Lexical DataBases and Textual Corpora:

a trend of convergence between

Computational Linguistics and Literary and Linguistic Computing

Nicoletta Calzolari - Antonio Zampolli

Istituto di Linguistica Computazionale del CNR, Pisa, Italy
Dipartimento di Linguistica, Universita' di Pisa

1. Introduction

In this paper the development of lexical knowledge bases and textual corpora will be considered in the framework of the recent trend towards the creation of large repositories of linguistic information. This trend concerns both researchers who call their discipline "computational linguistics", and researchers who identify their activities as "literary and linguistic computing". The two terms are often used in different ways. They are in fact sometimes considered to identify two different disciplines, other times they are considered to design two different orientations of one same discipline. In both cases, their relationships have not been the object of adequate theoretical reflection. However, it seems uncontroversial that the two terms identify two largely disjoint groups of researchers. We shall consider briefly, first, how these two groups developed, in the past, as separate entities, with a very limited overlapping membership, and why they are now beginning to consider possible cooperations in the development of large linguistic knowledge bases.

2. Some historical and terminological remarks

When the use of electronic data processing techniques (1) on linguistics data began at the end of the '40s, two main lines of research were, quite independently, activated:
- Machine Translation (*Traduction automatique*) (MT).
- Lexical Text Analysis (*Depouillement electronique de textes*) (LTA: production of indices, concordances, frequency counts, etc.).

While MT was promoted mainly in 'hard-science' departments, LTA was developed mainly in humanities departments and, probably also for this reason, the two lines had very few contacts (2).
At the beginning of the 1960s, the perception of a possible reciprocal interest was explicitly manifested, in particular through the invitation of MT researchers to the first LTA conferences, like Tubingen (1960), and Besancon (1961) (3).
The topics more often quoted for possible convergence of interest were, in particular, text encoding systems for different alphabets, frequency-count of linguistic elements in large corpora, automated dictionaries. But, in effect, real cooperation was very rare if not totally absent (4).
The year 1966 has been particularly important for both lines of research, but for opposing reasons. The Prague International Conference *'Les Machines dans la Linguistique'* ratified the international acceptance of the LTA as an autonomous disciplinary field, and its extension to a broader area, which included new dimensions of processing (phonology, historical linguistics, dialectology, etc.), called **Literary and Linguistic Computing (LLC)**, whereas the well-known ALPAC-Report (1966) brought about an abrupt arrest in the majority of MT projects throughout the world, and marked the beginning of the so-called 'dark ages' of MT.

Following, de facto, the recommendations of the ALPAC report (5), basic research on natural language processing slowly occupied the area characterized so far by MT activities, and **Computational Linguistics (CL)** emerged as a new disciplinary activity (6).

However, in spite of ALPAC recommendations for researches in large-scale grammars, dictionaries, corpora (7), CL focused mainly on the development of methods for the utilization of formal linguistic models in the analysis and generation of isolated sentences, in an almost exclusively monolingual framework, at the grammatical level.

The CL activities, which came after MT, almost completely neglected the development of lexica, practically restricted to small toy-lexicons of a few dozen words (8). A distorted (we believe) interpretation of the Chomskyan paradigm led to an almost complete disinterest in corpora and quantitative data, which, on the other hand, were attracting much attention in the LLC area due, among other things, to projects for national historical dictionaries (9) and for frequency dictionaries (10).

On the other hand, also the LLC delayed taking advantage of the know-how, methodology, and tools produced from the very beginning by MT in the field of automatic lexica. Not only had MT developed research on specialized hardware (11), storage, access techniques, inflectional and derivational morphological analysis, but certain projects had already begun the collection of large sets of monolingual and bilingual lexical and terminological data.

Very few exceptions can be reported in the LLC field, all primarily motivated by attempts to automatize the lemmatization of texts for the production of lemmatized indices and concordances. To our knowledge, the first experiments are related to Latin (CAAL, Gallarate and LASLA, Liege) (12).

For several years practically no relationship has existed between LLC and CL. As local organizer of the 1973 Pisa COLING, Zampolli endeavoured to include in the call for papers, and to promote in the Conference, sections explicitly dedicated to topics which could delineate the areas of common interest. The attempt was successful in terms of joint participation, and it was probably not just by chance that J. Smith presented there, at an international level, the newly founded ALLC (Smith, 1973).

But in those years a (so to speak) 'puristic' approach characterized the general reflections of CL, which was searching for a definitional and a disciplinary identity (13), focussing on problems of computation and on the nature of the algorithmic procedures, rather than on the nature of the results and on linguistic, in particular textual, data.

The variety of points of view is exemplified in the *Foreword* by Karlgren, and in the *Introduction* by Zampolli, to the *Proceedings* of COLING 1973 (Zampolli, Calzolari 1973) (14).

The development of CL, in the following years, has been influenced by the interest for Natural Language Processing (NLP) shown by large sectors of Artificial Intelligence. Many efforts have been directed towards the study of methods and tools for prototypes performing a "deep understanding" of natural language, necessarily limited to restricted linguistic fragments and to "miniature" pragmatic subdomains, thus enlarging the gap between CL and LLC activities.

In the LLC framework, the attention of a large part of the research community has been captured by the new technological developments, and efforts have been directed towards mastering new hardware and software facilities: the increasing variety of rich sets of characters, OCR, photocomposition, large database techniques, personal computers, new storage media, general purpose editors and word-processors, standardised concordance packages, etc.

Only in the last two years has a variety of contributing factors started to rouse the reciprocal interest of people working both in CL and LLC. Increasing contacts and exchanges; joint organization of conferences or conference sections; cooperative projects formulated at the international level are external signs of this process.

This convergence is, partly, due to the activities of some Institutes, programmatically oriented to perform researches in both fields, and thus naturally operating to construct a bridge and to promote sinergies (15). However, in our opinion, the key fact is that both fields are recognizing that an important aspect of their development depends on the capability of processing, at least at some level of linguistic analysis, large quantities of "real" texts of various types.

2.1 Computational Linguistics

CL has always considered as a main task the construction of computational components for the automatic generation and analysis of natural language sentences. However, only very recently has CL truly faced the problem of constructing components suitable for the treatment of large, real texts. This trend has been largely originated by the increasing interest of several national and supranational authorities for the potentials of the so-called **"language industries" (LI)**.

This expression, coined on the occasion of a Congress sponsored by the Council of Europe in Tours, February 1986 (16), is used to indicate activities based on computational systems, oriented to practical industrial and commercial applications, which contain, as an essential part, natural language processing components. Examples of typical applications include, within the domain of speech technology: access control, command and control to data entry, driver stations, document creation, telephone enquiries, transaction processing by telephone, data base enquiry, environmental control, voice messaging, announcement systems, augmented communication for handicapped people, etc. For written texts, we can quote: spelling checkers, computer-assisted lexicography and terminology, natural language interfaces, machine translation, information retrieval, computer-assisted language learning and teaching, computer-assisted consultation of reference works, translator workstations, etc.

A set of different factors and conditions are requiring today the promotion and development of LI. The keyword is, in our opinion, the advent of the so-called 'information society'. The global dimension of the economy conceived as a worldwide system (17), together with the technological development of telecommunications systems, entails a growing information flow. The principal information vehicles are still the natural languages, both for the production and the storing tasks. Furthermore, the major part of the information in natural language is nowadays produced directly through computer use, and recorded on machine readable supports: word-processors, office automation, electronic mail, photocomposition, databases, etc. Various countries are considering the possibility of progressively recording entire libraries in MRF.

This situation puts an obvious pressure for the creation of new products and services for the various economic activities primarily involved in information handling. The following passage of Makoto Nagao (1989, p. 4) seems particularly relevant to us: "Computers are a fusion with and unification of communications technology at both the hardware and the software levels, and computer systems will undoubtedly enter every corner of future society. When that day arrives, the most important technology will be specifically concerned with neither hardware nor software, but with what I have been advocating for many years: 'informationware'. In other words, the central problem will regard the ways in which the information signals sent by human beings will be mechanically processed, transmitted, stored, and then recalled in a form which can be interpreted by other human beings. The essence of informationware is therefore how information can be efficiently stored in a computer and activated in response to the various demands of its users. Information can in fact take different forms, including writing, speech and visual images, but objectively, the most accurate means for transmitting and receiving information is writing. For this reason, of the various aspects of informationware, linguistic information and its processing technique will be the primary technology at the heart of the information society. Such technology might be called 'language engineering', and the industry which it will span will be the 'language industry'".

A central aspect of the LI is **multilinguism**. Only an 'elite' minority in the world can operate today in a foreign language, without sacrificing its performance (Perschke, 1988). Furthermore, the conservation of national languages, principle adopted from the beginning, for example, by the EEC, is an important condition for the preservation of the national cultural identities (18).

The need for monolingual and multilingual natural language processing systems, to be used in products for information handling in the LI framework, is uncontroversial. Some studies are carried out in order to narrow down and focus the most urgent tasks and targets, identifying the principal sectors of activities and their economical dimension.

However, the major problem consists in evaluating: - which products can be created on the basis of existing technologies; - which applications can be envisaged at short and medium terms; - which are the priority areas and tasks for linguistic basic and applied research; - which can be an appropriate research and development strategy; - by which measures, at the organisational

level, the public Authorities and professional scientific Associations can stimulate progress in the field (19).

In this framework, one of the priority needs, recognized by several researchers in various countries, is the description, in a form which is suitable for computer use, of the natural languages, performed as far as possible exhaustively, at least for the linguistic aspects which can be treated at the present state-of-the-art of linguistics and of natural language processing. Such extended descriptions are considered the bases for the construction of components capable of dealing with the various types of large real texts which are the typical objects of a wide range of LI applications already possible or foreseeable at short and medium term.

These descriptions concern, first of all, grammars and lexica, and can take the form of repositories of grammatical and lexical knowledge bases. Large corpora of textual material in the form of textual databases are considered essential sources of information (20).

The construction of such large structured collections of linguistic data is very expensive. The availability of such extended linguistic knowledge bases is essential for the feasibility of various industrial applications. Therefore, they are often considered as precompetitive resources. Different categories of partners from the academic, industrial and publishing sectors must co-operate in their creation. To ensure reusability, the creation - as far as possible - of standards, is very important. Cooperation and coordination of efforts is required not only at the national but also the international level, if the monolingual linguistic knowledge bases are to converge in a multilingual network, both for the creation of bilingual systems and for the use of similar components in monolingual applications on different languages.

2.2 Literary and Linguistic Computing

The quantity of texts available in machine readable form is increasing very rapidly. Not only is there a progressive cumulation of texts directly encoded by various categories of humanists for electronic processing, but also the most part of texts nowadays is produced and (re)published through computers. Given the diffusion of individual workstations - with computational power and memory size adequate to the typical humanistic tasks - the distribution of the texts directly in MRF for the interactive use of individual researchers has become possible and more and more attractive.

As a consequence, the adoption of standards, for text representation, which ensure the exchangeability and reusability of texts for various users, has become very urgent (21).

LLC has always been interested in the process of large real texts, but the computational treatment has been performed on units identified, mainly if not exclusively, at the graphical level. Frequency counts, concordance production, interactive textual access usually operate essentially on the graphical forms, roughly defined as sequences of characters between two spaces or separators.

However, several operations on the texts, which enter in the performance of various scholarly humanistic activities, are based on the identification, in the text, of linguistic units at various levels, both as direct objects of linguistic, philological, literary research, and as referential units representing factual information. An exemplification list contains, among other units, phonemes, metrical schemata, syntagmatic patterns, rhymes, lemmata, lexemes, phrases, morphosyntactic categories, terminological units, conceptual units and their relations, etc.

The intrinsic complexity of the analysis, and the time required to perform it, are very high. The large diffusion of personal workstations enables more and more individual researchers to directly perform a variety of analyses on the increasing number of available texts. Therefore, LLC is obliged to consider the possibility of constructing or importing tools for automating, at least in part, the operations of analysis, or at least for assisting the humanists in its performance.

Roughly speaking, considering the present state-of-the-art in natural language processing and in knowledge acquisition and representation methods, we can distinguish two major categories of computational tools for computer-assisted humanistic text analysis.

- Robust parsers, supported by large computational lexica, conceived for identifying, in real texts, linguistic units, at certain levels of analysis: syllabic, metrical, syntagmatic patterns; lemmata; parts of speech; phrases; verbal arguments; superficial sentential structures; etc. The components constructed in the framework of CL, if they were adequate to process real texts, would supply the identification and the representation of such units and their relations (22).

- "Intelligent" access tools which, through the consultation of various kinds of knowledge sources, assist the researcher in the interaction with the texts. For example, appropriately structured reference sources, such as encyclopedias and dictionaries, can make explicit, and eventually complement, the linguistic and conceptual researcher's knowledges, in such a way that they can be used by the programs for text browsing. We shall briefly illustrate later examples of dictionaries which can also be used, for instance, to expand a user's query, searching in the texts the occurrences of "families" of words connected by particular semantic or conceptual relations: taxonomy, synonymy, etc.

2.3 Convergence between CL and LLC

Summing up, both CL and LLC are led by various factors, and in particular by the framework created by the expansion of the 'information society', to consider the creation of tools and resource systems for the processing of large real texts, as a major task in their present state of development.

From this, we are not arguing that CL and LLC aim at the construction of computational systems of the same nature, nor that they have to solve exactly the same range of linguistic problems. We notice only that both fields are now recognizing that the development of these systems require the availability of extended repositories of linguistic knowledges.

Our thesis is that the basic knowledges required are in large part the same. It is therefore important that the information encoded can be reused in both fields through appropriate interfaces. Cooperation must be promoted, in order to combine the efforts and the specific know-how of the two categories of researchers, who are for several aspects complementary. For example, CL has developed grammatical formalisms and parser models; LLC has developed knowledges and methods for corpora collection and treatment, statistical linguistic analysis, sublanguage description and identification.

In the following we shall describe our work in Pisa in the field of lexical knowledge bases, and of their interaction with textual corpora. This research work is explicitly intended to the creation of resources both for CL and LLC, in the present framework of their trend to convergence.

3. Trends in Computational Lexicography and Lexicology

We have already noticed the tendency inside CL in the last years to a shift in interest from almost only the grammatical aspects of the language, to the lexicon also, and, only quite recently, also to large corpora of texts. We are in the presence of a somewhat parallel evolution from the implementation of so-called 'toy-systems' (the prototypical is Winograd's block-world), to the development of 'expert systems' (more powerful, but acting within a limited domain, and therefore with a restricted vocabulary), and recently to 'very large NLP systems', such as Machine or Machine-aided Translation Systems, or products for Office Automation, where a strong need is felt for a real-size vocabulary and a general world knowledge.

Taking for granted these two main trends, both from the theoretical and the applicative viewpoint, it follows that dealing with the lexicon has become trendy, and dealing with textual corpora is becoming even more trendy.

There is a need not only for very large computerized lexicons or Lexical Databases (LDB), but also for lexicons where even the semantic information is made explicit, i.e. for large Lexical Knowledge Bases (LKB). The evolution within Computational Lexicography and Lexicology over the past few years can thus be outlined as follows:

i) from Machine Readable Dictionaries (**MRD**) in the '70s (simple sequential objects well exemplified by photocomposition tapes),

ii) to **LDBs** in the early '80s (more structured objects, provided with multipath access to the data, interactive in nature, and often with explicit taxonomies or IS-A hierarchies),

iii) to **LKBs** in the late '80s (where not only IS-A links but also many other types of lexical/semantic relations among conceptual categories are formalized, where therefore new access paths to the data are constructed, where inferential and deductive mechanisms are built in, and which are usually in the form of a conceptual network).

The main prioritary goal is thus today the creation of a vast 'reservoir' of linguistic knowledge, in the form of as complete as possible and reusable linguistic descriptions, structured in a large LKB or in various kinds of interconnected linguistic bases (grammatical, lexical, textual, knowledge bases).

Given what already stated about the current trends in the different areas, i.e. the request in the CL community of large scale NLP systems, and the fundamental importance that a CL system is able to deal with tens of thousands of lexical items for real world applications, in addition to the fact that lexicography, as a 'language industry' profession, has a very long tradition, and that the creation of a LBD of adequate content and dimension is very time-consuming and expensive, and duplication of efforts may be a very 'sad' fact, one of the key-words in the field of LDBs has recently become the word "**reusability**". This word is to be intended in two main senses: one towards the past, i.e. with respect to existing information, and one towards the future, i.e. with respect to future applications.

In the first case, the meaning is that of reusing lexical information implicitly or explicitly present in preexisting lexical resources (e.g. MRDs, terminological DBs, corpora of texts, etc.) as an aid to construct a LKB. In the second case, it is meant to construct a LKB so as to allow various users (procedural: e.g. different NLP systems; and possibly human: e.g. lexicographers or translators or normal dictionary users) to extract - with appropriate interfaces - relevant information to their different purposes.

With regard to the first meaning, these ideas in a sense originated the proposal for the ESPRIT Project "Acquisition of Lexical Knowledge for Natural Language Processing Systems" (AQUILEX) where groups of researchers in Cambridge, Amsterdam, Dublin, Paris, Barcelona, and Pisa (coordinator) are involved. The main goal is to develop techniques and methodologies for the use of existing MRDs in the construction of lexical components for NLP systems. The extraction of lexical information is carried out moreover from multiple MRD sources and in a multilingual context, with the overall purpose of the creation of a single multilingual LKB. "The knowledge base will be rooted in a common conceptual/semantic structure which is linked to, and defines, the individual word senses of the languages covered and which is rich enough to be able to support a 'deep' knowledge-intensive model of language processing. The knowledge base will contain substantial general vocabulary with associated phonological, morphological, syntactic and semantic/pragmatic information capable of deployment in the lexical components of a wide variety of practical NLP systems" (Boguraev et al. 1988).

If we look at the second meaning of the term reusability, it is strongly linked to two other properties which we consider essential in a LDB.

The first property of a LDB is that of being "**multifunctional**", and has essentially to do with the applicative viewpoint. The LDB must be a central repository of data which can be reused for several purposes and in many applications, through different interfaces, both for procedural and for human use.

The lexicon is obviously an essential component in any NLP system (for parsing, generating, machine translation question-answering, information retrieval, lemmatization, artificial intelligence, etc.). The usual practice is to construct an ad-hoc lexical component for each natural language NLP project. It is necessary to move towards large (both in extension and in depth of representation) lexicons, where information is represented in such a way that it can be easily interfaced by different application procedures according to the different applicative needs. This means that the same set of data can be shared by the various applications. Each interface will only project on the specific application that view on the data which is relevant for the particular requirements.

From this viewpoint, another essential property of a LDB is to be easily extendable, i.e. it must be possible for different researchers to add their own idiosyncratic information consistently with the actual content of the LDB.

The second property of a LDB has to do with the theoretical viewpoint, and consists in its being "**polytheoretical**", i.e."multifunctional" with respect to different linguistic theories. A large amount of work in CL has been carried out until now, as said above, on experimental lines, with consequently small-sized lexical prototype systems. Furthermore, emphasis was traditionally placed on the representation, organization and use of linguistic knowledge as encapsulated and expressed by linguistic rules and procedures. Lexical data seemed to be considered of secondary importance or, at least, easy to be handled.

It is a well recognized fact that different linguistic theories and different computational organizations may have important consequences on the grammar contruction. Less attention has been paid to the consequence on the lexicon. However, we have the intuition that lexicons designed for different linguistic theories may contain information which from a certain point of view is identical, as it describes the same linguistic facts. We have to assess the validity of this intuition before starting to implement in an LDB the information required by the NLP systems.

This characteristics of being polytheoretical is not without problems and difficulties, and a feasibility study is now underway to assess: i) the possibility of achieving a certain degree of consensus among different theories aimed at sharing the same bulk of lexical information, and if so ii) up to which level of linguistic analysis a "neutral" or "polytheoretical" representation of liguistic properties can be designed.

We have promoted a working group which involves outstanding representatives of the major current "linguistic schools". The group will investigate in detail the possibility of representing the linguistic information frequently used in parsers and generators (e.g. the major syntactic categories, subcategorization and complementation, verb classes, nominal taxonomies, etc.), in such a way that they can be reutilized in the following theoretical frameworks: government and binding, generalized phrase structure grammar, lexical functional grammar, relational grammar, systemic grammar, categorial grammar. This group will work on various languages. We shall start by examining in detail the treatment which the foregoing theories will assign to a representative sample of English and Italian verbs. If a polytheoretical lexicon appears to be feasible it should be possible for the lexical data to be reused within the framework of different linguistic theories (e.g. GB, LFG, GPSG, RG, etc.) and also of lexicographic practice, by appropriate interfaces translating the data in the relevant notation/representation (see Walker, Zampolli, Calzolari 1987).

4. Reusability of preexisting data in the form of MRDs

A large number of articles and books have already been written on this topic (see e.g. Amsler, Boguraev, Briscoe, Byrd, Calzolari, Nagao, Picchi, Walker, Zampolli, etc.). We wish to stress in particular what we consider as the natural evolution of all the work done so far in the field, i.e. the possibility of a procedural exploitation of the "full range" of semantic information implicitly contained in MRDs.

In this framework the dictionary is considered as a primary source of basic general knowledge, and many projects nowadays have as their main objectives word-sense acquisition from MRDs, and knowledge organization in a LKB. The method is inductive and the strategy adopted is heuristic: through progressive generalization from the common elements found in natural language definitions we tend to formalize the basic general knowledge implicitly contained in dictionary definitions, mainly in the attempt to extract the most basic concepts and the semantic relations between them. This means that we are going well beyond the extraction and organization of taxonomies, whose methodology of acquisition is now well established (Chodorow et al. 1985, Calzolari 1982, 1984). We simply have to process the first part of the definition, in order to identify the 'genus' term. This can be done by taking into account the fact that the definitions are NPs when the definiendum is a Noun, are VPs for Verbs, and AdjPs for Adjectives. The procedure has thus to look for the head/s of the NP, VP, AdjP, which are respectively a N, V, or Adj. These are the 'genus' terms and are connected by an IS-A link to the definiendum.

When we reorganize a MRD in a taxonomical structure, with only IS-A hierarchies made explicit, we use the MRD as a source of knowledge, but in only one of the possible ways of acquiring from it (in an inductive form) a concept, by linking this concept to all its instances, i.e. all the instances of the same category/class are extracted and connected together pointing to their immediate hypernym.

In the LKB approach the dictionary is seen as a much more powerful "classificatory device", i.e. as an empirical means of instantiating concepts and many types of lexical/semantic relationships among them (see Calzolari, Picchi, 1988).

The methodological approach that we follow can be summarized in these points:
a) to start from free-text definitions, in natural language and in linear form, usually formed by a 'genus term' and a 'differentia' part;

b) to analyze their structure and content from a linguistic and a computational point of view;
c) to convert and reorganize them into informationally equivalent structured formats made up by nodes and relations linking them.

Point b) in its turn can be subdivided, for the computational part, into the following steps:
1) to "parse" the dictionary entry, in the sense of "parsing a dictionary tape" which essentially means recognizing the various relevant fields in the lexical entry;
2) to produce a tree-structured lexical entry;
3) to perform a morphological analysis and a homograph disambiguation, i.e. to tag the definitions for POS;
4) after the above preliminary steps, we have adopted the technique of producing a very simple syntactic parse which roughly recognizes NPs and PPs;
5) the most powerful tool is then a "pattern-matching" mechanism, which is fed by: i) the results obtained by browsing dictionary data in the LDB (as outlined in the few examples presented below) in view of discovering the most interesting words and word-associations, ii) frequency counts on definitions words and syntagms, and obviously iii) the linguist's intuition.

Let us illustrate with some examples the process of analysing the definitions. In the figures we try to simulate the process of browsing the Italian LDB and of navigating the dictionary while searching for particular words, structures, patterns, etc. We can see some of the semantic data it is possible to search and find in a MRD if appropriately structured. Fig. 1 shows part of the taxonomy for the Italian word *libro* (book), i.e. a set of words defined as being "types of" books (we see them together with their definitions).

But there is something more that is said about books in a dictionary. It is also possible to extract the set of the Italian Verbs related to books (see Fig. 2), and the set of Adjectives and of other Nouns having to do with books (Fig. 3 and 4). In section 4.2 we shall come back to "books", stressing the type of information which, lacking in dictionaries, can instead be found in texts.

Our present work is devoted to the formalization also of the other kind of relations - not as simple as the taxonomical ones - which do hold between words, or between words and concepts, and for whose extraction we must analyze and process the whole definition and not only its 'genus' part.

Let us give some examples of the types of relations that it is possible to extract from MRDs. In Fig. 5 we find the first of the about 300 words linked in our LDB by a taxonomical link to the word *strumento* (instrument). The word *attrezzo* (tool) appears in this list. Fig. 6 shows the first hyponyms of this second word together with their definitions. From these definitions it is rather simple to extract semantic relations which we could label **USED FOR, USED IN, SHAPE, MADE OF**, etc. They are extracted by means of a pattern-matching procedure acting on the 'differentia' part of the definitions, where the different ways in which each relation is actually lexicalized in the definitions is associated with the relation-label. The relation **USED FOR**, for example, comes from lexical patterns like: *per, usato per, atto a, che serve a, utile a,* (for, used for, apt to, which serves to, useful to); these lexical patterns acquire this particolar relational meaning when found in particular positions in the definition of hyponyms of the word *strumento*. They can also acquire different meanings in other contexts. The result of this analysis of the definitional content will be restructured in a part of a conceptual network which is sketched in Fig. 7.

Other types of semantic relations rather easily and straightforwardly extractable from the definitions can be illustrated with some examples.

One is the relation **SET OF**, which can be further specified as to the type of its members. We have examples of words denoting **SET OF** *persone* (people) (Fig. 8), *oggetti* (objects) (Fig. 9), etc.

Other types of useful data concern information on selection restrictions for Verbs or for Adjectives and mainly derives from the lexical pattern *detto di* (said of), after which the type of Nouns is found of which an Adjective or a Verb can be typically predicated. See Fig. 10 for Adjectives and Verbs used for nouns denoting *persone* (people), Fig. 11 for Adjectives which collocate with names of colours, either generic colour names, or specific ones such as *giallo* (yellow), *rosso* (red), etc.

An interesting type of relational data which can be extracted for certain types of actions is the information on the words in the lexicon which are lexicalizations of the typical thematic roles of the action itself. Let us clarify what we mean by two examples. In Fig. 12 we find the result of querying the Italian LDB for all the entries in whose definitions the word-form *vende* (sells) appears (not in genus position). The result of the query is the following: we retrieve 242 entries of which well 221 are names of people who "typically sell" something, i.e. of typical **AGENTS** with respect to the action of selling. These entries represent lexicalized case/role fillers in the case-frame of *vendere* (to sell). This is obviously due to the defining pattern used, i.e. *chi vende* (who sells). Some interesting observations can be made with regard to this example.

The first concerns the fact that the same type of result was obtained by making a similar search on an English dictionary. After being shown the Italian example, the IBM Yorktown group repeated the experiment with the same kind of result (see Byrd 1989) for the English data. This shows that there is in fact a correspondence between the definitional patterns used in lexicographical practice independently from the language. This similarity in lexicographical conventions appears in many other examples and will be exploited for the creation of the multilingual LKB which is the ultimate goal of the already mentioned ESPRIT project.

Another observation regards the co-occurrence in these definitions of this kind of verb ("to sell") with another one ("to make", lexicalized in Italian as *fabbricare, fare, preparare*, etc.). Many of these Agent names also apply to the action of "making", and therefore belong to two portions of the resulting conceptual network.

We can also notice that the Noun Phrase following the verb denotes the type of object which is typically sold (or also made) by these Agents.

It is obviously possible to obtain the same type of information on Agents' names for the action of selling if we search for all the nouns whose 'genus term' is the word *venditore* (seller): from this query we retrieve other 131 Agent nouns (see some of them in Fig. 13). Here again some of the nouns are related also with the action of "making", while the PP introduced by the preposition *di* (of) expresses the object which is sold.

This example shows the way in which exactly the same information can be retrieved by browsing the dictionary in different ways, by exploiting the knowledge of its structure (in particular the internal structure of the definitions). In the final LKB all this data will be merged in a single piece of network, independently of the different ways of lexicalizing some concepts and relations.

With a slightly different type of query we can very easily retrieve also the names of the **LOCATIONS** where the action of "selling" is typically performed. Fig. 14 shows the result of the search for the entries in whose definitions the word *vendono* (they sell) is present. Again the fact that names of places are found in this way is due to the following 'defining formula' used by the lexicographers: *dove/in cui si vendono* (where ... are sold). All of the 33 entries retrieved share this definitional pattern: this query is completely without 'noise'.

We can observe that the genus terms are either the generic name *luogo* (place), or those of its hyponyms which are the generic names for the places where something is sold, i.e. *negozio, bottega, bancarella* (shop, store, stall). These are in turn hypernyms of the defined entries. This kind of hierarchical information is already formally coded in the taxonomies stored in the LDB.

What interests us here is the possibility of formalizing and implementing in the LKB the other types of semantic relations, such as **LOCATION** and **THEME** with respect to the actions of "selling" and "making". The Theme relation, i.e. the objects which are typically sold in the defined places are again expressed by the NP object of the verb.

Also in this case similar data are retrieved also by querying for the hyponyms of *negozio, bottega, etc.*. Our aim is to formalize all this information in a semantic network, like the piece sketched in Fig.15.

