

ACQUISITION OF SEMANTIC INFORMATION

FROM AN ON-LINE DICTIONARY

Nicoletta CALZOLARI - Eugenio PICCHI

*Dipartimento di Linguistica, Università di Pisa
Istituto di Linguistica Computazionale, CNR, Pisa
Via della Faggiola 32
I-56100 PISA - ITALY*

Abstract

After the first work on machine-readable dictionaries (MRDs) in the seventies, and with the recent development of the concept of a lexical database (LDB) in which interaction, flexibility and multidimensionality can be achieved, but everything must be explicitly stated in advance, a new possibility which is now emerging is that of a procedural exploitation of the full range of semantic information implicitly contained in MRDs. The dictionary is considered in this framework as a primary source of basic general knowledge. In the paper we describe a project to develop a system which has word-sense acquisition from information contained in computerized dictionaries and knowledge organization as its main objectives. The approach consists in a discovery procedure technique operating on natural language definitions, which is recursively applied and refined. We start from free-text definitions, in natural language linear form, analyzing and converting them into informationally equivalent structured forms. This new approach, which aims at reorganizing free text into elaborately structured information, could be called the Lexical Knowledge Base (LKB) approach.

1. Background

For a considerable period in theoretical and computational linguistics, there was a predominant lack of interest in lexical problems, which were regarded as being of minor importance with respect to "core" issues concerning linguistic phenomena, mainly of a syntactic nature. During the last few years, however, this trend has been almost reversed. The role of the lexicon in both linguistic theories and computational applications is now being greatly revalued and one aspect on which a number of research groups are now focussing their attention is the possibility of reusing the large quantity of data contained in already existing machine-readable lexical sources, mainly dictionaries prepared for photocomposition, as a short cut in the construction of extensive NLP-oriented lexicons. This position was formulated very clearly in a number of papers presented at a recent workshop organized in Grosseto (Italy) and sponsored by the European Community (see Walker, Zampolli, Calzolari, forthcoming), and can be found in the set

of recommendations which was one of the results of this workshop (Zampolli 1987, pp.332-335).

After the first work on machine-readable dictionaries (MRDs) in the seventies (see Olney 1972, Sherman 1974), and with the recent development of the concept of a lexical database (LDB) in which interaction, flexibility and multidimensionality can be achieved, but everything must be explicitly stated in advance (see e.g. Amsler 1980, Byrd 1983, Calzolari 1982, Michiels 1980), a new possibility which is now emerging is that of a procedural exploitation of the full range of semantic information implicitly contained in MRDs (see Wilks 1987, Binot 1987, Alshawi forthcoming, Calzolari forthcoming).

The dictionary is now considered as a primary source not only of lexical knowledge but also of basic general knowledge (ranging over the entire "world"), and some of the dictionary systems which are being developed have knowledge acquisition and knowledge organization as their principal objectives (see also Lenat and Feigenbaum 1987). In this paper we describe a project which we are now conducting on the acquisition of semantic information from computerized dictionaries.

2. Data and established methods for hierarchical semantic classifying

The data we use in our research include the lexical information contained in the Italian Machine Dictionary (DMI), which is already structured as a LDB and is mainly based on the Zingarelli Italian dictionary (1970); the DMI-DB has different types of linguistic information already accessible on-line. A morphological module generates and analyzes the inflected word-forms: approximately 1 million from 120,000 lemmas. Lemmas, word-forms, derivatives/suffixes, POS, usage codes, and specialized terminology codes, can be used as direct access search keys through which the user can query the database dictionary. On the semantic side, synonyms, hyponyms, and hypernyms constitute already implemented access paths covering all of the approximately 200,000 definitions contained in the dictionary. Examples of possible queries are the following: give me all the nouns defined as names of vehicles, of sounds, of games, all the verbs defined by a particular genus term, for example 'MUOVERE' (to move), 'TAGLIARE' (to cut), etc. The procedures used to find

hypernyms in definitions and to create taxonomies are similar to those used by other groups (see Chodorow 1985, Calzolari 1983, Amsler 1981).

