MULTIFUNCTIONAL LEXICAL DATABASES

Antonio Zampolli*

Automation has revolutionized the concept of dictionary both in the formal aspect and its potential use. Computerized dictionaries have now been recognized as being complex data bases with multiple access points which can have a variety of data structures, the full extent of which is as yet unexplored. Simultaneously the possible uses of dictionaries have multiplied so that the production of printed dictionaries has now become a matter of development rather than research. Dictionaries are now considered to be valuable sources, capable of providing sophisticated information for uses in word processing, text editing, machine translation, intelligent knowledge based systems and research in artificial intelligence.

While it is important to provide a favourable climate for innovation by supporting experimentation, it is nevertheless equally important to consolidate collective experience and good practice into norms and guidelines for methods, procedures and data structures on a regular basis. Only in this way can the growing information industry meet the expectations of users and at the same time both fund research and have a clear indication of areas where applied research is most wanted.

An ongoing dialogue is therefore desirable between all parties concerned in the design, production and use of dictionaries for the purpose of assessing progress, identifying areas where efforts directed to harmonization and regulation are likely to succeed, to stimulate cooperation and the dissemination of new ideas and techniques.

In this framework, it may be considered useful to promote activities on specialized aspects of this field, with some of the following tasks :

- to define the various type of computerized lexical and terminological resources,
- to establish the state of the art in automated dictionaries,
- to investigate the "dictionary entry" from the standpoints of individual form, sources in both text and data banks, links among the structural elements of the entries within a dictionary, and the implications of multilingualism for entry design ;
- to consider the possibility of exchanges of lexical data ;
- to assess the possible utilizations of lexical databases for different applications, including computer-aided translation;
- to consider the impact of new technology on dictionary design;
- to identify areas susceptible to harmonization;
- to identify new trends in dictionary research;
- to make recommendations concerning priorities in research and development.

* *Professeur de linguistique à l'Université de Pise et professeur de linguistique générale et appliquée à l'Université de Genève.*

The concept of multifunctional linguistic database system (MLDBs)

From automated dictionaries to multifunctional data bases

The characteristic of "multifunctionalism" is given by the capability of the linguistic data base (LDB) to be used for several purposes either in different applications or by different types of users. By developing the notion of "dictionary server", one should tend towards having a single central core of "neutral" dictionary data, which can be properly accessed by many different interfaces according to the needs of all the range of possible applications. Different external procedures must be able to use different portions of the dictionary content in specific applications, and each user (human or procedural) views only the data which are the relevance for him/it, while almost ignoring anything concerning the internal physical organization.

The characteristic of "multidimensionalism" is essentially given as a result of the property of multiple access. By implementing multiple access points one can search different word aspects by simply following different paths within the LDB. When the original data can be viewed under a variety of different perspectives, the important effect of "multiplying" the information offered by the same set of source data is obtained.

Moreover, it is possible to create, as by-products, many virtual secondary sub-lexicons, comprised of specifically selected portions of the dictionary, as terminological sub-dictionaries, synonyms, dictionaries, thesauri, etc. They in fact, in a well-structured and comprehensive LDB, differ only in the way the original data are selected, sorted and interrelated.

Multifunctional linguistics data bases in theoretical and applied research

MLDBs should tend towards reaching structured and finalized information at many descriptive levels. They make effective use of increasingly sophisticated computational technology, and they themselves appear to be the most promising field of research since they aim at considerably extending and developing the scopes of early machine dictionary projects, owing to their wide range of applications areas.

LDBs have a very large spectrum of potential users:

- human users (specialists and lexicographers, lexicologists, linguists, or normal users for everyday dictionaries look-up) ;
- "procedural users" (*i. e.* other programs or complex systems for which the LDB is one of the components)

A MLDB must therefore be designed as flexible as possible, both from the computational and the linguistic viewpoint. Its implementation must facilitate utilisation in the many applications in the field of "language industry" requiring the use of a lexicon : *i. e.* practically in all cases, because whatever task we are accomplish where natural language plays a role, we have at some point to cope with words and therefore to face the problems of lexical access.

The possible areas of application for MLDBs range from simple spelling verification to lemmatization, or to lexicological research and lexicographical practice (e.g. to improve the coherency and consistency in dictionary making), or to a number of computational linguistics applications, such as parsers, question-answering systems, man-machine communication, machine-(aided) translation, language teaching, etc.

