International Society for Knowledge Organization

Documentary Languages and Databases

Antonio Zampolli - Nicoletta Calzolari

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

LINGUISTIC TOOLS FOR INFORMATION RETRIEVAL

1. The activities of the Institute for Computational Linguistics (Istituto di Linguistica Computazionale) within the framework of the project "Transfer of the Technologies of the Mission Oriented Projects" (PF)

1.1 The participation of the Institute for Computational Linguistics of the National Research Council (CNR) in the Project "Transfer of the Technologies of the Mission Oriented Projects", mainly consists of studying and experimenting with the use of computational linguistics methods and tools as an aid to the operations of access to the information contained in natural language texts. Methods of this type could be applied to the textual parts of the documents which constitute the database of the **PF**.
These methods and tools should essentially make it possible to work on texts for two distinct, although complementary, purposes:

- to identify linguistic units of different levels and their syntagmatic relations, associating them with the knowledge on linguistic properties, paradigmatic relations, conceptual structures, etc., so as to make them available to research algorithms and information retrieval;
- to study, analyze, describe different qualitative and quantitative aspects of the sublanguages of the documents, to facilitate the construction of the tools and the improvement of linguistic methods on the basis of the specific features of each sublanguage.

The following are some of the methods we intend to use:

- Morphosyntactic analysis

Purpose of the analysis is to allow the user to "explore" the words he is looking for in the texts, so that the system can also check all the morphologically connected words, whether belonging to the same lemma (inflections, conjugations), or connected through derivation mechanisms.The analyst should also define, through appropriate rules, and/or statistical methods based on the analysis of textual corpora, some classes of grammatical and lexical ambiguity, and suggest the treatment of neologisms created by more productive mechanisms in the sectorial terminologies (suffixes, affixoids, compounds, etc.).

- Analysis of the paradigmatic relationships

The researches currently underway seem to show that by using semiautomatic procedures of analysis, it is possible to draw from the definitions of the dictionaries available in machine-readable form, semantic relations which structure and link in various ways the elements of the lexicon. For instance, it is possible to obtain conceptual taxonomies which reorganize the lexicon in the form of a "thesaurus". Other kinds of relations can be drawn from the analysis of textual corpora. Some experiments have been carried out which show the usefulness of using semantic relations thus constructed in the interrogation of different types of texts: literary, journalistic, political, etc. The systen practically "expands" the words the user wants to look up in the texts, searchinc in the data base lexical units connected by semantic relations of various type (synonyms, hyperonyms, etc.). In perspective, the relations extracted fror lexicons and corpora could meet in a knowledge base to which specific methods anc tools of the sector "knowledge representation" could be applied.

- Multilingual aspects of the querying

We also intend to explore the possibility of creating multilingual tools whicl will simplify the access to the texts for languages users of other languages This part of the project however has not yet been started.

1.2 We describe here briefly some recent trends of computational linguistics an some international projects, in which our Institute is participating, whic intend to create basic linguistic resources, suitable to the needs of th treatment of "real world texts", such as those which appear in the data bank o the National Research Council (CNR).
The availability of such resources appears today as an essential condition fo the development of applied systems based on the treatment of natural language (NLP).
It is a well-known fact that the majority of the current transformations, at political, social, economic, industrial level, is partly cause and partly effec of the growing flow of multilingual information of the "information society which characterizes the so-called "global village".Natural languages are stil privileged vehicles to produce, codify, transmit, retrieve the informatior Furthermore, the information is at present, in large part produced, memorizec distributed through computers and computer networks. The full quantity c information and the need for communicating rapidly and efficiently are such thε it seems natural to try to use the computers to reduce the problem to reasonable dimension and to limit the costs, which are very high, of the humε operations required for the production and processing of the texts.
It has been claimed that, if the problem of natural language processing is nc solved, the opportunity of the evolution of our society could be reduced c compromised.
Natural language processing, potentially, concerns not only the variou "traditional linguistic professions" (such as translation, documentatioı lexicography, publishing (editing), language teaching, etc.), but also new typ of relevant applications: production of documents, machines for dictation, ma machine interaction, electronic-mail, vocal input for restricted domain: increased communication for the handicapped, etc.
Many of these applications, in particular in Europe, must include multilingu functions, in order to assist the communication between speakers of differe languages: generation of multilingual messages, production of documents several languages, multilingual access to databases, automatic classification electronic mail, etc.
The capability of including linguistic components to produce, memorize, acce the information represented in natural language, is critical for a large numb of industrial systems. The potential market is on the whole quite relevan However, systems and applications can be produced only if an appropriate "kno how" and linguistic "technology" are available.
The multilingual character of Europe constitutes an additional dimension to t complexity of the problem. Multilingualism can create difficulties in t development of the European Common Market. It has been observed, however, th the impulse to develop a technology able to overcome the linguistic differenc could give Europe a specific know-how superior vis-à-vis the other two gre

economic blocks, USA and Japan.

The multilinguistic problems, the evident potentialities of the market of the language industry, some strategic implications of natural language processing, have raised, in the last period, an increasing interest by the main international organizations (EEC, Council of Europe) and by several national institutions (DARPA, NSF, ICOT, CNRS, DCI,etc.). These organizations, with the cooperation of the main scientific Associations (ACL, ALLC, ACH, etc.), have assigned, in the availability of the linguistic resources for NLP, a priority need to developing capabilities in this sector, and have promoted activities and projects for building such resources. The Institute for Computational Linguistics, with the cooperation of the Linguistic Department of the Pisa University, has played an important role in conceiving, planning and coordinating these activities at the international level.

We feel it is necessary to underline that, in the framework of the investigations concerning the possibility of starting a cooperation between DARPA and the various research activities of the EEC, the subject of linguistic resources has been considered as the most important. In 1991 we shall organize a workshop on this subject in Pisa, which will bring together representatives of NSF, DARPA, EEC.

We shall describe briefly here, as example, some projects whose results could be interesting for the realization of information retrieval systems applied to large textual databases.

2. Creation of lexical databases for NLP

2.1 Computational lexicography and lexicology and the concept of "Reusability"

The main purpose today is to create a large "repository" of linguistic knowledge, in the form of reusable linguistic descriptions, the most complete possible, structured in a vast "Lexical Knowledge Base" (LKB) or in different types of interconnected linguistic bases (grammatical, lexical, textual, knowledge bases). Given the request of large scale NLP systems, 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, 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 practicalNLP 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 characteristic 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 is investigating 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 is working on various languages (see Walker, Zampolli, Calzolari 1987).
These researches, on the one side, are being developed with the cooperation of the Universities of Stanford, MIT, Princeton, Cambridge, Heidelberg, Pisa, and research groups of IBM USA, BBN, Bell Research Lab; on the other, they have originated a feasibility study promoted by the EEC, in which we are participating.