We have now begun work on restructuring another dictionary available in MRF, the Garzanti Italian dictionary (1984). A parser has been implemented which, on the basis of the typesetting codes for photocomposition, identifies the rough structure of each lexical entry. Fig. 1 displays the output of a parsed entry of the Garzanti dictionary. Fig. 2 represents the provisional model for a monolingual lexical entry as we have defined it so far. Fig. 3 gives the projection of the first interpretation of the typesetting codes into this model; other kinds of information will be added afterwards (for example, that obtained by the inductive procedures described in the paper).

```

-----
[1] = arnese
[3] = [-ne²] (s.m.)
[4] = 1
[3] = utensile; attrezzo o strumento da lavoro:
      {gli arnesi del falegname}
[4] = 2
[3] = qualsiasi oggetto che non si sappia o non si voglia
      determinare: {a che serve quell'-?} / {quell' uomo
      e' un pessimo} -, e' un tipo poco raccomandabile
[4] = 3
[3] = abito, vestimento; maniera di vestire (anche {fig.})
      / {essere bene}, {male in} -, trovarsi in buone, cattive
      condizioni fisiche o economiche.
-----

```

Fig. 1 - Output of the photocomposition codes.
(the number in the first column identifies the type of data)

```

-----
Entry #
Homograph #
Pronunciation
Paradigm Label
POS
Syntactic Codes
Usage Label
Pointers to the base-lemma and/or to all derivatives
Pointers to graphical variants
Sense#
Field Label
Synctactic Codes
Figurative , extended, etc.
Definitions
Pointers to Synonyms
Pointers to Antonyms
Pointers to Hyponyms, Hyperonyms
Pointers to other Entries through other Relations
Semantic (inherent) Features
Formalized Word-sense Representation
Examples #
Example
Figurative, rare , ..
Definitions of a particular contextual usage
Idioms
Citations
Proverbs
-----

```

Fig. 2 - Provisional structure of a monolingual entry.

```

-----
001 Entry = arnese
005 PoS = s.m.
003 Pron = -ne²-
006 Sense = 1
007 Def = utensile
007 Def = attrezzo o strumento da lavoro
008 Examp = gli arnesi del falegname
006 Sense = 2
007 Def = qualsiasi oggetto che non si sappia o
          non si voglia determinare
008 Examp = a che serve quell'-
012 Idiom = quell'uomo e' un pessimo -
013 Expl = e' un tipo poco raccomandabile
006 Sense = 3
007 Def = abito, vestimento
014 Field = anche {fig.}
007 Def = maniera di vestire
012 Idiom = essere bene, male in -
013 Expl = trovarsi in buone, cattive condizioni
          fisiche o economiche.
-----

```

Fig. 3 - Example of a parsed Entry.

The merging of part of the data of the DMI and the Garzanti dictionary into a single LDB has already been completed, e.g. for lemmas, POSs, usage codes, etc. We now have to tackle the problem of reorganizing the semantic data (definitions and examples). Here our strategy is to design a new procedural system which is able to gradually "learn" and acquire semantic information from dictionary definitions, going well beyond the IS-A hierarchies constructed so far, in order to attempt to also capture what is present in the "differentia" part of the definition. This can be achieved with some success given the particular nature of lexicographic definitions, with: a) a generic (and perhaps over simplistic) description of the "world"; b) a rather lexically and syntactically constrained and a somewhat regular natural language text (Calzolari 1984, Wilks 1987).

After having mapped the codes for photocomposition into linguistically relevant codes, all the preliminarily parsed data of the Garzanti have been organized on a PC in the form of a Textual Database (DBT), a full-text Information Retrieval (IR) system in which all occurrences of any word-form or lemma can be directly accessed (Picchi 1983). The DBT has been found to be a very powerful tool in evidencing lexical units and particular syntagms which can then be exploited in our "pattern-matching" procedure. With the text in DBT form it is possible to search occurrences of single word-forms in definitions and examples, lemmas, codes of various types (POS, specialized languages, usage labels, etc.), and also cooccurrences of any of these items throughout the entire dictionary. In addition, structures composed of combinations of the above elements connected by the logical operators "and" and "or" to any degree of complexity can also be searched. The results of such queries are returned together with the pertinent dictionary entries. Obviously frequencies can also be obtained. All this information can be retrieved with fast interactive access.