The MLDBs can in fact be conceived as the point of conjunction of the different types of information to which any natural language processing system must have access : morphological, syntactic, semantic, pragmatic, conceptual.

Criteria and methods for accessing already existent lexical material.

The implementation of multiple access points in the restructuring of simple machine-readable dictionaries as MLDBs allows us to exploit at best the information provided by natural language definitions of standard dictionaries.

The products of the lexicographic tradition, *i.e.* printed dictionaries, are now widely recognized not only as one

of the main sources of data and information about language, but also as important banks of general knowledge, with a relevant cognitive role. By using methods and techniques of DBs, we can obtain from already existing dictionaries in magnetic readable form (also for the simple purpose of photocomposition) usable and explicit information on the lexicon, and we are able to discover and implement a number of lexical, morphological, semantic and syntactic relations between dictionary entries.

It is possible, for example, trough appropriate analyses and procedures, to perform a systematic translation of definitions into formalized structures which summarize the definitional informational content. It is possible to use an inductive method, through progressive generalization from the common elements.

Problems to be tackled are :

- is it possible to integrate different lexical resources ?
- is it possible to envisage LBDs which are at least partially independent of different linguistic theories ?
- which is the best way to fully exploit existing dictionaries ?

MLDBs uses : examples, problems, possible actions

The interest for MLDBs appears today in different sectors of research and applications. We shall now report a number of examples and mention some of the problems connected with the possible activities of research and development to be considered.

Computational parsing and generating systems

A major problem is if, to what extent, and under which conditions it is possible to build a MLDB in such a way that the linguistic information can be directly or indirectly used by computational systems (for parser, generation and in general, processing of natural langage), which utilize different theoretical and computational frameworks. This problem must not be underestimated, because the realization of the lexical component - in a real-word computational system - is the most time-consuming and money-demanding task.

Parsers have been written in a number of different theoretical and computational frameworks. Since these frameworks often propose different linguistic elements and operations, the consequences of such differences for the construction of grammars for parsing are usually immediately obvious. Less obvious, but not less real, are the consequences for the lexicon. Nevertheless, information that is, in some sense, "identical", is often represented in lexicons that differ considerably. There seems to be a continuum of possibilities for sharing lexical information between two systems :

1. The information is represented identically in the systems, e. g. both use features with the identical meanings and names.

2. The information is represented in similar ways in the systems, e. g. both use features with the same meanings, but with different names.

3. Linguistic phenomena are analyzed in similar ways in the systems, but the lexical representations are not comparable, e. g. a given syntactic construction is represented by a single feature in one systems, but by the concatenation of several features in another, as in the Raising to Object/Object Control example presented.

4. One system makes finer distinctions in its analyses than the other does, e. g. one system distinguishes raising to Object from Object Control, while the other does not.

5. Linguistic phemomena are analyzed in non-comparable ways in the systems, e. g. one uses a syntactic base for its analyses, while the other uses a semantic base.

Often two lexicons do not differ in any one of these ways, but rather in a mixture of them, i. e. some linguistic phenomena are analyzed and represented identically, others have the same analysis, but different representations, while still others have different analyses and representations. Differences (1) through (3) are amenable to unaided machine translation between lexicons, while (4) would require, at the minimum, human intervention in translating from the coarser grained system to the finer.

From a pre-theoretical point of view, it is possible to imagine several aspects of lexical specification which are crucial to the ideal text generation system, but less crucial to parsing (at least in the case of a parser which assumes well-formed, cooperative, idiomatic input). Lexical selection needs to be handled right on both the paradigmatic and the syntagmatic axes.

Paradigmatically, a term must be selected which is at least both morphologically and semantically appropriate. The latter requirement entails that the term should not only be denotationally appropriate but also carry the appropriate amount of information - that is, it must be at the right level of generality, and not carry presuppositions or implications which are inconsistent with the system's goals. Ideally, considerations such as speech register should be respected at this level too.

Syntagmatically, co-occurrence restrictions of various level of subtlety must be respected. At the coarsest level, the system must produce syntactically well-formed complement structures, while at more subtle levels, the system should be able to enforce semantic restrictions on those complements, use idioms (in the sense of noncompositional phrases), maintain a consistent speech register, and even (in an ideal system) respect the kinds of collocational preferences that certain lexical items have for certain other lexical items and certain syntactic patterns.