2.2 Reusability of preexisting data in the form of MRDs

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).
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 1988).
The methodological approach consists in converting and reorganizing the definitions into informationally equivalent structured formats made up by nodes and relations linking them.
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 3.2.4 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 also devoted to the formalization 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 particular 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 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 is obtained by making a similar search on an English dictionary. This shows that there is 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 the 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 also related to 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 lexiçographers: *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.
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 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 for 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 the implementation of a part of a LKB.

2.3 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 an
These dictionaries can be processed with a twofold purpose, as on the o
they too are a source of interesting 'monolingual' information, on the otł
they are obviously exploited as a source of links between two monolingu
(see Calzolari, Picchi 1986, and Picchi, Peters, Calzolari 1990).
One of the objectives is to integrate the different types of info
traditionally contained in monolingual and bilingual dictionaries, sc
expand the informational content of the single components in the new int
system. Bilingual dictionaries contain more information about examples of
fixed expressions or idioms. This kind of information can obviously ł
integrated in the monolingual dictionary, and can also be made easy to
We can envisage the original monolingual lexical entries, augmented w:
different types of information coming from the corresponding bilingual
different sense discriminations, other examples,syntactic infor
collocations, idioms, etc. We can also reverse the perspective, and look
bilingual entries supplied with the information traditionally contai
monolingual entries: mostly definitions.One of the two different viewpoint
virtually present in the integrated bilingual system, will be simply ac
and made available to the user by the first manner of access to the
bilingual lexical data base. We would like therefore to maintain in a
structure both the independent features of the monolingual and bilingual
dictionaries and the integration of the two with different views on the
The overall picture of the bilingual LDB system we have in mind is sketc
Fig. 16. Also with regard to bilingual dictionaries, the method we are ac
consists of reusing available data in machine-readable form by analyzi
transforming the information already contained in common dictionari
procedure of processing the bilingual MRD is rather similar to the one ou
above for monolingual dictionaries (i.e. parsing of the lexical entry, de
a new structure, computational reorganization, etc.). After this prelimina
comes out again the utility of browsing the bilingual LDB, taking advant
the structural elements already formalized in the LDB, with the purp
discovering properties and structures not immediately visible in the p
dictionary, but useful for further exploitation in the computational dicti
After the first processing phases that we have envisaged on the bil
dictionary data, it will make no difference which of the two languages are
as a starting point. In a certain sense, we would no longer have a
language and a target language, since the look-up and access procedur
independent and neutral with respect to direction (the object b
bidirectional). Bidirectional cross-references will also be automat
generated for the information contained at each sense level as se
indicators, i.e. synonyms/hyperonyms or contextual indicators.
Another possibility is the use of the monolingual lexical data base as a t
expand the information given as a single word to the whole set of words to
it actually refers. For example, the entry *vivido* has different transl
according to the contextual indicators referring to the subject (in brac

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

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

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 of either structural or 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-automated.

The end result may 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 databases 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.

3.3. Textual Reference Corpora

3.1. Textual Corpora and NLP Systems

The main reason for constructing and analyzing a textual corpus can be summarized and simplified as follows.

In order to describe a language, it is impossible to enter all the texts produced in a specific period ("population"). For this reason, a collection of texts appropriately chosen (corpus) is analyzed, and these texts are considered as a "representative sample", with the expectation of finding, in others of the same population, the same events, behaviour, distribution observed in the sample texts.

An NLP system, whose purpose is to work effectively on the written texts of a language, must be based on the evidence how this language is used in real texts.

The analysis of "representative" corpora is an irreplaceable means for obtaining this evidence. In particular, recent experimental work in the fields of the understanding of spoken language, information retrieval, classification of strategic messages, has shown that it is very useful - and perhaps necessary - to draw advantage from the features and specific properties of the various "sublanguages": different uses of the same language in different communicative contexts, with different information channels, to direct the information towards specific domains, etc. The differences between various sublanguages, concerning the variety and distribution of several classes of linguistic phenomena, can be expoited to reduce the number of linguistic situations which cannot "be treated" automatically, thus improving the efficiency of the systems, with increased satisfaction on the part of the users, and increasing the number of possible applications. It is possible, for example, to reduce the number of syntactic phenomena, or the selection restrictions concerning the arguments of the verbs, in order to reduce the number of ambiguities to be solved in a text.
The analysis of suitable texts is the only means known so far to describe the different sublanguages. Textual corpora are very useful for the contrastive description of different languages, and to define methods and construct systems for the assessment of components and systems for NLP.
Some of the recent and most successful NLP systems are essentially based on the use of statistical methods: for example, the Markovian models which are constructed by introducing probabilities derived from the study of frequencies in textual corpora. Some examples of extraction of knowledge from textual corpora will be briefly discussed below, in 3.2.4.
A number of international (EEC, Council of Europe) and national (DCI, DARPA, ERDI, etc.) organizations which acknowledge the creation of textual corpora as a priority need for the development of NLP applications, have launched projects of different types.
Let us describe briefly an initiative of the EEC in which we are now participating.

3.2 Project for a European Network of Reference Corpora

Following a proposal launched by our Institute, the EEC has promoted a study whose aim is to define the methods for the creation of a European Network of textual Reference Corpora (NERC). The study, entrusted to six European Institues, is coordinated by the University of Pisa and the Institute for Computational Linguistics from both a scientific and organizational point of view. The program of work is divided as follows.

3.2.1 Typology and composition of the corpus

When collecting the first corpora, even before the use of the computer, the authors have always tried to define a set of parameters and conditions for the choise of textual material to be included in the sample corpus, in order to ensure the maximum representativity, with respect to the "population", unaccessible in its totality. In particular, the following were discussed:

stratification of the corpus: how many and which subsets (sublanguages, text types, subjects, modes of communication, etc.), should be identified in the population and in what proportion they should be included in the corpus;

dimension of the corpus: what is the minimum dimension of the corpus and of its subsets which can ensure a suitable representativity;

sampling criteria: how many texts should be included for each subset and which criteria should be used to choose them; what can be the minimal textual units: sentences, paragraphs, chapters, whole texts, etc.
The aim of our study is to give an answer to these problems, in order to ensure on the one hand, that the corpora constructed for the different European languages are homogeneous and are comparable; on the other hand, to ensure that they are suitable to the needs of NLP.
Different possibilities will be examined and compared for this purpose, e.g.:
- to create an organizational structure for the constant collection of texts in MRF, thereby creating continually updated national archives. The texts would be collected mainly taking into account their availability. The archive would be at the disposal of researchers who should use the texts for their research whenever necessary.
- To define a typology of sublanguages; to assess the priorities in terms of possible applications which are specific to NLP systems; to consequently promote the construction of independent specialized corpora.
- To establish a design, based on scientific criteria, for a general, multifunctional, balanced corpus for each language.The development of the corpus could start with the most important subsets, within a general framework, or otherwise constructing initially a balanced multifunctional nucleus of suitable dimension. This nucleus could serve as a reference to assess progressively the composition criteria and to guide the extension of the corpus. It could be used to test methods, procedures, programs. It could also provide a first significant set of textual material for a first reply to the needs of different categories of users.