We have therefore already implemented two types of organization for dictionary data:

- 1) DB-type organization with the DMI (we have not used a standard DBMS, but an ad hoc designed relational DB system);
- 2) a full-text IR system for the Garzanti dictionary.

Although both types of organization have proved to be very powerful tools for different scopes, at the same time each presents certain drawbacks and difficulties, due to the particular nature of dictionary data which in neither case has it been possible to fully exploit. Dictionary data is in fact of a very particular nature, consisting of a combination of free text in a highly organized structure. The DB approach copes well with the second characteristic, while the IR approach is successful in handling free text. However neither is capable of fully exploiting the two features in combination. A new method must be envisaged, capable of reorganizing free text into elaborately structured information: this could be called the Lexical Knowledge Base (LKB) approach, and is the aim of the project described here.

3. Techniques for word-sense acquisition

Discovery procedure techniques prove to be useful in extracting semantic information from definition texts. In general, our approach consists in starting from free-text definitions, in natural language linear form, analyzing and converting them into informationally equivalent structured forms. The preliminary step of the work consisted in applying the morphological analyzer to the definitions; the result of this process for one definition appears in Fig. 4. A program designed for homograph disambiguation was then run on the output produced by this morphological processor. The disambiguator consists partly in rules generally valid for Italian, based on the immediate right and left context, and partly in ad hoc rules written for the particular syntax used in lexicographic definitions. Fig. 5 shows the result of applying this disambiguation procedure to all the homographs shown in the preceding example. We then had to implement a set of discovery procedures acting on dictionary definitions.

```

-----
Entry ( EDITORE )
Def ( che o chi stampa e pubblica libri, periodici
o musica, a scopo commerciale)
F ( che )
L (che, ['PR' ,['NN' ]], ['PT' ,['NS' ]],
['DT' ,['NN' ]], ['DE' ,['NN' ]],
['PI' ,['MS' ]], ['C' ,[' ' ]])
F ( o )
L (o, ['SN' ,['NS' ]], ['C' ,[' ' ]],
['I' ,[' ' ]])
F ( chi )
L (chi, ['PR' ,['NS' ]])
F ( stampa )
L (stampa, ['SF' ,['FS' ]])
L (stampare, ['VTP' ,['S3IP' ,['S2MP' ]])
F ( e )
L (e, ['SN' ,['NS' ]], ['CC' ,[' ' ]])
F ( pubblica )
L (pubblico, ['A' ,['FS' ' ]])
L (pubblicare, ['VT' ,['S3IP' ,['S2MP' ]])
F ( libri )
L (libro, ['SM' ,['MP' ]])
L (librare, ['VTR' ,['S2IP' ,['S1CP' ,
['S2CP' ,['S3CP' ]])
P ( , )
F ( periodici )
L (periodico, ['A' ,['MP' ]], ['SM' ,['MP' ]])
F ( o )
L (o, ['SN' ,['NS' ]], ['C' ,[' ' ]],
['I' ,[' ' ]])
F ( musica )
L (musica, ['SF' ,['FS' ]])
L (musicare, ['VTI' ,['S3IP' ,['S2MP' ]])
P ( , )
F ( a )
L (a, ['SN' ,['NN' ]], ['E' ,[' ' ]])
F ( scopo )
L (scopo, ['SM' ,['MS' ]])
L (scopare, ['VT' ,['S1IP' ]])
F ( commerciale )
L (commerciale, ['A' ,['NS' ' ]])
-----

