Meta
Journal des traducteurs
Translators' Journal

META

An Intelligent Computational Environment for Terminology and Text Handling

Anthony F. Hartley and Emmanuel J. Yannakoudakis

Volume 32, Number 2, juin 1987

Vers l'an 2000. La terminotique, bilan et prospectives
Objectives: Year 2000 Terminotics. State of the Art, Prospects for the Future

URI: https://id.erudit.org/iderudit/003036ar
DOI: https://doi.org/10.7202/003036ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)
1492-1421 (digital)

Explore this journal

Cite this article

Hartley, A. F. & Yannakoudakis, E. J. (1987). An Intelligent Computational Environment for Terminology and Text Handling. *Meta*, *32*(2), 139–148. https://doi.org/10.7202/003036ar

AN INTELLIGENT COMPUTATIONAL ENVIRONMENT FOR TERMINOLOGY AND TEXT HANDLING

ANTHONY F. HARTLEY AND EMMANUEL J. YANNAKOUDAKIS
University of Bradford, Bradford, U.K.

1. INTRODUCTION

The model we propose here for the termbank of the year 2000 is a sophisticated relational database integrated into a cluster of computerized modules for text analysis and text creation. It is designed to meet the varied needs of a wide range of professional linguists. Text analysis programs and expert system interfaces upstream of the termbank automate much of the time-consuming work of data capture at present performed manually. Downstream the bank is able to interface with such workstations as word-processors with terminology windows, controlled-language writing aids, "talkwriters", MT systems and other intelligent knowledge based systems.

At the hub of the network the bank contains in addition to the fields commonly held in existing linguistic data banks semantic information which serves to relate the linguistic entities stored and to enhance the number of modes of access and retrieval. Within our framework these relations can be invested with a statistical significance. Since a majority of termbank users are concerned with the comprehension or production of *texts* rather than of isolated words, the database enables the building not only of formal and conceptual links between terms but also of links between terms and texts.

The starting point for our thinking is a brief review of the requirements of likely users. We then describe the main thrusts of current computational research and outline the environment which will be available by the turn of the century for managing terminological and textual data structures. Finally, we point to some implications for linguistic research and development.

2. END USERS AND END USES – AN OVERVIEW

2.1 OUTPUT

End users can be characterized as having a monolingual or a multilingual focus. In the first category come, principally, standardization specialists and technical authors. A frequent request will be for *sets* of terms and definitions in order to establish controlled vocabularies designed to promote consistency and eliminate ambiguity in standards and texts. Consequently their preferred mode of interrogation of the database will be by systematic searches. These may be broadly classed as thematic (by subject field) or conceptual.

Multilingual users are, essentially, translators. They share with monolingual users the need to retrieve sets of terms, although the most frequent search path is likely to be that which leads from a SL term to one or more TL equivalents. They may be interested also in comparing definitions for candidate equivalent terms[1]. Increasingly the transla-

tor is interested in retrieving terms in context in anticipation that both the term unit itself and some of the surrounding textual matter will be utilizable in the TL document.

Another translating activity is that practised by subject specialists wishing to understand the gist of foreign language texts in their field. This goal can be achieved by the interlinear listing of translated terms within the original document — the Canisius system[2].

A further likely growth area is the interfacing of linguistic data banks with MT systems. Of the fields is the traditional termbank record only those of "syntactic class" and "foreign language equivalent(s)" are ultimately relevant within a MT dictionary. The latter contains additionally much richer morphosyntactic information and, increasingly, semantic information. The present model incorporates for the purpose of optimising storage and retrieval searches descriptive and classificatory data of this kind which would be available for exploitation by MT system managers also.

Terminologists, as members of a team of database administrators (DBAs) operating at the conceptual level of organisation of the termbank, are concerned directly with all the end uses described above. As language planners they are interested in the formal properties of term units. They expect also the possibility of searching from concept towards (possible) designation and of retrieving synonyms, tasks which can be accomplished only if a pertinent conceptual access key is provided.

