

Lecture Notes in Artificial Intelligence

Edited by J. Siekmann

Subseries of Lecture Notes in Computer Science

476

M. Filgueiras L. Damas N. Moreira
A.P. Tomás (Eds.)

Natural Language Processing

EAIA '90, 2nd Advanced School in Artificial Intelligence
Guarda, Portugal, October 1990
Proceedings



Springer-Verlag

Methods and Tools for Lexical Acquisition*

Nicoletta Calzolari, António Zampolli

Istituto di Linguistica Computazionale del CNR, Pisa

Dipartimento di Linguistica, Università di Pisa

Via de la Faggiola, 32 — 56100 Pisa — Italy

1 Introduction

In the very recent years, Computational Linguistics has shown an increasing interest in the development of vast 'reservoirs' of linguistic knowledge, in the form of as complete as possible and reusable linguistic descriptions, structured in various kinds of interconnected linguistic bases.

Several projects are underway or are going to begin, promoted by large international or national organizations, aiming at the creation of large lexical knowledge bases (LKB), grammatical knowledge bases, reference textual corpora.

One of the key-words in the field has recently become the word *reusability*. This word is to be intended in two main senses. The first meaning is that of constructing the linguistic resources in such a way that it is allowed to various users (procedural, e.g. different NLP systems, and possibly human, e.g. lexicographers, translators, normal dictionary users) to extract — with appropriate interfaces — relevant information for their different purposes. The second meaning is that of reusing lexical information implicitly or explicitly present in pre-existing lexical resources.

In this paper we will describe some aspects of our work aiming at reusing (in its second meaning) existing machine readable dictionaries (MRDs) for the construction of LKBs¹.

*Part of this article is reprinted, thanks to the permission of the Oxford University Press, from the article: N. Calzolari, A. Zampolli, "Lexical Databases and Textual Corpora: a trend of convergence between Computational Linguistics and Literary and Linguistic Computing", to appear in S. Hockey, N. Ide (eds.), *Literary and Linguistic Computing*, 1989, Oxford University Press, in print.

¹Our work in this field is now included in the ESPRIT Project "Acquisition of Lexical Knowledge for Natural Language Processing Systems" (AQUILEX) in which groups of researchers in Cambridge, Amsterdam, Dublin, Paris, Barcelona, and Pisa (coordinator) are involved.

2 Reusability of Preexisting Data in the Form of MRDs

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

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

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

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

The methodological approach that we follow can be summarized in these points:

- a) to start from free-text definitions, in natural language and in linear form, usually formed by a 'genus term' and a 'differentia' part;
- b) to analyse their structure and content from a linguistic and a computational point of view;
- c) to convert and reorganize them into informationally equivalent structured formats made up by nodes and relations linking them.

Point b) in its turn can be subdivided, for the computational part, into the following steps:

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

Let us illustrate with some examples the process of analysing the definitions. In the figures we try to simulate the process of browsing the Italian LDB and of navigating the dictionary while searching for particular words, structures, patterns, etc.. We can see some of the semantic data it is possible to search for and find in a MRD if appropriately structured. Figure 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 Figure 2), and the set of Adjectives and of other Nouns having to do with books (Figures 3 and 4). In section 2.2 we shall come back to “books”, stressing the type of information which, lacking in dictionaries, can instead be found in texts.

Our present work is 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 analyse 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 Figure 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. Figure 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,

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

Figure 1: Some of the hyponyms of *libro* (book)

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

Figure 2: Verbs related to *libri* (books)

ADESPOTA	1A	3ANONIMO/DETTO DI LIBRO,CODICE,MANOSCRITTO DI AUTORE IGNOTO
ADESPOTO	1A	ANONIMO/DETTO DI LIBRO,CODICE,MANOSCRITTO DI AUTORE IGNOTO
APOCRIFO	1A	DETTO DI LIBRO NON RICONOSCIUTO COME CANONICO
CARTOLIBRARIO	1A	DI COMMERCIO DI LIBRI E OGGETTI DA CANCELLERIA
CIRCOLANTE	1A	CHE DA' LIBRI A PRESTITO AGLI ABBONATI A TURNO
COMMERCIALE	1A	DETTO DI LIBRO,FILM CHE MIRA SOLO A OTTENERE BUONI INCASSI
COPERTINATO	1A	DETTO DI LIBRO O FASCICOLO CON COPERTINA
DEUTEROCANONICO	1A	DEI LIBRI DELL'ANTICO TESTAMENTO RESPINTI COME APOCRIFI
EDITORE	1A	CHI PUBBLICA LIBRI,RIVISTE
ERUDITO	1A	LIBRO ERUDITO
INTESTATO	1A	FORNITO DI TITOLO O INTESTAZIONE,DETTO DI LIBRO,LETTERA
INTONSO	1A	3DI LIBRO CUI NON SONO ANCORA STATE TAGLIATE LE PAGINE
LIBERIANO	3A	CHE RIGUARDA IL LIBRO
LIBRARIO	1A	DI,RELATIVO A LIBRO
LIBRESCO	1A	CHE DERIVA DAI LIBRI E NON DALLA VIVA ESPERIENZA
MASTRO	2A	LIBRO MASTRO
MOSAICO	2A	RELATIVO AI LIBRI BIBLICI
PAGA	4A	LIBRO PAGA
POSTUMO	1A	DI LIBRO PUBBLICATO DOPO LA MORTE DELL'AUTORE
PROTOCANONICO	1A	DETTO DI CIASCUN LIBRO BIBLICO INSERITO PER PRIMO NEL CANONE
SAPIENZIALE	1A	CHE SI RIFERISCE AI LIBRI SAPIENZIALI