```

Fig. 4 - Output of the morphological analyzer

The first analysis of the definitional data was performed manually for single definitions, and quantitatively for the most frequently occurring words and syntagms. From this analysis we have established a number of broadly defined and simplified Categories of knowledge and Relations, which on the one hand intuitively reflect basic "conceptual categories" and on the other represent attested lexicographic definitional categories. They also rely on past experience of similar work (both on Italian and on English), or of AI research. In order to allow the inductive

```

-----
Entry ( EDITORE )
Def ( che o chi stampa e pubblica libri, periodici
o musica, a scopo commerciale)
F ( che )
L (che, ['PR' ,['NN' ]])
F ( o )
L (o, ['C' ,[' ' ]])
F ( chi )
L (chi, ['PR' ,['NS' ]])
F ( stampa )
L (stampare, ['VTP' ,['S3IP' ]])
F ( e )
L (e, ['CC' ,[' ' ]])
F ( pubblica )
L (pubblicare, ['VT' ,['S3IP' ]])
F ( libri )
L (libro, ['SM' ,['MP' ]])
P ( , )
F ( periodici )
L (periodico, ['SM' ,['MP' ]])
F ( o )
L (o, ['C' ,[' ' ]])
F ( musica )
L (musica, ['SF' ,['FS' ]])
P ( , )
F ( a )
L (a, ['E' ,[' ' ]])
F ( scopo )
L (scopo, ['SM' ,['MS' ]])
F ( commerciale )
L (commerciale, ['A' ,['NS' ' ]])
-----

```

Fig. 5 - Output of the disambiguation procedure

pattern-matching rules to perform the successive phases correctly and so that more coherent retrieval operations are possible, a "basic vocabulary" has been established (both for the "Categories" and for the "Relations") mainly on the basis of quantitative and intuitive considerations, and is constituted by words acting as Labels. As an example, the following lemmas: 'arnese, attrezzo, dispositivo, strumento, congegno', which altogether appear in dictionary definitions 761 times, have been grouped under the Label 'INSTRUMENT'. Other examples of Labels belonging to the basic vocabulary which have been established for hypernyms are the following: SET, PART, SCIENCE, HUMAN, ANIMAL, PLACE, ACT, EFFECT, LIQUID, PLANT, INHABITANT, SOUND, GAME, TEXTILE, MOVE, BECOME, LOSE, etc.

This is, therefore, our approach. We begin with a system which has simple and general purpose pattern-matching capabilities, designing it as an incremental system. To cope with the fact that there are variations in the way the same conceptual category or the same relation is linguistically (lexically and/or syntactically) rendered in natural language definitions, each such category or relation is associated with a list of specified lexical units and/or syntactic features which give the variant forms. The search is then driven by these lists of patterns to handle the grammatical and lexical variations.

The "pattern-matching" strategy has been obviously integrated with the Italian morphological analyzer to handle inflectional variation. The patterns may contain either Labels, or Lemmas, or Word-forms. For the Labels, the system searches for all the associated lemmas and all their word-forms (unless otherwise specified); in the same way Lemmas are automatically expanded to cover their inflected word-forms.

Generally, we look for recurring patterns in the definitions and attempt to associate them with corresponding relations or conceptual categories. Fig. 6 lists some of the entries and definitions obtained when querying the dictionary in DBT form for cooccurrences of items such as 'science, discipline, branch,...' together with 'studies, concerns, ...'. Analyzing the

results of similar queries to the dictionary we are able to better identify a number of patterns to be used in the semantic scanning of the definitions.

Textual Data Base	Dizionario Garzanti

searching for ... SCIENZA & STUDIA	
3)	ANATOMIA : PoS s.f. S#1 scienza che mediante la dissezione e altri metodi di ricerca <i>studia</i> gli organismi viventi nella loro forma esteriore e ...
6)	ARALDICA : PoS s.f. scienza del blasone, che <i>studia</i> e regola la composizione degli stemmi gentilizi.
9)	ASTROFISICA : PoS s.f. scienza che <i>studia</i> la natura fisica degli astri.
18)	BIOLOGIA : PoS s.f. scienza che <i>studia</i> i fenomeni della vita e le leggi che li governano.
35)	ETIMOLOGIA : PoS s.f. S#1 scienza che <i>studia</i> le origini delle parole di una lingua.
37)	FISICA : PoS s.f. scienza teorico-sperimentale che <i>studia</i> i fenomeni naturali e le leggi relative
56)	MERCEOLOGIA : PoS s.f. scienza applicata che <i>studia</i> le merci secondo la loro origine, i caratteri fisici, gli usi, la produzione e ...