The above examples show that the LDB facilities can be usefully exploited to analyze and extract linguistic data which must then be restructured and represented in the LKB. In the LKB these types of concepts and of relations, and the interdependencies between word-senses will be explicitly spelled out. When we move beyond taxonomies in the LKB, we establish many different types of associations which are usefully represented in a conceptual network, and when we move from a "monolingual" to a "multilingual" environment, we also establish associations among different languages. These associations are obtained (for those parts of the languages which can be reduced to a common set of concepts and relations) through the common

conceptual network constructed by working on different languages but within the same "research template", i.e. trying to accomodate in the semantic network:
- the "same" world-knowledge,
- for the "same" purposes (NLP, Text Processing, etc.),
- with the "same" methodology,
- from the "same" type of sources (MRDs),
- into the "same" kind of representation.

The common semantic network will thus become the point of convergence of the results of the knowledge aquisition strategies applied on a number of different but homogeneous sources, and the multilingual environment will constitute a valid testbed to evaluate this strategy of design and implementation of a part of a LKB.

4.1 Reusability of bilingual dictionaries

Not only MR monolingual dictionaries, but also bilingual MRDs can be usefully exploited as sources of lexical information for the creation of LDBs and LKBs. These dictionaries can be processed with a twofold purpose, as on the one hand they too are a source of interesting 'monolingual' information, on the other hand they are obviously exploited as a source of links between two monolingual LDBs (see Calzolari, Picchi 1986, and Picchi, Peters, Calzolari, forthcoming).

One of the objectives is to integrate the different types of information traditionally contained in monolingual and bilingual dictionaries, so as to expand the informational content of the single components in the new integrated system. Bilingual dictionaries contain more information about examples of usage, fixed expressions or idioms. This kind of information can obviously be well integrated in the monolingual dictionary, and also made easy to access.

We can envisage the original monolingual lexical entries, augmented with the different types of information coming from the corresponding bilingual entry: different sense discriminations, other examples, syntactic information, collocations, idioms, etc. We can also reverse the perspective, and look at the bilingual entries provided with the information traditionally contained in monolingual entries: mostly definitions. One of the two different viewpoints, both virtually present in the integrated bilingual system, will be simply activated and made available to the user by the first manner of access to the on-line bilingual lexical data base. We would like therefore to maintain in a unique structure both the independent features of the source monolingual and bilingual dictionaries and the integration of the two with different views on the data.

The overall picture of the bilingual LDB system we have in mind is sketched in Fig. 16. Also with regard to bilingual dictionaries, the method we are adopting consists of reusing available data in machine-readable form by analyzing and transforming the information already contained in common dictionaries. The procedure of processing the bilingual MRD is rather similar to the one outlined above for monolingual dictionaries (i.e. parsing of the lexical entry, design of a new structure, computational reorganization, etc.). After this preliminary part again comes out the utility of browsing the bilingual LDB, taking advantage of the structural elements already formalized in the LDB, with the purpose of discovering properties and structures not immediately visible in the printed dictionary, but useful for further exploitation in the computational dictionary.

After the first processing phases that we have envisaged on the bilingual dictionary data, it will make no difference which of the two languages are taken as a starting point. In a certain sense, we would no longer have a source language and a target language, since the look-up and access procedures are independent and neutral with respect to direction (the object becomes bidirectional). Bidirectional cross-references will also be automatically generated for the information contained at each sense level as semantic indicators, i.e. synonyms/hyperonyms or contextual indicators.

One of the parts of the bilingual dictionary we are processing that can be partially made explicit in all its different meanings, is the field of the so-called *semantic indicators*. These provide the constraints for selecting one translation equivalent or the other. The problem is that these constraints are of a different nature, being either i) synonyms or hyponyms of the entry, or ii) contextual indicators such as typical subjects or objects of verbs, typical nouns of which an adjective can be predicated, etc. It is possible to semi-automatize the process of

disambiguation between the different values, after analyzing all the different possibilities and designing a typology of what can appear in this field.

Another possibility is the use of the monolingual lexical data base as a tool to expand the information given as a single word to the whole set of words to which it actually refers. For example, the entry *vivido* has different translations according to the contextual indicators referring to the subject (in brackets):

vivido ..... (***colori***) bright, vivid

In some cases the generic semantic restrictions on the possible object can be taken as a semantic feature, and can be procedurally expanded by the monolingual thesaurus to all the possible hyponyms (at the query moment) so that the appropriate translation can be chosen in any context where a specific name of *colore* (colour) is found (and this is already possible in our monolingual LDB). The information that can be formalized at the semantic level in a monolingual dictionary - which serves to discriminate among the different word-senses - should be in principle of the same type that is given in bilingual dictionaries in the form of "semantic indicators" or "selective conditions" to constrain the choice of a particular translation.

In the same way we can work on other fields in order to make explicit hidden information or to introduce new information on the basis either of structural or of content clues.

After the re-organization of the bilingual MRD in a well-structured LDB, we face the difficult task of using its data to build links between two monolingual LDBs. The difficulty obviously derives from the ambiguity of the words used both as entries and as translations. We never know which word-sense is meant in a particular situation. We shall try to solve this problem as much as possible in the above mentioned ESPRIT project, mostly by exploiting the semantic indicators in the bilingual and the taxonomies and other conceptual information in the monolingual LDBs.

Mapping between word-senses in monolingual dictionaries and different translations in a bilingual dictionary is one of the most interesting of the problems concerning the connection of these different types of dictionaries. As one of the main problems in translation is the correct choice among the various meanings of lexically ambiguous words, we feel that it is absolutely necessary also for a Machine Translation or a Machine Assisted Translation system to be linked to a linguistic data base, i.e. a source of lexical information organized in the form of a thesaurus by multi-dimensional taxonomies, where the possibility of disambiguating lexical items is at least semi-automatized.

One of the main uses of the system should be that of machine-aided translation (MAT), as a powerful aid for translators. The end result may in fact be viewed as a 'translator workstation', where access is provided to many types of dictionaries and other lexical resources, and where the power and the functions of lexical data bases and of textual data bases is exploited at best.

Other purposes of a Bilingual System like the one which appears in Fig. 16 are the following:
- a tool for lexicographers;
- a tool for lexicological-contrastive studies;
- a means for improving monolingual LDBs;
- an aid to construct Machine Translation dictionaries;
- a tool for language teaching;
- a computerized dictionary for "normal" users.

In our opinion, one of the main advantages of a bilingual LDB is the completely different type of "navigation" within its data, made possible both by the multiple access to its data and by its links to the monolingual LDB. In particular, it is not only possible to create links between couples of words in L1 and L2, as in the printed dictionary, but mainly between groups or families of semantically connected words, which we think is an essential property for a true bilingual dictionary and for all the purposes we have listed above.

4.2 Reusability of textual corpora and their integration into LKBs

We have seen that MRDs are very valuable sources of lexical and also of semantic information, but unfortunately not all what is needed to know about the lexicon is there. There are very important pieces of information which in MRDs are completely missing, or incomplete,

or simply are not very good or reliable or easily recoverable. For this type of information, we have to resort to different types of sources (see also Calzolari, 1989a).

Certain kinds of data can probably be acquired only after theoretical investigation of lexical facts, and their source can be seen in the typical linguists' work, mainly based on introspection and native speaker's intuition. In this paper we do not deal with this data, but we must be aware of its existence.

We want to stress here that there are many types of data which can be usefully extracted, more or less directly, by processing very large corpora of textual data. The results of this processing have also to be analysed and evaluated by the linguist and/or the lexicographer, but it is important to realize that for certain types of linguistic phenomena the study made through corpus analysis is 'favoured' with respect to introspection: typical examples are collocations and fixed phrases. A tentative, but not exhaustive, list of lexical information for which we can find data in textual corpora, with various degrees of difficulty and at various levels of completeness, is the following:

- frequency data (at the level of word, word-form, word-sense, word associations, etc.);
- subcategorization;
- collocations, fixed phrases, idioms;
- thematic roles, valency;
- semantic constraints on arguments;
- typical Subject, Object, Modifier, etc. (these are different from the types of thematic roles, being in fact their fillers; in a certain sense they are the same information but given "by example");
- aspectual information;
- proper nouns.

Let us take for example the verb *dividere* (to divide), and look at its occurrences and contexts in our Corpus of about 10 million words. From a total of 840 concordances, we obtain the most frequent syntactic patterns which are as follows:

dividere	NP **in** NP	**268**
"	NP	**175**
"	(NP) **tra** NP , NP , ...	**80**
"	NP **con** NP	**78**
		601

while the remaining 239 contexts are distributed in about 10 other subcategorization frames. If we analyze the contexts by hand, we see that each subcategorization frame can very often be correlated with one or more word-senses, so that we can think of using these frames as a very useful aid in a meaning disambiguation task. By analyzing concordances we can thus obtain data concerning:

a) syntactic frames;
b) their frequency ordering, and therefore their respective relevance for the user;
c) co-occurrences with other words and word classes (at the syntactic and semantic levels);
d) main word-senses;
e) correlation between word-senses and syntactic frames.

We must notice here that it is essential to pay attention to different types of texts, and therefore it is important a good balancing in a reference corpus, because frequency data (at any level: lexical, syntactical, semantic, collocational, etc.) can be very different for different text types.

Let us now consider again the word *libro* (book) for another example of information obtained from texts. If we look at the verbs related to books in the Italian dictionary we can notice that neither *leggere* (to read) nor *scrivere, pubblicare, etc.* (to write, publish) are among them. Again, the same observation has been made with regard to English dictionaries (see Boguraev et al., 1989), which is not by chance, but is again a clear indication of the similarity even between dictionaries of different languages.

In the definitions of these verbs we usually find more generic words related with printed things, such as *scrittura, parole, segni, lettere, scritto, opera, volume, giornale* (writing, words, signs, letters, script, work, volume, journal). The word "book" appears instead in some examples. The link could only be established indirectly, given that the word *libro* is defined in terms of words such as *volume, opera, scritti, stampati, ...*, the same words that appear in the definitions of the above verbs.

These verbs are instead directly associated with *libro* in the corpus of texts. Here, in fact, out of 3,222 concordances of the lemma *libro*, we find these figures for the above-mentioned verbs in the same contexts with *libro*:

leggere	187
scrivere	196
pubblicare	107

It is the analysis of large textual corpora that makes it possible to find this type of collocational information. We are also implementing some statistical/quantitative tools to allow semi-automatic extraction of this and other types of data from our corpus (see Bindi, Calzolari, forthcoming).

When analyzing a large corpus with millions of words in context, we are in a sense compelled to discover and describe:
- usages which are not described in commercial dictionaries;
- relative frequencies of the different word-senses, and of the different syntactic frames/patterns;
- and, above all, the grammatical/syntactic clues by which semantic disambiguation can be at least partially achieved, given the fact that i) in the presence of different syntactic constituency word-sense usually changes, ii) while, vice-versa, we do not necessarily have only one word-sense with the same syntactic frame.

When collecting this type of data for a number of words, we often realize that the data should be reorganized in a different way from how they are presently found in standard dictionaries, if they are to conform to the actual usage of the language.

In order to automatize the retrieval of this type of information directly from the corpus we should first be able to tag the corpus for the different POSs. For this task many systems already exist (see e.g. Hindle 1989, Webster, Marcus 1989). It should then be possible, even without a complete parser, to apply to the text corpus some pattern-matching procedures (as those we are presently using with dictionary definitions). These pattern-matching procedures should be explicitly geared to the extraction of the type of data we are searching (i.e. prepositional phrases, that-clauses, infinitives, etc.).

The same strategy of looking for syntactic (and collocational) clues for semantic disambiguation (to be used for different translations of the same word) is now evaluated in a pilot project we are carrying out in a multilingual context.

5. The lexicographer's workstation as a model of integration of tools and data from different environments and expertises

The importance of a collaboration between researchers working in the fields of CL/NLP and LLC/TP (as already said in more general terms in the first sections of this paper) is evident when we consider that it is necessary to process large textual corpora in order to achieve better LKBs. The design of these large integrated LKBs can really become the purpose of cooperative projects, where the "typical" data, tools, procedures, knowledge, expertise, results, etc., of the two areas of CL/NLP and LLC/TP "must" work in parallel and cooperate and interact with each other.

In order to achieve at least some of the results outlined so far, we can summarize the needs as follows:
- design and implementation of powerful tools;
- large sets of lexical and textual data;
- very modular systems;
- possibility of sharing resources, data and procedures;

- large cooperation among traditionally different research or industrial communities.

A model of the type of integration we have in mind can be seen in the lexicographer's workstation (LW) we are designing in Pisa (see Calzolari, Picchi, Zampolli 1987). It is conceived as a very modular system, where different types of data and of procedures are integrated. At the level of data the LW contains, or will contain: a textual data base, one or more monolingual lexical databases, a thesaurus with taxonomic information, bilingual lexical databases, a reference corpus, etc., while at the level of procedures, it contains: a morphological tool, dictionary parsers, a hyponym finder, an information retrieval system, a lemmatization package, a pattern-matching procedure for dictionary definitions, a redaction tool, etc.

This complex and various set of components reflects our view of the need for an integration and interaction between data and tools traditionally pertinent and pertaining either to CL or to LLC only. It appears therefore important the realization of a factive cooperation among many different groups of researchers (meaning here 'groups' as 'types'), with the aim of linking together worlds which up until now have not been so strongly related to each other, especially perhaps in the American tradition.

PASSIONARIO	1SM	ANTICO LIBRO LITURGICO CATTOLICO	3		
OMILIARIO	1SM	ANTICO LIBRO LITURGICO CONTENENTE OMELIE	1		
EPISTOLARIO	1SM	LIBRO CHE CONTENEVA BRANI DI EPISTOLE E VANGELO	3		
ORA	1SF	LIBRO CHE CONTENEVA LE OPERAZIONI PROPRIE DELLE VARIE ORE	9		
SALTERIO	2SM	LIBRO CHE CONTIENE I SALMI	3		
RITUALE	2SM	LIBRO CHE CONTIENE LE NORME CHE REGOLANO UN RITO	3		
UFFICIOLO	1SM	LIBRO CHE CONTIENE LE PREGHIERE IN ONORE DELLA VERGINE	3		
UFIZIOLO	1SM	LIBRO CHE CONTIENE LE PREGHIERE IN ONORE DELLA VERGINE	3		
CANTORINO	1SM	LIBRO CHE CONTIENE LE REGOLE DEL CANTO FERMO	3		
PORTULANO	1SM	1LIBRO CHE DESCRIVE MINUTAMENTE LA COSTA	34Z		
GUIDA	1SF	LIBRO CHE INSEGNA PRIMI ELEMENTI DI ARTE O TECNICA	3		
GRADUALE	2SM	LIBRO CHE RACCOGLIE I GRADUALI DELL'ANNO LITURGICO	3		
GIORNALMASTRO	1SM	LIBRO CHE RIUNISCE IL GIORNALE E IL MASTRO,PER CONTABILITA'	3		
ANNUARIO	1SM	LIBRO CHE SI PUBBLICA ANNUALMENTE	3		
....					
EFEMERIDE	1SF	LIBRO IN CUI ERANO ANNOTATI I FATTI CHE ACCADEVANO OGNI GIOR	3		
EFFEMERIDE	1SF	LIBRO IN CUI ERANO ANNOTATI I FATTI CHE ACCADEVANO OGNI GIOR	3		
COPIAFATTURE	1SM	LIBRO IN CUI SI COPIANO LE FATTURE	3		
SALDACONTI	1SM	LIBRO IN CUI SONO REGISTRATI I CREDITI E I DEBITI	3		
TASCABILE	2SM	LIBRO IN EDIZIONE ECONOMICA E PICCOLO FORMATO	3		
PERGAMENO	1SM	1LIBRO IN PERGAMENA	3	1	E
BENEDIZIONALE	1SM	LIBRO LITURGICO	3		
MESSALE	1SM	LIBRO LITURGICO CATTOLICO	3		
LEZIONARIO	1SM	LIBRO LITURGICO CON LE#LEZIONI(LEZIONE)DI UFFICI DIVINI	3		
CORALE	2SM	LIBRO LITURGICO CONTENENTE GLI UFFICI DEL#CORO()	1		
EVANGELIARIO	1SM	LIBRO LITURGICO CONTENENTE PASSI DELL' EVANGELO	1		
INNARIO	1SM	LIBRO LITURGICO,NEL CATTOLICESIMO E NELLE CHIESE ORIENTALI	3		
....					
CORANO	1SM	LIBRO SACRO DEI MUSSULMANI	3		
AVESTA	1SM	LIBRO SACRO DELLA RELIGIONE ZOROASTRIANA	3		
GENESI	1SF	PRIMO LIBRO DEL PENTATEUCO NELLA BIBBIA	3		
ALBO	2SM	SPECIE DI LIBRO CONTENENTE FOTOGRAFIE,DISCHI,FRANCOBOLLI	3		
LEVITICO	2SM	TERZO LIBRO BIBLICO DEL PENTATEUCO	9		
SAPIENZA	1SF	UNO DEI LIBRI DELL'ANTICO TESTAMENTO	3		
SAPIENZIA	1SF	1UNO DEI LIBRI DELL'ANTICO TESTAMENTO	3		

Fig. 1. Some of the hyponyms of *libro* (book).

ALLIBRARE	1VT	REGISTRARE SU UN LIBRO DI CONTI	1		
CARTOLINARE	1VT	RILEGARE UN LIBRO ALLA RUSTICA	3		
CIRCOLARE	1VIT	PASSARE DALL'UNA ALL'ALTRA PERSONA,DI DANARO,LIBRI	3		E
DISTRIBUIRE	1VT	DIFFONDERE TRA TUTTI I RIVENDITORI LIBRI,GIORNALI	3		
DIVOLGARE	1VTP	1RENDERE FINANZIARIAMENTE DISPONIBILI LIBRI,SAGGI	3		E
DIVULGARE	1VTP	RENDERE FINANZIARIAMENTE DISPONIBILI LIBRI,SAGGI	3		E
INTERFOGLIARE	1VT	INTERPORRE,CUCIRE TRA I FOGLI DI UN LIBRO FOGLI BIANCHI	3		
INTESTARE	1VTP	FORNIRE DI INTESTAZIONE O TITOLO UN LIBRO	1		
RITONDARE	1VT	1PAREGGIARE,TAGLIANDO LE SPORGENZE,DETTO DI LIBRI,TESSUTI	3	1	
SCARTABELLARE	1VT	SCORRERE IN FRETTA E DISORDINATAMENTE LE PAGINE D'UN LIBRO	3		
SCOMPAGINARE	1VTP	DISFARE,ROVINARE LA LEGATURA DI LIBRI	3		
SCRITTURARE	1VT	ANNOTARE,REGISTRARE SU LIBRI O SCRITTURE CONTABILI	3		
SFASCICOLARE	1VT	SCOMPORRE UN LIBRO,UN QUADERNO NEI FASCICOLI DI CUI E' FATTO	3		
SFOGLIARE	2VTP	SCORRERE UN LIBRO RAPIDAMENTE	3		
SFOGLIARE	2VTP	TAGLIARE LE PAGINE DI UN LIBRO	3	3	
SQUADERNARE	1VTP	3VOLTARE E RIVOLTARE PAGINE DI LIBRI,QUADERNI	3	3	
TOSARE	1VT	PAREGGIARE I FOGLI DEI LIBRI NEL RILEGARLI	3	3	E

Fig. 2. Verbs related to *libri* (books).

```
ADESPOTA          1A    3ANONIMO/DETTO DI LIBRO,CODICE,MANOSCRITTO DI AUTORE IGNOTO    5
ADESPOTO          1A     ANONIMO/DETTO DI LIBRO,CODICE,MANOSCRITTO DI AUTORE IGNOTO    5
APOCRIFO          1A     DETTO DI LIBRO NON RICONOSCIUTO COME CANONICO                 3
CARTOLIBRARIO     1A     DI COMMERCIO DI LIBRI E OGGETTI DA CANCELLERIA               3
CIRCOLANTE        1A     CHE DA' LIBRI A PRESTITO AGLI ABBONATI A TURNO               9
COMMERCIALE       1A     DETTO DI LIBRO,FILM CHE MIRA SOLO A OTTENERE BUONI INCASSI    3      F
COPERTINATO       1A     DETTO DI LIBRO O FASCICOLO CON COPERTINA                     1
DEUTEROCANONICO   1A     DEI LIBRI DELL'ANTICO TESTAMENTO RESPINTI COME APOCRIFI      3
EDITORE           1A     CHI PUBBLICA LIBRI,RIVISTE                                   3
ERUDITO           1A     LIBRO ERUDITO                                                   T
INTESTATO         1A     FORNITO DI TITOLO O INTESTAZIONE,DETTO DI LIBRO,LETTERA-     3
INTONSO           1A    3DI LIBRO CUI NON SONO ANCORA STATE TAGLIATE LE PAGINE        3      F
LIBERIANO         3A     CHE RIGUARDA IL LIBRO                                       36K
LIBRARIO          1A     DI,RELATIVO A LIBRO                                          1
LIBRESCO          1A     CHE DERIVA DAI LIBRI E NON DALLA VIVA ESPERIENZA              1      P
MASTRO            2A     LIBRO MASTRO                                                    L
MOSAICO           2A     RELATIVO AI LIBRI BIBLICI                                    3
PAGA              4A     LIBRO PAGA                                                      L
POSTUMO           1A     DI LIBRO PUBBLICATO DOPO LA MORTE DELL'AUTORE                3
PROTOCANONICO     1A     DETTO DI CIASCUN LIBRO BIBLICO INSERITO PER PRIMO NEL CANONE 3
SAPIENZIALE       1A     CHE SI RIFERISCE AI LIBRI SAPIENZIALI                        3          E
```

Fig. 3. Adjectives related to *libri* (books).

```
RISVOLTO          1SM     ALETTA/ PARTE DELLA SOPRACOPERTA DI LIBRO RIPIEGATA        5
BIBLIOFILO        1SG     AMATORE,RICERCATORE,COLLEZIONISTA DI LIBRI                 3
BIBLIOFILIA       1SF     AMORE PER I LIBRI                                          3
REGGILIBRI        1SM     ARNESE PIEGATO AD ANGOLO RETTO PER REGGERE IN PIEDI LIBRI  3
BIBLIOIATRICA     1SF    3ARTE DEL RESTAURO DEI LIBRI                                3   3
ERMENEUTICA       1SF     ARTE DI INTERPRETARE MONUMENTI,LIBRI ANTICHI               3
SFOGLIATA         2SF     ATTO DELLO SCORRERE UN LIBRO E SIMILI                      1
PUBBLICAZIONE     1SF     ATTO EFFETTO DEL RENDERE PUBBLICO O DEL PUBBLICARE LIBRI   1
BANCHEROZZO       1SM    1BANCARELLA DI LIBRI ALL' APERTO                            3   1
ZAZZERA           1SF     BARBA,RICCIO/ PARTE RUVIDA INTONSA DEI LIBRI               5
PORTACARTE        1SM     BORSA PER METTERVI CARTE,DOCUMENTI,LIBRI                   3
BOTTELLO          1SM    3CARTELLINO CHE SI METTE SU LIBRI E BOTTIGLIE               3   3
CARTOLIBRERIA     1SF     CARTOLERIA AUTORIZZATA ALLA VENDITA DI LIBRI               3
CANONE            1SM     CATALOGO DEI LIBRI SACRI RICONOSCIUTI AUTENTICI            3
REDATTORE         1SN     CHI CURA FASI PER PUBBLICAZIONE DI LIBRI IN CASE EDITRICI  3
CARRETTINISTA     1SM     CHI ESPONE O VENDE LIBRI SU UN CARRETTINO                  1
BIBLIOTECA        1SF     COLLEZIONE DI LIBRI SIMILI PER FORMATO ARGOMENTO EDITORE   3
LIBRATA           1SF     COLPO DATO CON UN LIBRO                                    1
....
BIBLIOTECA        1SF     EDIFICIO CON RACCOLTE DI LIBRI A DISPOSIZIONE DEL PUBBLICO 3
BIBLIOGRAFIA      1SF     ELENCO DI LIBRI CONSULTATI PER COMPILAZIONE DI OPERE       3
INDICE            1SM     ELENCO ORDINATO DI CAPITOLI O PARTI DI LIBRO               3
BIBLIOLATRIA      1SF     FEDE CIECA NEI LIBRI STAMPATI                              3
....                                                                                39Q
LIBRERIA          1SF     LUOGO O MOBILE IN CUI SONO ACCOLTI E CUSTODITI I LIBRI     3          C
BIBLIOTECA        1SF     LUOGO OVE SONO RACCOLTI E CONSERVATI LIBRI                 3
BIBLIOMANIA       1SF     MANIA DI RICERCARE E COLLEZIONARE LIBRI                    3
BIBLIOTECA        1SF     MOBILE A MURO CON SCAFFALI PER LIBRI                       3
CLASSIFICATORE    1SN     MOBILE PER CONTENERE LIBRI DOCUMENTI                       3
LIBRERIA          1SF     NEGOZIO O EMPORIO DI LIBRI
FRONTISPIZIO      1SM     PAGINA ALL' INIZIO DI UN LIBRO CON TITOLO NOTE TIPOGRAFICHE 3
ANTIPORTA         1SF     PAGINA CON TITOLO PRECEDENTE FRONTESPIZIO DI LIBRO          3
TAVOLA            1SF     PAGINA FOGLIO DI LIBRO CON ILLUSTRAZIONI                   3
INTERFOGLIO       1SM     PAGINA INTERPOSTA TRA I FOGLI DI UN LIBRO                  3
LIBRERIA          1SF     RACCOLTA DI LIBRI LIBRO                                    1
BIBLIOLOGIA       1SF     SCIENZA DEI LIBRI                                          3
LIBRAIO           1SN     VENDITORE DI LIBRI                                         1
LIBRARO           1SN    1VENDITORE DI LIBRI
VERSO             3SM     VERSETTO/SUDDIVISIONE IN FRASI DELLE PARTI DI LIBRI SACRI   5          E
```

Fig. 4. Some of the nouns related to *libri* (books).

```
STRUMENTO                 ---->>ABBASSALINGUA                1SM    00
                                ABERROMETRO                  1SM    00
                                ACCELEROGRAFO                1SM    00
                                ACCELEROMETRO                1SM    00
                                ACCHIAPPAMOSCHE              1SN    00
                                ACCIAINO                     1SM    00
                                AEROFONO                     1SM    00
                                AEROMETRO                    1SM    00
                                AEROSCOPIO                   1SM    00
                                AFFILATOIO                   1SM    00
                                AGGUAGLIATOIO                1SM    00
                                AGO                          1SM    0A
                                ALCOOLIMETRO                 1SM    00
                                ALGESIMETRO                  1SM    00
                                AMMOSTATOIO                  1SM    00
                                AMPEROMETRO                  1SM    00
                                ANALIZZATORE                 1SN    00
                                ANCORA                       1SF    10
                                ANEMOMETRO                   1SM    00
                                ANEMOSCOPIO                  1SM    00
                                ANGELICA                     1SF    00
                                APRIBOCCA                    1SM    00
                                APRICASSE                    1SM    00
                                ARCHIPENDOLO                 1SM    00
                                ARMA                         1SF    00
                                ARMONICA                     1SF    00
                                ARMONIO                      1SM    00
                                ARMONIUM                     1SM    00
                                ARPA                         1SF    10
                                ARPEGGIONE                   1SM    00
                                ARRIDATOIO                   1SM    00
                                ASPERSORIO                   1SM    00
                                ASPIRATORE                   1SM    00
                                ASSIOMETRO                   1SM    00
                                ASTIGMOMETRO                 1SM    00
                                ASTROFOTOMETRO               1SM    00
                                ASTROGRAFO                   1SM    00
                                ASTROLABIO                   1SM    00
                                ATTINOMETRO                  1SM    00
                                ATTREZZO                     1SM    0A
                                AUDIOMETRO                   1SM    00
                                AULOS                        1SM    00
                                AVENA                        1SF    00
                                BADILE                       1SM    00
```

Fig. 5. The first hyponyms of *strumento* (instrument).

```
AFFOSSATORE          1SN     ATTREZZO AGRICOLO PER SCAVARE FOSSI                                3
ALLARGATESE          1SM     ATTREZZO USATO PER ALLARGARE LE TESE DEI CAPPELLI                  3
ALLISCIATOIO         1SM     ATTREZZO USATO IN FONDERIA PER PREPARARE LE FORME                  3
ANELLO               1SM     ATTREZZO GEMELLARE IN GINNASTICA                                   3
APISCAMPO            1SM     ATTREZZO PER IMPEDIRE L' ASCESA DELLE API AL MELARIO               3
APPOGGIO             1SM     ATTREZZO GINNICO FORMATO DA BLOCCHETTI RETTANGOLARI DI LEGNO       3
ARATRO               1SM     ATTREZZO AGRICOLO ATTO A ROMPERE,DISSODARE IL TERRENO              3
ARNESE               1SM     ATTREZZO DA LAVORO                                                 3
ASPO                 1SM     ASPA,ANNASPO,NASPO/ ATTREZZO CHE SERVE AD ESEGUIRE L'ASPATURA      54E
ASTA                 1SF     ATTREZZO DI FORMA TUBOLARE NELL' ATLETICA                          3
BACCHETTA            1SF     ATTREZZO PER ESERCIZI GINNICI COLLETTIVI                           3
BARRAMINA            1SF     ATTREZZO PER LA PERFORAZIONE DELLE ROCCE                           3
BASTONCINO           1SM     ATTREZZO DEGLI SCIATORI CON RACCHETTA CIRCOLARE                    3
BASTONE              1SM     MAZZA/ ATTREZZO SPORTIVO                                           5
CACCIAVITE           1SM     ATTREZZO PER STRINGERE O ALLENTARE LE VITI                         3
CAVALLINA            1SF     ATTREZZO PER ESERCIZI DI VOLTEGGIO NELLA GINNASTICA                3
CAVALLO              1SD     ATTREZZO PER ESERCIZI DI VOLTEGGIO NELLA GINNASTICA                3   5
CERCHIO              1SM     ATTREZZO STRUTTURA FIGURA A FORMA DI CERCHIO                       3
CESTA                1SF     CHISTERA/ ATTREZZO DI VIMINI USATO NELLA PELOTA BASCA              5
CHIAVE               1SF     ATTREZZO METALLICO PER PROVOCARE CONTATTI                          3
CHIAVE               1SF     ATTREZZO METALLICO PER METTERE IN MOTO MECCANISMI                  3
CHIAVE               1SF     ATTREZZO METALLICO PER ALLENTARE E STRINGERE VITI O DADI           3
CHIODO               1SM     ATTREZZO IN METALLO DEGLI ALPINISTI                                3
CHIOVO               1SM    1ATTREZZO IN METALLO DEGLI ALPINISTI                                3   1
CILINDRO             1SM     ATTREZZO CILINDRICO NELLA GINNASTICA                               3
CLAVA                1SF     ATTREZZO IN LEGNO USATO PER ESERCIZI GINNICI                       3
COLTIVATORE          2SN     ATTREZZO PER SMUOVERE E SMINUZZARE LA SUPERFICIE DEL TERRENO       3
CORDA                1SF     ATTREZZO DA ALPINISMO O GINNASTICA                                 39L
CUCCHIAIA            1SF     ATTREZZO PER ESTRARRE DETRITI DI ROCCIA                            3
CUCITRICE            2SF     ATTREZZO USATO NEGLI UFFICI PER UNIRE FOGLI                        3
DISCO                1SM     ATTREZZO CIRCOLARE CHE SI LANCIA IN GARE SPORTIVE                  3
ERPICE               1SM     ATTREZZO DI FERRO PER LAVORARE IL TERRENO                          3
ESTENSORE            2SI     ATTREZZO GINNICO                                                   3
ESTIRPATORE          3SM     ATTREZZO PER SMUOVERE O LIBERARE IL TERRENO DA ERBACCE             3
FALCE                1SF     ATTREZZO PER TAGLIARE A MANO CEREALI ED ERBE                       3
FIOCINA              1SF     ATTREZZO CON TRE O PIU' DENTI FISSI PER CATTURARE PESCI            3
....
UTENSILE             2SM     OGNI ATTREZZO PER LAVORARE LEGNO,PIETRE,MATERIALI                  3
VANGHETTA            1SF     ATTREZZO LEGGERO DI SOLDATO PER PICCOLI LAVORI DI STERRO           3
VOGADORE             1SI    1ATTREZZO GINNICO PER MOVIMENTO DA REMATORE                         3
VOGATORE             1SN     ATTREZZO GINNICO PER MOVIMENTO DA REMATORE                         3
VOLTARISO            1SM     ATTREZZO PER RIVOLTARE SULL'AIA MODESTE QUANTITA' DI RISO          3
ZAPPA                1SF     ATTREZZO MANUALE PER LAVORARE IL TERRENO                           3
```

Fig. 6. Some of the hyponyms of *attrezzo* (tool) with their definitions.

