

The British Term Bank Prototype: user needs, modes of access, and co-operation in development

Lisa E. Price

*Centre for Computational Linguistics, University of
Manchester Institute of Science and Technology, UK*

INTRODUCTION

In 1980 the Centre for Computational Linguistics (CCL), at the University of Manchester Institute of Science and Technology, conducted an in-depth study into the feasibility of establishing a termbank for the UK which concluded with the specifications for a full-scale termbank. Since then interest and involvement in this field has remained strong at CCL, encouraged by the continued positive reactions of potential users who have received news of all major developments, and supported by the firm conviction that the UK is not only in need of a termbank but will indeed reap many benefits, provided such a termbank is constructed on the basis of sound systems design and regular consultation with future users.

I do not intend to extol the virtues of termbanks, as I believe their justification as a tool for translators, and indeed for other users, is presented comprehensively in existing literature. This paper describes the results already achieved in the development of a microcomputer-based termbank prototype, initiated (without any major financial backing) in October 1983 and due for completion in spring 1985, in order to demonstrate to British users the range of facilities that a system designed specifically for their requirements can and will provide.

I shall concentrate primarily on 'user considerations' that have been incorporated into the design, which I refer to in the context of this paper as the user interface, i.e. the means by which the user communicates with the system. Moreover I shall look in detail at certain aspects of the

interface which have been designed to accommodate the many and varied needs of distinct user groups. Although concerned mainly with software issues, which are of extreme importance as it is the software which allows the user to retrieve data, I would like to note that the processing of terminology for the prototype, although not discussed in the present paper, has also received a considerable amount of attention.

This paper has been written in three sections. The first outlines the database structure and software, whilst the other two consider the user interface, and future improvements to the prototype.

GENERAL OVERVIEW OF THE DATABASE

I implied above that the aim of the prototype is to demonstrate the capabilities and limitations of a termbank. Since a user need not be aware of the internal organisation of the database I do not intend to dwell on technical aspects of database structure; however, it is my view that the general explanation which follows provides the user with an initial insight into the system's capabilities and the degree of flexibility that has been built into the system.

Database Management System (DBMS)

The DBMS represents the nucleus of the database and serves to create, organise and manage the data. Its structure determines subsequent access to and use of the data stored within the termbank. Figure 1 presents the file structure of this system.

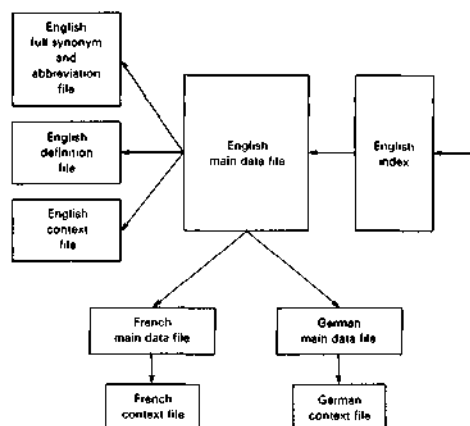


Figure 1. Connections between data files

These files, subdivided according to language, contain information which corresponds to an input fiche, a monolingual term record, representing a term-concept equation in English with foreign language equivalents. This record, outlined in Table 1, has been designed in conjunction with potential users with a view to satisfying their various requirements.

The DBMS developed for the prototype has been designed with a view to modification and flexibility. For example, although the existing model permits access to the database only via an English index, once indexes have been created for other languages included in the database it will permit searching via any language into any other arbitrary language.

Table 1. The term record format

HOUSEKEEPING DATA:	- Record identifier
	- Originator
	- Date of record
ENGLISH DATA:	- Language/country code
	- Entry term
	- Source
	- Note on usage
	- Scope note
	- Full synonym(s)
	- Abbreviated form
	- Synonym(s)
	- Conceptual link(s)
	- Subject field
	- Definition
	- Source
	- Context
	- Source
FOREIGN LANGUAGE EQUIVALENTS 1-7:	- Language/country code
	- Foreign language equivalent
	- Source
	- Note on usage
	- Scope note
	- Foreign language equivalent context
	- Source

Applications programs

Two applications programs have been designed and implemented in the prototype. The first, the data acquisition software, enables the terminologist to create, modify and delete term records. The second, the user interface, represents the language of communication between the user and the termbank, and differs from the first application in that a user may only retrieve information by this means, and cannot alter data in any way.