Ergonomically, querying the database must be enabled by transparent commands which approximate to natural language. The linguistic data called up by the user must be displayable simultaneously with any textual material it is related to, and transfer between the two screen files must be simple. We take it as axiomatic that by 2 000 all texts to be processed within our integrated system will either have been created on an electronic medium or have been read onto one by means of an intelligent optical character and image recognition (OCIR) device.

2.2 INPUT

We maintain that the practice of preparing new entries manually or at least on some device which is external to the database itself prior to updating the base in batch mode is outdated and inefficient. Working on-line, the terminologist can be prompted and guided by the systems itself on the basis of the data and information already resident within it.

Input has tended to be the preserve of the DBAs, who mediate any input proposals originating from end users. However, these necessary quality controls create a bottleneck which retards the growth of a bank, a drawback which some private users of EURODICAUTOM, for example, are seeking to remedy by the addition of an on-line direct input system[3]. Ways must, therefore, be found to accelerate growth while maintaining the quality, consistency, integrity and accessibility of the data. The increased complexity of the data which we propose to store makes a solution at this level all the more imperative.

In this context we recognize that one cannot afford either to jettison or to duplicate the holdings of existing linguistic data banks. Thus an issue which must be addressed by the architects of any new development is compatibility of data formats. Compatibility is not identity ; exchanges between banks will inevitably result in the presence of incomplete attribute values within entities.

3. COMPUTING — STATE OF THE ART

Computer science and technology are undergoing a continuous evolution and this has become particularly obvious during the last three years, especially with the introduction of the ESPRIT (European Strategic Programme for Research and Develop-

ment in Information Technology), and the Alvey program of research of the British government.

There is talk of the 5th generation of computer systems and although this has not even crystallized in the minds of the researchers themselves, certain research teams have already started discussing the follow up or what we can now refer to as the 6th generation.

Current research falls under five major areas which we can only characterize as transitional realms to a more advanced approach to storing and retrieving information.

3.1 OFFICE SYSTEMS

The focus of attention here becomes the traditional office as we know it with its paper work, primarily involving the preparation of documents/letters, communications, filing, retrieval, etc. Current research aims to analyse these functions and to design efficient structures and procedures for carrying out more or less the same functions but with the aid of the computer.

3.2 MAN-MACHINE COMMUNICATION

The aim here is to develop a more natural method for man-machine interaction and generally to overcome the traditional and clumsy tool of the typewriter or Visual Display Unit (VDU). Examples of alternative interfaces are voice input and output, touch, smell, vision and therefore sign language. The latter category is otherwise referred to as "image processing".

3.3 ARTIFICIAL INTELLIGENCE AND KNOWLEDGE BASES

Current research gives emphasis to the design of "local" rather than "global" intelligent systems aiming to transfer human knowledge upon electronic devices and thereby manipulate this further. Thus, individual areas of human knowledge, such as haematological disorders, mineral exploration, detection and correction of misspellings, are dealt with in isolation and appropriate knowledge bases are designed for general use. We do not, as yet, have a global and outline methodology for analysing nor, therefore, for operationalising intelligent activities.

3.4 SOFTWARE ENGINEERING

Generally speaking, the end product, whatever this may be, is as good as the tools and techniques which have been employed during its design, implementation and testing phases. We refer here to the high level languages and system software around which all "software packages" materialize and evolve. Research in this area aims to eliminate *ad hoc* techniques for the design of systems and to adopt an engineering approach in order to increase reliability, conformity with requirements, and ability to evolve with new encyclopedic knowledge.

3.5 HARDWARE ARCHITECTURES

Current research is heavily dependent on advances in Very Large Scale Integration (VLSI) of circuits on silicon. There is an ever increasing need for faster computers with more logic elements capable of storing an ever increasing amount of data. It is therefore the size of the databases as well as the processing complexity which in turn necessitate faster computers. Also, the ability to process data in parallel, rather than in sequence as has been the case so far, as well as concurrently, form important areas for the next generation of computers.

4. THE BIOCHIP AND FUTURE COMPUTER ARCHITECTURES

The above five areas together are usually termed as "the 5th generation". We claim that the results will not produce a breakthrough but rather an important optimization of current technology. A number of problems we face today, such as the need for larger and more intelligent primary memories, will not be dealt with satisfactorily.