searching for ... BRANCA & STUDIA	
3)	DIETETICA : PoS s.f. branca della medicina che <i>studia</i> la composizione dei cibi necessari a un'alimentazione razionale.
8)	FARMACOLOGIA : PoS s.f. branca della medicina che <i>studia</i> i farmaci e la loro azione terapeutica sull'organismo.
21)	TOSSICOLOGIA : PoS s.f. branca della medicina che <i>studia</i> la natura e gli effetti delle sostanze velenose e dei loro antidoti.

searching for ... SPECIALITA' & STUDIA	
1)	CARDIOLOGIA : PoS s.f. (med.) la specialità che <i>studia</i> le funzioni e le malattie del cuore.

searching for ... RAMO & STUDIA	
3)	ONOMASTICA : PoS s.f. ramo della linguistica che <i>studia</i> i nomi propri di persona o di luogo.

searching for ... SCIENZA & OCCUPA	
4)	PAPIROLOGIA : PoS s.f. scienza che si <i>occupa</i> dello studio e dell'interpretazione degli antichi papiri.

searching for ... DISCIPLINA & OCCUPA	
1)	AUXOLOGIA : PoS s.f. disciplina delle scienze biologiche che si <i>occupa</i> dell'accrescimento degli organismi, in particolare di quello umano.
2)	NEUROPSICHIATRIA : PoS s.f. disciplina medica che si <i>occupa</i> delle malattie nervose e mentali.

searching for ... DISCIPLINA & STUDIA	
1)	ALGOLOGIA : PoS s.f. disciplina medica che <i>studia</i> le cause e le terapie del dolore.
13)	IMMUNOLOGIA : PoS s.f. disciplina biologica che <i>studia</i> i fenomeni immunitari.

Fig. 6 - Same examples of queries to the dictionary in DBT form.

This is an example of a pattern where the Labels SCIENCE and STUDY appear:

$$\left[\text{Det/Adj} \right] \text{SCIENCE} \left[\text{di NP} / * \text{Adj/e NP} \right] \text{'che' (mediante NP)} \\ \text{STUDY NP-OBJ}$$

where the following are the lemmas associated to the Labels:

SCIENCE = (scienza, disciplina, specialita', branca, ramo, parte)

STUDY = (studia, si occupa di).

NP-OBJ is the subject matter of the science.

The results of a first run through the whole dictionary using an initial set of patterns can afterwards be recursively revised when new data are acquired. Our practical global research strategy is to develop a system which at the beginning has only a generalized expertise. This system obviously breaks down at many points on its first run; we can then evaluate all these

points, and consider when and where measures must be taken to overcome specific difficulties. In this way, new capabilities can be added incrementally to the system so that gradually it is able to cope with increasingly difficult data. Thus we systematically add new "knowledge" to the system, prompted each time by a failure to cope with the given data. It seems to us that this is a practical research strategy for eliciting and modelling vague and fuzzy knowledge.

Even though the methodological approach has been deliberately simplified at the beginning (in order to introduce problems gradually, a few at a time), the dimensions of the data have not been limited in any way.

4. The knowledge organization.

Although the body of knowledge with which we are dealing is at least partly based on intuition, on vague and not even coherent data (as lexicographic definitions often are), and on inductive empirical strategies, we must attempt to model the knowledge as the system acquires it. The formalism for the representation of word-senses is as follows.

Each element is defined as a Function characterized by a Type and Arguments. The Type qualifies the function. The main types include: Hypernym, Relation, Qualifier, etc.