3.2.2 Harmonization of the linguistic annotation of the texts

The majority of corpora available or underway contains no linguistic analysis ("raw texts"). A few corpora contain the indication of the morphosyntactic classification of the words (parts-of-speech and inflexional categories: "tagged texts"). Large corpora marked at a syntactic and semantic level ("parsed texts") are practically absent. This situation is certainly not due to the lack of potential users, who are a large majority. It is due to:
- the prohibitive cost of a completely manual analysis;
- the inadequacy of the parsers constructed so far, which are not sufficiently "robust", for the treatment of real corpora;
- the lack of common schemes of analysis accepted by the different types of possible users.

The aim of our study is, first of all, to define the feasibility of a common annotation scheme.
In the scientific community, there are two clarly distinct positions. Some researchers think that a common annotation scheme would be unable to satisfy the needs of different users, and that a "neutral" scheme with respect to the different theories would be impossible. Consequently, they suggest that one should concentrate on the creation of flexible, semiautomatic procedures of analysis, leaving it to each single researcher to decide his own specific scheme of analysis through the definition of the rules of the parsers.
Other researchers on the other hand feel that it is necessary to try to define a scheme of common annotation, and to apply it to the annotation of appropriately chosen corpora by using procedures aimed at optimizing the manual interventions which are unavoidable.
Our study will try to examine whether, up to which point and for which linguistic levels it is possible to design a multifunctional scheme of analysis, so that the different categories of users can derive from the common markup, at least in

part, the linguistic information they need, using appropriate interfaces.This requires a comparative analysis of the presently used schemes, both in the corpora and in the different NLP systems, and of the schemes proposed more or less explicitely, by the different linguistic theories. This also requires the identification of the specific needs of the diffeent categories of potential users.

3.2.3 Software Design and common procedures

The possibility of sharing among the different institutes the construction of the software necessary for the creation, handling, access, analysis, processing, distribution of corpora, is important considering that, whereas for some procedures it is essentially a matter of optimizing and standardizing already known methods, for other functions a considerable research effort is necessary. This is the case, for example, of the construction of "robust" parsers, able to analyse the variety of phenomena which appear in real texts, using large lexicons and grammars including widespread linguistic subsets. These parsers should also be able to:

- "failing gracefully", i.e. still operating in those cases in which they cannot obtain the desired level of analysis, but still providing results at an inferior level, and resorting to interactive human aid when necessary;
- exploiting the specificity of the different sublanguages: limitation of the vocabulary, reduced syntax (in terms of complexity and extension), use of patterns of specific selection restriction, etc.
- using, and possibly combining, traditional parsing methods based on grammar rules, and probabilistic systems, based on transition frequencies among adjacent categories and structures.

A very important task is that of planning, constructing, experimenting methods for the extraction of various types of knowledge from corpora.

3.2.4 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).

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 need 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 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 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 1990).

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 already exist many systems (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 supported by the Council of Europe that we are carrying out in a multilingual context.

3.2.5 Legal and organizational aspects

In order to prepare the creation of a European Network, our project also aims at clarifying the following aspects:

- legal problems connected with the inclusion in the corpus of texts protected by copyright;
- protection of the rights derived from the "added" value to the text by operations of memorization, structurization, analysis, etc.;
- assessment of the cost required by the different alternatives and phases of work;
- identification of possible sources of textual material in machine-readable form, and evaluation of their utilizability at both the technical and juridical level;
- identification of possible partners and conditions for their involvement (national authorities, industries, research agencies, etc.);
- description and classification of potential users;
- organizational scenarios for the periodical updating of the corpora, and for the management of services.

4. International Initiatives as a support to the creation of Linguistic Resources for NLP

4.1 Survey of the linguistic resources in machine readable form

All projects aimed at studying the feasibility of suitable linguistic resources for the different languages must take into account the already existing resources and the possibility of reutilising through their total or partial inclusion. Recognizing this need, A. Zampolli and D. Walker had promoted a survey (distributed, with the aid of the EEC, to the members of more than 20 scientific and professional associations and to the industries of this sector), which aims at the creation of a database of the resources already available, or underway, with regard to:

- collections of texts and textual corpora
- computer readable dictionaries
- lexical knowledge bases for NLP
- terminological data banks
- oral data bases for the processing of speech.

This information is essentially concerned with:

- the nature, the composition, the source of the data
- the representation system
- the acquisition system
- possible preediting interventions
- types of uses and users the data are created for
- level of analysis and annotation systems
- software for management, access, processing
- conditions for the utilization of the data by the researchers and/or the industries.

4.2 Towards an international standard: the "Text Encoding Initiative"

The possibility of exchanging the corpora among the various centers of the European Network, of distributing the text to the external users of the network, to share the cost of the researches and of the implementation of specialized software for the access and processing of the corpora, requires, as a necessary condition, that the textual material is represented in "machine readable form" according to a common encoding scheme. In order to satisfy this need, the NERC project has decided to use the standard proposed by the so called "Text Encoding Initiative". The use of computers for the study of texts has spread among the various classical disciplines (philology, (the history of) literature, lexicography, philosophy, anthropology, history, etc.) from the early '50s. In all these years the scientific communities have been unable to develop common schemes for the "mark-up" of "machine readable texts", and the situation has been described as a "virtual chaos".

The exchange of texts and their processing by common software are very difficult, while the recent technological developments and the widespread diffusion of computer tools promise to increase by the order of one the number of texts available in machine readable form.

In this situation the three major scientific associations of the sector (ACH, ACL, ALLC) have promoted the "Text Encoding Initiative", an international project aimed at formulating and diffusing the guidelines for the encoding and exchange of texts in MRF.

The project is promoted and directed by a Steering Committee formed by two representatives of each of the promoting Associations. Four committees, composed in equal part by European and North-American researchers, are responsible, respectively, for:

- Defining the metalanguage to be used in the mark-up of texts. The SGML was chosen for this purpose, also in consideration of the analogous choice made by the major international and national bodies, among which the American Publishers' Association, which cooperate in the TEI.
- Studying and defining the standards relative to the documentation of texts in MRF. This documentation includes a variety of information, ranging from the traditional bibliographical references, to the specification of interventions made in the texts during the preediting phase, to the choice of the coded textual elements, etc.
- Identifying the textual elements which can appear in the different types of texts in the various languages and are represented in the typographical tradition, describing them in their structure and function, and proposing a set of standardized "tags".
- Studying the most frequent types of analysis performed to enrich the texts with tags of various types (linguistic, literary, etc.), identyfing the descriptive categories used, and proposing a common representation system which takes into account the structures and concurrent levels of description.