INSTRUMENT <--IS-A--	*attrezzo*	**--USED FOR-->**	*tagliare ...*	= *FALCE*
	(tool)		...	= ...
		--USED IN-->	*ginnastica*	= *ANELLO*
			...	= ...
		--SHAPE-->	*tubolare*	= *ASTA*
		"	*circolare*	= *DISCO*
		--MADE OF-->	*vimini*	= *CESTA*
			metallo	= *CHIODO*

Fig. 7. Sketch of a piece of network for *attrezzo* (tool).

FORMICAIO	SM	MOLTITUDINE DI	PERSONE
GREGGE	SN	MOLTITUDINE DI	PERSONE
STORMO	SM	MOLTITUDINE DI	PERSONE
MANO	SF	GRUPPO DI	PERSONE
ROSA	SF	CERCHIA/ GRUPPO INSIEME DI	PERSONE
BRANCO	SM	INSIEME DI	PERSONE
CIRCOLO	SM	CENACOLO,SODALIZIO/INSIEME DI	PERSONE
COMMISSIONE	SF	GRUPPO DI	PERSONE A CUI E' AFFIDATO UN UNCARICO PUBBLICO
POPOLAZIONE	SF	INSIEME DELLE	PERSONE ABITANTI IN UN LUOGO
ORGANICO	SM	COMPLESSO DI	PERSONE ADDETTE A CERTE ATTIVITA'
SEGRETERIA	SF	INSIEME DELLE	PERSONE ADDETTE A UNA SEGRETERIA
SQUADRA	SF	COMPLESSO DI	PERSONE ADDETTE A UNO STESSO LAVORO
CIURMA	SF	INSIEME DELLE	PERSONE ADDETTE AI LAVORI DELLA TONNARA
NAZIONE	SF	INSIEME DI	PERSONE APPARTENENTI A STESSA STIRPE
FAMIGLIA	SF	COMPLESSO DI	PERSONE AVENTI UN ASCENDENTE DIRETTO COMUNE
VICINATO	SM	INSIEME DI	PERSONE CHE ABITANO UNA STESSA CASA
CORTE	SF	GRUPPO DI	PERSONE CHE ACCOMPAGNA UN PERSONAGGIO IMPORTANTE
LEGA	SF	INSIEME DI	PERSONE CHE AGISCONO PER UTILE PROPRIO
AUDITORIO	SM	UDITORIO/COMPLESSO DI	PERSONE CHE ASCOLTANO
UDIENZA	SF	UDITORIO/INSIEME DI	PERSONE CHE ASCOLTANO
CAROVANA	SF	GRUPPO DI	PERSONE CHE ATTRAVERSANO CON CARRI LUOGHI DESERTI
CORO	SM	GRUPPO DI	PERSONE CHE CANTANO INSIEME
MALAVITA	SF	L'INSIEME DELLE	PERSONE CHE CONDUCONO VITA DISSOLUTA
CROCCHIO	SM	GRUPPO DI	PERSONE CHE CONVERSANO
CORO	SM	GRUPPO DI	PERSONE CHE DICONO,GRIDANO Q.C. CONTEMPORANEAMENTE
CONCISTORO	SM	GRUPPO DI	PERSONE CHE DISCUTONO
FINANZA	SF	COMPLESSO DI	PERSONE CHE ESPLICANO ATTIVITA' BANCARIA
....			
FRONTE	SN	COMPLESSO DI	PERSONE OMOGENEO PER FINALITA' CONSUETUDINI
ARISTOCRAZIA	SF	COMPLESSO DI	PERSONE PIU' QUALIFICATE PER UNA ATTIVITA'
CHIESA	SF	INSIEME DI	PERSONE PROFESSANTI LA MEDESIMA DOTTRINA
DRAPPELLO	SM	GRUPPO DI	PERSONE RACCOLTE INSIEME
COMPAGNIA	SF	COMPLESSO DI	PERSONE RIUNITE INSIEME PER ATTIVITA' COMUNI
GRUPPO	SM	INSIEME DI	PERSONE UNITE DA VINCOLI NATURALI O DI INTERESSE

Fig. 8. Some of the nouns denoting **SET OF** *persone* (people).

ARCIPELAGO	SM	GRUPPO INSIEME DI	OGGETTI
ANTIQUARIATO	SM	COMMERCIO O RACCOLTA DI	OGGETTI ANTICHI
SERVIZIO	SM	INSIEME DI	OGGETTI CHE SERVONO A UN DETERMINATO SCOPO
TROFEO	SM	INSIEME DI	OGGETTI CHE TESTIMONIANO SUCCESSI E VITTORIE
AFFARDELLAMENTO	SM	COMPLESSO DEGLI	OGGETTI CONTENUTI NELLO ZAINO DEL SOLDATO
ARGENTERIA	SF	COMPLESSO DI	OGGETTI D'ARGENTO
ORERIA	SF	COMPLESSO DI	OGGETTI D'ORO
COLLEZIONE	SF	RACCOLTA DI	OGGETTI DELLA STESSA SPECIE
CRISTALLERIA	SF	INSIEME DEGLI	OGGETTI DI CRISTALLO DA TAVOLA
CIANFRUSAGLIA	SF	CHINCAGLIERIA/INSIEME DI	OGGETTI DI POCO PREGIO
CIANFRUSCAGLIA	SF	CHINCAGLIERIA/INSIEME DI	OGGETTI DI POCO PREGIO
ASSORTIMENTO	SM	INSIEME DI	OGGETTI DI STESSO GENERE DIVERSI NEI PARTICOLARI
ARSENALE	SM	INSIEME DI	OGGETTI DIVERSI
SUPPELLETTILE	SF	OGGETTO O INSIEME DI	OGGETTI IN UNA SCUOLA CHIESA E SIMILI
INTRECCIO	SM	COMPLESSO DI	OGGETTI INTRECCIATI
ATTREZZERIA	SF	INSIEME DI	OGGETTI NECESSARI PER UNA SCENA TEATRALE
SUPPELLETTILE	SF	OGGETTO O INSIEME DI	OGGETTI NELL'ARREDAMENTO DELLA CASA
ARREDO	SM	OGGETTO O COMPLESSO DI	OGGETTI PER GUARNIRE AMBIENTI
COMPLETO	SM	INSIEME DI	OGGETTI PER UN USO DETERMINATO
BAROCCUME	SM	INSIEME DI	OGGETTI PRETENZIOSI E DI CATTIVO GUSTO
GIOIELLERIA	SF	INSIEME DI	OGGETTI PREZIOSI
SUPPELLETTILE	SF	OGGETTO O INSIEME DI	OGGETTI RINVENUTI IN UNO SCAVO

Fig. 9. Nouns denoting **SET OF** *oggetti* (objects).

ASSESTATO	A	ASSENNATO,AVVEDUTO,DETTO DI	PERSONA
BARLACCIO	A	MALATICCIO,DEBOLE,DETTO DI	PERSONA
INSENSATO	A	STUPIDO,DEMENTE,DETTO DI	PERSONA
PRIMITIVO	A	C=INCIVILITO/SEMPLICE,ROZZO,CREDULONE,DETTO DI	PERSONA
PROVETTO	A	MATURO,DETTO DI	PERSONA
RIMESSO	A	LANGUIDO,LENTO,FIACCO,DETTO DI	PERSONA
RINCRESCIOSO	A	CHE SENTE RINCRESCIMENTO,DETTO DI	PERSONA
RIPOSANTE	A	CALMO,TRANQUILLO DETTO DI	PERSONA
RISPETTOSO	A	CHE HA,E' PIENO DI#RISPETTO(),DETTO DI	PERSONA
ROBUSTO	A	FORTE/CHE POSSIEDE FORZA,ENERGIA,DETTO DI	PERSONA
ROCO	A	RAUCO,DETTO DI	PERSONA
ROGNOSO	A	MISERO,MESCHINO,NOIOSO,DETTO DI	PERSONA
RUDE	A	ROZZO,GROSSOLANO,DETTO DI	PERSONA
RUGIADOSO	A	SANO,FLORIDO,DETTO DI	PERSONA
RUSTICO	A	NON MOLTO SOCIEVOLE NE' RAFFINATO,DETTO DI	PERSONA
RUVIDO	A	DI MANIERE ROZZE,DI CARATTERE ASPRO,DETTO DI	PERSONA
....			PERSONA
ADOMBRARE	VTE	INSOSPETTIRSI,TURBARSI,DETTO DI	PERSONA
ARRABBIARE	VIE	ESSERE PRESO DALL'IRA,DALLA COLLERA DETTO DI	PERSONA
CORVETTARE	VI	SALTARE,BALZARE,DETTO SPEC. DI	PERSONA
CUCCIARE	VET	GIACERSI/STARE A LETTO,DETTO DI	PERSONA
IMBIZZARRIRE	VET	INCOLLERIRE O DIVENTARE IRREQUIETO DETTO DI	PERSONA
IMPROSCIUTTIRE	VI	DIVENTARE ASCIUTTO COME UN PROSCIUTTO,DETTO DI	PERSONA
RABBRUSCARE	VEY	ADOMBRARSI/OFFUSCARSI IN VOLTO,DETTO DI	PERSONA
RICEVERE	VT	AMMETTERE,DETTO DI	PERSONA
RIDURRE	VT P	METTERE IN CONDIZIONI PEGGIORI,DETTO DI	PERSONA
RIMETTERE	VT PI	RISTABILIRSI,DETTO DI	PERSONA
RINFIERIRE	VI	INFIERIRE DI NUOVO O DI PIU',DETTO DI	PERSONA
RINSECCHIRE	VIT	DIVENTARE MAGRO,ASCIUTTO,DETTO DI	PERSONA
RINVENIRE	VI	RIANIMARSI,RIAVERSI/RICUPERARE I SENSI DETTO DI	PERSONA
RISALTARE	VNI	EMERGERE,DISTINGUERSI,DETTO DI	PERSONA
RISORGERE	VI T	SOLLEVARSI,RIAVERSI DETTO DI	PERSONA
RISPUNTARE	VIT	RIAPPARIRE,RICOMPARIRE,DETTO DI	PERSONA
RISURGERE	VI T	SOLLEVARSI,RIAVERSI,DETTO DI	PERSONA
RIUSCIRE	VI	RAGGIUNGERE IL FINE,LO SCOPO,DETTO DI	PERSONA
ROTOLARE	VTIR	GIRARSI SU DI SE',VOLTOLARSI,DETTO DI	PERSONA
ROVINARE	VITR	CADERE IN BASSO,DETTO DI	PERSONA
....			
CORDIALE	A	DETTO DI	PERSONA AFFABILE,GENTILE,APERTA
LONGO	A	CHE SI ESTENDE IN ALTEZZA,DETTO DI	PERSONA ALTA E MAGRA
LUNGO	A	CHE SI ESTENDE IN ALTEZZA,DETTO DI	PERSONA ALTA E MAGRA
PRODIGIO	A	DETTO DI	PERSONA CHE E' ECCEZIONALE
SUPINO	A	C=PRONO/DETTO DI	PERSONA CHE GIACE SUL DORSO
LACERO	A	CENCIOSO/DETTO DI	PERSONA CHE INDOSSA VESTITI LOGORI
SCIVOLOSO	A	DETTO DI	PERSONA CHE NASCONDE LE SUE VERE INTENZIONI
IMPREGIUDICATO	A	DETTO DI	PERSONA CHE NON HA AVUTO CONDANNE PENALI
IMPETTITO	A	DETTO DI	PERSONA CHE STA ERETTA E COL PETTO IN FUORI
ASOCIALE	A	DETTO DI	PERSONA CHIUSA INTROVERSA
....			
NAUFRAGARE	VI	ESSERE SUL BASTIMENTO CHE ROMPE IN MARE,DETTO DI	PERSONE
RICONGIUNGERE	VT D	CONGIUNGERSI DI NUOVO,RIUNIRSI,DETTO DI	PERSONE
RIMESCOLARE	VTP	INTROMETTERSI,MISCHIARSI A UN GRUPPO,DETTO DI	PERSONE
ROVESCIARE	VTP	ABBANDONARSI,DETTO DI	PERSONE
SBOCCARE	VIT	ARRIVARE IN UN DATO LUOGO,DETTO DI	PERSONE
SCHIAMAZZARE	VI	VOCIARE,STREPITARE,DETTO DI	PERSONE
SPELLICCIARE	VTB	PICCHIARSI,AZZUFFARSI RABBIOSAMENTE,DETTO DI	PERSONE
ULULARE	VI	EMETTERE PROLUNGATI,CUPI LAMENTI,DETTO DI	PERSONE

Fig. 10. Some of the adjectives and verbs which can be predicated of *persone* (people).

ACCESO	A	VIVO,INTENSO,DETTO DI	COLORE
CHIARO	A	C=SCURO/PALLIDO,TENUE,POCO INTENSO DETTO DI	COLORE
CUPO	A	DI TONALITA' SCURA DETTO DI	COLORE
SERPATO	A	CHE E' SCREZIATO,COME LA PELLE DEL SERPENTE,DETTO DI	COLORE
SQUILLANTE	A	VIVACE,INTENSO,DETTO DI	COLORE
STABILE	A	CHE NON SBIADISCE,DETTO DI	COLORE
TENUE	A	PALLIDO/NON MOLTO VIVO DETTO DI	COLORE
RISCHIARARE	VTE	FARSI CHIARO,LUMINOSO,DETTO DI	COLORE
SCARICARE	VTRIP	PERDERE VIVACITA',SBIADIRE,DETTO DI	COLORE
BERRETTINO	A	DETTO DI	COLORE AZZURRO CINEREO SU VASI DI MAIOLICA
CALCE	A	DETTO DI	COLORE BIANCO INTENSO
GIGLIACEO	A	DETTO DI	COLORE CHE RICORDA QUELLO DEL GIGLIO
SCURO	A	C=CHIARO/DETTO DI	COLORE CHE TENDE AL NERO
BRUNO	A	DETTO DEL	COLORE DEL MANTELLO DEI BOVINI
ALBICOCCA	A	DETTO DI	COLORE GIALLO ARANCIATO
ZAFFERANO	A	DETTO DI	COLORE GIALLO INTENSO
ISABELLA	A	DETTO DI	COLORE GIALLO TIPICO DI MANTELLO EQUINO
PERLA	A	DETTO DI	COLORE LATTIGINOSO E OPALESCENTE
TERRA	A	DETTO DI	COLORE MARRONE CHIARO SFUMATO AL GRIGIO
SUDICIO	A	DETTO DI	COLORE NON BRILLANTE,NON VIVO
DISUGUAGLIATO	A	DETTO DI	COLORE NON UNIFORME DI UNA TINTURA
NEGRO	A	DETTO DEL	COLORE PIU' SCURO
NERO	A	DETTO DEL	COLORE PIU' SCURO
GIACINTINO	A	DETTO DEL	COLORE ROSSASTRO,TIPICO DEL GIACINTO
TANGO	A	DETTO DI	COLORE ROSSO ASSAI BRILLANTE
GRANATA	A	DETTO DI	COLORE ROSSO SCURO
PULCE	A	DETTO DI	COLORE TRA GRIGIO E VERDE
RUGGINE	A	DETTO DI	COLORE TRA IL MARRONE E IL ROSSO SCURO
LILLA'	A	GRIGELLINO/DETTO DI	COLORE TRA ROSA E VIOLA
GIADA	A	DETTO DI	COLORE VERDAZZURRO CHIARO
SBIADATO	A	SBIADITO,TENUE,PALLIDO,DETTO DI	COLORI
ADDOLCIRE	VTP	AMMORBIDIRE,DETTO DI	COLORI
DISCORDARE	VE	STONARE/NON ARMONIZZARE,DETTO DI	COLORI
SBIADIRE	VET	SCOLORIRE,STINGERE/DIVENTARE PALLIDO,SMORTO,DETTO DI	COLORI
SGARGIARE	VI	ESSERE ECCESSIVAMENTE VIVACE E VISTOSO,DETTO DI	COLORI
SMONTARE	VTIP	SCHIARIRE,SCOLORIRE,STINGERE,DETTO DI	COLORI
TRIONFARE	VIT	RISALTARE/FARE SPICCO,DETTO DI	COLORI
USCIRE	VIT	RISALTARE DETTO DI	COLORI
SMORTO	A	CHE E' PRIVO DI SPLENDORE E VIVACITA' DETTO DI	COLORI E SIM.
ALLEGRO	A	VIVACE,BRIOSO DETTO DI	COLORI SUONI E SIMILI
RISALTARE	VNI	SPICCARE NITIDAMENTE,DETTO DI	COLORI,DISEGNI,PITTURE
TENDERE	VT IP	AVVICINARSI AD UNA GRADAZIONE DETTO DI	COLORI,SAPORI,ODORI

Fig. 11. Some of the adjectives and verbs which are typically predicated of *colori* (colours).

VENDE ---->>AGNELLAIO	1SI	CHI MACELLA O VENDE AGNELLI		1
AGORAIO	1SM	CHI FA O VENDE AGHI		
ALABASTRAIO	1SI	CHI VENDE OGGETTI DI ALABASTRO		
ARAZZIERE	1SI	CHI TESSE E VENDE ARAZZI		1
ARGENTIERE	1SI	CHI VENDE OGGETTI D'ARGENTO		
ARMAIOLO	1SI	CHI FABBRICA VENDE RIPARA ARMI		
ASTUCCIAIO	1SI	CHI FABBRICA O VENDE ASTUCCI		1
BABBUCCIAIO	1SI	CHI FA O VENDE BABBUCCE		1
BADILAIO	1SI	CHI FA O VENDE BADILI		1
BERRETTAIO	1SN	CHI FABBRICA O VENDE BERRETTI		1
BICCHIERAIO	1SI	CHI FABBRICA O VENDE BICCHIERI		1
BIGLIETTAIO	1SN	CHI VENDE I BIGLIETTI PER IL VIAGGIO		1
BILANCIAIO	1SI	STADERAIO/CHI FABBRICA E VENDE BILANCE		4
BILIARDAIO	1SI	CHI FABBRICA O VENDE BILIARDI		1
BIRRAIO	1SI	CHI FABBRICA O VENDE BIRRA		1
BOCCALAIO	1SI	CHI FABBRICA O VENDE BOCCALI		1
BORSAIO	1SG	CHI FABBRICA O VENDE BORSE		1
BOTTAIO	1SI	CHI FABBRICA,RIPARA O VENDE BOTTI		1
BOTTONAIO	1SN	CHI FABBRICA O VENDE BOTTONI		1
BUSTAIA	1SF	DONNA CHE CONFEZIONA O VENDE BUSTI		1
CALZETTAIO	1SN	CHI VENDE O FABBRICA CALZE		1
CANESTRAIO	1SI	CHI FA O VENDE CANESTRI		1
CARBONAIO	1SM	CHI VENDE CARBONE		1
....				
OROLOGIAIO	1SI	CHI FABBRICA,RIPARA O VENDE OROLOGI		1
ORTOPEDICO	2SI	CHI FABBRICA O VENDE APPARECCHI ORTOPEDICI		3
OTTICO	2SI	CHI CONFEZIONA E VENDE OCCHIALI E LENTI		3
PADELLAIO	1SI	CHI FA O VENDE PADELLE		1
PANETTIERE	1SN	FORNAIO/CHI FA O VENDE PANE		
PANIERAIO	1SG	CHI FA O VENDE PANIERI		
PANTOFOLAIO	1SN	CHI CONFEZIONA O VENDE PANTOFOLE		1
PASTAIO	1SN	CHI FABBRICA O VENDE PASTE ALIMENTARI		1
PASTICCERE	1SN	CHI FA O VENDE DOLCIUMI		
PASTICCIERE	1SN	CHI FA O VENDE DOLCIUMI		
PATACCARO	1SI	2CHI VENDE MONETE OD OGGETTI FALSI		
PELLETTIERE	1SG	CHI PRODUCE O VENDE OGGETTI DI PELLETTERIA		1
PELLICCIAIO	1SN	CHI LAVORA O VENDE PELLICCE		1
....				
VENDITORE	2SI	CHI VENDE		1
VETRAIO	1SI	CHI VENDE TAGLIA APPLICA LASTRE DI VETRO		
VINATTIERE	1SM	1CHE VENDE O COMMERCIA VINO	1	5
VIOLINAIO	1SI	LIUTAIO/CHI FABBRICA O VENDE VIOLINI	4	
ZOCCOLAIO	1SI	CHI FA O VENDE ZOCCOLI	1	

Fig. 12. Nouns of **AGENTS** for the action of "selling".

VENDITORE	---->>ABBACCHIARO	1SI	2VENDITORE DI ABBACCHI	1	2	
	ACQUAVITAIO	1SI	VENDITORE DI ACQUAVITE	1		
	ARCHIBUGIERE	1SM	FABBRICANTE O VENDITORE DI ARMI	3	1	
	....					
	BIBITARO	1SI	2VENDITORE DI BIBITE	1	2	
	BORSETTAIO	1SG	FABBRICANTE O VENDITORE DI BORSE E BORSETTE	1		
	BRONZISTA	1SN	VENDITORE DI OGGETTI ARTISTICI IN BRONZO			
	BURATTINAIO	1SI	FABBRICANTE O VENDITORE DI BURATTINI			
	CALCOGRAFO	1SI	VENDITORE DI INCISIONI	3		
	CALDARROSTAIO	1SN	VENDITORE DI CALDARROSTE	1		
	CAMICIAIO	1SD	FABBRICANTE O VENDITORE DI CAMICIE	1		
	CAPPELLAIO	1SN	FABBRICANTE O VENDITORE DI CAPPELLI DA UOMO	3		
	CARAMELLAIO	1SN	FABBRICANTE O VENDITORE DI CARAMELLE	1		
	....					
	FRUTTIVENDOLO	1SN	VENDITORE DI FRUTTA E ORTAGGI	3		
	LATTAIO	1SN	VENDITORE DI LATTE	1		
	LIBRAIO	1SN	VENDITORE DI LIBRI			
	MACELLAIO	1SN	VENDITORE DI CARNE MACELLATA	3		
	....					
	PROFUMIERE	1SN	FABBRICANTE O VENDITORE DI PROFUMI E COSMETICI	1		
	SALUMIERE	1SN	VENDITORE DI SALUMI	1		
	SPEZIALE	2SI	VENDITORE DI SPEZIE	1	1	
	STRILLONE	1SN	VENDITORE AMBULANTE DI GIORNALI	3		
	VALIGIAIO	1SN	FABBRICANTE O VENDITORE DI VALIGIE BAULI,BORSE	1		
	VINAIO	1SN	VENDITORE FORNITORE DI VINO	1		

Fig. 13. Nouns of **AGENTS** for the action of "selling".

VENDONO	---->>APPALTO	1SM	LUOGO DOVE SI VENDONO PRODOTTI DI MONOPOLIO DELLO STATO	3	2	
	BANCO	1SM	LOCALE DOVE SI VENDONO O SCAMBIANO BENI SERVIZI	3		
	BIGIOTTERIA	1SF	NEGOZIO DOVE SI VENDONO OGGETTI DECORATIVI NON PREZIOSI	3		E
	BIGLIETTERIA	1SF	LUOGO IN CUI SI VENDONO BIGLIETTI	1		
	BISCOTTERIA	1SF	NEGOZIO DOVE SI VENDONO I BISCOTTI			
	BOTTIGLIERIA	1SF	NEGOZIO DOVE SI VENDONO VINO LIQUORI IN BOTTIGLIA	3		
	BRICABRAC	1	NEGOZIO,BANCARELLA OVE SI VENDONO TALI ANTICAGLIE	3		E
	CALZETTERIA	1SF	NEGOZIO IN CUI SI VENDONO CALZE			
	CALZOLERIA	1SF	BOTTEGA IN CUI SI FABBRICANO O VENDONO SCARPE			
	CAMICERIA	1SF	NEGOZIO IN CUI SI VENDONO CAMICIE			
	CAPPELLERIA	1SF	NEGOZIO DOVE SI VENDONO CAPPELLI MASCHILI	1		
	CERERIA	1SF	LUOGO DOVE SI FABBRICANO E VENDONO CANDELE	3		
	CHINCAGLIERIA	1SF	NEGOZIO IN CUI SI VENDONO CHINCAGLIE			
	CONFETTURERIA	1SF	LUOGO OVE SI PREPARANO,VENDONO CONFETTURE	1		
	CREMERIA	1SF	2LATTERIA IN CUI SI VENDONO ANCHE GELATI DOLCI E SIM.	3		
	DIACCIATINO	2SN	2BOTTEGA DOVE SI VENDONO SORBETTI	3	1	
	DROGHERIA	1SF	BOTTEGA DOVE SI VENDONO DROGHE	1		
	FERRAMENTA	1SF	NEGOZIO IN CUI SI VENDONO OGGETTI DI FERRO	3		
	GELATERIA	1SF	SORBETTERIA/NEGOZIO OVE SI FANNO O VENDONO GELATI	4		
	MAGLIERIA	1SF	BOTTEGA NEGOZIO IN CUI VENDONO INDUMENTI DI MAGLIA			
	MESCITA	1SF	BOTTEGA IN CUI SI VENDONO VINO LIQUORI	3	2	
	MESTICHERIA	1SF	2BOTTEGA IN CUI SI VENDONO COLORI MESTICATI	3	2	
	NEGOZIO	1SM	BOTTEGA/ LOCALE DOVE SI ESPONGONO E VENDONO MERCI	5		
	NORCINERIA	1SF	2BOTTEGA IN CUI SI VENDONO SOLO CARNI DI MAIALE	3	2	
	OCCHIALERIA	1SF	NEGOZIO IN CUI SI VENDONO O SI RIPARANO OCCHIALI			
	OROLOGERIA	1SF	NEGOZIO DOVE SI VENDONO OROLOGI	3		
	PANTOFOLERIA	1SF	LUOGO IN CUI SI VENDONO PANTOFOLE			
	PELLETTERIA	1SF	NEGOZIO IN CUI SI VENDONO OGGETTI DI PELLE LAVORATA	3		
	PIATTERIA	1SF	BOTTEGA DOVE SI VENDONO I PIATTI	3		
	ROSTICCERIA	1SF	BOTTEGA DOVE SI PREPARANO O VENDONO ARROSTI	3		
	SALUMERIA	1SF	BOTTEGA,NEGOZIO,IN CUI SI VENDONO I SALUMI	3		
	SPACCIO	1SM	LOCALE DELLE CASERME DOVE SI VENDONO GENERI ALIMENTARI VARI	3		
	UTENSILERIA	1SF	BOTTEGA IN CUI SI VENDONO UTENSILI			

Fig. 14. Nouns of **PLACES** related to the action of "selling".