Examples of the Type.Relation arc: USED, PRODUCED, IN-THE-FORM, STUDY, LACK, etc. The type Hypernym can be instantiated by: Hypernym proper, PART, SET, etc. Arguments may be either Terms, or Terms plus Function, or Functions. A Term can be a Label, a Word, or a combination of these with the logical operators 'and/or'. A Word can be either a Word-form, or a Lemma plus Grammatical Information (e.g. [Np] means plural Noun).

The following definitions:

Batterio, s.m., microrganismo vegetale unicellulare privo di clorofilla.

Batteriologia, s.f., parte della microbiologia che studia i batteri.

are now represented as:

Batterio --def--> f(T.HYP,|N|microrganismo, f(T.QUAL,|A|vegetale, |A|unicellulare), f(T.REL-JACK,|N|clorofilla))

Batteriologia --def--> f(T.HYP-PART, f(T.REL-SPEC,|N|microbiologia), f(T.REL-STUD,|Np|batterio)).

As the metalanguage and the rules are declared separately from the pattern-matching parser, the system is incremental, flexible, portable (it can be used with other languages or other dictionaries), and testable. In fact, the system has been designed so that it is easy to test alternative strategies or sets of rules or constraints.

This kind of organization will allow us to draw inferences, using part of the formal structure associated to an entry and inserting it in other structures in which that entry appears as

an Argument. For example, 'microbiologia' present in the second definition above is defined in its turn as 'parte della biologia che studia i microrganismi...', translated as (T.HYP-PART,(T.REL-SPEC,|N|biologia)), and 'biologia' which is 'scienza che studia i fenomeni della vita...' is finally defined as T.HYP-SCIENCE. This last Label SCIENCE is obviously also inherited by 'Batteriologia' and by 'Microbiologia'.

5. Some experimental results

Already after just one run, by looking at cooccurrences of hypernyms and particular relations, we can identify those environments in which certain relations are most likely to appear, or in which certain ambiguous lexical and/or syntactic cues (e.g. the prepositions PER 'for', DI 'of', A 'to', etc.) can be disambiguated as referring to only one relation, or in which certain relations are never found, and so on.

A set of constraining rules can be associated to certain conceptual units (Hypernyms or Relations, expanded automatically to all the pertinent lexical realizations) in order to disambiguate their immediate context. Some units therefore activate particular subroutines for an ad-hoc interpretation of what follows. These rules explicitly look for items to which a determined meaning is associated. In the following pattern, we have a rule which, after an USED relation, links the word "in" to a 'place' relation, the words "per/a" (for/to) to the purpose, "da" (by) to the agent, and "come" (as) to the way of usage. Other kinds of relations are not activated by a particular rule, but have a meaning in themselves, e.g. CONSTITUTED BY, SIMILAR TO, etc.

HYPER. ... USED	{	come	NP	(= way)
		per/a	Vinf./NP	(= purpose)
		in	NP	(= place)
		da	NP	(= agent)

The analysis in some cases is therefore purposely delayed until more relevant information has been acquired, and will eventually be based on the results of definitions already successfully handled. This analysis of the first results will lead to an improvement of the system, adding other patterns or other surface realizations of already existing patterns to the first simple list of patterns, and also imposing constraints on given hypernyms or on given relations. Therefore, after the first stage, the system consists of patterns augmented with conditioning rules which will then drive subsequent runnings of the procedure (for those cases which are lexically or grammatically conditioned). In this way, the system can be gradually refined. The analysis procedure is envisaged as a series of cycles which find relevant cooccurrences of categories and relations that can then be set as conditioning rules to further guide successive searches. An interactive phase is also foreseen so that, when necessary, definitions can be modified for a normalization in accordance to acceptable analysis structures. From successive passes through the data, applying different and increasingly refined sets of patterns and rules, the procedure

builds up, as completely as possible with this methodology, a formal description of the structure of the lexical definitions.

At the end, from a comparison of the different formalized structures generated, we will be able to associate structures which differ for only one element (a conceptual category or relation). In this way, we can construct something like "minimal pairs" of sense-definitions, which only differ in one conceptual or relational feature. It can be reasonably supposed that this feature is related or realizes one of the differences between these words. It will also be possible to build hierarchies not only for hypernyms, but also, and more interestingly, for complex conceptual structures considered as a whole.