Thirty among the most important international scientific and professional associations, organized in an advisory board, have engaged to promote among their members the Guidelines produced by the TEI. The project is financed by the National Endowment for the Humanities for the American participation, while the European participation is financed by the EEC through the coordination of the University of Pisa and our Institute. The first version of the Guidelines appeared in June 1990. The final version is envisaged by July 1992.

```
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
```

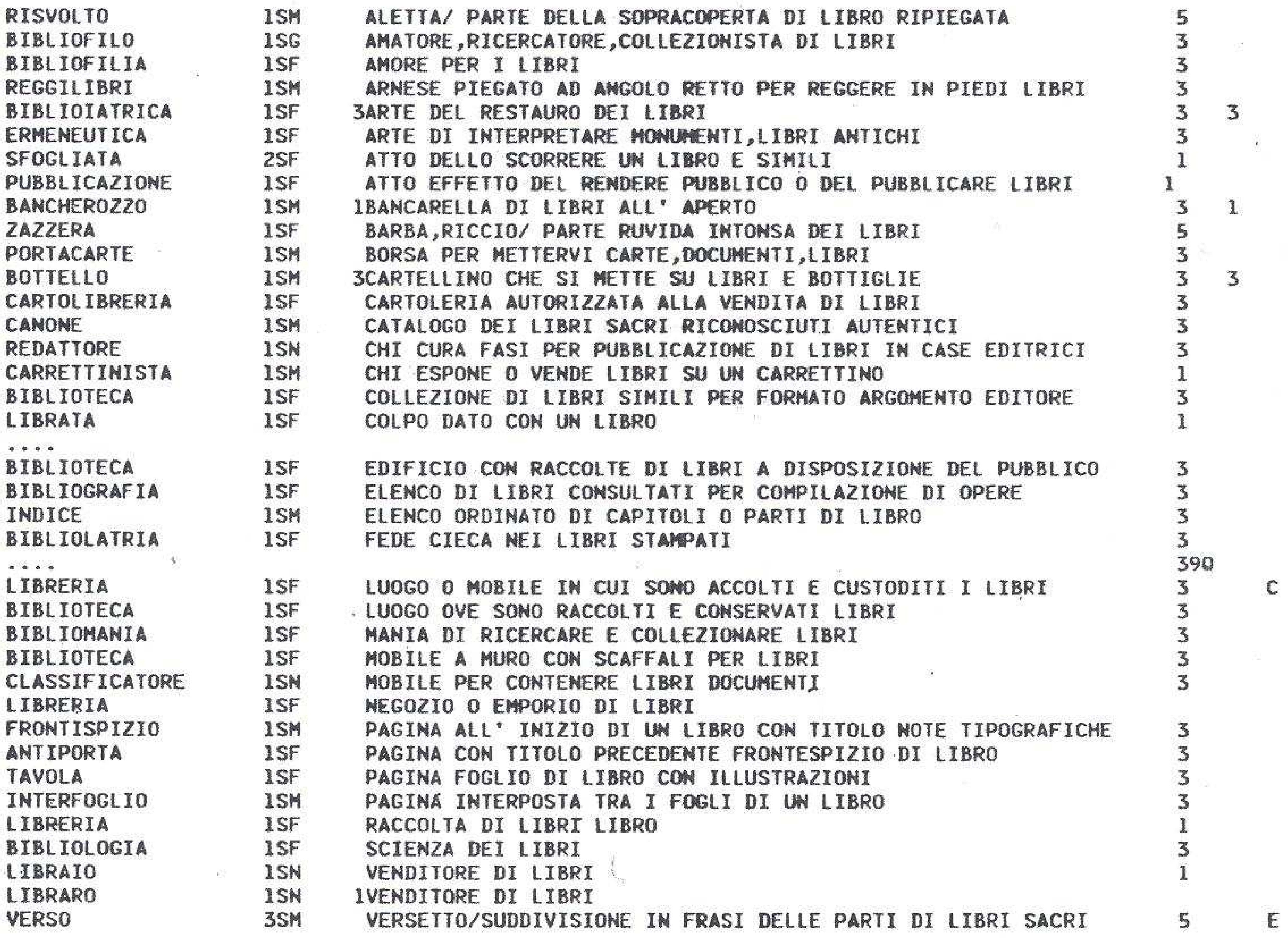
Fig. 1. Some of the hyponyms of *libro* (book).

```
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).

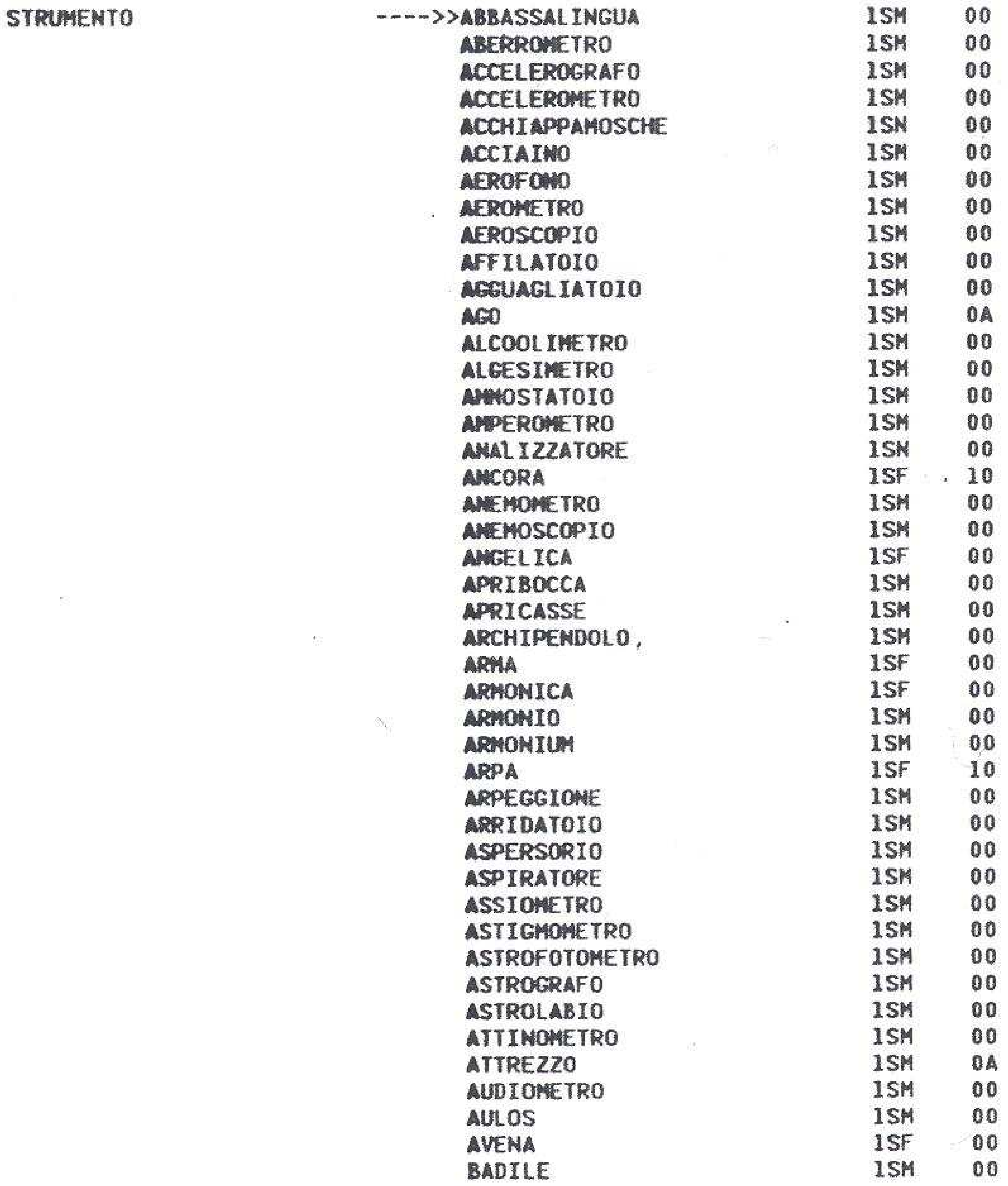
```
ADESPOTA            1A      3ANONIMO/DETTO DI LIBRO,CODICE,MANOSCRITTO DI AUTORE IGNOTO    5
ADESPOTO            1A       ANONIMO/DETTO DI LIBRO,CODICE,MANOSCRITTO DI AUTORE IGNOTO    5
APOCRIFO            1A       DETTO DI LIBRO NON RICONOSCIUTO COME CANONICO                 3
CARTOLIBRARIO       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
....                                                                                     390
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       1SM    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       2SM    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
CUÇITRICE         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.