While all of the above may appropriately be termed "lexical considerations", the information used to make these kinds of choices is frequently contained in components of text generator other than the lexicon. Existing systems differ widely, both as to which phenomena are handled, and also where in the grammar they are dealt with. This is hardly surprising, since systems have been built to perform various dissimilar tasks and are based on various different theories of communication, knowledge and grammar. As a result, information which one system stores in the lexicon may not be represented at all in another system, or it may be distributed (explicitly or implicitly) troughout any of the components which deal with semantics, world knowledge, text palns, and syntax.

After having assessed which lexical informaiton has been and could be distributed in computational text generation systems, an effort is perhaps necessary to examine possibilities of harmonization and reusability of sources.

Translation system, bilingual dictionaries.

The relaunching of research projects in the fields of machine translation and computer-assisted translation has imposed to the attention of the researchers and of the component financing bodies the problem of the construction of very large multilingual dictionaries. The necessity of research in contrastive linguistics became apparent, obviously connected with the need for multilingual reference and technical corpora. The cost and the effort in the construction of adequate bilingual dictionaries and of their connection with (existing) terminological DB is such that the European Economic Community has promoted a study on the reusability of lexical sources.

Computer lexicons and machine-readable dictionaries are used to ensure translation accuracy and terminological consistency. The differential utilization of such materials, in various forms of translation, by both human and machines is a very important topic. The central basis of discussion will be the different kinds of information required by men vs. machines, and how this relates to the form of translation. For example, terminological data banks accessed directly by humans, whether on-line or off-line, contain information useful to humans during translation, and this is not usually employable by computers for much of anything.

At the other extreme, the dictionaries accessed by fully-automatic machine translation systems contain information useful to the computer during translation, and this is not usually employable by humans for much of anything. In between lie dictionaries for use in machine-assisted human activities ; optimally, these might contain both kinds of information. The question arises, then : is there a synthesis of the two kinds of information, useful for both forms of translation ?

Issues to be explored are the following :

- are printed bilingual dictionaries a good or useful source of information for the implementation of bilingual LDBs, as ascertained for monolingual printed dictionaries ?
- or is there need for other sources of information as, e. g., textual (bi or plurilingual) data ?
- is it possible to extend the computational methods used to implement monolingual DBs also to bilingual ones ?
- are bilingual dictionaries useful in order to connect two monolingual lexical Data Bases ?
- which is the best structure to generate correct and valuable links between not simply lexical words, but lexical concepts ?

Knowledge bases

A Lexical Knowledge Base (LKB) is a computational representation of the information we have available to us about the meanings of concepts and their interrelationships. It provides computational systems with their counterpart to human lexical understanding. Fundamental to the operation of an LKB is a self-understanding of its data. LKBs are more than data bases by virtue of their data being non-redundantly stored, data values being checked for correctness, both in form and content, and in general because they assimilate new information rather than merely record it, and generate responses to information inquiries rather than merely retrieve them.

An LKB does not contain information for display for people. It contains information for the use of other programs ; programs which expect and can assume correct values are stored. A knowledge base knows what the conceptual class or ISA relationships are for all its data values.

Existing machine-readable texts, whether the highly structured entries of dictionaries or other reference books, or the descriptive text of news wire stories, can serve as the sources for information for lexical knowledge bases.

In order to extract information from these sources it is necessary to process them through a variety of stages, talking raw phototypesetting text and transliterating it into standard computer character sets, then reverse engineering text formatting commands to derive the syntactic and semantic information they embody into a list structured attribute-value format and from that performing additional processing steps to produce data bases from which the lexical knowledege base can be asssembled.

Using these techniques, it is possible to derive ISA hierarchies and subject domain categorizations from dictionary entries, grammatical and semantic rules from dictionary codes, as well as other useful information about the properties of words such as their degree of lexical ambiguity and redefinability. Almanacs can yield information about the classes of proper nouns as well as critical relationships between members of these classes. Encyclopedias extend these facts further and give basic word knowledge about concepts. From encyclopedias one can derive definitions of the concepts in their entries. Finally, from descriptive text as it appears in new wires one can derive updates and extensions to all of the above reference sources, both modifying changeable values for almanacs and encyclopedias, as well as extending the lexicon of dictionaries.

The techniques for extractinng information from text sources include both conventional text processing and more advanced computational linguistic text understanding techniques. While it is impossible to write general parsers for unrestricted textual input, it is relatively easy to write specific text expert parsers for given syntactic patterns. Such tools easily recognize given classes of information in text and over time and through processing millions of words of text, can derive signifiant amounts of knowledge about the world. Such text experts have been written for extracting geographic and biographic facts, for recognizing neologistic expressions, and in general for supporting the collection of proper nouns and compound nouns for addition to a lexical knowledge base.