A rather more promising area of research appears to be the use of "biochips", that is living organisms which act as memories and processors at the same time. Regardless of what happens with the 5th generation, the 6th generation should and could employ molecular structures instead of chips as it is only through these that speeds of 1 000 000 times faster then chip-based technology can be achieved. Moreover, a cubic centimeter of melecules can store 10 000 000 times more data than the largest chip-based component conceivably possible.

Experiments are already being carried out, usually behind closed doors, and various molecules are being investigated for use as switches to replace chips. Examples are the haemoglobin and the haem, which can be controlled effectively. Another example is the solaton, which exemplifies a peculiar structure as electrical current passes through its body. Generally speaking, protein-based molecules are ideal for use as memories because they tend to organize themselves rhythmically and consistently.

We envisage an architecture which integrates chip technology with biochip technology in order to establish "associative storage devices" otherwise referred to as "cellular logic devices" along the lines of database machines[4]. Here, we foresee a bio-processor for each storage slot (an equivalent term of present day technology is the "track" of direct access devices) enabling concurrent and parallel processing. For example, the request "find related terms to DEVICE" activates the same search operation in parallel, so that the entire storage device is searched simultaneously. This concept is certainly not new and the term Content Addressable File Store (CAFS) is already common in the literature.

In conclusion, we strongly believe that the "biochip" is the only way to a breakthrough and therefore to the design of intelligent structures which can store, organise, classify, correlate, cluster, associate, and finally present information in different forms. By the year 2000 the biochip will be available for use not only as a component for advanced information processing machines (computers) but also as a component for human extensions and replacements.

Expert systems is another exciting area of computing and this must be explored with the possibility of providing a more user-friendly interface[5]. Generally speaking, an expert system emulates and in most cases outperforms the human experts in a given area of knowledge, in our case term analysis and retrieval. Moreover, an expert system can explain its decision, or rather the path it has followed before coming to the stated conclusion.

There appear to be three stages to the creation and use of an expert system : (a) *knowledge elicitation* — the extraction of knowledge from human experts in the form of rules, or through deduction, or through induction ; (b) *knowledge representation* — the use of appropriate data structures and languages (*e.g.* PROLOG) to store knowledge ; (c) *inference* — the process whereby we infer new knowledge from old through forward or backward reasoning.

In conclusion, the need to store human knowledge and to process it intelligently and with reasonable response times necessitates research beyond chip technology. The latter appears to have reached its upper limits and further research can result only in minor refinements. Biochip technology is the only way forward.

5. THE RELATIONAL APPROACH TO DATA MANAGEMENT

Understandably, the user wishes to remain oblivious to the physical organization of the data and to be informed rather than bemused by its conceptual organization. This means, then, that the organizational burden must be carried by the system. The relational approach promises simplicity at both user and system design levels.

Relational technology in its broadest sense, implies the use of commonly available features or items of knowledge as a means to link different entities together. Relationships are thus establised by default rather than by artificial links or pointers as has been the case with traditional information systems. This principle is particularly relevant for the termbank system we envisage whereby attribute-value, rather than attribute-name, forms the necessary link. Term associations and relationships thus are established and identified through concurrent associative mechanisms[6] which traverse the inherent network.

A number of criticisms have been levelled at the idea of using a database management system (DBMS) and, by implication, a relational approach to manipulating terminological data[7]. Admittedly these comments are aimed at existing DBMS software and at the 1984 conception of termbanks, but they are serious enough to merit an argued response.

The first claim is that the data held in current banks is not sophisticated enough to warrant a DBMS approach. The complexity of our proposed network implemented through relational principles surely answers this objection. Besides, it is not the size of each "relation" in an DBMS which determines its viability or otherwise. What is more important is the complexity of operations or associations established among the attributes and indeed these are vital for a terminological database.

A second remark maintains that a DBMS realizes its true potential only when there is a significant proportion of data to be modified, that is a high volume of dynamic data. Since, it is advanced, the data within a termbank is overwhelmingly static in nature, to manage it with a DBMS is again inappropriate. Within our model, however, every new entry has an incremental effect on statistical weightings associated with terms and their attributes, which increases the proportion of dynamic data.