OROLOGERIA = <--**LOC**-- ***"selling"*** --**THEME**--> orologi --**IS-A**--> OBJECT

OROLOGIAIO = <--**AGENT**-- " " " " "

Fig. 15. Sketch of a piece of network for the action of *" **selling**"*.

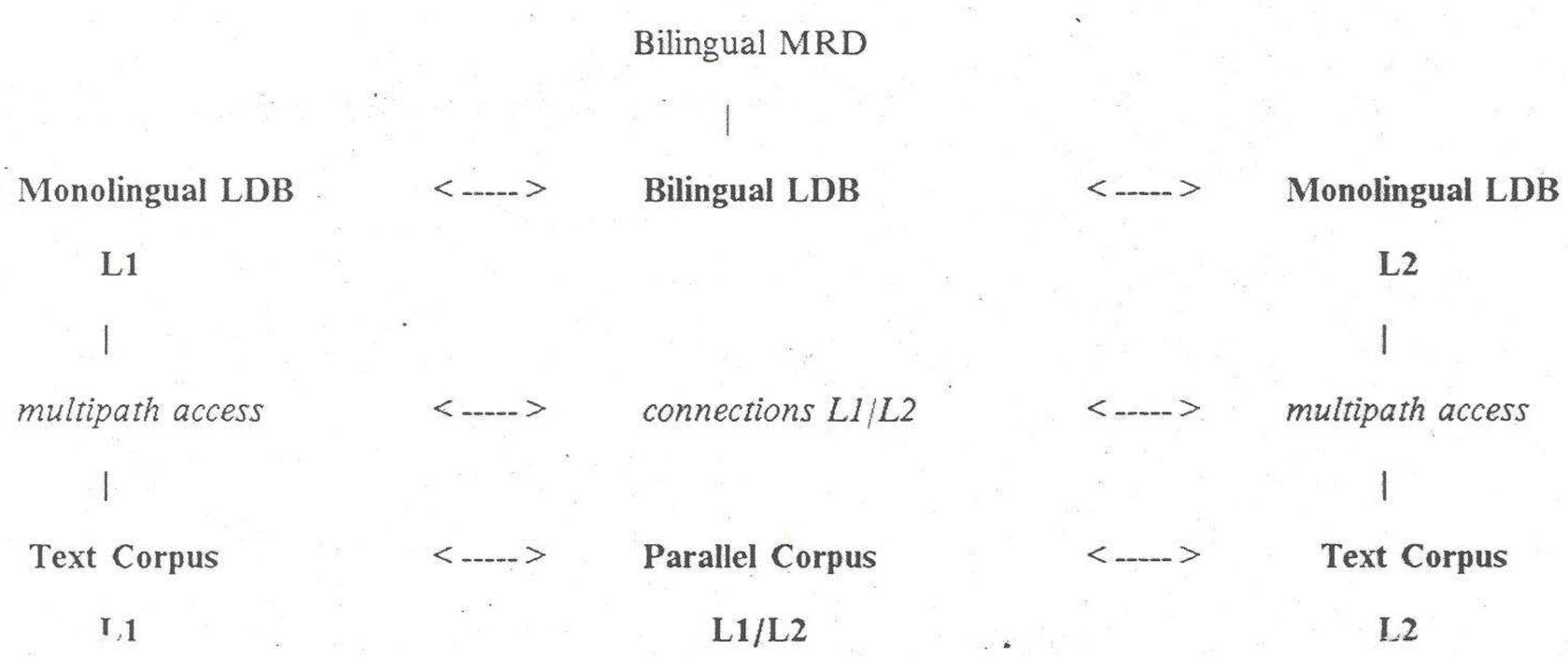

Fig. 16. A model of a Bilingual LDB System.

NOTES

(1) In fact, the first experiments of concordances and indices production were performed not with ′electronic machines′, but with ′punched card electrical accounting machines′ (Busa (1951), 22).

(2) For the history of the first years of MT, see Locke and Booth (1955), 1-23; Booth, Cleave and Brandwood (1958), 1-7; Vauquois (1975), 14-32; Nagao (1989), chapters 1-2.

(3) In the Introduction to the *"Actes du Colloque International sur la Mechanisation des Recherches Lexicologiques"* held in 1961 in Besancon, B. Quemada says: ′Un des buts de ce Colloque sera aussi de mettre en contact des chercheurs qui sans s′ignorer tout a fait, n′echangent guere d′informations alors qu′ils travaillent sur une matiere commune: la langue, et plus particulierement, le lexique dans diverses disciplines. Nous avons la chance d′accueillir ici a cote des lexicologues et des lexicographes francais et etrangers, des specialistes de la traduction automatique (vocabulaire de base, terminologies scientifiques, speciales, dictionnaires automatiques, homographes, synonymes), de la traduction ″artisanale″ (...) de la documentation automatique (...) de la pedagogie des langues vivantes′.
And R. Busa (in an article with a very significant title, given the period: *L′analisi linguistica nell′evoluzione mondiale dei mezzi di informazione* - ″the linguistic analysis in the world evolution of information tools)″ - published as a contribution to a debate on the fracture between sciences and humanities) says that the ′development of linguistic automation is triangular: lexical analysis, information retrieval, mechanical translation′, Busa (1961), 117.

(4) M. Kay (Kay, 1964), reporting on an informal meeting on ″Formats for Machine Readable Texts″ at the end of the IBM-sponsored Literary Data Processing Conference (Yorktown Heights, 1964), and in an article in the fifth issue of the *Computers and the Humanities* (Kay, 1967), explicitly stressed the common interest of MT and humanities researchers on this topic. But it is interesting to note that, in the very same issue, only two MT projects, both directed by well-known linguists, B. Pottier and W.P. Lehmann, are reported in the Directory of Scholars Active, of a total of 120 projects in the section Language and Literature.

(5) But not, we think, directly inspired by it.

(6) At page 2 of the ALPAC Report, the Chairman of the Committee on Science and Public Policy, in a letter to the President of the National Academy of Science, stated that ″the support needs for computational linguistics are distinct from automatic language translation″. At page 29, one reads ″work toward machine translation, together with computational linguistics work that has grown out of it″.

(7) We quote from the Recommendation: ′Small scale experiments and work with miniature models of language have proven seriously deceptive in the past, and one can come to grips with real problems only above a certain scale of grammar size, dictionary size, and available corpora′ (ALPAC, p. iv).

(8) This situation is still true today, ′A recent workshop on linguistic theory and computer applications (Withelock et al., 1987) reports an informal poll to establish the average size of the lexicon used by the prototypes discussed ...; the average size was about 25 words′ (Boguraev and Briscoe (1989) 10).

(9) See the Proceedings of the *Table Ronde sur les Grandes Dictionnaires Historiques* (Firenze, 1973).

(10) See, for example, the series of frequency dictionaries of romance languages coordinated by Juilland, published by Mouton in 1961 (Spanish), 1965 (Rumanian), 1970 (French), 1973 (Italian).

(11) A well-known example is the IBM development of specialized optical support for storing large dictionaries in early ′60.

(12) These two systems were presented and compared at the Pisa 1968 meeting *′De lexico electronico latino′*, during which was also presented the first proposal for a multifunctional lexicon (Italian Machine Dictionary: DMI), conceived as a repository of lexical knowledge both for computer programs (parsers, generators, phonological transcription, lemmatization, etc.) and human uses (qualitative and quantative researches on the structure of the Italian lexical system). The Gallarate Latin machine dictionary was made up of an alphabetical list of forms, progressively accumulated from processing the texts of St. Thomas Aquinus. The *Liege Dictionary* was based on a list of stems, extracted from the Forcellini lemmas, and an associated morphological analyser (See Busa, 1968).

(13) The article, ″The Field and Scope of Computational Linguistics′, of D. Hays in the *Proceedings of the Budapest COLING 1971* is particularly relevant, and it is interesting to observe the evolution towards a

'puristic' definition' of CL in the opinion of the author in respect to his chapter on 'computational linguistics' in the *Encyclopaedia of Linguistics Information and Control* (1969).

(14) The following passages seem to us to be very revealing.
On the one hand H. Karlgren (1973, XIII and XIX-XXI) - from the puristic point of view - wrote: "The characteristic feature of Computational Linguistics is a focus on computation, on the derivation of results by a "mechanical" procedure, operating according to rules, according to an "algorithm". A good tool for computation is, in many cases, a computer, but computational linguistics is not the same as Computer-based Linguistics or Linguistic Data Processing *(Linguistische Datenverarbeitung)*. (...) Linguistic research, like investigation in so many other fields, is often aided by the services of a computer without being, on that account, directed towards problems of computation. Thus lexicographic work is neither more or less computational because the clerical part of it has become easier - or possibly more complicated - thanks to new equipment. The data processing performed in linguistic institutes of various kinds is certainly worth studying in its own right - preferably together with experts of economy, organisation and office rationalization - but does not constitute a separate branch of scientific research. Again, the distinction is often vague in practice ".
On the other hand, A. Zampolli suggested the term *automated language processing* (ALP) to indicate "all the activities, theoretical or applied, encompassing "the use of computers or computational techniques in the processing of natural language". The area of ALP contains both computational linguistics (CL) and literary and linguistic computing (indicated with the abbreviation TP, from text processing, considered as the nucleus of LLC): "CL activities, which are focused on linguistic algorithms, are principally directed towards the study of linguistic models, and in general, towards the formalization, representation, and calculus of linguistic structures. TP activities are mainly concerned with the processing of collections of language data, usually large, very often for purposes of reorganization, extraction, summarization, etc. of some linguistic elements of the text, designated at the 'surface' level, i.e. distinguished by shape or code pattern. (.....). From a theoretical point of view, it must be remembered that many research projects currently in progress in TP are aimed at extracting, from linguistic facts, data and information which constitute the primary material that must be considered in theories and models of CL. At times, information obtained on the statistical and lexical composition of specific corpora is also used in the construction of algorithms and in the choice of working strategies for systems in CL; reference can be made, for example, to the use of statistical methods in several speech understanding systems or in some projects for machine translation. From an operational point of view, typical TP procedures include some crucial operations on the texts or data which are substantially the same as some of those requested from some components of typical systems in CL. Two of the more obvious examples are morphological analysis and the distinguishing of homographs for lemmatization".

(15) For example: the Istituto di Linguistica Computazionale, Pisa (Zampolli, 1983); The Institut fur Deutsche Sprache, Mannheim; Sprakdata, Goteborg; etc.

(16) The Proceedings of this Conference (*Les Industries de la Langue, Enjeux pour l'Europe*), Tours, 28 February - 1 March 1986) are published in number 16 (1986) of the revue *"Encrages"*. In the allocution pronounced on the occasion of the 350th anniversary of the Academie Francaise, the 12 November 1985, published in the same issue, the President F. Mitterrand said: "Nous nous trouvons a un point fort important de l'histoire de notre langue: ou bien elle saura maitriser l'informatique, ou bien, en peu d'annees, elle cessera d'etre l'un des grands moyens de communication dans le monde" (Allocution ... 1986, 145).

(17) Along with the dramatic advancement of the new information technologies, the world economy is undergoing a profound transformation. "It is estimated that the traditional sectors of economic activities - agriculture and manufacturing - constitute at present no more than 40% of the total, while already 60% of the workforce are concerned with 'immaterial' activities, principally information handling. This development goes in parallel with a trend towards a worldwide concept of the economy". (Perschke, 1988).

(18) "La diversite' linguistique se situe au coeur meme de l'identite' culturelle de l'Europe. Une langue n'est pas uniquement un vehicule de communication. Elle reflete une histoire, une civilisation, un systeme de valeurs ... et, comme le disait Gramsci, elle 'contient les elements d'une conception du monde et d'une culture'" (Vidal-Beneyto, 1986, 5). "The EC and its direct competitors, Japan and the USA, are confronted with the challenge of mastering our principal information medium: natural language. For the EC this challenge is more important, as unlike Japan and the USA, its internal market is linguistically not homogeneous: there are nine official languages, and several more regional languages currently used" (Perschke, 1988). It has been suggested that the obstacle of the 'linguistic barriers' created for the European economic activities by this diversity, could ultimately produce a potential advantage, forcing the Europeans to acquire a know-how in the sector of multilingual LI activities, which could be exported to other countries, and facilitate the relationships with them.

(19) Several initiatives have already been promoted. As an example, we can quote, at a national level, the Japanese Electronic Dictionary Research Institute, set up by the Japanese Government in cooperation with 8 major Japanese electronic industries, which aims at producing national Japanese and Japanese-English lexical databases (Japanese Electronic Dictionary Research Institute, 1988), and two national strategic research projects of the Italian National Research Council (Zampolli 1987, 1989).
At the EC level, we can quote the machine translation project EUROTRA (Maegaard, 1988), several ESPRIT projects (AQUILEX, see Boguraev et al., 1988), and the activities, in the framework research programme 1987-1991, which include lexical reusability and lexical and terminological standards. The Council of Europe has set up an ′ad hoc′ programme for language industries, with four activity lines: lexica (Gross), corpora (Zampolli, Cignoni, Rossi, 1987) terminology, common European doctorate in computational linguistics.

(20) Corpora analysis will give information on linguistic phenomena occurring in real texts, and on their frequency in specific sublanguages. Discussing the role of language corpora in linguistic technology, H.S. Thompson (1989) stresses that the access to large amounts of speech or text data is essential for the development of the technologies in question, regardless of whether they are self-organising (i.e. based on neuronal nets or Markov models or other similar stochastic approaches) or not (i.e. based on explicit representational knowledge bases). The self-organising approaches require large bodies of examples of the required input and output to provide the basis for the training process. It is well known that some recent successful NLP systems are almost entirely based on statistical probabilities derived from the analysis of textual samples: parts of speech taggers (Church, 1988), corpus-oriented parsers (Hindle, 1988), speech recognizers (Brown et al., 1988). ″Since explicit knowledge-based systems for the forseeable future will be specialised for specific application domains, the ability to derive linguistic knowledge bases from a corpus of linguistic material which exemplifies and in a sense defines such a domain will be crucial. Furthermore, one can anticipate that a sensible route to the required domain specific knowledge bases will be to develop a set of reasonably broad coverage knowledge bases, which can be specialized for specific domains. Finally, even the most rigorously knowledge-based approach will often require tuning to reflect the distribution of phenomena in the targetted linguistic tasks, which once again means processing large amounts of appropriately annotated linguistic data″ (Thompson, 1989, 2).
Lexicographers, in particular historical lexicographers, have always used corpora as sources of information for the description of the properties of the lexical units, in addition to their linguistic competence. In certain cases, for example collocations and phraseology, it is with great difficulty that linguistic competence can be made explicit without the evidence supplied by corpus analysis (cf. Smadja, 1989).

(21) The Text Encoding Initiative can be considered as an answer to this need, and it seems relevant to stress that it constitutes a paradigmatic example of cooperation between various kinds of partners. The Text Encoding Initiative is a cooperative undertaking of the textual research community to formulate and disseminate guidelines for the encoding and interchange of machine-readable texts intended for literary, linguistic, historical, or other textual research. It is sponsored by the Association for Computers and the Humanities (ACH), the Association for Computational Linguistics (ACL), and the Association for Literary and Linguistic Computing (ALLC). A number of other learned societies and professional associations support the project by their participation in the Initiative′s Advisory Board. The project is funded in part by the U.S. National Endowment for the Humanities, and in part by the EC, through Pisa University.

(22) The situation, unfortunately, has not changed much since 1973, when A. Zampolli (1973, XXI-XXII) wrote:
″The fact that these operations in TP are still performed manually is partly because of the inadequacies of the components of the CL systems in analysing, in a satisfactory manner, the variety and complexity of the texts and data usually processed in TP, but it is also a result of the lack of exchange of information and collaboration among reserachers in the two fields. Those who have worked for some time in TP, however, are well aware of the fact that the development of applications according to the ′classic′ methods and techniques of the 1950s and 1960s has reached saturation point. If we continue to use current methods, according to the current rules of the game (for example: processing, at a simple graphemic level, millions of running words, in order to produce frequency counts, concordances, lexical cards, etc., without any linguistic analysis), real prospects of development do not exist. Although the speed of the computer is continually being increased and programs are becoming more sophisticated, lexicographers and linguists are not able to profit from these facts proportionally because current methodology already produces much more data than any reasonably sized team of linguists could probably analyse, working according to current procedures. If the analysing operations are left to a successive phase, this would not alleviate the problem as it is not clear how we can resolve the enormous operational difficulties which are due to the sheer quantity of the documentation and material gathered″.

References

Actes du Colloque International sur la Mecanisation des Recherches Lexicologiques, Besancon, *Cahiers de Lexicologie*, 3, 1961.

Allocution prononcee par M. Francois Mitterrand, President de la Republique, lors de la seance solomnelle a' l'Academie Francaise a' l'occasion du 350me Anniversaire de l'Institut, *Encrages*, 16(1986), 144-147.

Ahlswede, T., Evens, M., Parsing vs. Text Processing in the Analysis of Dictionary Definitions, *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, Buffalo, New York, 1988, 217-224.

Almanacco Letterario Bompiani 1961, Milano, 1961.

ALPAC Report, Automatic Language Processing, Advisory Committee, Language and Machine-Computers in Translation and Linguistics, Washington, 1966.

Alshawi, H., Analyzing the Dictionary Definitions, in B. Boguraev, E. Briscoe (eds.), 1989, 153-170.

Amsler, R. A., A Taxonomy for English Nouns and Verbs, *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, Stanford, California, 1981, 133-138.

Atkins B.T., The Uses of Large Text Databases, Semantic ID Tags: Corpus Evidence for Dictionary Senses, *Third Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary*, Waterloo, Canada, 1987, 17-36.

Atkins, B.T., Kegl, J., Levin, B., Explicit and Implicit Information in Dictionaries, in *Proceedings of the Conference on Advances in Lexicology*, Waterloo, 1986.

Bindi, R., Calzolari, N., Statistical analysis of a large textual Italian Corpus in search of lexical information, presented for *EURALEX 1990*, Malaga, forthcoming.

Boguraev, B., Briscoe E.J. (eds.), *Computational Lexicography for Natural Language Processing*, Longman, London, 1989.

Boguraev, B., Briscoe, E.J., Calzolari, N., Cater, A., Meijs, W., Zampolli, A., Acquisition of Lexical Knowledge for Natural Language Processing Systems, (AQUILEX), Technical Annex, ESPRIT Basic Research Action No. 3030, Cambridge, 1988.

Boguraev, B., Byrd. R., Klavans, J., Neff, M., From structural analysis of lexical resources to semantics in a Lexical Knowledge Base, in *Proceedings of the First International Lexical Acquisition Workshop*. Detroit (Michigan), 1989.

Booth, A.D., Cleave, J.P., Brandwood, B.A., *Mechanical Resolution of Linguistic Problems*, London, 1958.

Brown, P., Cocke, J., Della Pietra, S., Della Pietra, V., Jelinek, F., Mercer, R., Roossin, P., A Statistical Approach to Language Translation, *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, 1988.

Busa, R., *Sancti Thomae Aquinitatis Hymnorum Ritualium Varia Specimina Concordantiarum*, Milano, 1951.

Busa, R., L'evoluzione linguistica dei mezzi di informazione, in *Almanacco Letterario Bompiani 1961*, Milano, 1961, 103-117.

Busa, R., Actes du Seminaire International sur le dictionnaire latin de machine, *Calcolo*, Supplemento n. 2 al vol. V., 1968.

Byrd, R.J., Discovering Relationships among Word Senses, *Dictionaries in the Electronic Age*, Fifth Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary, Oxford, 1989.

Byrd, R.J., Calzolari, N., Chodorow, M., Klavans, J., Neff, M., Rizk, O., Tools and Methods for Computational Lexicology, *Computational Linguistics*, 1987, vol. 13(3-4), 219-240.

Calzolari, N., Towards the organization of lexical definitions on a data base structure, *COLING82*, ed. by E. Hajicova, Prague, Charles University, 1982, pp.61-64.

Calzolari, N., Detecting Patterns in a Lexical Database, *Proceedings of the 10th International Conference on Computational Linguistics*, Stanford, California, 1984, 170-173.

Calzolari,N., The dictionary and the thesaurus can be combined, in *Relational Models of the Lexicon* , (Studies in Natural Language Processing series), ed. by M.Evens, Cambridge (Mass.), Cambridge University Press, 1988, 75-96.

Calzolari,N., Lexical Databases and Text Corpora: perspectives of integration for a Lexical Knowledge Base, in *Proceedings of the First International Lexical Acquisition Workshop*. Detroit (Michigan), 1989a, n.28.

Calzolari, N., Computer-aided lexicography: dictionaries and word databases, *Computational Linguistics*, edited by I.S. Batori, W. Lenders, W. Putschke, Berlin: Walter de Gruyter, 1989b, 510-519.

Calzolari, N., Structure and Access in an automated Lexicon and Related Issues, in D. Walker, A.Zampolli, N.Calzolari (eds.), forthcoming.

Calzolari, N., Picchi, E., A Project for a Bilingual Lexical Database System, *Advances in Lexicology, Second Annual Conference of the UW Centre for the New Oxford English Dictionary*, Waterloo, Ontario, 1986, 79-92.

Calzolari, N., Picchi, E., Acquisition of Semantic Information from an On-Line Dictionary, *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, 1988, 87-92.

Calzolari,N., E.Picchi, A.Zampolli, The use of computers in lexicography and lexicology, in *The Dictionary and the Language Learner*, ed. by A.Cowie, Lexicographica Series Maior 17, Tubingen, Niemayer, 1987, 55-77.

Chodorow, M.S., Byrd, R.J., Heidorn, G.E., Extracting Semantic Hierarchies from a Large On-line Dictionary, *Proceedings of the Association for Computational Linguistics*, Chicago, Illinois, 1985, 299-304.

Church, K.W., A Stochastic parts program and noun phrase parser for unrestricted text, *ACL, Second Conference on Applied Natural Language Processing*, 1988, 136-143.

Church, K., Hanks, P., Word Association Norms, Mutual Information and Lexicography, *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, British Columbia, 1989, 76-83.

Cumming, S., The Lexicon in Text Generation, in D. Walker, A.Zampolli, N.Calzolari (eds.), forthcoming.

Fox, E., Nutter, T., Ahlswede, T., Evens, M., Markowitz, J., Building a Large Thesaurus for Information Retrieval, *Proceedings of the Second Conference on Applied Natural Language Processing*, Austin, Texas, 1988, 101-108.

Goetschalckx, J., Rolling, L. (eds.), *Lexicography in the Electronic Age*, Amsterdam, North-Holland, 1982.

Gruppo di Pisa, Il Dizionario di Macchina dell'Italiano, in *Linguaggi e Formalizzazioni*, ed. by Gambarara, D., Lo Piparo, F., Ruggiero, G., Roma, Bulzoni, 1979, pp.683-707.

Hays, D.G., *Computational Linguistics: Introduction*, in Meetham and Hudson (eds.), 1969, 49-51.

Hays, D.G., *The Field and Scope of Computational Linguistics: Introduction*, in Papp and Szepe (eds.), 1976, 21-26.

Hindle, D., Acquiring Disambiguation Rules from Text, *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Morristown (NJ), 1988, 118-125.

Ingria, R., Lexical Information for parsing Systems: Points of Convergence and Divergence, in D. Walker, A.Zampolli, N.Calzolari (eds.), forthcoming.

Kay, M., The Dictionary of the Future and the Future of the Dictionary, in Zampolli, Cappelli (eds.), 1983, pp.161-174.

Japanese Electronic Dictionary Research Institute, *Electronic Dictionary Project*, Tokyo, 1988.

Locke, W.N., Booth, A.D., *Machine Translation of Languages*, MIT Press, 1955.

Katz, B., Levin, B., Exploiting Lexical Regularities in Designing Natural Language Systems, *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, 1988, 316-323.

Klavans, J.L., Building a Computational Lexicon using Machine Readable Dictionaries, paper presented at the Third Congress of the European Association for Lexicography, Budapest, 1988.

Kucera, H., Francis, W.N., *Computational Analysis of Present-Day American English*, Brown University Press, Providence, Rhode Island, 1967.

Maegaard, B., EUROTRA, The Machine Translation Project of the European Communities, *Literary and Linguistic Computing*, 3, no. 2, 1988, 61-65.

Meetham, A.R., Hudson, R.A., *Encyclopaedia of Linguistics, Information and Control*, Pergamon Press, 1969.

Nagao, M., *Machine Translation - How far can it go?*, OUP, 1989.

Nagao, M., Nakamura, J., Extraction of Semantic Information from an Ordinary English Dictionary and its Evaluation, *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, 1988, 459-464.

Neff, M., Boguraev, B., Dictionaries, Dictionary Grammars and Dictionary Entry Parsing, *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, British Columbia, 1989, 91-101.

Papp, F. Szepe, G. (eds.), *Papers in Computational Linguistics, Proceedings of the 3rd International Meeting on Computational Linguistics*, 1976.

Perschke, S., Hearing on the language industry in the European Community. Questions put to the participants. (Background Paper for Discussion), 1988.

Picchi,E., N.Calzolari, Textual perspectives through an automatized lexicon, in *Methodes quantitatives et informatiques dans l'etude des textes.* Geneve: Slatkine, 1986, 705-715.

Picchi,E., C.Peters, N.Calzolari, A tool for the second language learner: organizing bilingual dictionary data in an interactive workstation, in *Proceedings of the XX ALLC Conference,* Jerusalem, 1988, forthcoming.

Pustejovsky, J., Current Issues in Computational Lexical Semantics, Invited Lecture, *Proceedings of the Fourth Conference of the European Chapter of the ACL*, Manchester, England, 1989, xvii-xxv.

Quemada, B., Introduction, *Actes du Colloque International sur la Mecanisation des Recherches Lexicologiques*, Besancon, 1961, 13-18.

Smadja, F., Macrocoding the Lexicon with Co-occurrence Knowledge, paper presented at the First Lexical Acquisition Workshop, Detroit, 1989.

Smith, J., Ideals versus Practicalities in Linguistic Data Processing, in A. Zampolli, N. Calzolari (eds.), 1973, 895-8.

Table Ronde sur les grandes dictionnaires historiques, Firenze, 1973.

Talmy, L., Lexicalization Patterns: Semantic Structure in Lexical Forms, in T. Shopen (ed.), *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon*, Cambridge University Press, Cambridge, 1985.

Thompson, H., Linguistic Corpora for the Language Industry (Background paper), 1989.

Van der Steen, G.J., A Treatment of Queries in Large Text Corpora, in S. Johansson (ed.), *Computer Corpora in English Language Research*, Norwegian Computing Centre for the Humanities, Bergen, 1982, 49-65.

Vidal-Beneyto J., Presentation, *Encrages*, 16(1986), 15-7.

Vauquois, B., *La Traduction Automatique a' Grenoble*, Paris, 1975.

Vossen, P., Meijs, W., den Broeder, M., Meaning and Structure in Dictionary Definitions, in B. Boguraev and E. Briscoe (eds.), 1989, 171-192.

Walker, D., Zampolli, A., Foreword, in B. Boguraev, T. Briscoe (eds.), 1989, xiii-xiv.

Walker,D., A.Zampolli, N.Calzolari (eds.), *Towards a polytheoretical lexical database.* Pisa: ILC, 1987.

Walker, D., A.Zampolli, N.Calzolari (eds.), Special Issue of the *Journal of Computational Linguistics*, 13(1987)3-4, 193.

Walker, D., Zampolli, A., Calzolari, N. (eds.), *Automating the Lexicon: Research and Practice in a Multilingual Environment*, OUP, forthcoming.

Webster, M., M. Marcus, Automatic acquisition of the lexical semantics of verbs from sentence frames, in *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, British Columbia, 1989, 177-184.

Whitelock, P., Wood, M., Somers, H., Johnson, R., Bennett, P. (eds.), *Linguistic Theory and Computer Applications*, Academic Press, New York, 1987.

Wilks, Y., Fass, D, Guo, C.-M., McDonald J., Plate, T., Slator, B., A Tractable Machine Dictionary as a Resource for Computational Semantics, in B. Boguraev and E. Briscoe (eds.), 1989, 193-228.

Zampolli, A., Projet pour un lexique electronique de l'italien, in Busa (ed.), 1968, 109-26.

Zampolli, A., Lexicological and Lexicographical Activities at the Istituto di Linguistica Computazionale, in Zampolli, Cappelli (eds.), 1983, pp.237-278.

Zampolli, A., Multifunctional Lexical Databases, *Encrages*, 16(1986), 56-65.

Zampolli, A., Progetto Strategico "Metodi e strumenti per l'industria delle lingue nella cooperazione internazionale", Pisa, 1987.

Zampolli, A., Progetto Speciale "Aquisizione di una base di conoscenze lessicali per il trattamento automatico dell'Italiano: obiettivi nazionali e cooperazione internazionale", Pisa, 1989.

Zampolli, A., Calzolari, N., (eds.), *Computational and Mathematical Linguistics, Proceedings of the International Conference on Computational Linguistics 1973*, 2 Volumes, Firenze, 1973 and 1977.

Zampolli, A., Calzolari, N., Computational Lexicography and Lexicology, *AILA Bulletin*, 1985, 59-78.

Zampolli, A., Cappelli, A., (eds.), The Possibilities and Limits of the Computer in producing and publishing Dictionaries, *Linguistica Computazionale*, Pisa, III, 1983.

Zampolli, A., Cignoni, L., Rossi, S., Problems of Textual Corpora, ILC-9-2, Pisa, 1985.