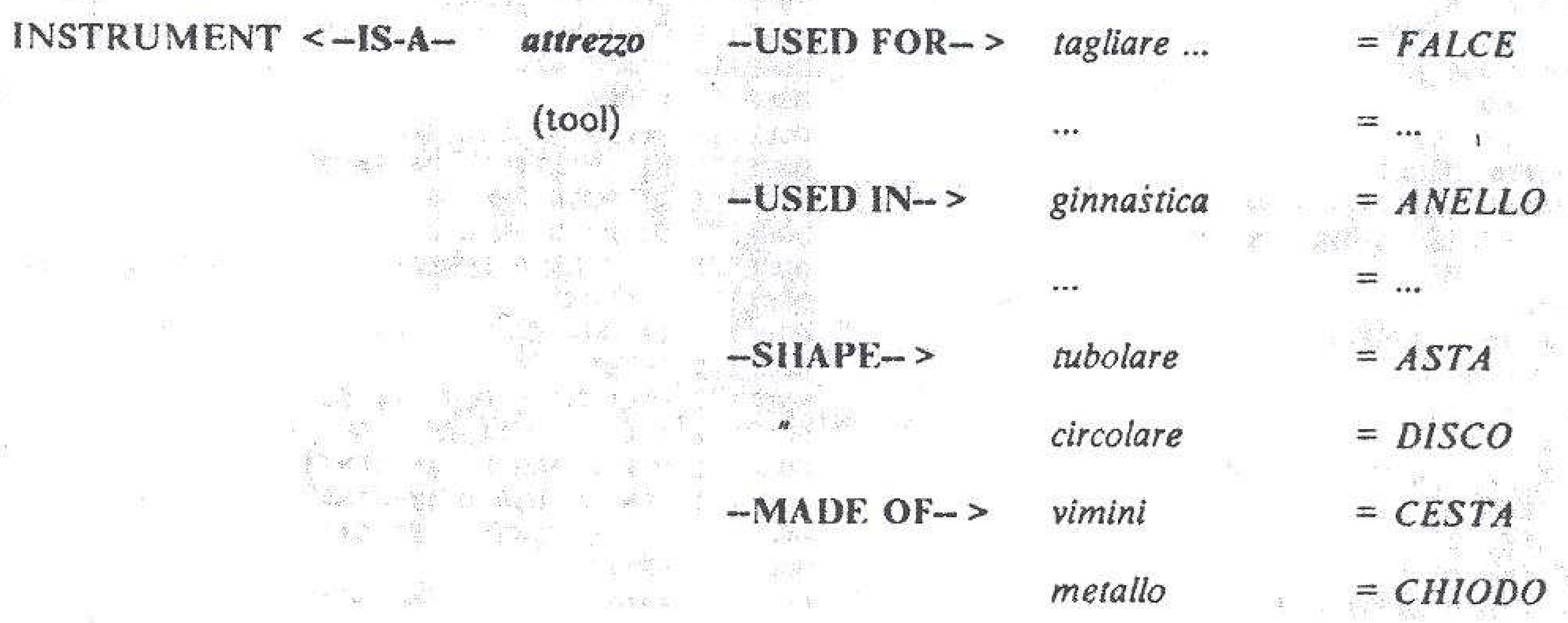


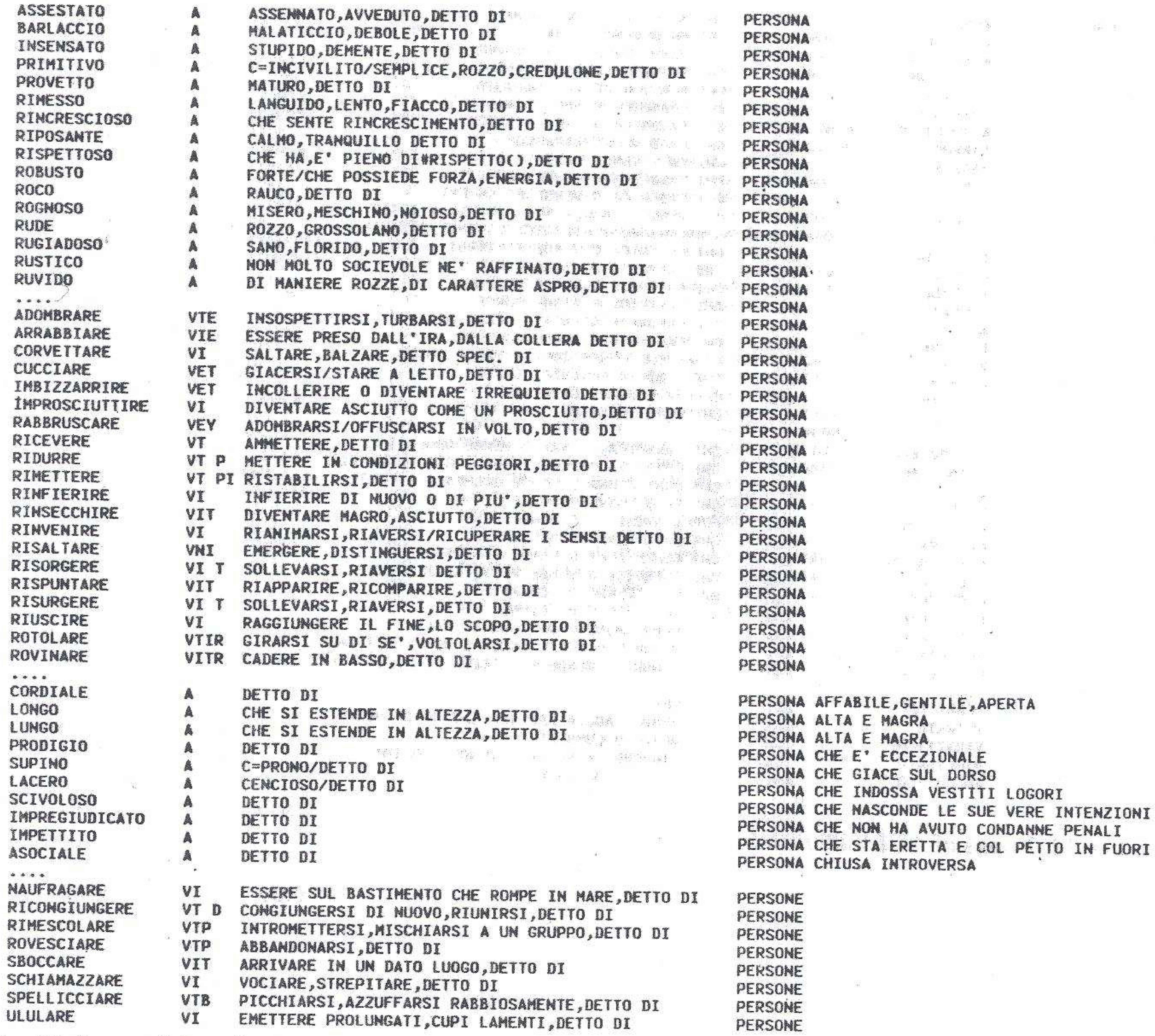
Fig. 7. Sketch of a piece of network for *attrezzo* (tool).