THE USER INTERFACE

The user interface designed for this system is based on a menu-driven system, a form of directed graph in which the actions of the user are largely controlled by the system. This technique has been chosen for two reasons.

Firstly, a menu system can offer a search language which, although not allowing the users to express a request in natural language, is relatively easy to learn and is less restrictive than having to formulate coded input. In such a system the user is confronted with a display offering a number of alternative actions. The choice of a particular action will lead to a further display containing a more refined list of possibilities. Each level of such a menu system provides the user with an option to exit, either to the next level or from the system itself, depending on how far the user has travelled down the menu network. With the exception of the outer level, no direct exit from the system is permitted. This design criterion has been chosen partly to avoid accidental LOGGING OFF, even though a check could be introduced at all levels and not just at the outer command level as is presently the case, and partly due to the fact that the menu is made up of only two main command levels; consequently, exit via all higher levels should pose no great inconvenience to the user. The menu technique has therefore been chosen to provide user-friendly access to the termbank and overcome an initial hurdle in demonstrating to potential users the services a termbank can offer by avoiding any unnecessary distractions.

Secondly, a well-designed menu structure can incorporate a wide range of facilities and still avoid a rigid and restrictive structure. Moreover, if well-structured, a directed network of choices of this kind lends itself to modification in future developments.

The menu system developed has taken this into account and, by avoiding deeply-nested structures, offers users a flexible service appropriate to their specific needs. This latter point made above is best exemplified and it is this