Lexical DataBases and Textual Corpora:

a trend of convergence between

Computational Linguistics and Literary and Linguistic Computing

Nicoletta Calzolari - Antonio Zampolli

Istituto di Linguistica Computazionale del CNR, Pisa, Italy
Dipartimento di Linguistica, Universita' di Pisa

1. Introduction

In this paper the development of lexical knowledge bases and textual corpora will be considered in the framework of the recent trend towards the creation of large repositories of linguistic information. This trend concerns both researchers who call their discipline "computational linguistics", and researchers who identify their activities as "literary and linguistic computing". The two terms are often used in different ways. They are in fact sometimes considered to identify two different disciplines, other times they are considered to design two different orientations of one same discipline. In both cases, their relationships have not been the object of adequate theoretical reflection. However, it seems uncontroversial that the two terms identify two largely disjoint groups of researchers. We shall consider briefly, first, how these two groups developed, in the past, as separate entities, with a very limited overlapping membership, and why they are now beginning to consider possible cooperations in the development of large linguistic knowledge bases.

2. Some historical and terminological remarks

When the use of electronic data processing techniques (1) on linguistics data began at the end of the '40s, two main lines of research were, quite independently, activated:
- Machine Translation (*Traduction automatique*) (MT).
- Lexical Text Analysis (*Depouillement electronique de textes*) (LTA: production of indices, concordances, frequency counts, etc.).

While MT was promoted mainly in 'hard-science' departments, LTA was developed mainly in humanities departments and, probably also for this reason, the two lines had very few contacts (2).
At the beginning of the 1960s, the perception of a possible reciprocal interest was explicitly manifested, in particular through the invitation of MT researchers to the first LTA conferences, like Tubingen (1960), and Besancon (1961) (3).
The topics more often quoted for possible convergence of interest were, in particular, text encoding systems for different alphabets, frequency-count of linguistic elements in large corpora, automated dictionaries. But, in effect, real cooperation was very rare if not totally absent (4).
The year 1966 has been particularly important for both lines of research, but for opposing reasons. The Prague International Conference *'Les Machines dans la Linguistique'* ratified the international acceptance of the LTA as an autonomous disciplinary field, and its extension to a broader area, which included new dimensions of processing (phonology, historical linguistics, dialectology, etc.), called **Literary and Linguistic Computing (LLC)**, whereas the well-known ALPAC-Report (1966) brought about an abrupt arrest in the majority of MT projects throughout the world, and marked the beginning of the so-called 'dark ages' of MT.

Following, de facto, the recommendations of the ALPAC report (5), basic research on natural language processing slowly occupied the area characterized so far by MT activities, and **Computational Linguistics (CL)** emerged as a new disciplinary activity (6).

However, in spite of ALPAC recommendations for researches in large-scale grammars, dictionaries, corpora (7), CL focused mainly on the development of methods for the utilization of formal linguistic models in the analysis and generation of isolated sentences, in an almost exclusively monolingual framework, at the grammatical level.

The CL activities, which came after MT, almost completely neglected the development of lexica, practically restricted to small toy-lexicons of a few dozen words (8). A distorted (we believe) interpretation of the Chomskyan paradigm led to an almost complete disinterest in corpora and quantitative data, which, on the other hand, were attracting much attention in the LLC area due, among other things, to projects for national historical dictionaries (9) and for frequency dictionaries (10).

On the other hand, also the LLC delayed taking advantage of the know-how, methodology, and tools produced from the very beginning by MT in the field of automatic lexica. Not only had MT developed research on specialized hardware (11), storage, access techniques, inflectional and derivational morphological analysis, but certain projects had already begun the collection of large sets of monolingual and bilingual lexical and terminological data.

Very few exceptions can be reported in the LLC field, all primarily motivated by attempts to automatize the lemmatization of texts for the production of lemmatized indices and concordances. To our knowledge, the first experiments are related to Latin (CAAL, Gallarate and LASLA, Liege) (12).

For several years practically no relationship has existed between LLC and CL. As local organizer of the 1973 Pisa COLING, Zampolli endeavoured to include in the call for papers, and to promote in the Conference, sections explicitly dedicated to topics which could delineate the areas of common interest. The attempt was successful in terms of joint participation, and it was probably not just by chance that J. Smith presented there, at an international level, the newly founded ALLC (Smith, 1973).

But in those years a (so to speak) 'puristic' approach characterized the general reflections of CL, which was searching for a definitional and a disciplinary identity (13), focussing on problems of computation and on the nature of the algorithmic procedures, rather than on the nature of the results and on linguistic, in particular textual, data.

The variety of points of view is exemplified in the *Foreword* by Karlgren, and in the *Introduction* by Zampolli, to the *Proceedings* of COLING 1973 (Zampolli, Calzolari 1973) (14).

The development of CL, in the following years, has been influenced by the interest for Natural Language Processing (NLP) shown by large sectors of Artificial Intelligence. Many efforts have been directed towards the study of methods and tools for prototypes performing a "deep understanding" of natural language, necessarily limited to restricted linguistic fragments and to "miniature" pragmatic subdomains, thus enlarging the gap between CL and LLC activities.

In the LLC framework, the attention of a large part of the research community has been captured by the new technological developments, and efforts have been directed towards mastering new hardware and software facilities: the increasing variety of rich sets of characters, OCR, photocomposition, large database techniques, personal computers, new storage media, general purpose editors and word-processors, standardised concordance packages, etc.

Only in the last two years has a variety of contributing factors started to rouse the reciprocal interest of people working both in CL and LLC. Increasing contacts and exchanges; joint organization of conferences or conference sections; cooperative projects formulated at the international level are external signs of this process.

This convergence is, partly, due to the activities of some Institutes, programmatically oriented to perform researches in both fields, and thus naturally operating to construct a bridge and to promote sinergies (15). However, in our opinion, the key fact is that both fields are recognizing that an important aspect of their development depends on the capability of processing, at least at some level of linguistic analysis, large quantities of "real" texts of various types.

2.1 Computational Linguistics

CL has always considered as a main task the construction of computational components for the automatic generation and analysis of natural language sentences. However, only very recently has CL truly faced the problem of constructing components suitable for the treatment of large, real texts. This trend has been largely originated by the increasing interest of several national and supranational authorities for the potentials of the so-called **"language industries" (LI)**.

This expression, coined on the occasion of a Congress sponsored by the Council of Europe in Tours, February 1986 (16), is used to indicate activities based on computational systems, oriented to practical industrial and commercial applications, which contain, as an essential part, natural language processing components. Examples of typical applications include, within the domain of speech technology: access control, command and control to data entry, driver stations, document creation, telephone enquiries, transaction processing by telephone, data base enquiry, environmental control, voice messaging, announcement systems, augmented communication for handicapped people, etc. For written texts, we can quote: spelling checkers, computer-assisted lexicography and terminology, natural language interfaces, machine translation, information retrieval, computer-assisted language learning and teaching, computer-assisted consultation of reference works, translator workstations, etc.

A set of different factors and conditions are requiring today the promotion and development of LI. The keyword is, in our opinion, the advent of the so-called 'information society'. The global dimension of the economy conceived as a worldwide system (17), together with the technological development of telecommunications systems, entails a growing information flow. The principal information vehicles are still the natural languages, both for the production and the storing tasks. Furthermore, the major part of the information in natural language is nowadays produced directly through computer use, and recorded on machine readable supports: word-processors, office automation, electronic mail, photocomposition, databases, etc. Various countries are considering the possibility of progressively recording entire libraries in MRF.

This situation puts an obvious pressure for the creation of new products and services for the various economic activities primarily involved in information handling. The following passage of Makoto Nagao (1989, p. 4) seems particularly relevant to us: "Computers are a fusion with and unification of communications technology at both the hardware and the software levels, and computer systems will undoubtedly enter every corner of future society. When that day arrives, the most important technology will be specifically concerned with neither hardware nor software, but with what I have been advocating for many years: 'informationware'. In other words, the central problem will regard the ways in which the information signals sent by human beings will be mechanically processed, transmitted, stored, and then recalled in a form which can be interpreted by other human beings. The essence of informationware is therefore how information can be efficiently stored in a computer and activated in response to the various demands of its users. Information can in fact take different forms, including writing, speech and visual images, but objectively, the most accurate means for transmitting and receiving information is writing. For this reason, of the various aspects of informationware, linguistic information and its processing technique will be the primary technology at the heart of the information society. Such technology might be called 'language engineering', and the industry which it will span will be the 'language industry'".

A central aspect of the LI is **multilinguism**. Only an 'elite' minority in the world can operate today in a foreign language, without sacrificing its performance (Perschke, 1988). Furthermore, the conservation of national languages, principle adopted from the beginning, for example, by the EEC, is an important condition for the preservation of the national cultural identities (18).

The need for monolingual and multilingual natural language processing systems, to be used in products for information handling in the LI framework, is uncontroversial. Some studies are carried out in order to narrow down and focus the most urgent tasks and targets, identifying the principal sectors of activities and their economical dimension.

However, the major problem consists in evaluating: - which products can be created on the basis of existing technologies; - which applications can be envisaged at short and medium terms; - which are the priority areas and tasks for linguistic basic and applied research; - which can be an appropriate research and development strategy; - by which measures, at the organisational

level, the public Authorities and professional scientific Associations can stimulate progress in the field (19).

In this framework, one of the priority needs, recognized by several researchers in various countries, is the description, in a form which is suitable for computer use, of the natural languages, performed as far as possible exhaustively, at least for the linguistic aspects which can be treated at the present state-of-the-art of linguistics and of natural language processing. Such extended descriptions are considered the bases for the construction of components capable of dealing with the various types of large real texts which are the typical objects of a wide range of LI applications already possible or foreseeable at short and medium term.

These descriptions concern, first of all, grammars and lexica, and can take the form of repositories of grammatical and lexical knowledge bases. Large corpora of textual material in the form of textual databases are considered essential sources of information (20).

The construction of such large structured collections of linguistic data is very expensive. The availability of such extended linguistic knowledge bases is essential for the feasibility of various industrial applications. Therefore, they are often considered as precompetitive resources. Different categories of partners from the academic, industrial and publishing sectors must co-operate in their creation. To ensure reusability, the creation - as far as possible - of standards, is very important. Cooperation and coordination of efforts is required not only at the national but also the international level, if the monolingual linguistic knowledge bases are to converge in a multilingual network, both for the creation of bilingual systems and for the use of similar components in monolingual applications on different languages.

2.2 Literary and Linguistic Computing

The quantity of texts available in machine readable form is increasing very rapidly. Not only is there a progressive cumulation of texts directly encoded by various categories of humanists for electronic processing, but also the most part of texts nowadays is produced and (re)published through computers. Given the diffusion of individual workstations - with computational power and memory size adequate to the typical humanistic tasks - the distribution of the texts directly in MRF for the interactive use of individual researchers has become possible and more and more attractive.

As a consequence, the adoption of standards, for text representation, which ensure the exchangeability and reusability of texts for various users, has become very urgent (21).

LLC has always been interested in the process of large real texts, but the computational treatment has been performed on units identified, mainly if not exclusively, at the graphical level. Frequency counts, concordance production, interactive textual access usually operate essentially on the graphical forms, roughly defined as sequences of characters between two spaces or separators.

However, several operations on the texts, which enter in the performance of various scholarly humanistic activities, are based on the identification, in the text, of linguistic units at various levels, both as direct objects of linguistic, philological, literary research, and as referential units representing factual information. An exemplification list contains, among other units, phonemes, metrical schemata, syntagmatic patterns, rhymes, lemmata, lexemes, phrases, morphosyntactic categories, terminological units, conceptual units and their relations, etc.

The intrinsic complexity of the analysis, and the time required to perform it, are very high. The large diffusion of personal workstations enables more and more individual researchers to directly perform a variety of analyses on the increasing number of available texts. Therefore, LLC is obliged to consider the possibility of constructing or importing tools for automating, at least in part, the operations of analysis, or at least for assisting the humanists in its performance.

Roughly speaking, considering the present state-of-the-art in natural language processing and in knowledge acquisition and representation methods, we can distinguish two major categories of computational tools for computer-assisted humanistic text analysis.

- Robust parsers, supported by large computational lexica, conceived for identifying, in real texts, linguistic units, at certain levels of analysis: syllabic, metrical, syntagmatic patterns; lemmata; parts of speech; phrases; verbal arguments; superficial sentential structures; etc. The components constructed in the framework of CL, if they were adequate to process real texts, would supply the identification and the representation of such units and their relations (22).

- "Intelligent" access tools which, through the consultation of various kinds of knowledge sources, assist the researcher in the interaction with the texts. For example, appropriately structured reference sources, such as encyclopedias and dictionaries, can make explicit, and eventually complement, the linguistic and conceptual researcher's knowledges, in such a way that they can be used by the programs for text browsing. We shall briefly illustrate later examples of dictionaries which can also be used, for instance, to expand a user's query, searching in the texts the occurrences of "families" of words connected by particular semantic or conceptual relations: taxonomy, synonymy, etc.

2.3 Convergence between CL and LLC

Summing up, both CL and LLC are led by various factors, and in particular by the framework created by the expansion of the 'information society', to consider the creation of tools and resource systems for the processing of large real texts, as a major task in their present state of development.

From this, we are not arguing that CL and LLC aim at the construction of computational systems of the same nature, nor that they have to solve exactly the same range of linguistic problems. We notice only that both fields are now recognizing that the development of these systems require the availability of extended repositories of linguistic knowledges.

Our thesis is that the basic knowledges required are in large part the same. It is therefore important that the information encoded can be reused in both fields through appropriate interfaces. Cooperation must be promoted, in order to combine the efforts and the specific know-how of the two categories of researchers, who are for several aspects complementary. For example, CL has developed grammatical formalisms and parser models; LLC has developed knowledges and methods for corpora collection and treatment, statistical linguistic analysis, sublanguage description and identification.

In the following we shall describe our work in Pisa in the field of lexical knowledge bases, and of their interaction with textual corpora. This research work is explicitly intended to the creation of resources both for CL and LLC, in the present framework of their trend to convergence.

3. Trends in Computational Lexicography and Lexicology

We have already noticed the tendency inside CL in the last years to a shift in interest from almost only the grammatical aspects of the language, to the lexicon also, and, only quite recently, also to large corpora of texts. We are in the presence of a somewhat parallel evolution from the implementation of so-called 'toy-systems' (the prototypical is Winograd's block-world), to the development of 'expert systems' (more powerful, but acting within a limited domain, and therefore with a restricted vocabulary), and recently to 'very large NLP systems', such as Machine or Machine-aided Translation Systems, or products for Office Automation, where a strong need is felt for a real-size vocabulary and a general world knowledge.

Taking for granted these two main trends, both from the theoretical and the applicative viewpoint, it follows that dealing with the lexicon has become trendy, and dealing with textual corpora is becoming even more trendy.

There is a need not only for very large computerized lexicons or Lexical Databases (LDB), but also for lexicons where even the semantic information is made explicit, i.e. for large Lexical Knowledge Bases (LKB). The evolution within Computational Lexicography and Lexicology over the past few years can thus be outlined as follows:

i) from Machine Readable Dictionaries (**MRD**) in the '70s (simple sequential objects well exemplified by photocomposition tapes),

ii) to **LDBs** in the early '80s (more structured objects, provided with multipath access to the data, interactive in nature, and often with explicit taxonomies or IS-A hierarchies),

iii) to **LKBs** in the late '80s (where not only IS-A links but also many other types of lexical/semantic relations among conceptual categories are formalized, where therefore new access paths to the data are constructed, where inferential and deductive mechanisms are built in, and which are usually in the form of a conceptual network).

The main prioritary goal is thus today the creation of a vast 'reservoir' of linguistic knowledge, in the form of as complete as possible and reusable linguistic descriptions, structured in a large LKB or in various kinds of interconnected linguistic bases (grammatical, lexical, textual, knowledge bases).

Given what already stated about the current trends in the different areas, i.e. the request in the CL community of large scale NLP systems, and the fundamental importance that a CL system is able to deal with tens of thousands of lexical items for real world applications, in addition to the fact that lexicography, as a 'language industry' profession, has a very long tradition, and that the creation of a LBD of adequate content and dimension is very time-consuming and expensive, and duplication of efforts may be a very 'sad' fact, one of the key-words in the field of LDBs has recently become the word "**reusability**". This word is to be intended in two main senses: one towards the past, i.e. with respect to existing information, and one towards the future, i.e. with respect to future applications.

In the first case, the meaning is that of reusing lexical information implicitly or explicitly present in preexisting lexical resources (e.g. MRDs, terminological DBs, corpora of texts, etc.) as an aid to construct a LKB. In the second case, it is meant to construct a LKB so as to allow various users (procedural: e.g. different NLP systems; and possibly human: e.g. lexicographers or translators or normal dictionary users) to extract - with appropriate interfaces - relevant information to their different purposes.

With regard to the first meaning, these ideas in a sense originated the proposal for the ESPRIT Project "Acquisition of Lexical Knowledge for Natural Language Processing Systems" (AQUILEX) where groups of researchers in Cambridge, Amsterdam, Dublin, Paris, Barcelona, and Pisa (coordinator) are involved. The main goal is to develop techniques and methodologies for the use of existing MRDs in the construction of lexical components for NLP systems. The extraction of lexical information is carried out moreover from multiple MRD sources and in a multilingual context, with the overall purpose of the creation of a single multilingual LKB. "The knowledge base will be rooted in a common conceptual/semantic structure which is linked to, and defines, the individual word senses of the languages covered and which is rich enough to be able to support a 'deep' knowledge-intensive model of language processing. The knowledge base will contain substantial general vocabulary with associated phonological, morphological, syntactic and semantic/pragmatic information capable of deployment in the lexical components of a wide variety of practical NLP systems" (Boguraev et al. 1988).

If we look at the second meaning of the term reusability, it is strongly linked to two other properties which we consider essential in a LDB.

The first property of a LDB is that of being "**multifunctional**", and has essentially to do with the applicative viewpoint. The LDB must be a central repository of data which can be reused for several purposes and in many applications, through different interfaces, both for procedural and for human use.

The lexicon is obviously an essential component in any NLP system (for parsing, generating, machine translation question-answering, information retrieval, lemmatization, artificial intelligence, etc.). The usual practice is to construct an ad-hoc lexical component for each natural language NLP project. It is necessary to move towards large (both in extension and in depth of representation) lexicons, where information is represented in such a way that it can be easily interfaced by different application procedures according to the different applicative needs. This means that the same set of data can be shared by the various applications. Each interface will only project on the specific application that view on the data which is relevant for the particular requirements.

From this viewpoint, another essential property of a LDB is to be easily extendable, i.e. it must be possible for different researchers to add their own idiosyncratic information consistently with the actual content of the LDB.

The second property of a LDB has to do with the theoretical viewpoint, and consists in its being "**polytheoretical**", i.e."multifunctional" with respect to different linguistic theories. A large amount of work in CL has been carried out until now, as said above, on experimental lines, with consequently small-sized lexical prototype systems. Furthermore, emphasis was traditionally placed on the representation, organization and use of linguistic knowledge as encapsulated and expressed by linguistic rules and procedures. Lexical data seemed to be considered of secondary importance or, at least, easy to be handled.

It is a well recognized fact that different linguistic theories and different computational organizations may have important consequences on the grammar contruction. Less attention has been paid to the consequence on the lexicon. However, we have the intuition that lexicons designed for different linguistic theories may contain information which from a certain point of view is identical, as it describes the same linguistic facts. We have to assess the validity of this intuition before starting to implement in an LDB the information required by the NLP systems.

This characteristics of being polytheoretical is not without problems and difficulties, and a feasibility study is now underway to assess: i) the possibility of achieving a certain degree of consensus among different theories aimed at sharing the same bulk of lexical information, and if so ii) up to which level of linguistic analysis a "neutral" or "polytheoretical" representation of liguistic properties can be designed.

We have promoted a working group which involves outstanding representatives of the major current "linguistic schools". The group will investigate in detail the possibility of representing the linguistic information frequently used in parsers and generators (e.g. the major syntactic categories, subcategorization and complementation, verb classes, nominal taxonomies, etc.), in such a way that they can be reutilized in the following theoretical frameworks: government and binding, generalized phrase structure grammar, lexical functional grammar, relational grammar, systemic grammar, categorial grammar. This group will work on various languages. We shall start by examining in detail the treatment which the foregoing theories will assign to a representative sample of English and Italian verbs. If a polytheoretical lexicon appears to be feasible it should be possible for the lexical data to be reused within the framework of different linguistic theories (e.g. GB, LFG, GPSG, RG, etc.) and also of lexicographic practice, by appropriate interfaces translating the data in the relevant notation/representation (see Walker, Zampolli, Calzolari 1987).

4. Reusability of preexisting data in the form of MRDs

A large number of articles and books have already been written on this topic (see e.g. Amsler, Boguraev, Briscoe, Byrd, Calzolari, Nagao, Picchi, Walker, Zampolli, etc.). We wish to stress in particular what we consider as the natural evolution of all the work done so far in the field, i.e. the possibility of a procedural exploitation of the "full range" of semantic information implicitly contained in MRDs.

In this framework the dictionary is considered as a primary source of basic general knowledge, and many projects nowadays have as their main objectives word-sense acquisition from MRDs, and knowledge organization in a LKB. The method is inductive and the strategy adopted is heuristic: through progressive generalization from the common elements found in natural language definitions we tend to formalize the basic general knowledge implicitly contained in dictionary definitions, mainly in the attempt to extract the most basic concepts and the semantic relations between them. This means that we are going well beyond the extraction and organization of taxonomies, whose methodology of acquisition is now well established (Chodorow et al. 1985, Calzolari 1982, 1984). We simply have to process the first part of the definition, in order to identify the 'genus' term. This can be done by taking into account the fact that the definitions are NPs when the definiendum is a Noun, are VPs for Verbs, and AdjPs for Adjectives. The procedure has thus to look for the head/s of the NP, VP, AdjP, which are respectively a N, V, or Adj. These are the 'genus' terms and are connected by an IS-A link to the definiendum.

When we reorganize a MRD in a taxonomical structure, with only IS-A hierarchies made explicit, we use the MRD as a source of knowledge, but in only one of the possible ways of acquiring from it (in an inductive form) a concept, by linking this concept to all its instances, i.e. all the instances of the same category/class are extracted and connected together pointing to their immediate hypernym.

In the LKB approach the dictionary is seen as a much more powerful "classificatory device", i.e. as an empirical means of instantiating concepts and many types of lexical/semantic relationships among them (see Calzolari, Picchi, 1988).

The methodological approach that we follow can be summarized in these points:
a) to start from free-text definitions, in natural language and in linear form, usually formed by a 'genus term' and a 'differentia' part;

b) to analyze their structure and content from a linguistic and a computational point of view;
c) to convert and reorganize them into informationally equivalent structured formats made up by nodes and relations linking them.

Point b) in its turn can be subdivided, for the computational part, into the following steps:
1) to "parse" the dictionary entry, in the sense of "parsing a dictionary tape" which essentially means recognizing the various relevant fields in the lexical entry;
2) to produce a tree-structured lexical entry;
3) to perform a morphological analysis and a homograph disambiguation, i.e. to tag the definitions for POS;
4) after the above preliminary steps, we have adopted the technique of producing a very simple syntactic parse which roughly recognizes NPs and PPs;
5) the most powerful tool is then a "pattern-matching" mechanism, which is fed by: i) the results obtained by browsing dictionary data in the LDB (as outlined in the few examples presented below) in view of discovering the most interesting words and word-associations, ii) frequency counts on definitions words and syntagms, and obviously iii) the linguist's intuition.

Let us illustrate with some examples the process of analysing the definitions. In the figures we try to simulate the process of browsing the Italian LDB and of navigating the dictionary while searching for particular words, structures, patterns, etc. We can see some of the semantic data it is possible to search and find in a MRD if appropriately structured. Fig. 1 shows part of the taxonomy for the Italian word *libro* (book), i.e. a set of words defined as being "types of" books (we see them together with their definitions).

But there is something more that is said about books in a dictionary. It is also possible to extract the set of the Italian Verbs related to books (see Fig. 2), and the set of Adjectives and of other Nouns having to do with books (Fig. 3 and 4). In section 4.2 we shall come back to "books", stressing the type of information which, lacking in dictionaries, can instead be found in texts.

Our present work is devoted to the formalization also of the other kind of relations - not as simple as the taxonomical ones - which do hold between words, or between words and concepts, and for whose extraction we must analyze and process the whole definition and not only its 'genus' part.

Let us give some examples of the types of relations that it is possible to extract from MRDs. In Fig. 5 we find the first of the about 300 words linked in our LDB by a taxonomical link to the word *strumento* (instrument). The word *attrezzo* (tool) appears in this list. Fig. 6 shows the first hyponyms of this second word together with their definitions. From these definitions it is rather simple to extract semantic relations which we could label **USED FOR, USED IN, SHAPE, MADE OF**, etc. They are extracted by means of a pattern-matching procedure acting on the 'differentia' part of the definitions, where the different ways in which each relation is actually lexicalized in the definitions is associated with the relation-label. The relation **USED FOR**, for example, comes from lexical patterns like: *per, usato per, atto a, che serve a, utile a,* (for, used for, apt to, which serves to, useful to); these lexical patterns acquire this particolar relational meaning when found in particular positions in the definition of hyponyms of the word *strumento*. They can also acquire different meanings in other contexts. The result of this analysis of the definitional content will be restructured in a part of a conceptual network which is sketched in Fig. 7.

Other types of semantic relations rather easily and straightforwardly extractable from the definitions can be illustrated with some examples.

One is the relation **SET OF**, which can be further specified as to the type of its members. We have examples of words denoting **SET OF** *persone* (people) (Fig. 8), *oggetti* (objects) (Fig. 9), etc.

Other types of useful data concern information on selection restrictions for Verbs or for Adjectives and mainly derives from the lexical pattern *detto di* (said of), after which the type of Nouns is found of which an Adjective or a Verb can be typically predicated. See Fig. 10 for Adjectives and Verbs used for nouns denoting *persone* (people), Fig. 11 for Adjectives which collocate with names of colours, either generic colour names, or specific ones such as *giallo* (yellow), *rosso* (red), etc.

An interesting type of relational data which can be extracted for certain types of actions is the information on the words in the lexicon which are lexicalizations of the typical thematic roles of the action itself. Let us clarify what we mean by two examples. In Fig. 12 we find the result of querying the Italian LDB for all the entries in whose definitions the word-form *vende* (sells) appears (not in genus position). The result of the query is the following: we retrieve 242 entries of which well 221 are names of people who "typically sell" something, i.e. of typical **AGENTS** with respect to the action of selling. These entries represent lexicalized case/role fillers in the case-frame of *vendere* (to sell). This is obviously due to the defining pattern used, i.e. *chi vende* (who sells). Some interesting observations can be made with regard to this example.

The first concerns the fact that the same type of result was obtained by making a similar search on an English dictionary. After being shown the Italian example, the IBM Yorktown group repeated the experiment with the same kind of result (see Byrd 1989) for the English data. This shows that there is in fact a correspondence between the definitional patterns used in lexicographical practice independently from the language. This similarity in lexicographical conventions appears in many other examples and will be exploited for the creation of the multilingual LKB which is the ultimate goal of the already mentioned ESPRIT project.

Another observation regards the co-occurrence in these definitions of this kind of verb ("to sell") with another one ("to make", lexicalized in Italian as *fabbricare, fare, preparare*, etc.). Many of these Agent names also apply to the action of "making", and therefore belong to two portions of the resulting conceptual network.

We can also notice that the Noun Phrase following the verb denotes the type of object which is typically sold (or also made) by these Agents.

It is obviously possible to obtain the same type of information on Agents' names for the action of selling if we search for all the nouns whose 'genus term' is the word *venditore* (seller): from this query we retrieve other 131 Agent nouns (see some of them in Fig. 13). Here again some of the nouns are related also with the action of "making", while the PP introduced by the preposition *di* (of) expresses the object which is sold.

This example shows the way in which exactly the same information can be retrieved by browsing the dictionary in different ways, by exploiting the knowledge of its structure (in particular the internal structure of the definitions). In the final LKB all this data will be merged in a single piece of network, independently of the different ways of lexicalizing some concepts and relations.

With a slightly different type of query we can very easily retrieve also the names of the **LOCATIONS** where the action of "selling" is typically performed. Fig. 14 shows the result of the search for the entries in whose definitions the word *vendono* (they sell) is present. Again the fact that names of places are found in this way is due to the following 'defining formula' used by the lexicographers: *dove/in cui si vendono* (where ... are sold). All of the 33 entries retrieved share this definitional pattern: this query is completely without 'noise'.

We can observe that the genus terms are either the generic name *luogo* (place), or those of its hyponyms which are the generic names for the places where something is sold, i.e. *negozio, bottega, bancarella* (shop, store, stall). These are in turn hypernyms of the defined entries. This kind of hierarchical information is already formally coded in the taxonomies stored in the LDB.

What interests us here is the possibility of formalizing and implementing in the LKB the other types of semantic relations, such as **LOCATION** and **THEME** with respect to the actions of "selling" and "making". The Theme relation, i.e. the objects which are typically sold in the defined places are again expressed by the NP object of the verb.

Also in this case similar data are retrieved also by querying for the hyponyms of *negozio, bottega, etc.*. Our aim is to formalize all this information in a semantic network, like the piece sketched in Fig.15.