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).

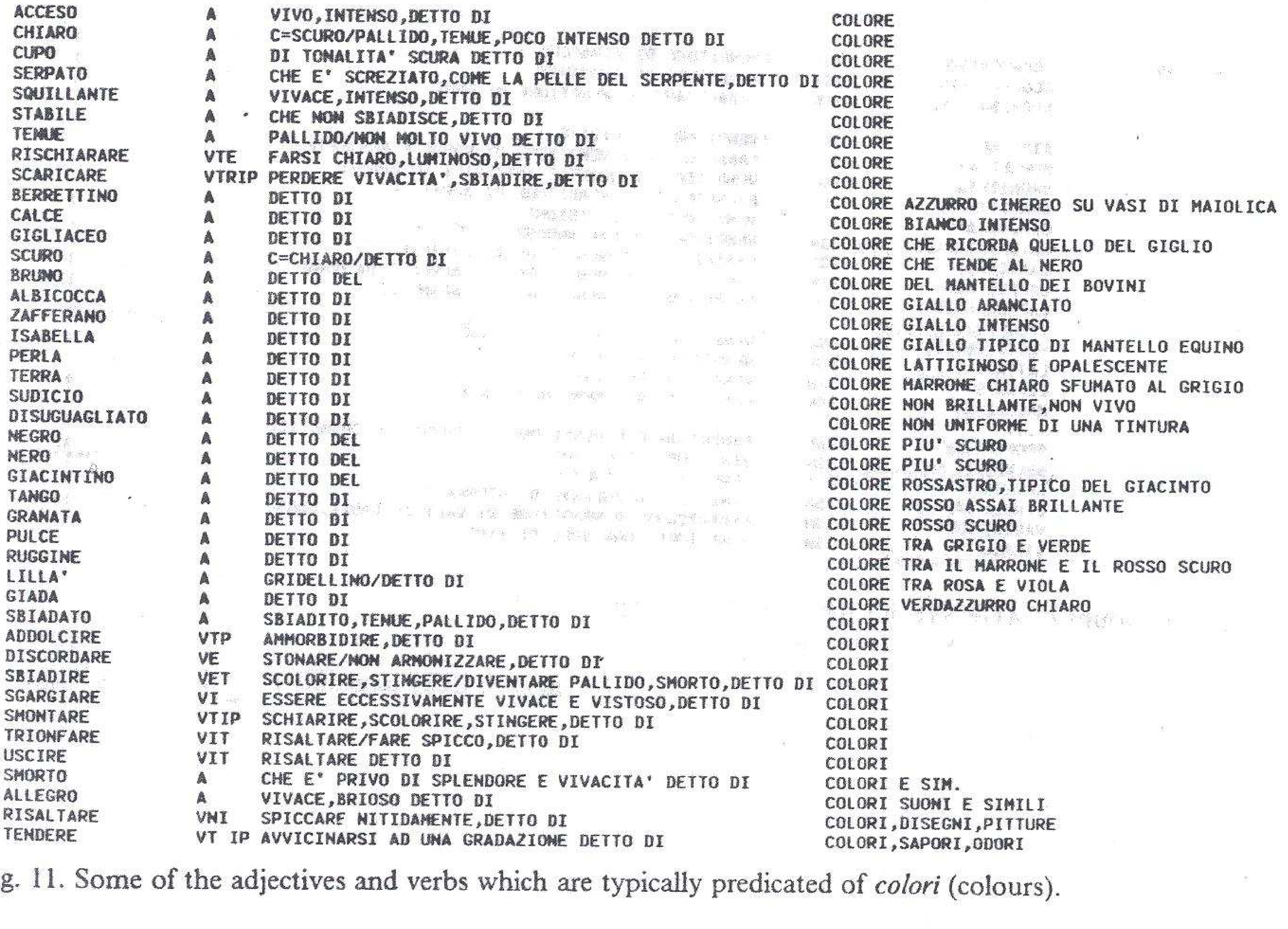
ARCIPELAGO	SM	GRUPPO INSIEME DI	OGGETTI
ANTIQUARIATO	SM	COMMERCIO O RACCOLTA DI	OGGETTI ANTICHI
SERVIZIO	SM	INSIEME DI	OGGETTI CHE SERVONO A UN DETERMINATO SCOPO
TROFEO	SM	INSIEME DI	OGGETTI CHE TESTIMONIANO SUCCESSI E VITTORIE
AFFARDELLAMENTO	SM	COMPLESSO DEGLI	OGGETTI CONTENUTI NELLO ZAINO DEL SOLDATO
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).



