb
AV :	adverb		

2.3.2.6. *Gender and number* :

M :	masculine	S :	singular
F :	feminine	P :	plural
X :	neuter		

The codes appearing in 2.3.2.2., 2.3.2.5., 2.3.2.6. are given only if terminologically relevant.

2.3.2.7. *Fields code,* explained in 1.6.1. This may be obtained by applying to the Bank. When more than one field is indicated, the indications are separated by blanks.

2.3.2.8. *Data unit author code*

2.3.2.8.1. Data unit authors can be either agencies or individuals.

2.3.2.8.2. For agencies, the Bank assigns a code of two capital letters to each (resembling as closely as possible the acronym already used by the agency), plus a chronological key, if appropriate, without a hyphen. Examples :

CN :	Canadian National Railways	CP :	Canadian Pacific
RC :	Canadian Broadcasting Corporation	CI :	Canadian Industries Limited
		DT :	Domtar

2.3.2.8.3. For individuals, the code is composed of their initials in small letters (2 letters plus chronological key without a hyphen). When the ten digits (1 to 0) have been used, the figure is replaced by capital letters (A to Z) or other combinations of these letters. Examples :

rd (1 to 0)	rD (1 to 0)
rd (A to Z)	rD (A to Z)
Rd (1 to 0)	RD (1 to 0)
Rd (A to Z)	RD (A to Z)

2.3.2.8.4. Every data unit author certified by the Bank registers by giving the following information : last name, usual first name, profession, home address, business address, telephone number, years of practice in the profession, specialization.

2.4. *Data unit logic design*

The information in the data unit is divided into two parts.

2.4.1. The first (primary data unit) contains essential terminological data, that is, the corresponding keywords and their parameters, coupled with references to fields of application, data unit author key, and the overall weighting index.

2.4.2. The second (secondary data unit) contains instances of use or definitions for each language. Since these data are of a documentary nature (as opposed to the information in the first part), it was thought appropriate only to convey them on request so as not to overburden the look-up process.

2.4.3. All essential information, therefore, is grouped into the primary data unit for storage purposes and can be graphically illustrated like this :

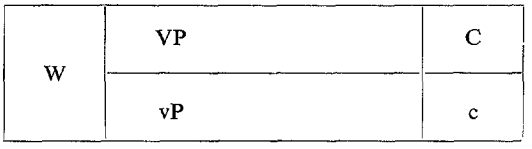


FIGURE 3

W : Data common to both keywords ;
VP : Keyword and parameters in first language ;
vP : Keyword and parameters in second language ;
C : Pointer to retrieve first language context ;
c : Pointer to retrieve second language context.

2.4.4. Information on context is recorded in the secondary data units and retrievable from the primary data unit pointers. This information can take the form of a context alone, a definition, a grammatical comment, semantic explanation, etc.

2.5. *Pseudo data unit*

2.5.1. The pseudo data unit is a way of broadening, where necessary, the bounds of the basic data unit upon storage. It can be used to record phrases within the keywords in the primary data units, extend contexts beyond the 225-character limit, or expand the indication on fields of application. For scientific terminologies, Latin and German names can also be included.

2.5.2. Every pseudo data unit bears a special PF key (PF standing for *pseudo-fiche*) and necessarily follows the basic data unit of which it is an extenstion. Its place in the system is determined by the kind of information it contains.

2.6. *Correction data unit*

Like the pseudo data unit, the correction data unit is a means of modifying information already in the memory. It also fits into the basic structure of the primary and secondary data unit system, according to the type of information it contains.

3.0. SYSTEM OPERATION

3.1. *General*

Now that we have described the design of the terminology data unit, we shall examine how the computer system may attain its objectives, despite inherent limitations.

3.2. *Limitations*

3.2.1. The first limitation is obviously due to the size of the project. Developing a bilingual computer file for scientific and technical fields in an area of rapid

discoveries and developments presents quantitative problems in the mere storage of the information. It was therefore important first to develop an « economical » system which would avoid the blockages and bottlenecks occasioned by a mass of information which can only become increasingly unwieldy.