Psycholinguistic research.

The word is a very important unit in the natural language system. It is not only the case that words are meaning carrying units, but also function as an interface between our cognitive system and the world. For these, and other reasons, words have an essential role in psycholinguistic research.

Words are used in psycholinguistic research as a unit of investigation themselves, or as subunit within larger units of interest. If certain properties of these (subunits are studied then typically different classes of units will be defined, each reflecting the property under study in a different way.

The logic of experimental research dictates then that these different classes of units should not differ in any other respect.

Therefore it is necessary to compare these classes on many of the properties that word have and that might play a role in that particular experiment.

The reasons for the need to know properties of words are thus twofold : one is an interest in these properties themselves and one has to do with the principles of experimental design.

A lexical data base is an extremely useful tool for experimental psycholinguistics as a way of having access to the many properties of the words of a language. The structure and the information of a LDB (word properties, like syntactic category, orthographic composition, morphological structure, phonological composition, syllabic structure, stress patterns, isolation point, frequency) must be investigated in relation to the questions that an experimental psycholinguist might ask a lexical data base.

Practical applications (e. g. in secong language learning) have to be discussed.

In the pursuit of psycholinguistic goals, some researchers have used LDBs to test the plausibility and applicability of certain concepts in theories of auditory word recognition and, in general, for answering certain crucial questions about human language processing.

Once and for all no "shopping list" can be given. There are, of course, some essential properties that a LDB should contain. These are largely common to those requested by linguists and by computational linguists. In particular, the psycholinguistic researchers seem to require adequate frequency counts, regularly updated.

Dictionary systems for mass market

The market for machine readable dictionaries (MRD) will cover three principal types of applications :

- Where the MRD is used as a tool in the computer processing of text.
- The scholarly tool for literary, linguistic, ot other academic research.
- General references - a more flexible alternative to the use of a printed dictionary.

MRDs in the office : Two trends look set to influence the use of MRDs in the office. the first is the transfer of typing work from the typing pool/secretary to the executive. This trend is a natural consequence in the number of work-stations on executive desks and the development of networks. The executive typing his electronic mail is going to want more than a word- processor - perhaps something along the lines of IBM's Epistle Project. The second - and related - trend is the advance in computer technology which means that neither processing nor storage capacity is likely to be a limiting factor.

This will perhaps be the first major market for MRDs. However, use requirements are unlikely to be homogeneous - at least initially. Undoubtedly such markets, with very specific and specialized requirements, will develop, but this is likely to be in the second wave. What will be required initially is an MRD offering a full range of lexical information with software that will enable the user to obtain the right type and level of information with the minimum of difficulty.

The fact that the MRD may contain large amounts of data unneeded by the user will not constitute a barrier to use in the way it would with a printed dictionary.

Undoubtedly it will be necessary to tailor both contents and software to executive needs. The types of information that will almost certainly be required are, for example, specialized vocabulary relating to specific

technical/occupational fields, new words, a thesaurus facility.

In the home : The accelerating rate of technical change is likely to foster a market for MRDs in the home also.

As a consequence of the greatly increased storage and processing capacity, computing in the home is likely to offer a full range of functions which will include not only stand-alone applications (like games and word-processing), but also communications facilities (like electronic mail and home banking) and information (like share prices, an encyclopedia and lexical information).

Such a facility would be not only a tool for recreation, as the home computer has largely been up to know. It would also be a tool for creation-writing, designing, composing music, etc. Value added by the availability of lexical information may, in many instances, help to justify the outlay required for a system of this type.

In education : This sector, and in general the utilization of the computer for a better and more convenient use of reference material, is only just beginning to be explored.

We are now starting to think of using MRDs in second language acquisition for the improvement of writing skills. We are quoting from a paper by Benbow. The development of the market for the native english speaker will, as in other spheres, be dictated by the availability of appropriate technology. One can envisage signifiant and wide-ranging roles MRDs in schools, vocational training, and higher education. It is unlikely, however, that the availability of MRDs alone will be sufficient to generate an important market.

If educational and training institutions are using equipement on which MRDs could also be mounted, for a modest outlay on MRD (or a series of MRDs) could add significant value to the system.