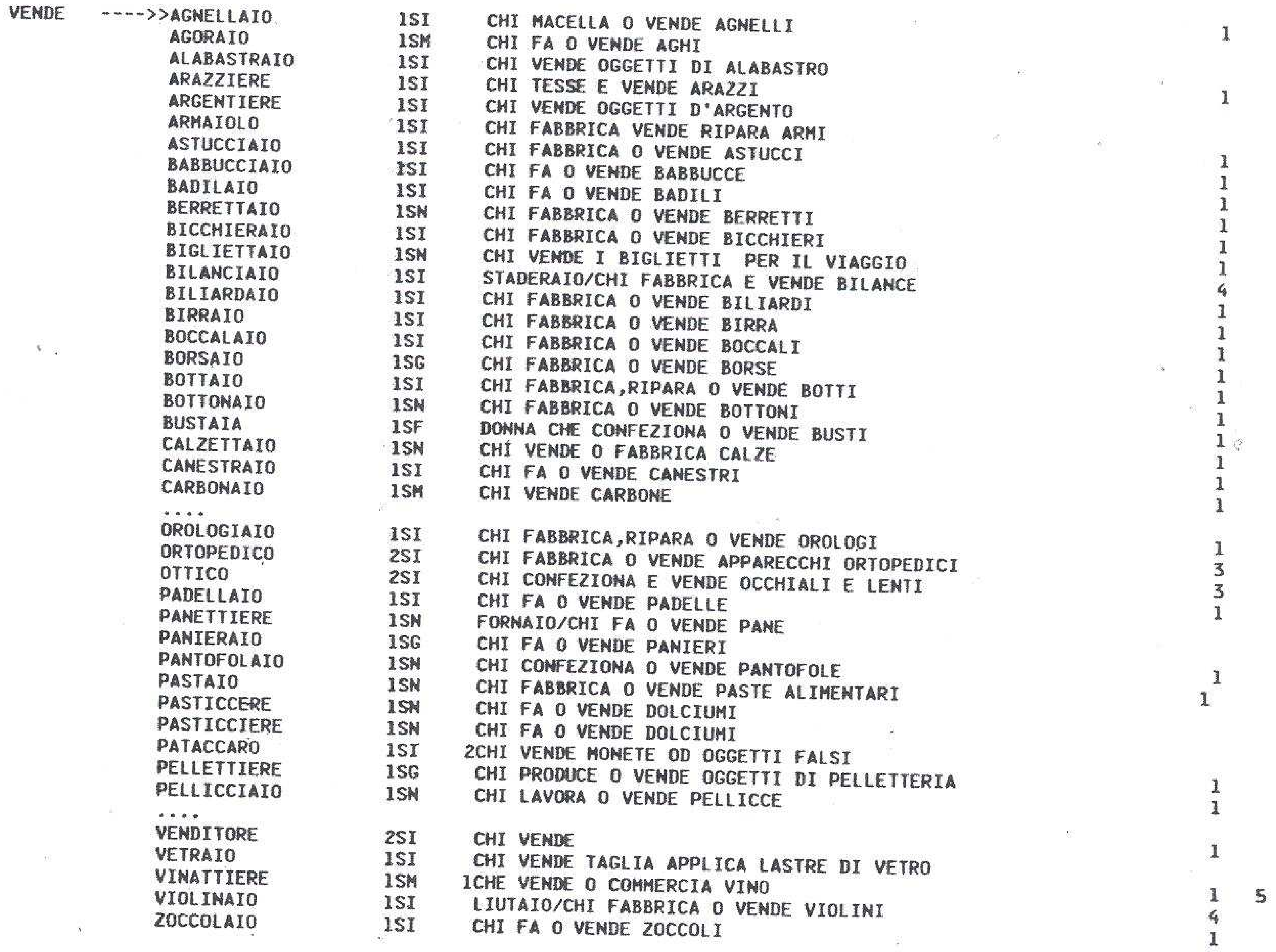
```
ASSESTATO          A      ASSENNATO,AVVEDUTO,DETTO DI                          PERSONA
BARLACCIO          A      MALATICCIO,DEBOLE,DETTO DI                           PERSONA
INSENSATO          A      STUPIDO,DEMENTE,DETTO DI                             PERSONA
PRIMITIVO          A      C=INCIVILITO/SEMPLICE,ROZZO,CREDULONE,DETTO 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).



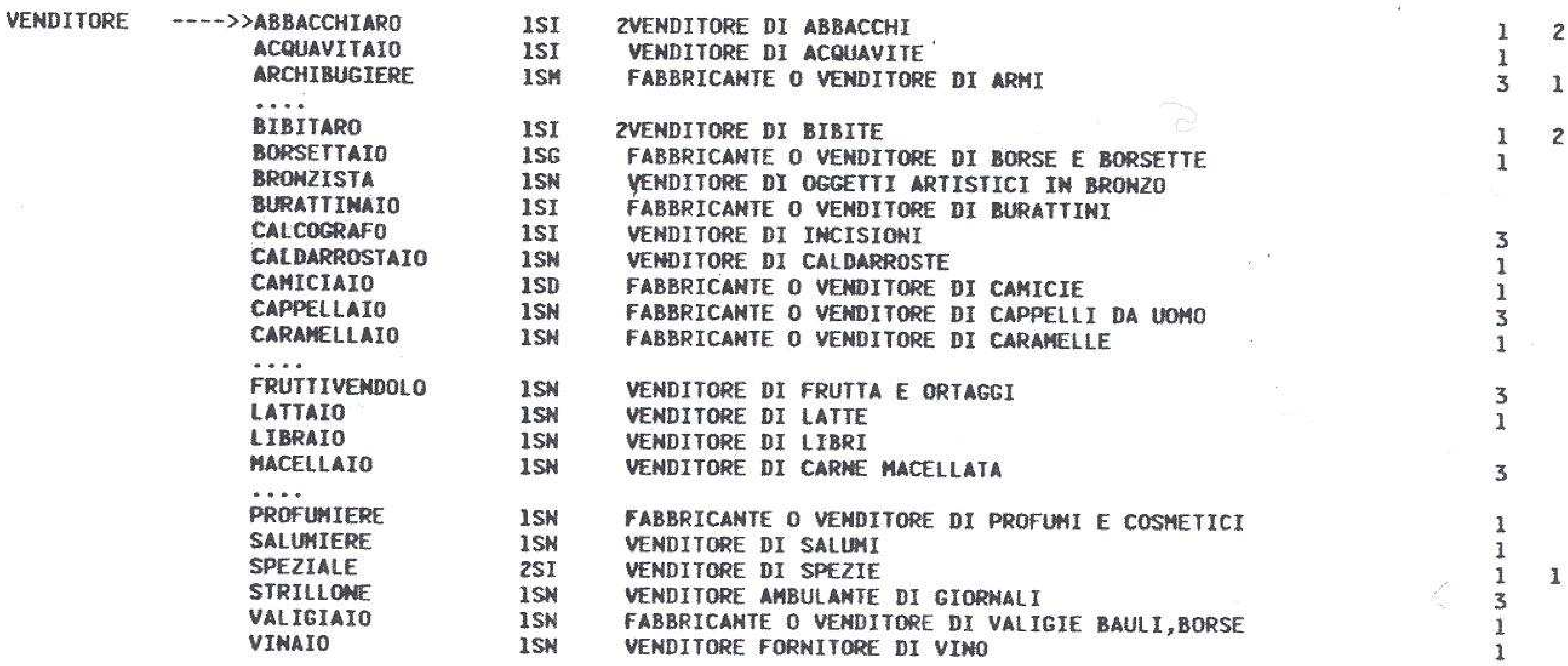
```
ACCESO             A      VIVO,INTENSO,DETTO DI                                COLORE
CHIARO             A      C=SCURO/PALLIDO,TENUE,POCO INTENSO DETTO DI          COLORE
CUPO               A      DI TONALITA' SCURA DETTO DI                          COLORE
SERPATO            A      CHE E' SCREZIATO,COME 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      GRIDELLINO/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    SPICCARF 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
               AGORAIO          1SM     CHI FA O VENDE AGHI                                1
               ALABASTRAIO      1SI     CHI VENDE OGGETTI DI ALABASTRO
               ARAZZIERE        1SI     CHI TESSE E VENDE ARAZZI
               ARGENTIERE       1SI     CHI VENDE OGGETTI D'ARGENTO                        1
               ARMAIOLO         1SI     CHI FABBRICA VENDE RIPARA ARMI
               ASTUCCIAIO       1SI     CHI FABBRICA O VENDE ASTUCCI
               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             1
               BILIARDAIO       1SI     CHI FABBRICA O VENDE BILIARDI                      4
               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     CHÍ VENDE O FABBRICA CALZE                         1
               CANESTRAIO       1SI     CHI FA O VENDE CANESTRI                            1
               CARBONAIO        1SM     CHI VENDE CARBONE                                  1
               ....                                                                        1
               OROLOGIAIO       1SI     CHI FABBRICA,RIPARA O VENDE OROLOGI
               ORTOPEDICO       2SI     CHI FABBRICA O VENDE APPARECCHI ORTOPEDICI         1
               OTTICO           2SI     CHI CONFEZIONA E VENDE OCCHIALI E LENTI            3
               PADELLAIO        1SI     CHI FA O VENDE PADELLE                             3
               PANETTIERE       1SN     FORNAIO/CHI FA O VENDE PANE                        1
               PANIERAIO        1SG     CHI FA O VENDE PANIERI
               PANTOFOLAIO      1SN     CHI CONFEZIONA O VENDE PANTOFOLE
               PASTAIO          1SN     CHI FABBRICA O VENDE PASTE ALIMENTARI              1
               PASTICCERE       1SN     CHI FA O VENDE DOLCIUMI                           1
               PASTICCIERE      1SN     CHI FA O VENDE DOLCIUMI
               PATACCARO        1SI    2CHI VENDE MONETE OD OGGETTI FALSI
               PELLETTIERE      1SG     CHI PRODUCE O VENDE OGGETTI DI PELLETTERIA
               PELLICCIAIO      1SN     CHI LAVORA O VENDE PELLICCE                        1
               ....                                                                        1
               VENDITORE        2SI     CHI VENDE
               VETRAIO          1SI     CHI VENDE TAGLIA APPLICA LASTRE DI VETRO           1
               VINATTIERE       1SM    1CHE VENDE O COMMERCIA VINO
               VIOLINAIO        1SI     LIUTAIO/CHI FABBRICA O VENDE VIOLINI               1   5
               ZOCCOLAIO        1SI     CHI FA O VENDE ZOCCOLI                             4
                                                                                           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"*.