3.2.2. Coupled with this quantitative limitation is a qualitative one ; in virtue of the important role as information disseminator which a terminology bank is expected to play, its information must always be as close to perfection as possible, and must in addition fulfill those formal requirements which clients have a right to expect. A system which could be easily updated therefore had to be developed, one that could make all appropriate corrections with a minimum of change and could present information in a satisfactory form. In French, for example, clients would certainly be unwilling to accept a record without accents or capitals because of their bearing on the quality of the information. Unfortunately, the computer telecommunications systems in North America do not generally have the necessary typographical marks to transmit information in French with all the appropriate diacritics.

3.3. *Objectives*

3.3.1. *Termium,* therefore, with due recognition of these limitations, has been designed to attain those objectives considered essential for the efficient operation of a terminology bank.

3.3.2. First of all, a technique for rapid, simple and inexpensive storage of data had to be designed which would also ensure maximum quality of the product and furthermore allow corrections and mass updating to be done with a minimum of change.

3.3.3. The system must also be quick in its output ; in other words, it must afford access to the information in the least possible time, in view of its customer requirements, the high cost of machine time and the sheer volume of the data in storage.

3.3.4. Auxiliary or secondary uses of the store of information must also be made possible ; for example, the arrangement of data by field of study, weighting index, data unit author, etc.

3.3.5. Similarly, to ensure that the system keeps pace with market demands, certain statistics must be kept on the nature and object of the queries submitted by clients.

3.4. *System*

In the light of these objectives and limitations, the Computer Centre experts have developed a truly tailor-made system to ensure maximum efficiency for the Banque de terminologie de l'Université de Montréal.

3.4.1. *Mass input generation.* — Mass input generation, that is, the mass storage of terminological data, is accomplished through the Flexowriter, whereby a hard copy of a record can be made at the same time as a coded version on perforated tape. The advantage of this system is that it is fast ; an experienced keypunch operator can easily punch 200 data units per working day. Since there is a hard

copy of each data unit, errors can easily be spotted and the correction process is thus greatly simplified. The Flexowriter is therefore a very adequate answer in dealing with the limitations and attaining the objectives of mass information storage.

3.4.1.1. The use of the Flexowriter has led to the development of three separate programmes : a programme for decoding the punched paper tape ; a programme governing input and storage of the taped information which will give a list, on a high-speed printer, of the entries declared or of the data units in storage, with identification number and an indication of acceptance or rejection ; finally, a corrections programme for making the changes indicated on the lists obtained on the high-speed printer.

3.4.1.2. *Updating and special input generation.* — The development of a mode of mass input generation does not preclude the possibility of on-line input of urgent corrections, additions or withdrawals, via the terminal. However, because of the safety and economy of storage by Flexowriter, this possibility is reserved for special cases.

3.5. *Operation*

3.5.1. To answer clients' needs and to keep pace with new developments in computer techniques, the Bank has made *Termium* a direct-access, conversational look-up system.

3.5.2. Because of the volume of data and the type of service to be provided, it was essential that the computer answer-time be virtually unaffected by the Bank's total volume or the length of the query. The *Termium* system provides the answer to this most difficult problem in the astute design of its files.