For the teaching of English as a secong foreign language, it is consiiderably easier to see the development of a stand-alone market for MRDs. English language teaching (ELT) has always been quick to adopt appropriate technology - such as the use of tape and cassette-recorders in language laboratories and the development of interactive video as a teaching tool.

Clearly the type of information required by a non-native speaker of English will differ considerably from that required by a native speaker : information on verb patterns and on contextual usage will be high in the priority list, for example. As this type of information is already available in printed ELT dictionaries, this specialized requirement should not pose a problem.

Consequences for lexicography.

1. Uses of an MRD will differ from those of a printed dictionary.

A recent OED survey certainly indicates that this will be the case as far as that particular dictionary is concerned. The difference in use, moreover, will not simply stem from the different media on which the text is held. The MRD user will in many instances be searching for information not readily extracted from a printed dictionary and the paths to that information will be infinitaly more diverse than those by which printed dictionaries can be explored. A prime example of such use would be reverse look-up (find a word from its meaning) and other semantic searches.

2. New types of dictionaries user will emerge with the availability of MRDs.

Use of the printed dictionary is practically limited to obtaining information about known lemmata. Through the MRD it will be possible to obtain information about word groups which are related by one or any number of factors ; by subject classification, by grammatical class, by pronunciation, by language of origin, and so on - the list is virtually limitless. The full implications of this are impossible to predict.

3. Printed dictionaries and their machine-readable counterparts will diverge.

With different uses, different users groups, and different natures, the initial correspondance between printed dictionaries and their machine-readable counterparts will soon disappear. If, as seems likely, MRDs do proliferate, the problem of maintaining standards will arise. The quantity of material may well replace quality as the prime concern of the MRD

supplier. the establishment of an acceptable balance should be an important concern to all those involved in the development of MRDs.

4. Software will be a key determinant of the usefulness of an MRD.
Much research and development work will be needed to provide software good enough to make really valuable MRDs.

5. The lexicographer's influence in the development of MRDs may well be eroded.
In the preparation of a printed dictionary the lexicographer's role is paramount. This may not be the case as far as the development of MRDs is concerned. The influence of the user, the computer scientist, and the marketing specialist is likely to be significantly stronger in the production of an MRD. This brings us back to the question of quality : the lexicographer is the most important factor in the control of quality. If the role is down-graded quality will almost certainly be under threat.

6. MRDs will not supersede dictionaries for all purposes.
It should not be overlooked that for most simple enquiries (for example, the meaning, spelling, or pronunciation, of a given word) the printed dictionary will remain the most convenient and efficient way of obtaining the information required (unless, of course, at the time the information is needed one is using a work-station with ready access to an MRD).

7. It is likely that for many dictionary users the availability of illustrations would be a very useful adjunct. The same may also be true of video and audio material which could be called up in support of verbal definitions. Additional technical features -such as the synthetic reproduction of the pronunciation of a given word form- will add further value to the MRD.

Reusability of machine-readable dictionaries

The number of machine-readable dictionaries is increasing, owing to the diffusion of photocomposition. The possibility offered by new technologies for the distribution of traditional or new lexicographical products will probably be another factor for the growing availability of lexical ressources in machine readable form.

There is no doubt that the traditional dictionaries do not contain all the information necessary for the natural langage processing systems.

On the other hand, however, they contain a lot or more or less organized and relevant information for the computational systems. Therefore, it seems useful to assess on the methodological and organizational level the possibilities of using the information available in MRD for different computational applications.

The reorganization of MRD in data bases form with a proper integration methodology may also open to the normal specialized users access to information which is difficult to retrieve in alphabetical order.

MRDs and research in computational linguistics.

Considering a great number of systems within the computational linguistics paradigm, one is struck by the very small number of dictionary entries available to these systems.

Admittedly, most of these are experimental systems ; nonetheless, the fact remains that for real word applications, especially given the recent advances in computational linguistics technology which make such applications feasible, considerably larger vocabularies are required. Given a number of real constraints, ranging across a wide variety of systems, the most critical of which are idiosyncratic formalisms for writing dictionary entries, different lexicon formats, and diverse views on what constitutes linguistically relevant (syntactic, semantic anf pragmatic) information whose proper place is in the lexicon, the task of generating a comprehensive and consistent vocabulary for any application looks, and indeed is, quite awsome. It is hardly surprising, then, that a number of researchers have been

looking at available machine-readable dictionaries (MRDs). They expect that information already digested, categorised, indexed and most importantly, available in machine-readable form, can be suitably used, if not to get a sizable lexicon "for free", then at least to construct automatically a substantial portion of such a lexicon, getting for free what is hoped to be an internally consistent (and coherent) object. A lexicon derived in such a way would save considerable effort and still produce the bulk of the target vocabulary which could then subsequently be expanded, if necessary, and tailored to the task and application at hand.