The terms "dynamic" and "static" databases warrant further clarification of their use in our environment. A physical structure must be able to cope with upgrades in both software and hardware and therefore adapt to the ever changing technology. Although disk technology (Winchester) appears to be the current favourite, laser bases disks and general light emitting devices are becoming very popular as computer peripherals. A database therefore which cannot integrate with new technology and evolve with it is bound to become obsolete quickly. Similarly, new developments in software engineering imply that substantial portions of current system software and application software particularly must change in order to improve performance.

It is true that most DBMS have so far been used for the design of formatted data, that is data which can be defined in logical units around the concept of record or entity. It is also true to say that string processing, and general document handling has been left to specialised packages, such as word processors. However, the power of the database approach lies not so much in the type of information it handles, as to the type of processing functions if offers, primarily, the dynamic associations between different units of data.

Recent research has proved that the DBMS approach is indeed equally applicable for string/document handling functions[8]. The techniques proposed offer integrated solutions for database managements and information retrieval in the traditional sense.

More specifically, the relational approach has been investigated and its suitability for document handling has been firmly established[9].

The relational approach therefore becomes the means for storing and retrieving information. But as a front-end to the database there is an expert system which helps the user interrogate and navigate through the knowledge base network.

5.1 CONCLUSION

The environment described here offers a realistic medium for creating and managing a multilingual database of several million "records", each with a variable number of attributes including textual data. Additionally, image data complements the linguistic data which is handled through commonly available control commands in order to amplify user-views or sections of the knowledge base. Moreover, the prospects are that, despite the multiplicity of views available to different classes and sub-classes of end user, response times will be acceptable.

The model caters for a multiplicity of user-views which can be controlled under a "distributed database environment" through local and remote network access operations. Individual sections of the database pertinent to a user can exist as virtual or permanent views. Also, the DBMS can collect statistics on usage of specific terms and contexts for individual user-views and thereby accumulate information and "learn" about user needs.

6. CREATING RELATIONS BETWEEN LINGUISTIC ENTITIES

The cursory survey above of the types of output requested by users was sufficient to demonstrate an important need to establish relations between term constituents on a purely formal basis, as well as between terms and contexts. These are the least problematic of the relations we envisage.

6.1 FORMAL

The ability to match the form of a query term to that of a likely corresponding entry in the linguistic database even where the two are not identical in form ("valves" *vs* "valve", for example) is a widely implemented facility.

The internal level, otherwise refered to as the "storage schema", can employ intelligent compression techniques in order to bring formally related terms together. For example, the terms "connect", "connection" and "connectivity" can share the common substring "connect-" in order to optimize in both storage and retrieval time. Research at the University of Bradford has already succeeded in achieving over 55% compression on over 93 000 words selected from the Shorter Oxford Dictionary.

Research on coding and compression for identification purposes has established a number of practical techniques for linking, not only formatted information, but also unformatted and highly unstructured database entities[10]. The aim here is to create a blueprint for each entity (the term in our case) which serves as an identifier and where the level of discrimination between different terms can also be varied in order to satisfy a given unique user-request.

Systems which can retrieve one or more constituents of a string which has not been terminologized have been shown as a potentially invaluable aid to translators, not least in that they can provide models for neologizing[11]. Any termbank software must incorporate this "near match" feature otherwise referred to as the search for "the nearest neighbourhood".

Our specification is that queries should be promptable direct from running text. Consider a SL translation text displayed in one zone of a VDU screen and in another, information such as a TL term which has been extracted from the database. Complete

zones or sections of them can of course be manipulated independently and merged with each other as and where necessary. Touch sensitive screens and appropriate commands can now be used to activate functions such as "search", "isolate", "transfer" and "concatenate" without recourse to the keyboard.

In the case of heavily inflected languages the pattern matching strategies require the adjunct of algorithms for morphological analysis in order to relate text forms to canonical forms. This represents one of the text analysis modules to be associated with the termbank of the future. It is also a prerequisite for the automated identification of term units.