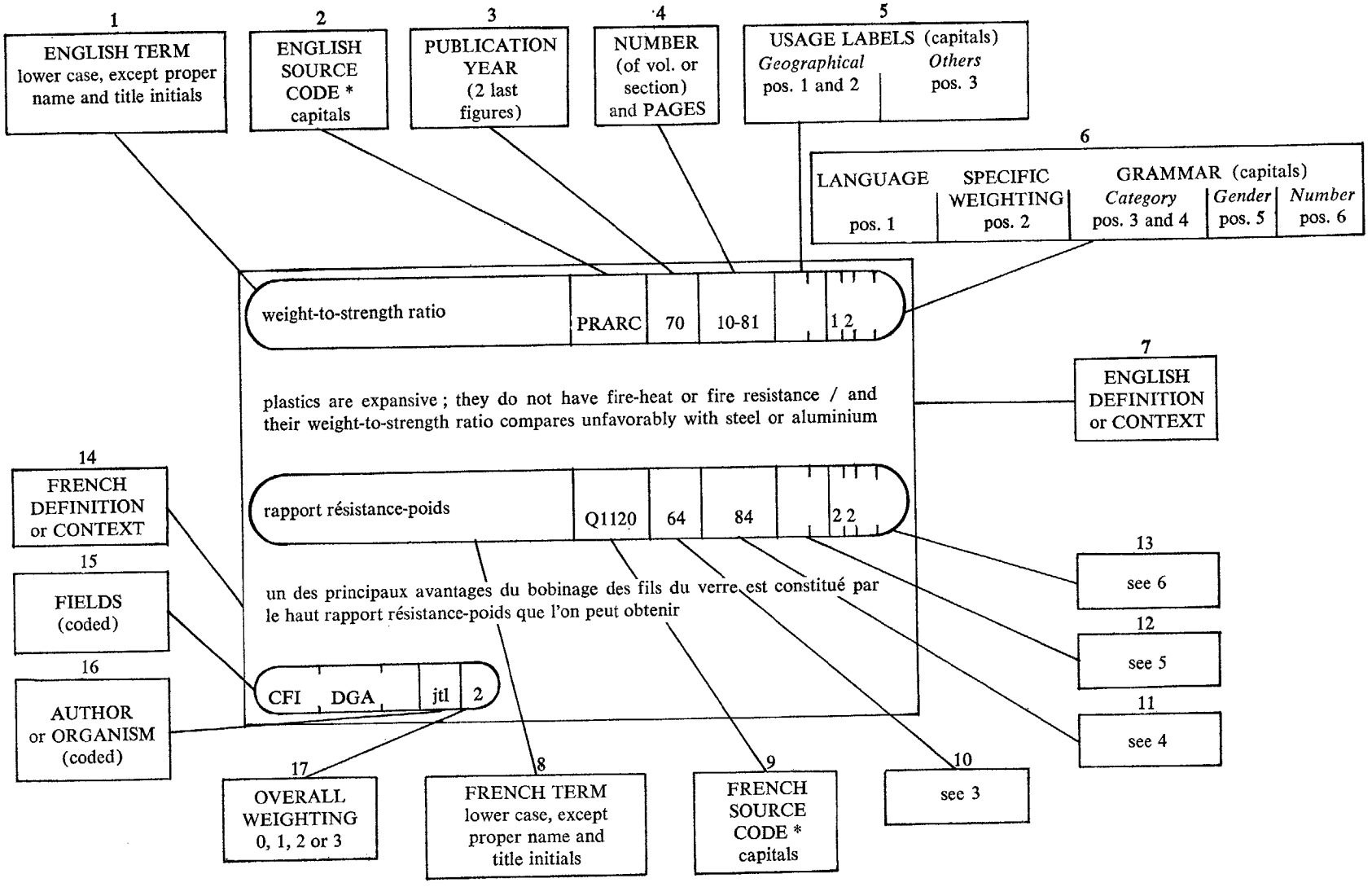
3.5.3. The logic design of the terminological data unit has led to the development of a system of three basic files.

3.5.3.1. The first phase of the first file contains the primary data units, each with a pair of keywords and their parameters, as well as the assignment of fields, record author code and overall weighting index. The second phase is composed of the secondary data units containing contextual information. For each keyword in the primary data units, there is a « hook » for retrieval of corresponding contexts. Each primary data unit, moreover, bears a numbered address for retrieval purposes.