The above examples show that the LDB facilities can be usefully exploited to analyze and extract linguistic data which must then be restructured and represented in the LKB. In the LKB these types of concepts and of relations, and the interdependencies between word-senses will be explicitly spelled out. When we move beyond taxonomies in the LKB, we establish many different types of associations which are usefully represented in a conceptual network, and when we move from a "monolingual" to a "multilingual" environment, we also establish associations among different languages. These associations are obtained (for those parts of the languages which can be reduced to a common set of concepts and relations) through the common

conceptual network constructed by working on different languages but within the same "research template", i.e. trying to accomodate in the semantic network:
- the "same" world-knowledge,
- for the "same" purposes (NLP, Text Processing, etc.),
- with the "same" methodology,
- from the "same" type of sources (MRDs),
- into the "same" kind of representation.

The common semantic network will thus become the point of convergence of the results of the knowledge aquisition strategies applied on a number of different but homogeneous sources, and the multilingual environment will constitute a valid testbed to evaluate this strategy of design and implementation of a part of a LKB.

4.1 Reusability of bilingual dictionaries

Not only MR monolingual dictionaries, but also bilingual MRDs can be usefully exploited as sources of lexical information for the creation of LDBs and LKBs. These dictionaries can be processed with a twofold purpose, as on the one hand they too are a source of interesting 'monolingual' information, on the other hand they are obviously exploited as a source of links between two monolingual LDBs (see Calzolari, Picchi 1986, and Picchi, Peters, Calzolari, forthcoming).

One of the objectives is to integrate the different types of information traditionally contained in monolingual and bilingual dictionaries, so as to expand the informational content of the single components in the new integrated system. Bilingual dictionaries contain more information about examples of usage, fixed expressions or idioms. This kind of information can obviously be well integrated in the monolingual dictionary, and also made easy to access.

We can envisage the original monolingual lexical entries, augmented with the different types of information coming from the corresponding bilingual entry: different sense discriminations, other examples, syntactic information, collocations, idioms, etc. We can also reverse the perspective, and look at the bilingual entries provided with the information traditionally contained in monolingual entries: mostly definitions. One of the two different viewpoints, both virtually present in the integrated bilingual system, will be simply activated and made available to the user by the first manner of access to the on-line bilingual lexical data base. We would like therefore to maintain in a unique structure both the independent features of the source monolingual and bilingual dictionaries and the integration of the two with different views on the data.

The overall picture of the bilingual LDB system we have in mind is sketched in Fig. 16. Also with regard to bilingual dictionaries, the method we are adopting consists of reusing available data in machine-readable form by analyzing and transforming the information already contained in common dictionaries. The procedure of processing the bilingual MRD is rather similar to the one outlined above for monolingual dictionaries (i.e. parsing of the lexical entry, design of a new structure, computational reorganization, etc.). After this preliminary part again comes out the utility of browsing the bilingual LDB, taking advantage of the structural elements already formalized in the LDB, with the purpose of discovering properties and structures not immediately visible in the printed dictionary, but useful for further exploitation in the computational dictionary.

After the first processing phases that we have envisaged on the bilingual dictionary data, it will make no difference which of the two languages are taken as a starting point. In a certain sense, we would no longer have a source language and a target language, since the look-up and access procedures are independent and neutral with respect to direction (the object becomes bidirectional). Bidirectional cross-references will also be automatically generated for the information contained at each sense level as semantic indicators, i.e. synonyms/hyperonyms or contextual indicators.

One of the parts of the bilingual dictionary we are processing that can be partially made explicit in all its different meanings, is the field of the so-called *semantic indicators*. These provide the constraints for selecting one translation equivalent or the other. The problem is that these constraints are of a different nature, being either i) synonyms or hyponyms of the entry, or ii) contextual indicators such as typical subjects or objects of verbs, typical nouns of which an adjective can be predicated, etc. It is possible to semi-automatize the process of

disambiguation between the different values, after analyzing all the different possibilities and designing a typology of what can appear in this field.

Another possibility is the use of the monolingual lexical data base as a tool to expand the information given as a single word to the whole set of words to which it actually refers. For example, the entry *vivido* has different translations according to the contextual indicators referring to the subject (in brackets):

vivido ..... (***colori***) bright, vivid

In some cases the generic semantic restrictions on the possible object can be taken as a semantic feature, and can be procedurally expanded by the monolingual thesaurus to all the possible hyponyms (at the query moment) so that the appropriate translation can be chosen in any context where a specific name of *colore* (colour) is found (and this is already possible in our monolingual LDB). The information that can be formalized at the semantic level in a monolingual dictionary - which serves to discriminate among the different word-senses - should be in principle of the same type that is given in bilingual dictionaries in the form of "semantic indicators" or "selective conditions" to constrain the choice of a particular translation.

In the same way we can work on other fields in order to make explicit hidden information or to introduce new information on the basis either of structural or of content clues.

After the re-organization of the bilingual MRD in a well-structured LDB, we face the difficult task of using its data to build links between two monolingual LDBs. The difficulty obviously derives from the ambiguity of the words used both as entries and as translations. We never know which word-sense is meant in a particular situation. We shall try to solve this problem as much as possible in the above mentioned ESPRIT project, mostly by exploiting the semantic indicators in the bilingual and the taxonomies and other conceptual information in the monolingual LDBs.

Mapping between word-senses in monolingual dictionaries and different translations in a bilingual dictionary is one of the most interesting of the problems concerning the connection of these different types of dictionaries. As one of the main problems in translation is the correct choice among the various meanings of lexically ambiguous words, we feel that it is absolutely necessary also for a Machine Translation or a Machine Assisted Translation system to be linked to a linguistic data base, i.e. a source of lexical information organized in the form of a thesaurus by multi-dimensional taxonomies, where the possibility of disambiguating lexical items is at least semi-automatized.

One of the main uses of the system should be that of machine-aided translation (MAT), as a powerful aid for translators. The end result may in fact be viewed as a 'translator workstation', where access is provided to many types of dictionaries and other lexical resources, and where the power and the functions of lexical data bases and of textual data bases is exploited at best.

Other purposes of a Bilingual System like the one which appears in Fig. 16 are the following:
- a tool for lexicographers;
- a tool for lexicological-contrastive studies;
- a means for improving monolingual LDBs;
- an aid to construct Machine Translation dictionaries;
- a tool for language teaching;
- a computerized dictionary for "normal" users.

In our opinion, one of the main advantages of a bilingual LDB is the completely different type of "navigation" within its data, made possible both by the multiple access to its data and by its links to the monolingual LDB. In particular, it is not only possible to create links between couples of words in L1 and L2, as in the printed dictionary, but mainly between groups or families of semantically connected words, which we think is an essential property for a true bilingual dictionary and for all the purposes we have listed above.

4.2 Reusability of textual corpora and their integration into LKBs

We have seen that MRDs are very valuable sources of lexical and also of semantic information, but unfortunately not all what is needed to know about the lexicon is there. There are very important pieces of information which in MRDs are completely missing, or incomplete,

or simply are not very good or reliable or easily recoverable. For this type of information, we have to resort to different types of sources (see also Calzolari, 1989a).

Certain kinds of data can probably be acquired only after theoretical investigation of lexical facts, and their source can be seen in the typical linguists' work, mainly based on introspection and native speaker's intuition. In this paper we do not deal with this data, but we must be aware of its existence.

We want to stress here that there are many types of data which can be usefully extracted, more or less directly, by processing very large corpora of textual data. The results of this processing have also to be analysed and evaluated by the linguist and/or the lexicographer, but it is important to realize that for certain types of linguistic phenomena the study made through corpus analysis is 'favoured' with respect to introspection: typical examples are collocations and fixed phrases. A tentative, but not exhaustive, list of lexical information for which we can find data in textual corpora, with various degrees of difficulty and at various levels of completeness, is the following:

- frequency data (at the level of word, word-form, word-sense, word associations, etc.);
- subcategorization;
- collocations, fixed phrases, idioms;
- thematic roles, valency;
- semantic constraints on arguments;
- typical Subject, Object, Modifier, etc. (these are different from the types of thematic roles, being in fact their fillers; in a certain sense they are the same information but given "by example");
- aspectual information;
- proper nouns.

Let us take for example the verb *dividere* (to divide), and look at its occurrences and contexts in our Corpus of about 10 million words. From a total of 840 concordances, we obtain the most frequent syntactic patterns which are as follows:

dividere	NP **in** NP	**268**
"	NP	**175**
"	(NP) **tra** NP , NP , ...	**80**
"	NP **con** NP	**78**
		601

while the remaining 239 contexts are distributed in about 10 other subcategorization frames. If we analyze the contexts by hand, we see that each subcategorization frame can very often be correlated with one or more word-senses, so that we can think of using these frames as a very useful aid in a meaning disambiguation task. By analyzing concordances we can thus obtain data concerning:

a) syntactic frames;
b) their frequency ordering, and therefore their respective relevance for the user;
c) co-occurrences with other words and word classes (at the syntactic and semantic levels);
d) main word-senses;
e) correlation between word-senses and syntactic frames.

We must notice here that it is essential to pay attention to different types of texts, and therefore it is important a good balancing in a reference corpus, because frequency data (at any level: lexical, syntactical, semantic, collocational, etc.) can be very different for different text types.

Let us now consider again the word *libro* (book) for another example of information obtained from texts. If we look at the verbs related to books in the Italian dictionary we can notice that neither *leggere* (to read) nor *scrivere, pubblicare, etc.* (to write, publish) are among them. Again, the same observation has been made with regard to English dictionaries (see Boguraev et al., 1989), which is not by chance, but is again a clear indication of the similarity even between dictionaries of different languages.

In the definitions of these verbs we usually find more generic words related with printed things, such as *scrittura, parole, segni, lettere, scritto, opera, volume, giornale* (writing, words, signs, letters, script, work, volume, journal). The word "book" appears instead in some examples. The link could only be established indirectly, given that the word *libro* is defined in terms of words such as *volume, opera, scritti, stampati, ...*, the same words that appear in the definitions of the above verbs.

These verbs are instead directly associated with *libro* in the corpus of texts. Here, in fact, out of 3,222 concordances of the lemma *libro*, we find these figures for the above-mentioned verbs in the same contexts with *libro*:

leggere	187
scrivere	196
pubblicare	107

It is the analysis of large textual corpora that makes it possible to find this type of collocational information. We are also implementing some statistical/quantitative tools to allow semi-automatic extraction of this and other types of data from our corpus (see Bindi, Calzolari, forthcoming).

When analyzing a large corpus with millions of words in context, we are in a sense compelled to discover and describe:
- usages which are not described in commercial dictionaries;
- relative frequencies of the different word-senses, and of the different syntactic frames/patterns;
- and, above all, the grammatical/syntactic clues by which semantic disambiguation can be at least partially achieved, given the fact that i) in the presence of different syntactic constituency word-sense usually changes, ii) while, vice-versa, we do not necessarily have only one word-sense with the same syntactic frame.

When collecting this type of data for a number of words, we often realize that the data should be reorganized in a different way from how they are presently found in standard dictionaries, if they are to conform to the actual usage of the language.

In order to automatize the retrieval of this type of information directly from the corpus we should first be able to tag the corpus for the different POSs. For this task many systems already exist (see e.g. Hindle 1989, Webster, Marcus 1989). It should then be possible, even without a complete parser, to apply to the text corpus some pattern-matching procedures (as those we are presently using with dictionary definitions). These pattern-matching procedures should be explicitly geared to the extraction of the type of data we are searching (i.e. prepositional phrases, that-clauses, infinitives, etc.).

The same strategy of looking for syntactic (and collocational) clues for semantic disambiguation (to be used for different translations of the same word) is now evaluated in a pilot project we are carrying out in a multilingual context.

5. The lexicographer's workstation as a model of integration of tools and data from different environments and expertises

The importance of a collaboration between researchers working in the fields of CL/NLP and LLC/TP (as already said in more general terms in the first sections of this paper) is evident when we consider that it is necessary to process large textual corpora in order to achieve better LKBs. The design of these large integrated LKBs can really become the purpose of cooperative projects, where the "typical" data, tools, procedures, knowledge, expertise, results, etc., of the two areas of CL/NLP and LLC/TP "must" work in parallel and cooperate and interact with each other.

In order to achieve at least some of the results outlined so far, we can summarize the needs as follows:
- design and implementation of powerful tools;
- large sets of lexical and textual data;
- very modular systems;
- possibility of sharing resources, data and procedures;

- large cooperation among traditionally different research or industrial communities.

A model of the type of integration we have in mind can be seen in the lexicographer's workstation (LW) we are designing in Pisa (see Calzolari, Picchi, Zampolli 1987). It is conceived as a very modular system, where different types of data and of procedures are integrated. At the level of data the LW contains, or will contain: a textual data base, one or more monolingual lexical databases, a thesaurus with taxonomic information, bilingual lexical databases, a reference corpus, etc., while at the level of procedures, it contains: a morphological tool, dictionary parsers, a hyponym finder, an information retrieval system, a lemmatization package, a pattern-matching procedure for dictionary definitions, a redaction tool, etc.

This complex and various set of components reflects our view of the need for an integration and interaction between data and tools traditionally pertinent and pertaining either to CL or to LLC only. It appears therefore important the realization of a factive cooperation among many different groups of researchers (meaning here 'groups' as 'types'), with the aim of linking together worlds which up until now have not been so strongly related to each other, especially perhaps in the American tradition.

PASSIONARIO	1SM	ANTICO LIBRO LITURGICO CATTOLICO	3		
OMILIARIO	1SM	ANTICO LIBRO LITURGICO CONTENENTE OMELIE	1		
EPISTOLARIO	1SM	LIBRO CHE CONTENEVA BRANI DI EPISTOLE E VANGELO	3		
ORA	1SF	LIBRO CHE CONTENEVA LE OPERAZIONI PROPRIE DELLE VARIE ORE	9		
SALTERIO	2SM	LIBRO CHE CONTIENE I SALMI	3		
RITUALE	2SM	LIBRO CHE CONTIENE LE NORME CHE REGOLANO UN RITO	3		
UFFICIOLO	1SM	LIBRO CHE CONTIENE LE PREGHIERE IN ONORE DELLA VERGINE	3		
UFIZIOLO	1SM	LIBRO CHE CONTIENE LE PREGHIERE IN ONORE DELLA VERGINE	3		
CANTORINO	1SM	LIBRO CHE CONTIENE LE REGOLE DEL CANTO FERMO	3		
PORTULANO	1SM	1LIBRO CHE DESCRIVE MINUTAMENTE LA COSTA	34Z		
GUIDA	1SF	LIBRO CHE INSEGNA PRIMI ELEMENTI DI ARTE O TECNICA	3		
GRADUALE	2SM	LIBRO CHE RACCOGLIE I GRADUALI DELL'ANNO LITURGICO	3		
GIORNALMASTRO	1SM	LIBRO CHE RIUNISCE IL GIORNALE E IL MASTRO,PER CONTABILITA'	3		
ANNUARIO	1SM	LIBRO CHE SI PUBBLICA ANNUALMENTE	3		
....					
EFEMERIDE	1SF	LIBRO IN CUI ERANO ANNOTATI I FATTI CHE ACCADEVANO OGNI GIOR	3		
EFFEMERIDE	1SF	LIBRO IN CUI ERANO ANNOTATI I FATTI CHE ACCADEVANO OGNI GIOR	3		
COPIAFATTURE	1SM	LIBRO IN CUI SI COPIANO LE FATTURE	3		
SALDACONTI	1SM	LIBRO IN CUI SONO REGISTRATI I CREDITI E I DEBITI	3		
TASCABILE	2SM	LIBRO IN EDIZIONE ECONOMICA E PICCOLO FORMATO	3		
PERGAMENO	1SM	1LIBRO IN PERGAMENA	3	1	E
BENEDIZIONALE	1SM	LIBRO LITURGICO	3		
MESSALE	1SM	LIBRO LITURGICO CATTOLICO	3		
LEZIONARIO	1SM	LIBRO LITURGICO CON LE#LEZIONI(LEZIONE)DI UFFICI DIVINI	3		
CORALE	2SM	LIBRO LITURGICO CONTENENTE GLI UFFICI DEL#CORO()	1		
EVANGELIARIO	1SM	LIBRO LITURGICO CONTENENTE PASSI DELL' EVANGELO	1		
INNARIO	1SM	LIBRO LITURGICO,NEL CATTOLICESIMO E NELLE CHIESE ORIENTALI	3		
....					
CORANO	1SM	LIBRO SACRO DEI MUSSULMANI	3		
AVESTA	1SM	LIBRO SACRO DELLA RELIGIONE ZOROASTRIANA	3		
GENESI	1SF	PRIMO LIBRO DEL PENTATEUCO NELLA BIBBIA	3		
ALBO	2SM	SPECIE DI LIBRO CONTENENTE FOTOGRAFIE,DISCHI,FRANCOBOLLI	3		
LEVITICO	2SM	TERZO LIBRO BIBLICO DEL PENTATEUCO	9		
SAPIENZA	1SF	UNO DEI LIBRI DELL'ANTICO TESTAMENTO	3		
SAPIENZIA	1SF	1UNO DEI LIBRI DELL'ANTICO TESTAMENTO	3		

Fig. 1. Some of the hyponyms of *libro* (book).

ALLIBRARE	1VT	REGISTRARE SU UN LIBRO DI CONTI	1		
CARTOLINARE	1VT	RILEGARE UN LIBRO ALLA RUSTICA	3		
CIRCOLARE	1VIT	PASSARE DALL'UNA ALL'ALTRA PERSONA,DI DANARO,LIBRI	3		E
DISTRIBUIRE	1VT	DIFFONDERE TRA TUTTI I RIVENDITORI LIBRI,GIORNALI	3		
DIVOLGARE	1VTP	1RENDERE FINANZIARIAMENTE DISPONIBILI LIBRI,SAGGI	3		E
DIVULGARE	1VTP	RENDERE FINANZIARIAMENTE DISPONIBILI LIBRI,SAGGI	3		E
INTERFOGLIARE	1VT	INTERPORRE,CUCIRE TRA I FOGLI DI UN LIBRO FOGLI BIANCHI	3		
INTESTARE	1VTP	FORNIRE DI INTESTAZIONE O TITOLO UN LIBRO	1		
RITONDARE	1VT	1PAREGGIARE,TAGLIANDO LE SPORGENZE,DETTO DI LIBRI,TESSUTI	3	1	
SCARTABELLARE	1VT	SCORRERE IN FRETTA E DISORDINATAMENTE LE PAGINE D'UN LIBRO	3		
SCOMPAGINARE	1VTP	DISFARE,ROVINARE LA LEGATURA DI LIBRI	3		
SCRITTURARE	1VT	ANNOTARE,REGISTRARE SU LIBRI O SCRITTURE CONTABILI	3		
SFASCICOLARE	1VT	SCOMPORRE UN LIBRO,UN QUADERNO NEI FASCICOLI DI CUI E' FATTO	3		
SFOGLIARE	2VTP	SCORRERE UN LIBRO RAPIDAMENTE	3		
SFOGLIARE	2VTP	TAGLIARE LE PAGINE DI UN LIBRO	3	3	
SQUADERNARE	1VTP	3VOLTARE E RIVOLTARE PAGINE DI LIBRI,QUADERNI	3	3	
TOSARE	1VT	PAREGGIARE I FOGLI DEI LIBRI NEL RILEGARLI	3	3	E

Fig. 2. Verbs related to *libri* (books).

```
ADESPOTA          1A    3ANONIMO/DETTO DI LIBRO,CODICE,MANOSCRITTO DI AUTORE IGNOTO    5
ADESPOTO          1A     ANONIMO/DETTO DI LIBRO,CODICE,MANOSCRITTO DI AUTORE IGNOTO    5
APOCRIFO          1A     DETTO DI LIBRO NON RICONOSCIUTO COME CANONICO                 3
CARTOLIBRARIO     1A     DI COMMERCIO DI LIBRI E OGGETTI DA CANCELLERIA               3
CIRCOLANTE        1A     CHE DA' LIBRI A PRESTITO AGLI ABBONATI A TURNO               9
COMMERCIALE       1A     DETTO DI LIBRO,FILM CHE MIRA SOLO A OTTENERE BUONI INCASSI    3      F
COPERTINATO       1A     DETTO DI LIBRO O FASCICOLO CON COPERTINA                     1
DEUTEROCANONICO   1A     DEI LIBRI DELL'ANTICO TESTAMENTO RESPINTI COME APOCRIFI      3
EDITORE           1A     CHI PUBBLICA LIBRI,RIVISTE                                   3
ERUDITO           1A     LIBRO ERUDITO                                                   T
INTESTATO         1A     FORNITO DI TITOLO O INTESTAZIONE,DETTO DI LIBRO,LETTERA-     3
INTONSO           1A    3DI LIBRO CUI NON SONO ANCORA STATE TAGLIATE LE PAGINE        3      F
LIBERIANO         3A     CHE RIGUARDA IL LIBRO                                       36K
LIBRARIO          1A     DI,RELATIVO A LIBRO                                          1
LIBRESCO          1A     CHE DERIVA DAI LIBRI E NON DALLA VIVA ESPERIENZA              1      P
MASTRO            2A     LIBRO MASTRO                                                    L
MOSAICO           2A     RELATIVO AI LIBRI BIBLICI                                    3
PAGA              4A     LIBRO PAGA                                                      L
POSTUMO           1A     DI LIBRO PUBBLICATO DOPO LA MORTE DELL'AUTORE                3
PROTOCANONICO     1A     DETTO DI CIASCUN LIBRO BIBLICO INSERITO PER PRIMO NEL CANONE 3
SAPIENZIALE       1A     CHE SI RIFERISCE AI LIBRI SAPIENZIALI                        3          E
```

Fig. 3. Adjectives related to *libri* (books).

```
RISVOLTO          1SM     ALETTA/ PARTE DELLA SOPRACOPERTA DI LIBRO RIPIEGATA        5
BIBLIOFILO        1SG     AMATORE,RICERCATORE,COLLEZIONISTA DI LIBRI                 3
BIBLIOFILIA       1SF     AMORE PER I LIBRI                                          3
REGGILIBRI        1SM     ARNESE PIEGATO AD ANGOLO RETTO PER REGGERE IN PIEDI LIBRI  3
BIBLIOIATRICA     1SF    3ARTE DEL RESTAURO DEI LIBRI                                3   3
ERMENEUTICA       1SF     ARTE DI INTERPRETARE MONUMENTI,LIBRI ANTICHI               3
SFOGLIATA         2SF     ATTO DELLO SCORRERE UN LIBRO E SIMILI                      1
PUBBLICAZIONE     1SF     ATTO EFFETTO DEL RENDERE PUBBLICO O DEL PUBBLICARE LIBRI   1
BANCHEROZZO       1SM    1BANCARELLA DI LIBRI ALL' APERTO                            3   1
ZAZZERA           1SF     BARBA,RICCIO/ PARTE RUVIDA INTONSA DEI LIBRI               5
PORTACARTE        1SM     BORSA PER METTERVI CARTE,DOCUMENTI,LIBRI                   3
BOTTELLO          1SM    3CARTELLINO CHE SI METTE SU LIBRI E BOTTIGLIE               3   3
CARTOLIBRERIA     1SF     CARTOLERIA AUTORIZZATA ALLA VENDITA DI LIBRI               3
CANONE            1SM     CATALOGO DEI LIBRI SACRI RICONOSCIUTI AUTENTICI            3
REDATTORE         1SN     CHI CURA FASI PER PUBBLICAZIONE DI LIBRI IN CASE EDITRICI  3
CARRETTINISTA     1SM     CHI ESPONE O VENDE LIBRI SU UN CARRETTINO                  1
BIBLIOTECA        1SF     COLLEZIONE DI LIBRI SIMILI PER FORMATO ARGOMENTO EDITORE   3
LIBRATA           1SF     COLPO DATO CON UN LIBRO                                    1
....
BIBLIOTECA        1SF     EDIFICIO CON RACCOLTE DI LIBRI A DISPOSIZIONE DEL PUBBLICO 3
BIBLIOGRAFIA      1SF     ELENCO DI LIBRI CONSULTATI PER COMPILAZIONE DI OPERE       3
INDICE            1SM     ELENCO ORDINATO DI CAPITOLI O PARTI DI LIBRO               3
BIBLIOLATRIA      1SF     FEDE CIECA NEI LIBRI STAMPATI                              3
....                                                                                39Q
LIBRERIA          1SF     LUOGO O MOBILE IN CUI SONO ACCOLTI E CUSTODITI I LIBRI     3          C
BIBLIOTECA        1SF     LUOGO OVE SONO RACCOLTI E CONSERVATI LIBRI                 3
BIBLIOMANIA       1SF     MANIA DI RICERCARE E COLLEZIONARE LIBRI                    3
BIBLIOTECA        1SF     MOBILE A MURO CON SCAFFALI PER LIBRI                       3
CLASSIFICATORE    1SN     MOBILE PER CONTENERE LIBRI DOCUMENTI                       3
LIBRERIA          1SF     NEGOZIO O EMPORIO DI LIBRI
FRONTISPIZIO      1SM     PAGINA ALL' INIZIO DI UN LIBRO CON TITOLO NOTE TIPOGRAFICHE 3
ANTIPORTA         1SF     PAGINA CON TITOLO PRECEDENTE FRONTESPIZIO DI LIBRO          3
TAVOLA            1SF     PAGINA FOGLIO DI LIBRO CON ILLUSTRAZIONI                   3
INTERFOGLIO       1SM     PAGINA INTERPOSTA TRA I FOGLI DI UN LIBRO                  3
LIBRERIA          1SF     RACCOLTA DI LIBRI LIBRO                                    1
BIBLIOLOGIA       1SF     SCIENZA DEI LIBRI                                          3
LIBRAIO           1SN     VENDITORE DI LIBRI                                         1
LIBRARO           1SN    1VENDITORE DI LIBRI
VERSO             3SM     VERSETTO/SUDDIVISIONE IN FRASI DELLE PARTI DI LIBRI SACRI   5          E
```

Fig. 4. Some of the nouns related to *libri* (books).

```
STRUMENTO                 ---->>ABBASSALINGUA                1SM    00
                                ABERROMETRO                  1SM    00
                                ACCELEROGRAFO                1SM    00
                                ACCELEROMETRO                1SM    00
                                ACCHIAPPAMOSCHE              1SN    00
                                ACCIAINO                     1SM    00
                                AEROFONO                     1SM    00
                                AEROMETRO                    1SM    00
                                AEROSCOPIO                   1SM    00
                                AFFILATOIO                   1SM    00
                                AGGUAGLIATOIO                1SM    00
                                AGO                          1SM    0A
                                ALCOOLIMETRO                 1SM    00
                                ALGESIMETRO                  1SM    00
                                AMMOSTATOIO                  1SM    00
                                AMPEROMETRO                  1SM    00
                                ANALIZZATORE                 1SN    00
                                ANCORA                       1SF    10
                                ANEMOMETRO                   1SM    00
                                ANEMOSCOPIO                  1SM    00
                                ANGELICA                     1SF    00
                                APRIBOCCA                    1SM    00
                                APRICASSE                    1SM    00
                                ARCHIPENDOLO                 1SM    00
                                ARMA                         1SF    00
                                ARMONICA                     1SF    00
                                ARMONIO                      1SM    00
                                ARMONIUM                     1SM    00
                                ARPA                         1SF    10
                                ARPEGGIONE                   1SM    00
                                ARRIDATOIO                   1SM    00
                                ASPERSORIO                   1SM    00
                                ASPIRATORE                   1SM    00
                                ASSIOMETRO                   1SM    00
                                ASTIGMOMETRO                 1SM    00
                                ASTROFOTOMETRO               1SM    00
                                ASTROGRAFO                   1SM    00
                                ASTROLABIO                   1SM    00
                                ATTINOMETRO                  1SM    00
                                ATTREZZO                     1SM    0A
                                AUDIOMETRO                   1SM    00
                                AULOS                        1SM    00
                                AVENA                        1SF    00
                                BADILE                       1SM    00
```

Fig. 5. The first hyponyms of *strumento* (instrument).

```
AFFOSSATORE          1SN     ATTREZZO AGRICOLO PER SCAVARE FOSSI                                3
ALLARGATESE          1SM     ATTREZZO USATO PER ALLARGARE LE TESE DEI CAPPELLI                  3
ALLISCIATOIO         1SM     ATTREZZO USATO IN FONDERIA PER PREPARARE LE FORME                  3
ANELLO               1SM     ATTREZZO GEMELLARE IN GINNASTICA                                   3
APISCAMPO            1SM     ATTREZZO PER IMPEDIRE L' ASCESA DELLE API AL MELARIO               3
APPOGGIO             1SM     ATTREZZO GINNICO FORMATO DA BLOCCHETTI RETTANGOLARI DI LEGNO       3
ARATRO               1SM     ATTREZZO AGRICOLO ATTO A ROMPERE,DISSODARE IL TERRENO              3
ARNESE               1SM     ATTREZZO DA LAVORO                                                 3
ASPO                 1SM     ASPA,ANNASPO,NASPO/ ATTREZZO CHE SERVE AD ESEGUIRE L'ASPATURA      54E
ASTA                 1SF     ATTREZZO DI FORMA TUBOLARE NELL' ATLETICA                          3
BACCHETTA            1SF     ATTREZZO PER ESERCIZI GINNICI COLLETTIVI                           3
BARRAMINA            1SF     ATTREZZO PER LA PERFORAZIONE DELLE ROCCE                           3
BASTONCINO           1SM     ATTREZZO DEGLI SCIATORI CON RACCHETTA CIRCOLARE                    3
BASTONE              1SM     MAZZA/ ATTREZZO SPORTIVO                                           5
CACCIAVITE           1SM     ATTREZZO PER STRINGERE O ALLENTARE LE VITI                         3
CAVALLINA            1SF     ATTREZZO PER ESERCIZI DI VOLTEGGIO NELLA GINNASTICA                3
CAVALLO              1SD     ATTREZZO PER ESERCIZI DI VOLTEGGIO NELLA GINNASTICA                3   5
CERCHIO              1SM     ATTREZZO STRUTTURA FIGURA A FORMA DI CERCHIO                       3
CESTA                1SF     CHISTERA/ ATTREZZO DI VIMINI USATO NELLA PELOTA BASCA              5
CHIAVE               1SF     ATTREZZO METALLICO PER PROVOCARE CONTATTI                          3
CHIAVE               1SF     ATTREZZO METALLICO PER METTERE IN MOTO MECCANISMI                  3
CHIAVE               1SF     ATTREZZO METALLICO PER ALLENTARE E STRINGERE VITI O DADI           3
CHIODO               1SM     ATTREZZO IN METALLO DEGLI ALPINISTI                                3
CHIOVO               1SM    1ATTREZZO IN METALLO DEGLI ALPINISTI                                3   1
CILINDRO             1SM     ATTREZZO CILINDRICO NELLA GINNASTICA                               3
CLAVA                1SF     ATTREZZO IN LEGNO USATO PER ESERCIZI GINNICI                       3
COLTIVATORE          2SN     ATTREZZO PER SMUOVERE E SMINUZZARE LA SUPERFICIE DEL TERRENO       3
CORDA                1SF     ATTREZZO DA ALPINISMO O GINNASTICA                                 39L
CUCCHIAIA            1SF     ATTREZZO PER ESTRARRE DETRITI DI ROCCIA                            3
CUCITRICE            2SF     ATTREZZO USATO NEGLI UFFICI PER UNIRE FOGLI                        3
DISCO                1SM     ATTREZZO CIRCOLARE CHE SI LANCIA IN GARE SPORTIVE                  3
ERPICE               1SM     ATTREZZO DI FERRO PER LAVORARE IL TERRENO                          3
ESTENSORE            2SI     ATTREZZO GINNICO                                                   3
ESTIRPATORE          3SM     ATTREZZO PER SMUOVERE O LIBERARE IL TERRENO DA ERBACCE             3
FALCE                1SF     ATTREZZO PER TAGLIARE A MANO CEREALI ED ERBE                       3
FIOCINA              1SF     ATTREZZO CON TRE O PIU' DENTI FISSI PER CATTURARE PESCI            3
....
UTENSILE             2SM     OGNI ATTREZZO PER LAVORARE LEGNO,PIETRE,MATERIALI                  3
VANGHETTA            1SF     ATTREZZO LEGGERO DI SOLDATO PER PICCOLI LAVORI DI STERRO           3
VOGADORE             1SI    1ATTREZZO GINNICO PER MOVIMENTO DA REMATORE                         3
VOGATORE             1SN     ATTREZZO GINNICO PER MOVIMENTO DA REMATORE                         3
VOLTARISO            1SM     ATTREZZO PER RIVOLTARE SULL'AIA MODESTE QUANTITA' DI RISO          3
ZAPPA                1SF     ATTREZZO MANUALE PER LAVORARE IL TERRENO                           3
```

Fig. 6. Some of the hyponyms of *attrezzo* (tool) with their definitions.