6.2 CONTEXTUAL

Translators often like to see a term in an actual context[12]. The technology alluded to immediately above offers the writer the possibility of "poaching" and "cannibalizing" whole stretches of text. The option of text composition by "text retrieval"[13] is probably more realistic in large institutions where there is much updating and quoting of existing documents. A refinement of this technique is the option of specifying the co-occurrence (or non-occurrence) of one or more other words in a given contextual span. Text storage in a large and commonly available pool is probably inappropriate for a database serving users from many different organizations with diverse interests.

Even if contextual data is restricted to one field per record, a given term may appear not only in the context field of its own record but also in that of other terms. This demands the facility of searching rapidly through possibly millions of words of running text. These conditions may still yield a large number of "hits" which need to be sorted into an order likely to correspond to the user's own priorities. One approach to this problem is through conceptual relations between terms.

7. CREATING RELATIONS BETWEEN CONCEPTUAL ENTITIES

The two previous types of relation are linguistic rather than terminological. They will capture the link between "lubricant", "lubrication" and "lubricator". But to establish the link between these "oil" and "grease" requires a truly terminological, that is concept driven database. We reject the model of a two- or three-tier classification as inadequate for representing the interrelationships of human knowledge and activity.

The originality of our approach lies in adopting instead a thesaural type approach which allows for polyhierarchical relations and is extendable, on the lines of the BSI ROOT thesaurus[14]. The price of generating conceptual sets is a more sophisticated input software[15] which prevents anarchy by requiring guided responses about all new data, which is then more richly characterizable by the system's own redundancy and inference rules.

From the network we propose it will be possible, of course, to create dynamically any hierarchic relations, which can be either virtual or permanent.

7.1 DATA CHARACTERIZATION

The necessary characterization of terms in the bank ranges along a single scale from the thematic via the extensional to the intensional. For each established domain a structure to a level of complexity analogous to that of ROOT must be established.

7.2 AN INTELLIGENT INPUT INTERFACE

We view the input module(s) as an expert system whose local knowledge consists of (*a*) rules about terminological relations in general, and (*b*) structural representations of particular realms. The system is an interactive tool for the construction of a knowledge base.

An example is the MICROSYNICS expert system which allows the user to set-up and conduct a dialogue in the form of a network of nodes. The structure of a "node" depends on the designer (for example, in our case it can become the term of the termbank), but we can identify three major components : (*a*) textual data pertinent to a node ; (*b*) calls to independent software tasks to cater for specialist needs, computations, or term associations ; (*c*) specification of successor nodes or exit points with terminal nodes, that is, nodes which can be presented to the user. The nodes are kept separate from the associated tasks and this makes it easy to modify the termbank or generally maintain its structure, consistency and integrity.

For data input the prestructured network is displayed on the screen, with a focus which narrows progressively until the term can be assigned to the most specific available node, or descriptor. At this stage the terminologist is shown other terms in the immediate vicinity and invited to declare, from a menu, one or more ontological relations (generic, partitive, etc.) between the new entry and at least one of its neighbours. At this level of description the system is able to generate outline extensional definitions of the form "Types of A are X, Y and Z" or "Parts of A are B and C", on a model already developed by McNaught[16].

The usefulness of this systematically formatted information is beyond question. However, advances in computer graphics raise the question whether it might not be more lucidly expressed as captions on 3-D images. The image representation of a term or group of terms can also be offered to the user on request and manipulated further (*e.g.* merged within a standard or a user manual). These need to be stored, but the termbank of 2000 can be expected to contain a large amount of image data in any case.

In order to approximate to the intension of a term, each descriptor in a network is further specified at a lower level from a set of *semantic* — or *terminological* — *primitives* of the type PHYSOBJ, ACTION, LOCATION, CONTAINER, INSTRUMENT, TOOL, GAS, etc., themselves organized in classes and related by redundancy rules. These primitives are automatically "inherited" as characteristics by all terms attached to a given descriptor.

In order to move towards a unique specification for each term, the terminologist is now guided by the system to assign restricting characteristics to the concept by selecting appropriate values for facets such as purpose, form or material. One can speculate whether the system could not operate *directly* on existing textual definitions of the concept. Certainly, given that genus and characteristics are known to the system, it will generate outline intensional definitions following a fixed format.