References

- H. Alshawi, Processing dictionary definitions with phrasal pattern hierarchies, in *Special Issue of CL on the Lexicon*, forthcoming.
- R. Amsler, A taxonomy for English nouns and verbs, in *Proceedings of the 19th Annual Meeting of the ACL*, Stanford (Ca), 1981, 133-138.
- J.L. Binot, K. Jensen, A semantic expert using an on-line standard dictionary, in *Proceedings of the 10th IJCAI*, Milano, 1987, 709-714.
- B.K. Boguraev, Machine-readable dictionaries in computational linguistics research, in D. Walker, A. Zampolli, N. Calzolari (eds.), forthcoming.
- R.J. Byrd, Word formation in natural language processing systems, in *Proceedings of the 8th IJCAI*, Karlsruhe, 1983, 704-706.
- R.J. Byrd, N. Calzolari, M.S. Chodorow, J.L. Klavans, M. Neff, O.A. Rizk, Tools and methods for Computational Lexicology, in *Journal of Computational Linguistics*, forthcoming.
- N. Calzolari, Towards the organization of lexical definitions on a database structure, in *COLING 82*, Prague, Charles University, 1982, 61-64.
- N. Calzolari, Lexical definitions in a computerized dictionary, in *Computers and Artificial Intelligence*, II(1983)3, 225-233.
- N. Calzolari, Detecting patterns in a lexical database, in *Proceedings of the 10th International Conference on Computational Linguistics*, Stanford (Ca), 1984, 170-173.
- N. Calzolari, Structure and access in an automated dictionary and related issues, in D. Walker, A. Zampolli, N. Calzolari (eds.), forthcoming.

M.S. Chodorow, R.J. Byrd, G.E. Heidorn, Extracting semantic hierarchies from a large on-line dictionary, in *Proceedings of the 23rd Annual Meeting of the ACL*, Chicago (Ill), 1985, 299-304.

Garzanti, *Il nuovo dizionario Italiano Garzanti*, Garzanti: Milano, 1984.

D.B. Lenat, E.A. Feigenbaum, On the thresholds of knowledge, in *Proceedings of the 10th IJCAI*, Milano, 1987, 1173-1182.

J. Markowitz, T. Ahlswede, M. Evens, Semantically significant patterns in dictionary definitions, in *Proceedings of the 24th Annual Meeting of the ACL*, New York, 1986, 112-119.

A. Michiels, Exploiting a large dictionary database, Ph.D. thesis, Liege, 1982.

J. Olney, D. Ramsey, From machine-readable dictionaries to a lexicon tester: progress, plans, and an offer, in *Computer Studies in the Humanities and Verbal Behavior*, 3(1972)4, 213-220.

E. Picchi, Textual Data Base, in *Proceedings of the International Conference on Data Bases in the Humanities and Social Sciences*, Rutgers University Library: New Brunswick, 1983.

E. Picchi, N. Calzolari, Textual perspectives through an automatized lexicon, in *Proceedings of the XII International ALLC Conference*, Slatkine: Geneve, 1986.

D. Sherman, A new computer format for Webster's Seventh Collegiate Dictionary, in *Computers and the Humanities*, VIII(1974), 21-26.

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

D. Walker, A. Zampolli, N. Calzolari (eds.), *Automating the Lexicon: Research and Practice in a Multilingual Environment*, Proceedings of a Workshop held in Grosseto, Cambridge University Press, forthcoming.

Y. Wilks, D. Fass, C.M. Guo, J.E. McDonald, T. Plate, B.M. Slator, A tractable machine dictionary as a resource for computational semantics, MCCS-87-105, New Mexico State University, 1987.

A. Zampolli, Perspectives for an Italian Multifunctional Lexical Database, in A. Zampolli (ed.), *Studies in honour of Roberto Busa S.J.*, Giardini: Pisa, 1987.

N. Zingarelli, *Vocabolario della Lingua Italiana*, Zanichelli: Bologna, 1970.