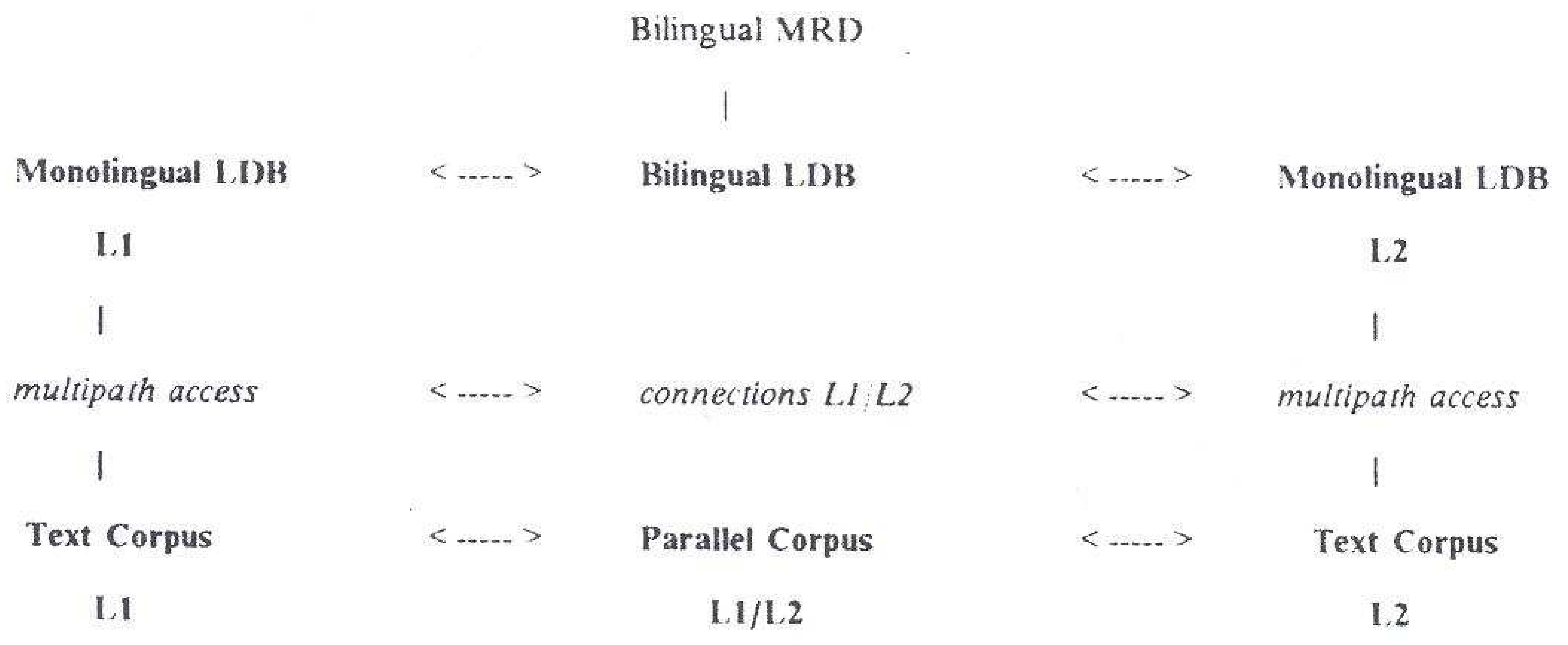


Fig. 16. A model of a Bilingual LDB System.

References

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

Allocution prononcee par M. Francois Mitterrand, President de la Republique, lorș de la seance solomnelle a′ l'Academie Francaise a′ l'occasion du 350mȩ 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* plemento n. 2 al vol. V., 1968.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Cumming, S., The Lexicon in Text Generation, in D. Walker, A.Zampolli, N.Calzolari 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.