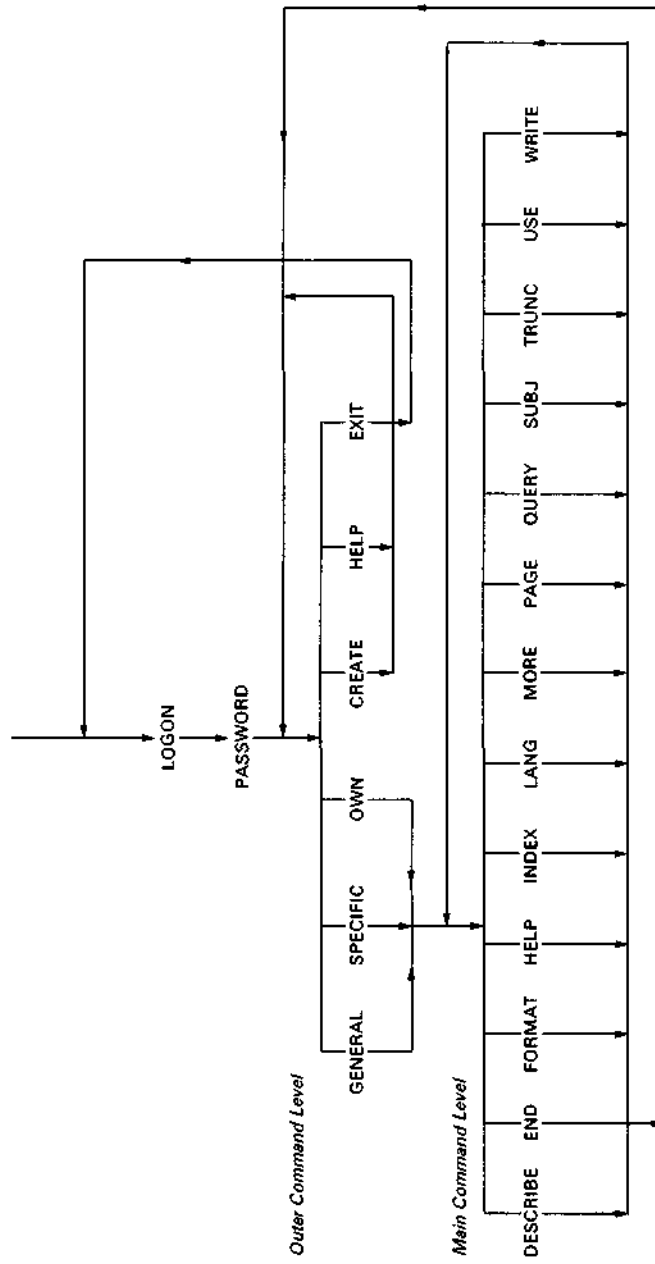


Figure 2. The menu interface

aspect of the termbank to which I will now turn my attention.

The Menu

The menu constructed for the user interface is presented in a simplified form in Figure 2. The interface operates on two main levels, the outer command level and the main command level, which correspond to two menu displays in Figures 3a and 3b respectively.

The outer command level (OCL) has been designed to allow users to specify their output needs during the initial stages of a session, thus saving time at the later stages of a search when they are interested in assessing the output and, equally important from the user's point of view, overcoming the need to enter a complicated output format code.

The user accesses the termbank by entering a user name and password. Provided the correct user name and password have been entered, the OCL menu is displayed. On entering Options 1, 2 or 3, the output formats associated with the specified user group, i.e. combinations of fields in the term record, are recorded by the system but displayed only when the user wishes to inspect the results of a search. Option 2 differs from 1 and 3 only by the fact that the user is 'led into' a submenu to indicate the specific user group to which the user belongs. This submenu, presented in Figure 4, is once again based on the feasibility study specifications. Each user name represents a number of user types, i.e. 'translator' includes translators, interpreters, technical writers and journalists, and the relevant information on these groupings should be consulted offline before use. Subsequent to entering any one of the above options, the user is directed to the main command level.

If users require an output format not currently offered by the system they may enter the fourth choice, CREATE OWN MODE, which initiates an input routine to create new formats. Users are at present limited to a maximum of four formats, each with a maximum six fields. These limits appear to be adequate as far as has been ascertained from user experience with other termbanks, but they may, of course, be altered. The user's password, entered at the beginning of the session, will now serve a dual purpose: it will be linked to the output formats specified in this subroutine, as well as being used for commercial and security purposes. On leaving this module, users are returned to the OCL where they may specify Option 3, OWN MODE. The connection to the new formats is then made for later use by recording those output formats associated with the user password in question.

```

* * * * *
GENERAL MODE ..... 1
SPECIFIC MODE ..... 2
OWN MODE ..... 3
CREATE OWN MODE ..... 4
HELP ..... 5
EXIT ..... 6
*****

```

Enter your choice:

Figure 3a. The outer command level

```

*****
GENERAL MODE
*****

```

```

D(escribe E(nd F(ormat H(elp I(ndex L(ang M(ore
P(age      Q(uey S(ubj T(runc U(se W(rite
Select command:

```

Figure 3b. The main command level

```

* * * * *
TRANSLATOR .....1
TERMINOLOGIST.....2
INFORMATION SCIENTIST .....3
STANDARDISATION EXPERT ...4
APPLIED LINGUIST .....5
PUBLISHER .....6
*****

```

Enter your choice:

Figure 4. The specific mode

The advantages of such a mechanism are twofold:

1. Users are offered flexibility since they may choose output formats tailored to their needs.
2. There is no intermediary involved in the creation of new formats which may slow down the process, frustrate users and reduce their feeling of responsibility.

An explanation option, HELP, is included at this level to provide general assistance, presenting information on the facilities offered by the termbank and an outline of the commands on both the outer and main command levels. Although intended to be unambiguous and easy to learn, a menu system may pose problems to the unfamiliar or experienced but infrequent user. Such HELP commands are therefore necessary to provide clarification of a system's facilities and modes of operation and are also included at the main command level.