Figure 3: Adjectives related to *libri* (books)

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

Figure 4: Some of the nouns related to *libri* (books)

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 Figure 7.

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

Figure 5: The first hyponyms of *strumento* (instrument)

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), (Figure 8), *oggetti* (objects) (Figure 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 Figure 10 for Adjectives and Verbs used for nouns denoting *persone* (people), Figure 11

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

Figure 6: Some of the hyponyms of *attrezzo* (tool) with their definitions

INSTRUMENT ← IS-A — <i>attrezzo</i>	— USED FOR →	<i>tagliare</i>	= <i>FALCE</i>
...			= ...
	— USED IN →	<i>ginnastica</i>	= <i>ANELLO</i>
...			= ...
	— SHAPE →	<i>tubolare</i>	= <i>ASTA</i>
		<i>circolare</i>	= <i>DISCO</i>
	— MADE OF →	<i>vimini</i>	= <i>CESTA</i>
		<i>metallo</i>	= <i>CHiodo</i>

Figure 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
BRANCO	SM	INSIEME DI	PERSONE
POPOLAZIONE	SF	INSIEME DELLE	PERSONE ABITANTI IN UN LUOCO
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
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
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
CONCISTORO	SM	GRUPPO DI	PERSONE CHE DISCUOTONO
FINANZA	SF	COMPLESSO DI	PERSONE CHE ESPLICANO ATTIVITA' BANCARIA
...			
FRONTE	SN	COMPLESSO DI	PERSONE OMOGENEO PER FINALITA' CONSUETUDINI
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

Figure 8: Some of the nouns denoting SET OF *persone* (people)

ARCIPELAGO	SM	GRUPPO INSIEME DI	OGGETTI
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
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
SUPPELLETTILE	SF	OGGETTO O INSIEME DI	OGGETTI NELL'ARREDAMENTO DELLA CASA
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

Figure 9: Nouns denoting SET OF *oggetti* (objects)

for Adjectives which collocate with names of colours, either generic colour names, or specific ones such as *giallo* (yellow), *rosso* (red), etc.).

ASSESTATO	A	ASSENATO, 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
...			
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
...			
NAUFRAGARE	VI	ESSERE SUL BASTIMENTO CHE ROMPE IN MARE, DETTO DI	PERSONA
RICONGIUNGERE	VT D	CONGIUNGERSI DI NUOVO, RIUNIRSI, DETTO DI	PERSONA
RIMESCOLARE	VTP	INTROMETTERSI, MISCHIARSI A UN GRUPPO, DETTO DI	PERSONA
ROVESCARE	VTP	ABBANDONARSI, DETTO DI	PERSONA
SBOCCARE	VIT	ARRIVARE IN UN DATO LUOGO, DETTO DI	PERSONA
SCHIAMAZZARE	VI	VOCIARE, STREPITARE, DETTO DI	PERSONA
SPELLICCIARE	VTB	PICCHIARSI, AZZUFFARSI RABBIOSAMENTE, DETTO DI	PERSONA
ULULARE	VI	EMETTERE PROLUNGATI, CUPI LAMENTI, DETTO DI	PERSONA
...			
CORDIALE	A	DETTO DI	PERSONA AFFABILE, GENTILE, APERTA
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

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

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 Figure 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 the 'genus' position). The result of the query is the following: we retrieve 242 entries of which well 221 are names of people who "typically sell" something, i.e. of typical AGENTS with respect to the action of selling. These entries represent lexicalized case/role fillers in the case-frame of *vendere*

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
SBIADATO	A	SBIADITO,TENUE,PALLIDO,DETTO DI	COLORE
ADDOLCIRE	VTP	AMMORBIDIRE,DETTO DI	COLORE
DISCORDARE	VE	STONARE/NON ARMONIZZARE,DETTO DI	COLORE
SBIADIRE	VET	SCOLORIRE,STINGERE/DIVENTARE PALLIDO,SMORTO,DETTO DI	COLORE
SGARGIARE	VI	ESSERE ECCESSIVAMENTE VIVACE E VISTOSO,DETTO DI	COLORE
SMONTARE	VTIP	SCHIARIRE,SCOLORIRE,STINGERE,DETTO DI	COLORE
TRIONFARE	VIT	RISALTARE/FARE SPICCO,DETTO DI	COLORE
USCIRE	VIT	RISALTARE DETTO DI	COLORE
...			
BERRETTINO	A	DETTO DI	COLORE AZZURO 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 VERDAZZURO CHIARO
...			
SMORTO	A	CHE E' PRIVO DI SPLENDORE E VIVACITA' DETTO DI	COLORE E SIM.
ALLEGRO	A	VIVACE,BRIOSO DETTO DI	COLORE SUONI E SIMILI
RISALTARE	VNI	SPICCARRE NITIDAMENTE,DETTO DI	COLORE,DISEGNI,PITTURE
TENDERE	VT IP	AVVICINARSI AD UNA GRADAZIONE DETTO DI	COLORE,SAPORI,ODORI

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

(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.