3.5.3.2. In the second file, all the entries declared (1.4.2.) are grouped in alphabetical order into blocks which are appropriately identified. For each of these entries there is a corresponding primary data unit address.

3.5.3.3. Access to the second file is achieved via a third alphabetical file which gives the first 7.5 characters of the first entry of each component block (block address). Information look-up, therefore, is begun by reading the index file for the block key containing the address of the primary data unit which will give the answer to the query.

Diagram of the basic data unit or record



* In areas 2 and 9, the sign @ is the code symbol for the Banque de terminologie.

3.6. *On-line operation*

This mode of operation involves special programming.

3.6.1. The programme for on-line look-up provides :

a) Verification for accounting purposes of the identity of the client and the extent of his privileges ;

b) Analysis of the query on the basis of which the computer will accept or reject it or, if necessary, signal an error ;

c) Execution of the queries via automatic look-up of the files and routing of the answers.

3.6.1.1. Because of its direct access operation, the Bank must be able to assess needs by compiling look-up statistics (how many times a term has been requested, which terms have been requested without any answer being obtained, and how often, etc.). The system must also be able to record automatically the number and length of queries per client.

3.6.1.2. The above operations are done by programmes governing another two-phase file which will give the account code of each subscriber together with a statement of his operations with the Bank over a given period, and also a statement of the operations of the whole system, which will make it possible for the Bank to define its deficiencies and the market's needs.

3.6.2. *Updating.* — In order to keep pace with the latest developments in terminologies, the Bank will be able to do on-line updating. Besides ensuring fast storage of information, this system also affords easy and efficient correction of the data. For this reason, there must be a special programme to :

3.6.2.1. correct (delete, add, withdraw, amend) all elements in the data unit (except keywords and supplementary entries) ;

3.6.2.2. update weighting indices for all keywords as well as for any data unit as a whole, without undue complication ;

3.6.2.3. make the accounting changes necessitated by :

i) admission of new clients ;

ii) changes in a client's privileges with the Bank (on the basis of data unit authorship as well as look-up, or look-up only) ;

iii) cancellation of a client's subscription.

3.6.3. On-line operation also necessitates the creation of an accessory file comprising a list of data sources by code and title (with bibliographical reference), a list of record authors by code and name, and finally the catalogue and the fields index provided to facilitate classification and retrieval.

3.7. *Sequential operation*

Although the first priority is direct-access operation, *Termium* nevertheless has provided for sequential operation ; where there is no urgency involved requiring ultrafast look-up, sequential processing (mass look-up) proves to be efficient enough and much less expensive.

3.7.1. *Termium* uses sequential access for billing subscribers, making lists of on-line-input data units with a weighting index of 1, and obtaining Flexowriter hard copies of one-line-input data units.

3.7.2. Similarly, the following alphabetical lists are accessible sequentially : terms by field or part of field, records by a single author, and records of equal weighting index ; the Bank's entire store of terminology ; unanswered queries. These factors can also be combined, as for example, to obtain an alphabetical list of all terms in the field of electronic acoustics with a weighted index of 1.

3.8. *Back-up files*

To provide for all contingencies, the system must have back-up files which will ensure that a complete copy of the Bank's basic files (3.5.3.1., 3.5.3.2., 3.5.3.3.), updated daily, are preserved on disks and reproduced on magnetic tape. These safeguards minimize the danger of accidental loss of data.

3.9. *Inquiry language*

The system has its own inquiry language which will greatly facilitate clients' access to the data by guiding them step by step towards answers to their terminology problems. The inquiry language, if it is not precise enough, could be costly and inefficient, besides being frustrating for the client. Thus, if the system requests it, a client can be called upon to clarify his query by identifying the field, so that he will not be swamped by a mass of useless information.

3.9.1. As regards the correction procedure, apart from the direct access method explained above, *Termium* has developed a special language for corrections which is related to the mass storage mode and has the same features of reliability and economy.

ROBERT DUBUC *

* English version : Patricia Claxton.