The main command level (MCL) comprises a set of commands based on those put forward in the Specifications of a linguistic databank for the UK(1) which are considered to be the minimum set necessary to offer an adequate range of facilities. For the purpose of description, these commands may be subdivided according to their function, a subdivision which may not differ greatly from the way in which a user interacts with the system during a session. A logical division to make is that between 'retrieval-related commands' and 'explanatory commands', i.e. those commands concerned directly with the retrieval process and those providing assistance to users. I will concentrate on the former and discuss the remaining explanatory commands in less detail.

The Retrieval Process

Restrictive or 'pre-search' commands are built into the system to enable users to produce greater immediate precision of output by imposing restrictions on certain variables. The variables operating in this system are language and subject field.

At present, the LANGUAGE command may only be used to alter or to verify the target language, since the only index allowing access to the holdings is the English index. Alteration and verification of the source language is, however, recognised as being necessary if the termbank is to offer realistic modes of translation work, for example, and this facility is to be included in subsequent developments.

The LANGUAGE submenu indicates the current target language, i.e. either the language previously chosen in the

current session or, if the user has just begun to conduct a search, a predefined language stored by the system. The display is presented in Figure 5. The user may retain the language indicated for the current search or select an alternative from those offered. Whatever the choice, the user is returned to the MCL.

The SUBJECT submenu operates in a similar manner to the language menu, displaying the current subject and available options. However, whereas the language option is necessary, the subject restriction is optional so that if desired the user may conduct a search on the entire database.

At present it is feasible to display the subject headings in a menu. However, in a fully operational database the number of subject fields will be enormous, making their display in this manner highly undesirable and impracticable. A facility to access a classification file containing a list of subject headings will therefore be incorporated into the system. The notion of a 'pool', i.e. a subset of the database which may be a topic, the terminology for a particular translation etc., may also be included in the interface as a restrictive command. This concept was originally used by Siemens, (2) and, although developed in the first place for administrative purposes, serves as a useful search restriction.

```

*****
The current target language is: French
The following target languages are available
Retain current language...0
FRENCH .....2
GERMAN .....3
SPANISH .....4
Enter your choice:
*****

```

Figure 5. The language submenu

SEARCH COMMANDS

There are two commands which enable the user to search the database, QUERY and TRUNCATE. The QUERY command initiates a search on a string which may be a term (uni- or multiterm), full synonym (i.e. spelling variation of the entry term in a record) or abbreviation. Once a string is entered, the database or subset of the database is searched and the number of hits, i.e. term records containing the string, is displayed. The user then has the option of viewing the results by returning to the MCL or continuing with another search, for example if no records have been retrieved.

The TRUNCATE command is essentially the same facility as QUERY but offers greater flexibility by truncating the search term to capture variants of a term or different but often related terms. The possibilities offered are left-hand and right-hand truncation, and internal truncation, examples of which are given in Table 2.

The advantages of truncation facilities include being able to express several spelling variants in one simple command and being able to search on compounds in which an internal constituent may be substituted by one of several terms and

Table 2. Truncation examples

TRUNCATION	INPUT STRING	EXAMPLES OF TERMS RETRIEVED
Left-hand	*chloro	dichloro trichloro tetrachloro
Right-hand	document*	documents documentation documentalist
Internal	alumin*m	aluminium aluminum
	di*methane	dichloromethane dibromomethane

retrieving all the resulting compounds relevant to a user's needs. Care must however be taken when using truncation facilities, as several unwanted terms included in the database may also be retrieved. Having conducted a search, the user may now wish to view a term record and there are two commands available for this: `FORMAT` and `MORE`. The significance of the user mode options at the OCL is realised in the `FORMAT` command, which presents a selection of appropriate output formats in a menu display. The user need only indicate the desired option. The results are displayed and when ready, the user may return to the MCL. An example display is presented in Figure 6. The `MORE` command is designed to offer flexibility in addition to that already provided by the predefined user mode packages and the `CREATE OWN MODE` facility.