This recourse to a thesaural "shell" in conjunction with classes of terminological primitives opens several possibilities. Operating at the upper echelons of the hierarchy, one can extract sets of terms which correspond to traditional thematic groupings by subject field, delimited at will. Within these parameters systematic, alphabetic and contextual presentations are possible. Working from the lowest level, one can conceive of searches to see if a designation exists for a given concept, prompted by questions of the form "Is there a tool for purpose X ?" It is also conceivable that the primitives associated to term units will be of direct relevance within MT systems.

Furthermore, it is hoped that this more detailed semantic characterization of terms will help overcome the difficulties inherent in automatic term recognition. The limitations of merely syntactic and statistical approaches to term formation are well known. The terminological primitives, being constructs designed to represent knowledge in circumscribed fields, may well provide the (artificial) "pragmatic" information needed to recognize the lexicalized status of strings of potential term constituents[17].

7.3 THE STATISTICAL DIMENSION OF THE NETWORK

We envisage also that this semantic and relational information can be brought to bear by associative scanning devices[18] designed to peruse text for the purposes of term identification and clustering prior to any look-up. They will build from known terms in the text a statistically weighted conceptual profile to aid the extraction of the likely realization of homographs and of candidate term units.

Our proposed network becomes the nervous system of the termbank which can also carry statistical data on levels of connectivity between different terms. Moreover, individual attributes can be linked together to form subnetworks for further association through decomposition. We do therefore envisage an environment whereby the connectivity level or strength of connection (weight) between terms or attributes forms the basis for all processing functions. This environment thus involves the following distinct stages :

a) calculate the weight automatically at the setup stage,
b) use the weight for associative functions,
c) use the weight to decompose the network into manageable units for storage, optimization or association functions,
d) recalculate and maintain the weight dynamically as the term bank is used.

Task (a) appears to be more difficult than the rest, but recent research has proved that a mathematical model[19] can assign connectivity weights between different entities and therefore proceed to decompose the network into appropriate groups with strongly related items (task (c) above). In fact, depending on the criteria used to establish the connections, the designer can employ the mathematical model in order to bring conceptually related terms together. The model[20] can of course be extended to perform task (d) above and this is currently being investigated at the University of Bradford using bibliographic attributes.

The overall design problem can now be expressed as a linear undirected graph G(M,L) where M is a set of m terms corresponding to m termbank units and L is a set of undirected and weighted links representing the existence or absence of relationships between pairs of terms. Take for example the case where $M = \{t_1,t_2,t_3,t_4,t_5,t_6\}$, $m = 6$ and $L = 4.2.4\ \{t_1\text{-}t_2, t_1\text{-}t_3, t_2\text{-}t_3, t_2\text{-}t_4, t_2\text{-}t_6, t_2\text{-}t_5, t_4\text{-}t_5\}$ t_i being a term. A binary random variable y_i can be associated with each term i, $i=1,2,...,m$ and the entropy function

$$H(M) = -\sum_{i=1}^{m} P(t_i) \log P(t_i)$$

can measure the amount of information contained by the termbank, where t_i is the state

$$\{y_1 = x_1, y_2, \ldots, y_m = x_m\},\ x_i = 0 \text{ or } 1\ ;\ i=1,2,...,m.$$

This criterion is expressed as the minimisation of an objective function representing the information transfer between subgraphs and is given by

$$\min \Sigma \rho^2_{ij}$$

where ρ_{ij} is the correlation coefficient between y_i and y_j expressing the weighted links, and the summation is taken over all those pairs of terms separated by an arbitrary partition π of M into subsets $S_1,S_2,...,S_n$ such that

$$\exists S_i,S_j \ \ni\ S_i \cap S_j \neq \emptyset \text{ and } M \subseteq \bigcup_{i=1}^{n} S_i$$

On the one hand this technique enables the isolation of sets of data. Given for example, the form "element" one can explore the different clusters of its attributes ant its behaviour in various fields of human activity — physics, meteorology, mathematics,

etc. On the other hand it is a means of identifying intersecting attributes between different terms.