INSTRUMENT <--IS-A--	*attrezzo*	**--USED FOR-->**	*tagliare ...*	= *FALCE*
	(tool)		...	= ...
		--USED IN-->	*ginnastica*	= *ANELLO*
			...	= ...
		--SHAPE-->	*tubolare*	= *ASTA*
		"	*circolare*	= *DISCO*
		--MADE OF-->	*vimini*	= *CESTA*
			metallo	= *CHIODO*

Fig. 7. Sketch of a piece of network for *attrezzo* (tool).

FORMICAIO	SM	MOLTITUDINE DI	PERSONE
GREGGE	SN	MOLTITUDINE DI	PERSONE
STORMO	SM	MOLTITUDINE DI	PERSONE
MANO	SF	GRUPPO DI	PERSONE
ROSA	SF	CERCHIA/ GRUPPO INSIEME DI	PERSONE
BRANCO	SM	INSIEME DI	PERSONE
CIRCOLO	SM	CENACOLO,SODALIZIO/INSIEME DI	PERSONE
COMMISSIONE	SF	GRUPPO DI	PERSONE A CUI E' AFFIDATO UN UNCARICO PUBBLICO
POPOLAZIONE	SF	INSIEME DELLE	PERSONE ABITANTI IN UN LUOGO
ORGANICO	SM	COMPLESSO DI	PERSONE ADDETTE A CERTE ATTIVITA'
SEGRETERIA	SF	INSIEME DELLE	PERSONE ADDETTE A UNA SEGRETERIA
SQUADRA	SF	COMPLESSO DI	PERSONE ADDETTE A UNO STESSO LAVORO
CIURMA	SF	INSIEME DELLE	PERSONE ADDETTE AI LAVORI DELLA TONNARA
NAZIONE	SF	INSIEME DI	PERSONE APPARTENENTI A STESSA STIRPE
FAMIGLIA	SF	COMPLESSO DI	PERSONE AVENTI UN ASCENDENTE DIRETTO COMUNE
VICINATO	SM	INSIEME DI	PERSONE CHE ABITANO UNA STESSA CASA
CORTE	SF	GRUPPO DI	PERSONE CHE ACCOMPAGNA UN PERSONAGGIO IMPORTANTE
LEGA	SF	INSIEME DI	PERSONE CHE AGISCONO PER UTILE PROPRIO
AUDITORIO	SM	UDITORIO/COMPLESSO DI	PERSONE CHE ASCOLTANO
UDIENZA	SF	UDITORIO/INSIEME DI	PERSONE CHE ASCOLTANO
CAROVANA	SF	GRUPPO DI	PERSONE CHE ATTRAVERSANO CON CARRI LUOGHI DESERTI
CORO	SM	GRUPPO DI	PERSONE CHE CANTANO INSIEME
MALAVITA	SF	L'INSIEME DELLE	PERSONE CHE CONDUCONO VITA DISSOLUTA
CROCCHIO	SM	GRUPPO DI	PERSONE CHE CONVERSANO
CORO	SM	GRUPPO DI	PERSONE CHE DICONO,GRIDANO Q.C. CONTEMPORANEAMENTE
CONCISTORO	SM	GRUPPO DI	PERSONE CHE DISCUTONO
FINANZA	SF	COMPLESSO DI	PERSONE CHE ESPLICANO ATTIVITA' BANCARIA
....			
FRONTE	SN	COMPLESSO DI	PERSONE OMOGENEO PER FINALITA' CONSUETUDINI
ARISTOCRAZIA	SF	COMPLESSO DI	PERSONE PIU' QUALIFICATE PER UNA ATTIVITA'
CHIESA	SF	INSIEME DI	PERSONE PROFESSANTI LA MEDESIMA DOTTRINA
DRAPPELLO	SM	GRUPPO DI	PERSONE RACCOLTE INSIEME
COMPAGNIA	SF	COMPLESSO DI	PERSONE RIUNITE INSIEME PER ATTIVITA' COMUNI
GRUPPO	SM	INSIEME DI	PERSONE UNITE DA VINCOLI NATURALI O DI INTERESSE

Fig. 8. Some of the nouns denoting **SET OF** *persone* (people).

ARCIPELAGO	SM	GRUPPO INSIEME DI	OGGETTI
ANTIQUARIATO	SM	COMMERCIO O RACCOLTA DI	OGGETTI ANTICHI
SERVIZIO	SM	INSIEME DI	OGGETTI CHE SERVONO A UN DETERMINATO SCOPO
TROFEO	SM	INSIEME DI	OGGETTI CHE TESTIMONIANO SUCCESSI E VITTORIE
AFFARDELLAMENTO	SM	COMPLESSO DEGLI	OGGETTI CONTENUTI NELLO ZAINO DEL SOLDATO
ARGENTERIA	SF	COMPLESSO DI	OGGETTI D'ARGENTO
ORERIA	SF	COMPLESSO DI	OGGETTI D'ORO
COLLEZIONE	SF	RACCOLTA DI	OGGETTI DELLA STESSA SPECIE
CRISTALLERIA	SF	INSIEME DEGLI	OGGETTI DI CRISTALLO DA TAVOLA
CIANFRUSAGLIA	SF	CHINCAGLIERIA/INSIEME DI	OGGETTI DI POCO PREGIO
CIANFRUSCAGLIA	SF	CHINCAGLIERIA/INSIEME DI	OGGETTI DI POCO PREGIO
ASSORTIMENTO	SM	INSIEME DI	OGGETTI DI STESSO GENERE DIVERSI NEI PARTICOLARI
ARSENALE	SM	INSIEME DI	OGGETTI DIVERSI
SUPPELLETTILE	SF	OGGETTO O INSIEME DI	OGGETTI IN UNA SCUOLA CHIESA E SIMILI
INTRECCIO	SM	COMPLESSO DI	OGGETTI INTRECCIATI
ATTREZZERIA	SF	INSIEME DI	OGGETTI NECESSARI PER UNA SCENA TEATRALE
SUPPELLETTILE	SF	OGGETTO O INSIEME DI	OGGETTI NELL'ARREDAMENTO DELLA CASA
ARREDO	SM	OGGETTO O COMPLESSO DI	OGGETTI PER GUARNIRE AMBIENTI
COMPLETO	SM	INSIEME DI	OGGETTI PER UN USO DETERMINATO
BAROCCUME	SM	INSIEME DI	OGGETTI PRETENZIOSI E DI CATTIVO GUSTO
GIOIELLERIA	SF	INSIEME DI	OGGETTI PREZIOSI
SUPPELLETTILE	SF	OGGETTO O INSIEME DI	OGGETTI RINVENUTI IN UNO SCAVO

Fig. 9. Nouns denoting **SET OF** *oggetti* (objects).

ASSESTATO	A	ASSENNATO,AVVEDUTO,DETTO DI	PERSONA
BARLACCIO	A	MALATICCIO,DEBOLE,DETTO DI	PERSONA
INSENSATO	A	STUPIDO,DEMENTE,DETTO DI	PERSONA
PRIMITIVO	A	C=INCIVILITO/SEMPLICE,ROZZO,CREDULONE,DETTO DI	PERSONA
PROVETTO	A	MATURO,DETTO DI	PERSONA
RIMESSO	A	LANGUIDO,LENTO,FIACCO,DETTO DI	PERSONA
RINCRESCIOSO	A	CHE SENTE RINCRESCIMENTO,DETTO DI	PERSONA
RIPOSANTE	A	CALMO,TRANQUILLO DETTO DI	PERSONA
RISPETTOSO	A	CHE HA,E' PIENO DI#RISPETTO(),DETTO DI	PERSONA
ROBUSTO	A	FORTE/CHE POSSIEDE FORZA,ENERGIA,DETTO DI	PERSONA
ROCO	A	RAUCO,DETTO DI	PERSONA
ROGNOSO	A	MISERO,MESCHINO,NOIOSO,DETTO DI	PERSONA
RUDE	A	ROZZO,GROSSOLANO,DETTO DI	PERSONA
RUGIADOSO	A	SANO,FLORIDO,DETTO DI	PERSONA
RUSTICO	A	NON MOLTO SOCIEVOLE NE' RAFFINATO,DETTO DI	PERSONA
RUVIDO	A	DI MANIERE ROZZE,DI CARATTERE ASPRO,DETTO DI	PERSONA
....			PERSONA
ADOMBRARE	VTE	INSOSPETTIRSI,TURBARSI,DETTO DI	PERSONA
ARRABBIARE	VIE	ESSERE PRESO DALL'IRA,DALLA COLLERA DETTO DI	PERSONA
CORVETTARE	VI	SALTARE,BALZARE,DETTO SPEC. DI	PERSONA
CUCCIARE	VET	GIACERSI/STARE A LETTO,DETTO DI	PERSONA
IMBIZZARRIRE	VET	INCOLLERIRE O DIVENTARE IRREQUIETO DETTO DI	PERSONA
IMPROSCIUTTIRE	VI	DIVENTARE ASCIUTTO COME UN PROSCIUTTO,DETTO DI	PERSONA
RABBRUSCARE	VEY	ADOMBRARSI/OFFUSCARSI IN VOLTO,DETTO DI	PERSONA
RICEVERE	VT	AMMETTERE,DETTO DI	PERSONA
RIDURRE	VT P	METTERE IN CONDIZIONI PEGGIORI,DETTO DI	PERSONA
RIMETTERE	VT PI	RISTABILIRSI,DETTO DI	PERSONA
RINFIERIRE	VI	INFIERIRE DI NUOVO O DI PIU',DETTO DI	PERSONA
RINSECCHIRE	VIT	DIVENTARE MAGRO,ASCIUTTO,DETTO DI	PERSONA
RINVENIRE	VI	RIANIMARSI,RIAVERSI/RICUPERARE I SENSI DETTO DI	PERSONA
RISALTARE	VNI	EMERGERE,DISTINGUERSI,DETTO DI	PERSONA
RISORGERE	VI T	SOLLEVARSI,RIAVERSI DETTO DI	PERSONA
RISPUNTARE	VIT	RIAPPARIRE,RICOMPARIRE,DETTO DI	PERSONA
RISURGERE	VI T	SOLLEVARSI,RIAVERSI,DETTO DI	PERSONA
RIUSCIRE	VI	RAGGIUNGERE IL FINE,LO SCOPO,DETTO DI	PERSONA
ROTOLARE	VTIR	GIRARSI SU DI SE',VOLTOLARSI,DETTO DI	PERSONA
ROVINARE	VITR	CADERE IN BASSO,DETTO DI	PERSONA
....			
CORDIALE	A	DETTO DI	PERSONA AFFABILE,GENTILE,APERTA
LONGO	A	CHE SI ESTENDE IN ALTEZZA,DETTO DI	PERSONA ALTA E MAGRA
LUNGO	A	CHE SI ESTENDE IN ALTEZZA,DETTO DI	PERSONA ALTA E MAGRA
PRODIGIO	A	DETTO DI	PERSONA CHE E' ECCEZIONALE
SUPINO	A	C=PRONO/DETTO DI	PERSONA CHE GIACE SUL DORSO
LACERO	A	CENCIOSO/DETTO DI	PERSONA CHE INDOSSA VESTITI LOGORI
SCIVOLOSO	A	DETTO DI	PERSONA CHE NASCONDE LE SUE VERE INTENZIONI
IMPREGIUDICATO	A	DETTO DI	PERSONA CHE NON HA AVUTO CONDANNE PENALI
IMPETTITO	A	DETTO DI	PERSONA CHE STA ERETTA E COL PETTO IN FUORI
ASOCIALE	A	DETTO DI	PERSONA CHIUSA INTROVERSA
....			
NAUFRAGARE	VI	ESSERE SUL BASTIMENTO CHE ROMPE IN MARE,DETTO DI	PERSONE
RICONGIUNGERE	VT D	CONGIUNGERSI DI NUOVO,RIUNIRSI,DETTO DI	PERSONE
RIMESCOLARE	VTP	INTROMETTERSI,MISCHIARSI A UN GRUPPO,DETTO DI	PERSONE
ROVESCIARE	VTP	ABBANDONARSI,DETTO DI	PERSONE
SBOCCARE	VIT	ARRIVARE IN UN DATO LUOGO,DETTO DI	PERSONE
SCHIAMAZZARE	VI	VOCIARE,STREPITARE,DETTO DI	PERSONE
SPELLICCIARE	VTB	PICCHIARSI,AZZUFFARSI RABBIOSAMENTE,DETTO DI	PERSONE
ULULARE	VI	EMETTERE PROLUNGATI,CUPI LAMENTI,DETTO DI	PERSONE

Fig. 10. Some of the adjectives and verbs which can be predicated of *persone* (people).

ACCESO	A	VIVO,INTENSO,DETTO DI	COLORE
CHIARO	A	C=SCURO/PALLIDO,TENUE,POCO INTENSO DETTO DI	COLORE
CUPO	A	DI TONALITA' SCURA DETTO DI	COLORE
SERPATO	A	CHE E' SCREZIATO,COME LA PELLE DEL SERPENTE,DETTO DI	COLORE
SQUILLANTE	A	VIVACE,INTENSO,DETTO DI	COLORE
STABILE	A	CHE NON SBIADISCE,DETTO DI	COLORE
TENUE	A	PALLIDO/NON MOLTO VIVO DETTO DI	COLORE
RISCHIARARE	VTE	FARSI CHIARO,LUMINOSO,DETTO DI	COLORE
SCARICARE	VTRIP	PERDERE VIVACITA',SBIADIRE,DETTO DI	COLORE
BERRETTINO	A	DETTO DI	COLORE AZZURRO CINEREO SU VASI DI MAIOLICA
CALCE	A	DETTO DI	COLORE BIANCO INTENSO
GIGLIACEO	A	DETTO DI	COLORE CHE RICORDA QUELLO DEL GIGLIO
SCURO	A	C=CHIARO/DETTO DI	COLORE CHE TENDE AL NERO
BRUNO	A	DETTO DEL	COLORE DEL MANTELLO DEI BOVINI
ALBICOCCA	A	DETTO DI	COLORE GIALLO ARANCIATO
ZAFFERANO	A	DETTO DI	COLORE GIALLO INTENSO
ISABELLA	A	DETTO DI	COLORE GIALLO TIPICO DI MANTELLO EQUINO
PERLA	A	DETTO DI	COLORE LATTIGINOSO E OPALESCENTE
TERRA	A	DETTO DI	COLORE MARRONE CHIARO SFUMATO AL GRIGIO
SUDICIO	A	DETTO DI	COLORE NON BRILLANTE,NON VIVO
DISUGUAGLIATO	A	DETTO DI	COLORE NON UNIFORME DI UNA TINTURA
NEGRO	A	DETTO DEL	COLORE PIU' SCURO
NERO	A	DETTO DEL	COLORE PIU' SCURO
GIACINTINO	A	DETTO DEL	COLORE ROSSASTRO,TIPICO DEL GIACINTO
TANGO	A	DETTO DI	COLORE ROSSO ASSAI BRILLANTE
GRANATA	A	DETTO DI	COLORE ROSSO SCURO
PULCE	A	DETTO DI	COLORE TRA GRIGIO E VERDE
RUGGINE	A	DETTO DI	COLORE TRA IL MARRONE E IL ROSSO SCURO
LILLA'	A	GRIGELLINO/DETTO DI	COLORE TRA ROSA E VIOLA
GIADA	A	DETTO DI	COLORE VERDAZZURRO CHIARO
SBIADATO	A	SBIADITO,TENUE,PALLIDO,DETTO DI	COLORI
ADDOLCIRE	VTP	AMMORBIDIRE,DETTO DI	COLORI
DISCORDARE	VE	STONARE/NON ARMONIZZARE,DETTO DI	COLORI
SBIADIRE	VET	SCOLORIRE,STINGERE/DIVENTARE PALLIDO,SMORTO,DETTO DI	COLORI
SGARGIARE	VI	ESSERE ECCESSIVAMENTE VIVACE E VISTOSO,DETTO DI	COLORI
SMONTARE	VTIP	SCHIARIRE,SCOLORIRE,STINGERE,DETTO DI	COLORI
TRIONFARE	VIT	RISALTARE/FARE SPICCO,DETTO DI	COLORI
USCIRE	VIT	RISALTARE DETTO DI	COLORI
SMORTO	A	CHE E' PRIVO DI SPLENDORE E VIVACITA' DETTO DI	COLORI E SIM.
ALLEGRO	A	VIVACE,BRIOSO DETTO DI	COLORI SUONI E SIMILI
RISALTARE	VNI	SPICCARE NITIDAMENTE,DETTO DI	COLORI,DISEGNI,PITTURE
TENDERE	VT IP	AVVICINARSI AD UNA GRADAZIONE DETTO DI	COLORI,SAPORI,ODORI

Fig. 11. Some of the adjectives and verbs which are typically predicated of *colori* (colours).

VENDE ---->>AGNELLAIO	1SI	CHI MACELLA O VENDE AGNELLI		1
AGORAIO	1SM	CHI FA O VENDE AGHI		
ALABASTRAIO	1SI	CHI VENDE OGGETTI DI ALABASTRO		
ARAZZIERE	1SI	CHI TESSE E VENDE ARAZZI		1
ARGENTIERE	1SI	CHI VENDE OGGETTI D'ARGENTO		
ARMAIOLO	1SI	CHI FABBRICA VENDE RIPARA ARMI		
ASTUCCIAIO	1SI	CHI FABBRICA O VENDE ASTUCCI		1
BABBUCCIAIO	1SI	CHI FA O VENDE BABBUCCE		1
BADILAIO	1SI	CHI FA O VENDE BADILI		1
BERRETTAIO	1SN	CHI FABBRICA O VENDE BERRETTI		1
BICCHIERAIO	1SI	CHI FABBRICA O VENDE BICCHIERI		1
BIGLIETTAIO	1SN	CHI VENDE I BIGLIETTI PER IL VIAGGIO		1
BILANCIAIO	1SI	STADERAIO/CHI FABBRICA E VENDE BILANCE		4
BILIARDAIO	1SI	CHI FABBRICA O VENDE BILIARDI		1
BIRRAIO	1SI	CHI FABBRICA O VENDE BIRRA		1
BOCCALAIO	1SI	CHI FABBRICA O VENDE BOCCALI		1
BORSAIO	1SG	CHI FABBRICA O VENDE BORSE		1
BOTTAIO	1SI	CHI FABBRICA,RIPARA O VENDE BOTTI		1
BOTTONAIO	1SN	CHI FABBRICA O VENDE BOTTONI		1
BUSTAIA	1SF	DONNA CHE CONFEZIONA O VENDE BUSTI		1
CALZETTAIO	1SN	CHI VENDE O FABBRICA CALZE		1
CANESTRAIO	1SI	CHI FA O VENDE CANESTRI		1
CARBONAIO	1SM	CHI VENDE CARBONE		1
....				
OROLOGIAIO	1SI	CHI FABBRICA,RIPARA O VENDE OROLOGI		1
ORTOPEDICO	2SI	CHI FABBRICA O VENDE APPARECCHI ORTOPEDICI		3
OTTICO	2SI	CHI CONFEZIONA E VENDE OCCHIALI E LENTI		3
PADELLAIO	1SI	CHI FA O VENDE PADELLE		1
PANETTIERE	1SN	FORNAIO/CHI FA O VENDE PANE		
PANIERAIO	1SG	CHI FA O VENDE PANIERI		
PANTOFOLAIO	1SN	CHI CONFEZIONA O VENDE PANTOFOLE		1
PASTAIO	1SN	CHI FABBRICA O VENDE PASTE ALIMENTARI		1
PASTICCERE	1SN	CHI FA O VENDE DOLCIUMI		
PASTICCIERE	1SN	CHI FA O VENDE DOLCIUMI		
PATACCARO	1SI	2CHI VENDE MONETE OD OGGETTI FALSI		
PELLETTIERE	1SG	CHI PRODUCE O VENDE OGGETTI DI PELLETTERIA		1
PELLICCIAIO	1SN	CHI LAVORA O VENDE PELLICCE		1
....				
VENDITORE	2SI	CHI VENDE		1
VETRAIO	1SI	CHI VENDE TAGLIA APPLICA LASTRE DI VETRO		
VINATTIERE	1SM	1CHE VENDE O COMMERCIA VINO	1	5
VIOLINAIO	1SI	LIUTAIO/CHI FABBRICA O VENDE VIOLINI	4	
ZOCCOLAIO	1SI	CHI FA O VENDE ZOCCOLI	1	

Fig. 12. Nouns of **AGENTS** for the action of "selling".

VENDITORE	---->>ABBACCHIARO	1SI	2VENDITORE DI ABBACCHI	1	2	
	ACQUAVITAIO	1SI	VENDITORE DI ACQUAVITE	1		
	ARCHIBUGIERE	1SM	FABBRICANTE O VENDITORE DI ARMI	3	1	
	....					
	BIBITARO	1SI	2VENDITORE DI BIBITE	1	2	
	BORSETTAIO	1SG	FABBRICANTE O VENDITORE DI BORSE E BORSETTE	1		
	BRONZISTA	1SN	VENDITORE DI OGGETTI ARTISTICI IN BRONZO			
	BURATTINAIO	1SI	FABBRICANTE O VENDITORE DI BURATTINI			
	CALCOGRAFO	1SI	VENDITORE DI INCISIONI	3		
	CALDARROSTAIO	1SN	VENDITORE DI CALDARROSTE	1		
	CAMICIAIO	1SD	FABBRICANTE O VENDITORE DI CAMICIE	1		
	CAPPELLAIO	1SN	FABBRICANTE O VENDITORE DI CAPPELLI DA UOMO	3		
	CARAMELLAIO	1SN	FABBRICANTE O VENDITORE DI CARAMELLE	1		
	....					
	FRUTTIVENDOLO	1SN	VENDITORE DI FRUTTA E ORTAGGI	3		
	LATTAIO	1SN	VENDITORE DI LATTE	1		
	LIBRAIO	1SN	VENDITORE DI LIBRI			
	MACELLAIO	1SN	VENDITORE DI CARNE MACELLATA	3		
	....					
	PROFUMIERE	1SN	FABBRICANTE O VENDITORE DI PROFUMI E COSMETICI	1		
	SALUMIERE	1SN	VENDITORE DI SALUMI	1		
	SPEZIALE	2SI	VENDITORE DI SPEZIE	1	1	
	STRILLONE	1SN	VENDITORE AMBULANTE DI GIORNALI	3		
	VALIGIAIO	1SN	FABBRICANTE O VENDITORE DI VALIGIE BAULI,BORSE	1		
	VINAIO	1SN	VENDITORE FORNITORE DI VINO	1		

Fig. 13. Nouns of **AGENTS** for the action of "selling".

VENDONO	---->>APPALTO	1SM	LUOGO DOVE SI VENDONO PRODOTTI DI MONOPOLIO DELLO STATO	3	2	
	BANCO	1SM	LOCALE DOVE SI VENDONO O SCAMBIANO BENI SERVIZI	3		
	BIGIOTTERIA	1SF	NEGOZIO DOVE SI VENDONO OGGETTI DECORATIVI NON PREZIOSI	3		E
	BIGLIETTERIA	1SF	LUOGO IN CUI SI VENDONO BIGLIETTI	1		
	BISCOTTERIA	1SF	NEGOZIO DOVE SI VENDONO I BISCOTTI			
	BOTTIGLIERIA	1SF	NEGOZIO DOVE SI VENDONO VINO LIQUORI IN BOTTIGLIA	3		
	BRICABRAC	1	NEGOZIO,BANCARELLA OVE SI VENDONO TALI ANTICAGLIE	3		E
	CALZETTERIA	1SF	NEGOZIO IN CUI SI VENDONO CALZE			
	CALZOLERIA	1SF	BOTTEGA IN CUI SI FABBRICANO O VENDONO SCARPE			
	CAMICERIA	1SF	NEGOZIO IN CUI SI VENDONO CAMICIE			
	CAPPELLERIA	1SF	NEGOZIO DOVE SI VENDONO CAPPELLI MASCHILI	1		
	CERERIA	1SF	LUOGO DOVE SI FABBRICANO E VENDONO CANDELE	3		
	CHINCAGLIERIA	1SF	NEGOZIO IN CUI SI VENDONO CHINCAGLIE			
	CONFETTURERIA	1SF	LUOGO OVE SI PREPARANO,VENDONO CONFETTURE	1		
	CREMERIA	1SF	2LATTERIA IN CUI SI VENDONO ANCHE GELATI DOLCI E SIM.	3		
	DIACCIATINO	2SN	2BOTTEGA DOVE SI VENDONO SORBETTI	3	1	
	DROGHERIA	1SF	BOTTEGA DOVE SI VENDONO DROGHE	1		
	FERRAMENTA	1SF	NEGOZIO IN CUI SI VENDONO OGGETTI DI FERRO	3		
	GELATERIA	1SF	SORBETTERIA/NEGOZIO OVE SI FANNO O VENDONO GELATI	4		
	MAGLIERIA	1SF	BOTTEGA NEGOZIO IN CUI VENDONO INDUMENTI DI MAGLIA			
	MESCITA	1SF	BOTTEGA IN CUI SI VENDONO VINO LIQUORI	3	2	
	MESTICHERIA	1SF	2BOTTEGA IN CUI SI VENDONO COLORI MESTICATI	3	2	
	NEGOZIO	1SM	BOTTEGA/ LOCALE DOVE SI ESPONGONO E VENDONO MERCI	5		
	NORCINERIA	1SF	2BOTTEGA IN CUI SI VENDONO SOLO CARNI DI MAIALE	3	2	
	OCCHIALERIA	1SF	NEGOZIO IN CUI SI VENDONO O SI RIPARANO OCCHIALI			
	OROLOGERIA	1SF	NEGOZIO DOVE SI VENDONO OROLOGI	3		
	PANTOFOLERIA	1SF	LUOGO IN CUI SI VENDONO PANTOFOLE			
	PELLETTERIA	1SF	NEGOZIO IN CUI SI VENDONO OGGETTI DI PELLE LAVORATA	3		
	PIATTERIA	1SF	BOTTEGA DOVE SI VENDONO I PIATTI	3		
	ROSTICCERIA	1SF	BOTTEGA DOVE SI PREPARANO O VENDONO ARROSTI	3		
	SALUMERIA	1SF	BOTTEGA,NEGOZIO,IN CUI SI VENDONO I SALUMI	3		
	SPACCIO	1SM	LOCALE DELLE CASERME DOVE SI VENDONO GENERI ALIMENTARI VARI	3		
	UTENSILERIA	1SF	BOTTEGA IN CUI SI VENDONO UTENSILI			

Fig. 14. Nouns of **PLACES** related to the action of "selling".