There is a parallel strand of work, adhering to the current trend in "knowledge-based programming", which addresses the problem of acquiring non-trivial amounts of structured knowledge about the real word. The hope here similarly pivots on the assumption that ways can be found to localise and extract some of the knowledge required for performing a class of (semi)intelligent tasks from suitable machine readable sources, e.g. dictionaries or encyclopaedias.

It will be useful to investigate some ways in which thes goals are being pursued by a varity of researchers, using different dictionaries and utilising different techniques to support different applications. The emphasis will be on establishing how information available in machine readable form has been and could be used in computational linguistics research.

While it is true that systems differ in organisation, structure and content of their lexicons, it is still possible to isolate certain types of information, to be determinated by the particuliar task and application, which ought to be made availabel at the lexical level. A categorisation of the lexical requirements, as they vary across a range of applications within the framework of computational linguistics, will provide a dimension for evaluating the utility of lexical information, that may be available from machine-readable form.

A machine-readable source can be viewed as a lexical data base, offering information about, for instance, word syllabification or idiosycratic syntactic behaviour. Such information might be relevant to a speech synthesis system or a parsing program, but is of very little utility to, say, an interpretation component of a natural language front end system. In the latter case, the intention is to regard the dictionary as a sort of a knowledge base, where information of more general semantic nature would be encoded, either in a free text format, or in some form of semantic tagging. Of general interest in both cases are issues like the kind of data that is likely to be available in machine-readable form, and the class of computational system to which this data may be useful.

A substantial amount of work has gone into attempts to provide computational solutions to a wide range of problems, which rely on information extracted from an MRD. Syntactic parsing, grammar development, word sense selection, speech synthesis, robust text interpretation, knowledge acquisistion, information management, phonetically guided lexical access, phrasal analysis, these are only a few of the tasks which have been substantially aided by the availability of large dictionaries on-line.

Collective experience makes it feasible, and indeed necessary, to examine some broader aspects of working with machine-readable dictionaries. It seems now possible to ask questions like, for example, what is the overall cost (the manpower effort spent on a project) of attempting to harness what sometimes can be a bulky and unwieldy object ? What type of information could be reasonably expected to be found in an MRD ? How reliable is this information ? What would be the best way to make use of it in any particular context ? A consensus on these topics, together with a feeling for how the range of available dictionaries score on such scale is going to be of enormous utility to the research community.

Structure and access in automated lexicons.

The problem of structure and access in relation to the organization of an automated lexicon can be considered under many perspectives :

- as problems of software and hardware ;
- as problems of conceptual or logical structure of the data, e.g. in a data base and/or in a network ;
- as problems of "lexical" structure and of relations among words, and therefore of the linguistics relevant link and concept ;
- as problems of a more psycholinguistic nature, if the mental process of lexical access is to be simulated ;

- as problems related to the different types of goal to be achieved ;
- as problem to be connected to the users (human or procedural) of the data.

The type of organization and consequent access required are obviously very different if we are to implement a simple machine-readable dictionary, or more "structured" lexical database or an even more complex knowledge base. In the first case, a "string" or "text format" is certainly sufficient, whereas in the other two a more structured/codified organization is very likely to give the best results.

The restructuring of MRD data into a data base organization can make it possible to achieve additional information with regard to the original data, multiplied by a factor given by the number of diffrent views established on the data, i.e. of the different ordering relations or different paths connecting to the data.

The lexicon will appear as virtually divided into as many subsets as the relations which have been determinated and formalized. By representing the lexicon by the set of these relations, we can access the dictionary either by lexical items, or by features, or by relations, search the network to see where it matches with the query, and retrieve different parts of the lexical content on the basis both of the access point and of the options activated on this point.

The concept of lexicon itself can be widely touched and changed in the computational framework, namely considered as a much more interrelated object, where words are composite at many levels and each kind of complexity must be analized and handled so that complexities on various layers can be captured either by general rules or by appropriate access functions.

N.B. Cet article a été rédigé à partir des résumés de plusieurs contributions écrites par MM. Ingria, Cumming, Slocum, Amsler, Schreuder, Bendow, Byrd, Boguraev et Calzolari.