8. CONCLUSION

We are proposing an integrated approach to document flow, from the analyses preceding data input to the generation of documents which exploit the database. It draws extensively on advances in office technology and man-machine communication. More importantly, the model we envisage can be described as intelligent in the manner in which it elicits knowledge about terms and subsequently applies this knowledge to textual analysis.

This apparently daunting ambition can be put in perspective by comparing it with a project already underway in Austin, Texas to represent in a computer the encapsulated knowledge of an encyclopaedia.

Software tools appropriate for manipulating terminological attributes — including the "primitives" we have described — exist and are being refined. The urgent problem is for the experts — specialists, terminologists and translators — to isolate the pertinent attributes which will characterize a particular domain. Sparck Jones describes an analogous task as "not just the computational linguist's challenge, but his nightmare"[21]. We believe that these troubled dreams will be calmed by the dawning of new research in terminotics.

REFERENCES

1. SAGER, J.C. (1982) : "New Approaches to Specialised Dictionary Consultation in the United Kingdom", *Lebende Sprachen*, 2, pp. 59-63.
2. SAGER, J.C. and J. McNAUGHT (1980) : "Feasibility Study of the Establishment of a Terminological Data Base in the U.K." (Stage I, Part 1), *Manchester, CCL/UMIST* (nr. 81/8), p. 3.
3. STELLBRINK, H.-J. (1985) : "Efficient Terminology Work in a Medium-Sized Translation Service", *Multilingua*, 4-3, pp. 157-164.
4. IEEE Computer Society (1979) : "Special Issue on Database Machines", *Computer*, 12-3.
5. HAYES-ROTH, F. (1984) : "Knowledge-Bases Expert Systems", *IEEE Computer*, 17-10, pp. 263-273.
6. IEEE, *op. cit.*
7. HENNING, J.M., J.C. PERRAUD, B. PEUCHOT and M. SCHNEIDER (1984) : *Logiciel de gestion de banques de données terminologiques*, Paris, MIDIST (nr. 83 3 94 0187), pp. 57-59.
8. SCHEK, H.J. and P. PISTOR (1982) : "Data Structures for an Integrated Database Management and Information Retrieval System", in *Proc. 8th Int. Conf. on Very Large Databases*, pp. 197-207.
9. STONEBRAKER, M., H. STETTNER, N. LYNN, J. KALASH and A. GUTTMAN (1983) : "Document Processing in a Relational Database System", *ACM Trans. Office Inf. Syst.*, 1-2, pp. 143-158.
10. YANNAKOUDAKIS, E.J. (1986) : "Intelligent Matching and Retrieval for Electronic Document Manipulation", *Int. Conf. on Text Processing and Document Manipulation*, Univ. of Nottingham.
11. BAUDOT, J., A. CLAS and M. GROSS (1981) : "Un modèle de mini-banque de terminologie bilingue", *META*, 26-4, pp. 315-331.
12. SAGER, *op. cit.*, p. 62.
13. ARTHERN, P.J. (1979) : "Machine Translation and Computerized Terminology Systems — A Translator's Viewpoint", in B.M. Snell (ed.), *Translating and the Computer*, Amsterdam, North-Holland.
14. B.S.I. (1981) : *ROOT Thesaurus*, Hemel Hempstead, British Standards Institution.
15. BAUDOT *et al., op. cit.*, p. 320.
16. McNAUGHT, J. (1982) : "The Role of Terminological Relationships", *Multilingua*, 1-1, pp. 53-54.
17. SPARCK JONES, K. (1985) : "Compound Noun Interpretation Problems", in F. Fallside and W.A. Woods, *Computer Speech Processing*, London, Prentice-Hall, pp. 363-381.
18. IEEE, *op. cit.*
19. KOUVATSOS, D.D. and E.J. YANNAKOUDAKIS (1982) : "A New Approach to the Design of Structured Bibliographic Records", *Information Technology- Research and Development*, 1:4, pp. 285-300.
20. *Ibid.*
21. *Op. cit.*, p. 380.