OROLOGERIA = <--**LOC**-- ***"selling"*** --**THEME**--> orologi --**IS-A**--> OBJECT

OROLOGIAIO = <--**AGENT**-- " " " " "

Fig. 15. Sketch of a piece of network for the action of *" **selling**"*.

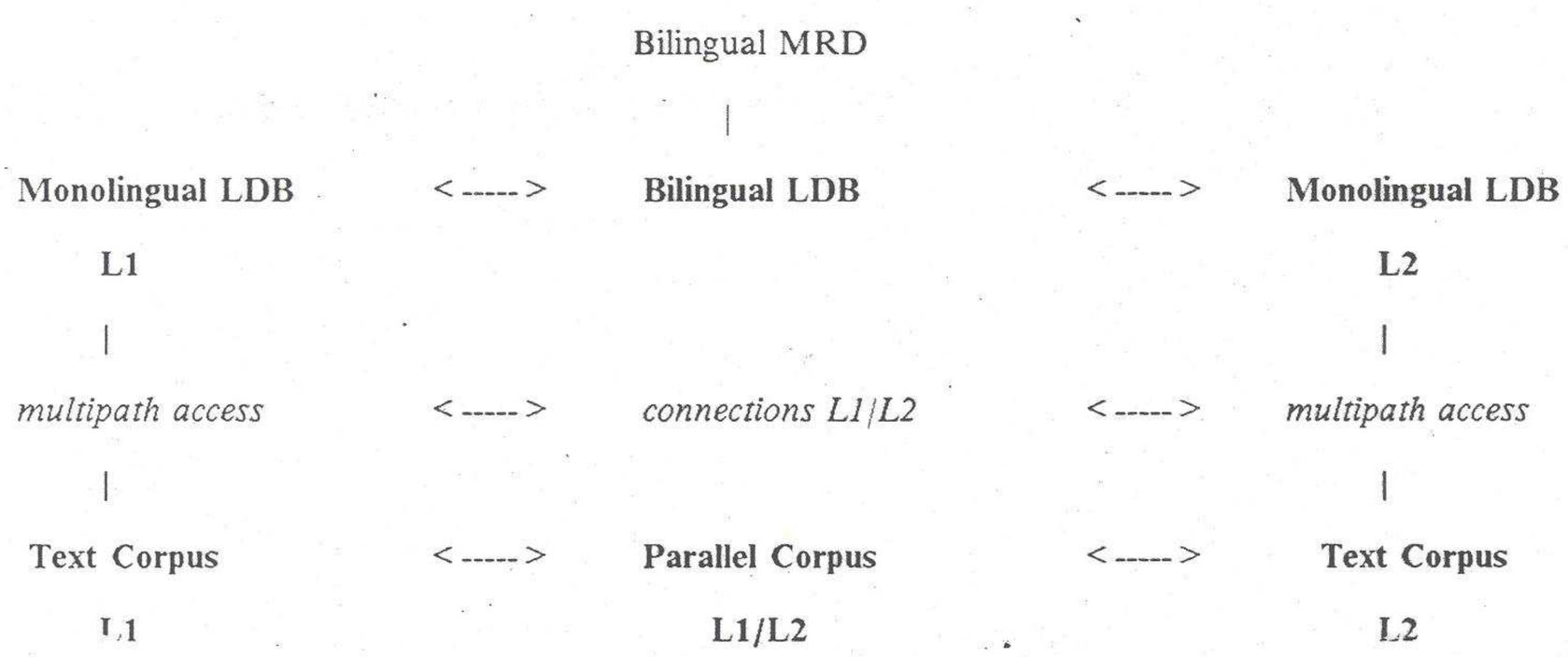

Fig. 16. A model of a Bilingual LDB System.

NOTES

(1) In fact, the first experiments of concordances and indices production were performed not with ′electronic machines′, but with ′punched card electrical accounting machines′ (Busa (1951), 22).

(2) For the history of the first years of MT, see Locke and Booth (1955), 1-23; Booth, Cleave and Brandwood (1958), 1-7; Vauquois (1975), 14-32; Nagao (1989), chapters 1-2.

(3) In the Introduction to the *"Actes du Colloque International sur la Mechanisation des Recherches Lexicologiques"* held in 1961 in Besancon, B. Quemada says: ′Un des buts de ce Colloque sera aussi de mettre en contact des chercheurs qui sans s′ignorer tout a fait, n′echangent guere d′informations alors qu′ils travaillent sur une matiere commune: la langue, et plus particulierement, le lexique dans diverses disciplines. Nous avons la chance d′accueillir ici a cote des lexicologues et des lexicographes francais et etrangers, des specialistes de la traduction automatique (vocabulaire de base, terminologies scientifiques, speciales, dictionnaires automatiques, homographes, synonymes), de la traduction ″artisanale″ (...) de la documentation automatique (...) de la pedagogie des langues vivantes′.
And R. Busa (in an article with a very significant title, given the period: *L′analisi linguistica nell′evoluzione mondiale dei mezzi di informazione* - ″the linguistic analysis in the world evolution of information tools)″ - published as a contribution to a debate on the fracture between sciences and humanities) says that the ′development of linguistic automation is triangular: lexical analysis, information retrieval, mechanical translation′, Busa (1961), 117.

(4) M. Kay (Kay, 1964), reporting on an informal meeting on ″Formats for Machine Readable Texts″ at the end of the IBM-sponsored Literary Data Processing Conference (Yorktown Heights, 1964), and in an article in the fifth issue of the *Computers and the Humanities* (Kay, 1967), explicitly stressed the common interest of MT and humanities researchers on this topic. But it is interesting to note that, in the very same issue, only two MT projects, both directed by well-known linguists, B. Pottier and W.P. Lehmann, are reported in the Directory of Scholars Active, of a total of 120 projects in the section Language and Literature.

(5) But not, we think, directly inspired by it.

(6) At page 2 of the ALPAC Report, the Chairman of the Committee on Science and Public Policy, in a letter to the President of the National Academy of Science, stated that ″the support needs for computational linguistics are distinct from automatic language translation″. At page 29, one reads ″work toward machine translation, together with computational linguistics work that has grown out of it″.

(7) We quote from the Recommendation: ′Small scale experiments and work with miniature models of language have proven seriously deceptive in the past, and one can come to grips with real problems only above a certain scale of grammar size, dictionary size, and available corpora′ (ALPAC, p. iv).

(8) This situation is still true today, ′A recent workshop on linguistic theory and computer applications (Withelock et al., 1987) reports an informal poll to establish the average size of the lexicon used by the prototypes discussed ...; the average size was about 25 words′ (Boguraev and Briscoe (1989) 10).

(9) See the Proceedings of the *Table Ronde sur les Grandes Dictionnaires Historiques* (Firenze, 1973).

(10) See, for example, the series of frequency dictionaries of romance languages coordinated by Juilland, published by Mouton in 1961 (Spanish), 1965 (Rumanian), 1970 (French), 1973 (Italian).

(11) A well-known example is the IBM development of specialized optical support for storing large dictionaries in early ′60.

(12) These two systems were presented and compared at the Pisa 1968 meeting *′De lexico electronico latino′*, during which was also presented the first proposal for a multifunctional lexicon (Italian Machine Dictionary: DMI), conceived as a repository of lexical knowledge both for computer programs (parsers, generators, phonological transcription, lemmatization, etc.) and human uses (qualitative and quantative researches on the structure of the Italian lexical system). The Gallarate Latin machine dictionary was made up of an alphabetical list of forms, progressively accumulated from processing the texts of St. Thomas Aquinus. The *Liege Dictionary* was based on a list of stems, extracted from the Forcellini lemmas, and an associated morphological analyser (See Busa, 1968).

(13) The article, ″The Field and Scope of Computational Linguistics′, of D. Hays in the *Proceedings of the Budapest COLING 1971* is particularly relevant, and it is interesting to observe the evolution towards a

'puristic' definition' of CL in the opinion of the author in respect to his chapter on 'computational linguistics' in the *Encyclopaedia of Linguistics Information and Control* (1969).

(14) The following passages seem to us to be very revealing.
On the one hand H. Karlgren (1973, XIII and XIX-XXI) - from the puristic point of view - wrote: "The characteristic feature of Computational Linguistics is a focus on computation, on the derivation of results by a "mechanical" procedure, operating according to rules, according to an "algorithm". A good tool for computation is, in many cases, a computer, but computational linguistics is not the same as Computer-based Linguistics or Linguistic Data Processing *(Linguistische Datenverarbeitung)*. (...) Linguistic research, like investigation in so many other fields, is often aided by the services of a computer without being, on that account, directed towards problems of computation. Thus lexicographic work is neither more or less computational because the clerical part of it has become easier - or possibly more complicated - thanks to new equipment. The data processing performed in linguistic institutes of various kinds is certainly worth studying in its own right - preferably together with experts of economy, organisation and office rationalization - but does not constitute a separate branch of scientific research. Again, the distinction is often vague in practice ".
On the other hand, A. Zampolli suggested the term *automated language processing* (ALP) to indicate "all the activities, theoretical or applied, encompassing "the use of computers or computational techniques in the processing of natural language". The area of ALP contains both computational linguistics (CL) and literary and linguistic computing (indicated with the abbreviation TP, from text processing, considered as the nucleus of LLC): "CL activities, which are focused on linguistic algorithms, are principally directed towards the study of linguistic models, and in general, towards the formalization, representation, and calculus of linguistic structures. TP activities are mainly concerned with the processing of collections of language data, usually large, very often for purposes of reorganization, extraction, summarization, etc. of some linguistic elements of the text, designated at the 'surface' level, i.e. distinguished by shape or code pattern. (.....). From a theoretical point of view, it must be remembered that many research projects currently in progress in TP are aimed at extracting, from linguistic facts, data and information which constitute the primary material that must be considered in theories and models of CL. At times, information obtained on the statistical and lexical composition of specific corpora is also used in the construction of algorithms and in the choice of working strategies for systems in CL; reference can be made, for example, to the use of statistical methods in several speech understanding systems or in some projects for machine translation. From an operational point of view, typical TP procedures include some crucial operations on the texts or data which are substantially the same as some of those requested from some components of typical systems in CL. Two of the more obvious examples are morphological analysis and the distinguishing of homographs for lemmatization".

(15) For example: the Istituto di Linguistica Computazionale, Pisa (Zampolli, 1983); The Institut fur Deutsche Sprache, Mannheim; Sprakdata, Goteborg; etc.

(16) The Proceedings of this Conference (*Les Industries de la Langue, Enjeux pour l'Europe*), Tours, 28 February - 1 March 1986) are published in number 16 (1986) of the revue *"Encrages"*. In the allocution pronounced on the occasion of the 350th anniversary of the Academie Francaise, the 12 November 1985, published in the same issue, the President F. Mitterrand said: "Nous nous trouvons a un point fort important de l'histoire de notre langue: ou bien elle saura maitriser l'informatique, ou bien, en peu d'annees, elle cessera d'etre l'un des grands moyens de communication dans le monde" (Allocution ... 1986, 145).

(17) Along with the dramatic advancement of the new information technologies, the world economy is undergoing a profound transformation. "It is estimated that the traditional sectors of economic activities - agriculture and manufacturing - constitute at present no more than 40% of the total, while already 60% of the workforce are concerned with 'immaterial' activities, principally information handling. This development goes in parallel with a trend towards a worldwide concept of the economy". (Perschke, 1988).

(18) "La diversite' linguistique se situe au coeur meme de l'identite' culturelle de l'Europe. Une langue n'est pas uniquement un vehicule de communication. Elle reflete une histoire, une civilisation, un systeme de valeurs ... et, comme le disait Gramsci, elle 'contient les elements d'une conception du monde et d'une culture'" (Vidal-Beneyto, 1986, 5). "The EC and its direct competitors, Japan and the USA, are confronted with the challenge of mastering our principal information medium: natural language. For the EC this challenge is more important, as unlike Japan and the USA, its internal market is linguistically not homogeneous: there are nine official languages, and several more regional languages currently used" (Perschke, 1988). It has been suggested that the obstacle of the 'linguistic barriers' created for the European economic activities by this diversity, could ultimately produce a potential advantage, forcing the Europeans to acquire a know-how in the sector of multilingual LI activities, which could be exported to other countries, and facilitate the relationships with them.

(19) Several initiatives have already been promoted. As an example, we can quote, at a national level, the Japanese Electronic Dictionary Research Institute, set up by the Japanese Government in cooperation with 8 major Japanese electronic industries, which aims at producing national Japanese and Japanese-English lexical databases (Japanese Electronic Dictionary Research Institute, 1988), and two national strategic research projects of the Italian National Research Council (Zampolli 1987, 1989).
At the EC level, we can quote the machine translation project EUROTRA (Maegaard, 1988), several ESPRIT projects (AQUILEX, see Boguraev et al., 1988), and the activities, in the framework research programme 1987-1991, which include lexical reusability and lexical and terminological standards. The Council of Europe has set up an ′ad hoc′ programme for language industries, with four activity lines: lexica (Gross), corpora (Zampolli, Cignoni, Rossi, 1987) terminology, common European doctorate in computational linguistics.

(20) Corpora analysis will give information on linguistic phenomena occurring in real texts, and on their frequency in specific sublanguages. Discussing the role of language corpora in linguistic technology, H.S. Thompson (1989) stresses that the access to large amounts of speech or text data is essential for the development of the technologies in question, regardless of whether they are self-organising (i.e. based on neuronal nets or Markov models or other similar stochastic approaches) or not (i.e. based on explicit representational knowledge bases). The self-organising approaches require large bodies of examples of the required input and output to provide the basis for the training process. It is well known that some recent successful NLP systems are almost entirely based on statistical probabilities derived from the analysis of textual samples: parts of speech taggers (Church, 1988), corpus-oriented parsers (Hindle, 1988), speech recognizers (Brown et al., 1988). ″Since explicit knowledge-based systems for the forseeable future will be specialised for specific application domains, the ability to derive linguistic knowledge bases from a corpus of linguistic material which exemplifies and in a sense defines such a domain will be crucial. Furthermore, one can anticipate that a sensible route to the required domain specific knowledge bases will be to develop a set of reasonably broad coverage knowledge bases, which can be specialized for specific domains. Finally, even the most rigorously knowledge-based approach will often require tuning to reflect the distribution of phenomena in the targetted linguistic tasks, which once again means processing large amounts of appropriately annotated linguistic data″ (Thompson, 1989, 2).
Lexicographers, in particular historical lexicographers, have always used corpora as sources of information for the description of the properties of the lexical units, in addition to their linguistic competence. In certain cases, for example collocations and phraseology, it is with great difficulty that linguistic competence can be made explicit without the evidence supplied by corpus analysis (cf. Smadja, 1989).

(21) The Text Encoding Initiative can be considered as an answer to this need, and it seems relevant to stress that it constitutes a paradigmatic example of cooperation between various kinds of partners. The Text Encoding Initiative is a cooperative undertaking of the textual research community to formulate and disseminate guidelines for the encoding and interchange of machine-readable texts intended for literary, linguistic, historical, or other textual research. It is sponsored by the Association for Computers and the Humanities (ACH), the Association for Computational Linguistics (ACL), and the Association for Literary and Linguistic Computing (ALLC). A number of other learned societies and professional associations support the project by their participation in the Initiative′s Advisory Board. The project is funded in part by the U.S. National Endowment for the Humanities, and in part by the EC, through Pisa University.

(22) The situation, unfortunately, has not changed much since 1973, when A. Zampolli (1973, XXI-XXII) wrote:
″The fact that these operations in TP are still performed manually is partly because of the inadequacies of the components of the CL systems in analysing, in a satisfactory manner, the variety and complexity of the texts and data usually processed in TP, but it is also a result of the lack of exchange of information and collaboration among reserachers in the two fields. Those who have worked for some time in TP, however, are well aware of the fact that the development of applications according to the ′classic′ methods and techniques of the 1950s and 1960s has reached saturation point. If we continue to use current methods, according to the current rules of the game (for example: processing, at a simple graphemic level, millions of running words, in order to produce frequency counts, concordances, lexical cards, etc., without any linguistic analysis), real prospects of development do not exist. Although the speed of the computer is continually being increased and programs are becoming more sophisticated, lexicographers and linguists are not able to profit from these facts proportionally because current methodology already produces much more data than any reasonably sized team of linguists could probably analyse, working according to current procedures. If the analysing operations are left to a successive phase, this would not alleviate the problem as it is not clear how we can resolve the enormous operational difficulties which are due to the sheer quantity of the documentation and material gathered″.

References

Actes du Colloque International sur la Mecanisation des Recherches Lexicologiques, Besancon, *Cahiers de Lexicologie*, 3, 1961.

Allocution prononcee par M. Francois Mitterrand, President de la Republique, lors de la seance solomnelle a' l'Academie Francaise a' l'occasion du 350me Anniversaire de l'Institut, *Encrages*, 16(1986), 144-147.

Ahlswede, T., Evens, M., Parsing vs. Text Processing in the Analysis of Dictionary Definitions, *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, Buffalo, New York, 1988, 217-224.

Almanacco Letterario Bompiani 1961, Milano, 1961.

ALPAC Report, Automatic Language Processing, Advisory Committee, Language and Machine-Computers in Translation and Linguistics, Washington, 1966.

Alshawi, H., Analyzing the Dictionary Definitions, in B. Boguraev, E. Briscoe (eds.), 1989, 153-170.

Amsler, R. A., A Taxonomy for English Nouns and Verbs, *Proceedings of the 19th Annual Meeting of the Association for Computational Linguistics*, Stanford, California, 1981, 133-138.

Atkins B.T., The Uses of Large Text Databases, Semantic ID Tags: Corpus Evidence for Dictionary Senses, *Third Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary*, Waterloo, Canada, 1987, 17-36.

Atkins, B.T., Kegl, J., Levin, B., Explicit and Implicit Information in Dictionaries, in *Proceedings of the Conference on Advances in Lexicology*, Waterloo, 1986.

Bindi, R., Calzolari, N., Statistical analysis of a large textual Italian Corpus in search of lexical information, presented for *EURALEX 1990*, Malaga, forthcoming.

Boguraev, B., Briscoe E.J. (eds.), *Computational Lexicography for Natural Language Processing*, Longman, London, 1989.

Boguraev, B., Briscoe, E.J., Calzolari, N., Cater, A., Meijs, W., Zampolli, A., Acquisition of Lexical Knowledge for Natural Language Processing Systems, (AQUILEX), Technical Annex, ESPRIT Basic Research Action No. 3030, Cambridge, 1988.

Boguraev, B., Byrd. R., Klavans, J., Neff, M., From structural analysis of lexical resources to semantics in a Lexical Knowledge Base, in *Proceedings of the First International Lexical Acquisition Workshop*. Detroit (Michigan), 1989.

Booth, A.D., Cleave, J.P., Brandwood, B.A., *Mechanical Resolution of Linguistic Problems*, London, 1958.

Brown, P., Cocke, J., Della Pietra, S., Della Pietra, V., Jelinek, F., Mercer, R., Roossin, P., A Statistical Approach to Language Translation, *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, 1988.

Busa, R., *Sancti Thomae Aquinitatis Hymnorum Ritualium Varia Specimina Concordantiarum*, Milano, 1951.

Busa, R., L'evoluzione linguistica dei mezzi di informazione, in *Almanacco Letterario Bompiani 1961*, Milano, 1961, 103-117.

Busa, R., Actes du Seminaire International sur le dictionnaire latin de machine, *Calcolo*, Supplemento n. 2 al vol. V., 1968.

Byrd, R.J., Discovering Relationships among Word Senses, *Dictionaries in the Electronic Age*, Fifth Annual Conference of the University of Waterloo Centre for the New Oxford English Dictionary, Oxford, 1989.

Byrd, R.J., Calzolari, N., Chodorow, M., Klavans, J., Neff, M., Rizk, O., Tools and Methods for Computational Lexicology, *Computational Linguistics*, 1987, vol. 13(3-4), 219-240.

Calzolari, N., Towards the organization of lexical definitions on a data base structure, *COLING82*, ed. by E. Hajicova, Prague, Charles University, 1982, pp.61-64.

Calzolari, N., Detecting Patterns in a Lexical Database, *Proceedings of the 10th International Conference on Computational Linguistics*, Stanford, California, 1984, 170-173.

Calzolari,N., The dictionary and the thesaurus can be combined, in *Relational Models of the Lexicon* , (Studies in Natural Language Processing series), ed. by M.Evens, Cambridge (Mass.), Cambridge University Press, 1988, 75-96.

Calzolari,N., Lexical Databases and Text Corpora: perspectives of integration for a Lexical Knowledge Base, in *Proceedings of the First International Lexical Acquisition Workshop*. Detroit (Michigan), 1989a, n.28.

Calzolari, N., Computer-aided lexicography: dictionaries and word databases, *Computational Linguistics*, edited by I.S. Batori, W. Lenders, W. Putschke, Berlin: Walter de Gruyter, 1989b, 510-519.

Calzolari, N., Structure and Access in an automated Lexicon and Related Issues, in D. Walker, A.Zampolli, N.Calzolari (eds.), forthcoming.

Calzolari, N., Picchi, E., A Project for a Bilingual Lexical Database System, *Advances in Lexicology, Second Annual Conference of the UW Centre for the New Oxford English Dictionary*, Waterloo, Ontario, 1986, 79-92.

Calzolari, N., Picchi, E., Acquisition of Semantic Information from an On-Line Dictionary, *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, 1988, 87-92.

Calzolari,N., E.Picchi, A.Zampolli, The use of computers in lexicography and lexicology, in *The Dictionary and the Language Learner*, ed. by A.Cowie, Lexicographica Series Maior 17, Tubingen, Niemayer, 1987, 55-77.

Chodorow, M.S., Byrd, R.J., Heidorn, G.E., Extracting Semantic Hierarchies from a Large On-line Dictionary, *Proceedings of the Association for Computational Linguistics*, Chicago, Illinois, 1985, 299-304.

Church, K.W., A Stochastic parts program and noun phrase parser for unrestricted text, *ACL, Second Conference on Applied Natural Language Processing*, 1988, 136-143.

Church, K., Hanks, P., Word Association Norms, Mutual Information and Lexicography, *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, British Columbia, 1989, 76-83.

Cumming, S., The Lexicon in Text Generation, in D. Walker, A.Zampolli, N.Calzolari (eds.), forthcoming.

Fox, E., Nutter, T., Ahlswede, T., Evens, M., Markowitz, J., Building a Large Thesaurus for Information Retrieval, *Proceedings of the Second Conference on Applied Natural Language Processing*, Austin, Texas, 1988, 101-108.

Goetschalckx, J., Rolling, L. (eds.), *Lexicography in the Electronic Age*, Amsterdam, North-Holland, 1982.

Gruppo di Pisa, Il Dizionario di Macchina dell'Italiano, in *Linguaggi e Formalizzazioni*, ed. by Gambarara, D., Lo Piparo, F., Ruggiero, G., Roma, Bulzoni, 1979, pp.683-707.

Hays, D.G., *Computational Linguistics: Introduction*, in Meetham and Hudson (eds.), 1969, 49-51.

Hays, D.G., *The Field and Scope of Computational Linguistics: Introduction*, in Papp and Szepe (eds.), 1976, 21-26.

Hindle, D., Acquiring Disambiguation Rules from Text, *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Morristown (NJ), 1988, 118-125.

Ingria, R., Lexical Information for parsing Systems: Points of Convergence and Divergence, in D. Walker, A.Zampolli, N.Calzolari (eds.), forthcoming.

Kay, M., The Dictionary of the Future and the Future of the Dictionary, in Zampolli, Cappelli (eds.), 1983, pp.161-174.

Japanese Electronic Dictionary Research Institute, *Electronic Dictionary Project*, Tokyo, 1988.

Locke, W.N., Booth, A.D., *Machine Translation of Languages*, MIT Press, 1955.

Katz, B., Levin, B., Exploiting Lexical Regularities in Designing Natural Language Systems, *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, 1988, 316-323.

Klavans, J.L., Building a Computational Lexicon using Machine Readable Dictionaries, paper presented at the Third Congress of the European Association for Lexicography, Budapest, 1988.

Kucera, H., Francis, W.N., *Computational Analysis of Present-Day American English*, Brown University Press, Providence, Rhode Island, 1967.

Maegaard, B., EUROTRA, The Machine Translation Project of the European Communities, *Literary and Linguistic Computing*, 3, no. 2, 1988, 61-65.

Meetham, A.R., Hudson, R.A., *Encyclopaedia of Linguistics, Information and Control*, Pergamon Press, 1969.

Nagao, M., *Machine Translation - How far can it go?*, OUP, 1989.

Nagao, M., Nakamura, J., Extraction of Semantic Information from an Ordinary English Dictionary and its Evaluation, *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, 1988, 459-464.

Neff, M., Boguraev, B., Dictionaries, Dictionary Grammars and Dictionary Entry Parsing, *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, British Columbia, 1989, 91-101.

Papp, F. Szepe, G. (eds.), *Papers in Computational Linguistics, Proceedings of the 3rd International Meeting on Computational Linguistics*, 1976.

Perschke, S., Hearing on the language industry in the European Community. Questions put to the participants. (Background Paper for Discussion), 1988.

Picchi,E., N.Calzolari, Textual perspectives through an automatized lexicon, in *Methodes quantitatives et informatiques dans l'etude des textes.* Geneve: Slatkine, 1986, 705-715.

Picchi,E., C.Peters, N.Calzolari, A tool for the second language learner: organizing bilingual dictionary data in an interactive workstation, in *Proceedings of the XX ALLC Conference,* Jerusalem, 1988, forthcoming.

Pustejovsky, J., Current Issues in Computational Lexical Semantics, Invited Lecture, *Proceedings of the Fourth Conference of the European Chapter of the ACL*, Manchester, England, 1989, xvii-xxv.

Quemada, B., Introduction, *Actes du Colloque International sur la Mecanisation des Recherches Lexicologiques*, Besancon, 1961, 13-18.

Smadja, F., Macrocoding the Lexicon with Co-occurrence Knowledge, paper presented at the First Lexical Acquisition Workshop, Detroit, 1989.

Smith, J., Ideals versus Practicalities in Linguistic Data Processing, in A. Zampolli, N. Calzolari (eds.), 1973, 895-8.

Table Ronde sur les grandes dictionnaires historiques, Firenze, 1973.

Talmy, L., Lexicalization Patterns: Semantic Structure in Lexical Forms, in T. Shopen (ed.), *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon*, Cambridge University Press, Cambridge, 1985.

Thompson, H., Linguistic Corpora for the Language Industry (Background paper), 1989.

Van der Steen, G.J., A Treatment of Queries in Large Text Corpora, in S. Johansson (ed.), *Computer Corpora in English Language Research*, Norwegian Computing Centre for the Humanities, Bergen, 1982, 49-65.

Vidal-Beneyto J., Presentation, *Encrages*, 16(1986), 15-7.

Vauquois, B., *La Traduction Automatique a' Grenoble*, Paris, 1975.

Vossen, P., Meijs, W., den Broeder, M., Meaning and Structure in Dictionary Definitions, in B. Boguraev and E. Briscoe (eds.), 1989, 171-192.

Walker, D., Zampolli, A., Foreword, in B. Boguraev, T. Briscoe (eds.), 1989, xiii-xiv.

Walker,D., A.Zampolli, N.Calzolari (eds.), *Towards a polytheoretical lexical database.* Pisa: ILC, 1987.

Walker, D., A.Zampolli, N.Calzolari (eds.), Special Issue of the *Journal of Computational Linguistics*, 13(1987)3-4, 193.

Walker, D., Zampolli, A., Calzolari, N. (eds.), *Automating the Lexicon: Research and Practice in a Multilingual Environment*, OUP, forthcoming.

Webster, M., M. Marcus, Automatic acquisition of the lexical semantics of verbs from sentence frames, in *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, Vancouver, British Columbia, 1989, 177-184.

Whitelock, P., Wood, M., Somers, H., Johnson, R., Bennett, P. (eds.), *Linguistic Theory and Computer Applications*, Academic Press, New York, 1987.

Wilks, Y., Fass, D, Guo, C.-M., McDonald J., Plate, T., Slator, B., A Tractable Machine Dictionary as a Resource for Computational Semantics, in B. Boguraev and E. Briscoe (eds.), 1989, 193-228.

Zampolli, A., Projet pour un lexique electronique de l'italien, in Busa (ed.), 1968, 109-26.

Zampolli, A., Lexicological and Lexicographical Activities at the Istituto di Linguistica Computazionale, in Zampolli, Cappelli (eds.), 1983, pp.237-278.

Zampolli, A., Multifunctional Lexical Databases, *Encrages*, 16(1986), 56-65.

Zampolli, A., Progetto Strategico "Metodi e strumenti per l'industria delle lingue nella cooperazione internazionale", Pisa, 1987.

Zampolli, A., Progetto Speciale "Aquisizione di una base di conoscenze lessicali per il trattamento automatico dell'Italiano: obiettivi nazionali e cooperazione internazionale", Pisa, 1989.

Zampolli, A., Calzolari, N., (eds.), *Computational and Mathematical Linguistics, Proceedings of the International Conference on Computational Linguistics 1973*, 2 Volumes, Firenze, 1973 and 1977.

Zampolli, A., Calzolari, N., Computational Lexicography and Lexicology, *AILA Bulletin*, 1985, 59-78.

Zampolli, A., Cappelli, A., (eds.), The Possibilities and Limits of the Computer in producing and publishing Dictionaries, *Linguistica Computazionale*, Pisa, III, 1983.

Zampolli, A., Cignoni, L., Rossi, S., Problems of Textual Corpora, ILC-9-2, Pisa, 1985.