The user receives results on a term in a graduated manner, the specification for which has been defined and stored in sections. The user may read the information section by section and may terminate this operation at the end of any section if so desired. As far as it is possible highly informative data are given in the initial sections, followed by supplementary details to those already presented and the sections for foreign language information.

```
*****
GENERAL MODE
*****
```

The following output formats are available:

```
TERM + TARGET LANGUAGE EQUIVALENT + SOURCE ..1
TERM + TARGET LANGUAGE EQUIVALENT + CONTEXT .2
TERM + SYNONYM(S) + DEFINITION .....3
TERM + FULL SYSTEMATIC DISPLAY .....4
```

Enter option:

Figure 6. General mode output formats

Another command, WRITE, enables the user to receive a hard-copy record of the results obtained.

EXPLANATORY COMMANDS

Two facilities are provided for browsing through the contents of the database online.

The INDEX command allows users to display the terms held in the bank to verify a term's existence in the database or to check its spelling. The index may also act as an aid to searching for multiterm compound terms in cases when only particular constituents of the term are stored in the bank. The user may simply check a term's existence in the bank, in which case a positive or negative response is displayed as appropriate, or choose to retrieve a term in the context of alphabetically adjacent terms.

The PAGE command allows the user to page through the database either alphabetically or systematically. Although this is flexible in a prototype containing only a few thousand terms, the inclusion of alphabetical paging becomes less attractive and more expensive for a full-scale termbank and in subsequent developments it will only be available in conjunction with subject field restriction. The facility of restricting by pools of data, referred to earlier under the SUBJECT command, may also be included in this command.

Online assistance of a practical nature is provided by two further commands. DESCRIBE provides a brief but concise description of all command features at the MCL, and HELP provides assistance of a more specific nature, giving clarification of the main operations according to the user mode specified at the OCL. The latter command differs from DESCRIBE on two counts:

1. It provides detailed assistance on commands used for restricting and conducting a search.
2. It provides information on the output formats selected according to the user mode specified at the OCL.

Assistance provided by all of the above explanatory commands is, of course, available in hardcopy for browsing offline.

The one remaining feature of the MCL is the USE command which provides details on connect time and gives a cost estimate for the current session so far. As the operating modes of a British Term Bank have not yet been finalised, it is not appropriate to comment further on charges applicable to users in relation to services received.

However, it is envisaged that certain commands providing online assistance will be offered free of charge.

FUTURE IMPROVEMENTS TO THE PROTOTYPE

The preceding discussion has concentrated on the facilities incorporated into the system designed to offer users an acceptable and easily accessible service, with emphasis on the flexibility available for retrieving terminological data. Whilst the prototype does allow the users a great degree of flexibility, areas where improvements can be made, in addition to those already referred to in the preceding section, still remain and these require a considerable amount of research before they can be implemented.

The implementation of theoretical research

Compound analysis is a major area for research. Currently, if a user searches for a multiterm compound which is not stored in the bank, a negative response is displayed and the user must revise the search string, for example by searching for a broader term or on a recognisable part of that compound. This may result in a time-consuming, iterative process. Truncation facilities do enable the user to reduce the iterative process to a certain extent but this operation may also produce irrelevant hits. The optimal solution would be a termbank which could think on behalf of the user and search for constituents within a multiterm compound if the term as a single entry has not been retrieved, thus producing a more sophisticated system capable of conducting 'intelligent' searches.

There is already a certain amount of theoretical research available which might be applicable to developments relating to the online retrieval of compounds. Although I am not in a position to comment further on this matter, it remains for me to say that research implementing aspects of theoretical analysis is promising. One research team in Canada has already produced a 'mini termbank' which incorporates the analysis of a multiterm expression into its constituents, storing both the term and its constituents in network form, as a branching tree which indicates the syntactic analysis of the expression or term for consistent retrieval of both term and groupings of its constituents.(3) A further development, connected to the above and already tested for small subsets of terminology, is the notion of a network to represent the conceptual links among terms of a particular subject field.(4) If incorporated into an operational termbank, such a network would allow, for instance, browsing through the database by traversing the conceptual network built into the system.

Representing the relations between the constituents of a multiword term or expression and recording the conceptual links between terms are two methods which can be applied successfully to small databases. These processes both depend, however, on the manual analysis of data for input, an obvious restriction for large banks as far as manpower is concerned. Other considerations, such as the computer facilities necessary to store this information, must also be taken into account for larger banks.

With respect to developments in which a termbank can automatically analyse and assimilate a compound and its constituents consistently, given some theory of compound analysis for the number of languages stored in a bank, results will be less forthcoming than those comparatively modest, although in reality innovative and already sophisticated, models outlined above.

The development of a translator's work station

A second area for research which has many implications for potential users in the UK is the design as a possible 'spin-off' of present developments of a translator's work station, allowing users to access the British Term Bank holdings and simultaneously work on a translation. A combination of online termbank facilities and word processing with split-screen facilities is one possible configuration, with perhaps an additional facility to enter glossaries and notes online.

Although concrete proposals have not yet been drawn up, it is envisaged that, after the completion of the prototype, modifications such as word processing, electronic diaries etc. will be incorporated in an attempt to create a more complete online working environment for the user.

The involvement of users in developments

The final point I would like to make is that there is a need for continued and even greater involvement of potential users in the design of the British Term Bank.

The prototype has been designed on the basis of a thorough investigation into the requirements of users in the UK. However, since it is necessary to remain constantly aware of any changes in the work modes of users, feedback on the facilities already incorporated into the system will be invaluable. An up-to-date, in-depth analysis of the users, their working environment and their resources is necessary in order to determine the most effective and feasible modes of operating a termbank, particularly within the UK. This is essential for further development and future marketing of the British Term Bank and has implications for developments

of a translator's work station and the incorporation of sophisticated search facilities. Such a survey is scheduled for early in 1985. The fact that the prototype is being developed on what is by all accounts a widely available microcomputer may also play a major role in the future marketing of the termbank.

CONCLUSION

This paper has presented those aspects of the British Term Bank prototype most relevant to users and their requirements. Particular attention has been paid to the user interface and the search facilities incorporated into the system as a means of providing a taste of what is to come. In producing a multifunctional system of this kind, it is a major task to cater for a wide range of user needs. However, the design specifications have from the start taken account of the requirements expressed by users in the UK, and the resulting system offers, we believe, a friendly and flexible system.

The prototype has yet to be demonstrated, an event which we hope to arrange early in 1985. Only then will we see the true results of our efforts, and the reactions of the users, which we trust will be favourable. The exact nature of subsequent developments is difficult to define. They may include more sophisticated design based on theoretical aspects of terminological research, but will undoubtedly include further modifications to refine existing features. However, perhaps most important of all, subsequent developments relating to the interface, modes of operation and other aspects of database design will involve continued interaction with the users. This will, we hope, result in greater communication and co-operation between ourselves as termbank designers and you, the users, forming a sound basis from which we may embark on other forms of co-operation. More immediately, however, we will have achieved our initial aim: the production of a British Term Bank prototype tailor-made for its users.

REFERENCES

1. SAGER, J.C., and McNAUGHT, J. Specifications of a linguistic data bank for the UK. British Library Research and Development Report No. 5644, 1981. Manchester, UK: Centre for Computational Linguistics, UMIST.

2. HOHNHOLD, I. Die Terminologie-Datenbank TEAM im Sprachendienst Siemens München. Mitteilungsblatt für Dolmetscher und Übersetzer, 30(3), Mai/Juni 1984, 1-10.
3. BAUDOT, J., and CLAS, A. A model for a bilingual terminology mini-bank. Lebende Sprachen, 29(2), 1984, 49-54.
4. McNAUGHT, J. The role of terminological relationships in encoding and accessing systematic multilingual specialised computer dictionaries. Multilingua, 1(1), 1982, 53-54.

AUTHOR

Lisa Price has now left CCL, but any enquiries regarding this research should be directed to the Centre for Computational Linguistics, Institute of Science and Technology, University of Manchester, PO Box 88, Sackville Street, Manchester M60 1QD, UK.