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

Figure 12: Names of AGENTS for the action of "selling"

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

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

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

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

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

Figure 13: Names of AGENTS for the action of "selling"

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 in 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. Figure 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 lexicographers: *dove/in cui si vendono* (where ... are sold). All of the 33 entries retrieved share this definitional pattern: this query is completely without 'noise'.

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

What interests us here is the possibility of formalizing and implementing in the LKB the other types of semantic relations, such as LOCATION and THEME with respect to the actions of "selling" and "making". The Theme relation, i.e. the objects which are

VENDONO	→BANCO	1SM	LOCALE DOVE SI VENDONO O SCAMBIANO BENI SERVIZI
	BIGLIETTERIA	1SF	LUOGO IN CUI SI VENDONO BIGLIETTI
	BISCOTTERIA	1SF	NEGOZIO DOVE SI VENDONO I BISCOTTI
	BOTTIGLIERIA	1SF	NEGOZIO DOVE SI VENDONO VINO LIQUORI IN BOTTIGLIA
	BRICABRAC	1	NEGOZIO,BANCARELLA OVE SI VENDONO TALI ANTICAGLIE
	CALZETTERIA	1SF	NEGOZIO IN CUI SI VENDONO CALZE
	CALZOLERIA	1SF	BOTTEGA IN CUI SI FABBRICANO O VENDO SCARPE
	CAMICERIA	1SF	NEGOZIO IN CUI SI VENDONO CAMICIE
	CAPPELLERIA	1SF	NEGOZIO DOVE SI VENDONO CAPPELLI MASCHILI
	CERERIA	1SF	LUOGO DOVE SI FABBRICANO E VENDONO CANDELE
	CREMERIA	1SF	2LATTERIA IN CUI SI VENDONO ANCHE GELATI DOLCI E SIM.
	DIACCIATINO	2SN	2BOTTEGA DOVE SI VENDONO SORBETTI
	DROGHERIA	1SF	BOTTEGA DOVE SI VENDONO DROGHE
	FERRAMENTA	1SF	NEGOZIO IN CUI SI VENDONO OGGETTI DI FERRO
	GELATERIA	1SF	SORBETTERIA/NEGOZIO OVE SI FANNO O VENDONO GELATI
	MAGLIERIA	1SF	BOTTEGA NEGOZIO IN CUI VENDONO INDUMENTI DI MAGLIA
	MESCITA	1SF	BOTTEGA IN CUI SI VENDONO VINO LIQUORI
	MESTICHERIA	1SF	2BOTTEGA IN CUI SI VENDONO COLORI MESTICATI
	NEGOZIO	1SM	BOTTEGA/ LOCALE DOVE SI ESPONGONO E VENDONO MERCI
	NORCINERIA	1SF	2BOTTEGA IN CUI SI VENDONO SOLO CARNI DI MAIALI
	OCCHIALERIA	1SF	NEGOZIO IN CUI SI VENDONO O SI RIPARANO OCCHIALI
	OROLOGERIA	1SF	NEGOZIO DOVE SI VENDONO OROLOGI
	PANTOFOLERIA	1SF	LUOGO IN CUI SI VENDONO PANTOFOLE
	PELLETTERIA	1SF	NEGOZIO IN CUI SI VENDONO OGGETTI DI LAVORATA
	PIATTERIA	1SF	BOTTEGA DOVE SI VENDONO I PIATTI
	ROSTICCERIA	1SF	BOTTEGA DOVE SI PREPARANO O VENDONO ARROSTI
	SALUMERIA	1SF	BOTTEGA,NEGOZIO,IN CUI SI VENDONO I SALUMI
	UTENSILERIA	1SF	BOTTEGA IN CUI SI VENDONO UTENSILI

Figure 14: Some names of PLACES related to the action of “selling”

typically sold in the defined places are again expressed by the NP object of the verb.

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

OROLOGERIA= ← LOC – *selling* – THEME → *orologi* – IS-A → OBJECT
OROLOGIAIO= ← AGENT – ” ” ” ” ”

Figure 15: Sketch of a piece of network for the action of “selling”

The above examples show that the LDB facilities can be usefully exploited to analyse and extract linguistic data which must then be restructured and represented in the LKB. In the LKB these types of concepts and of relations, and the interdependencies between word-senses will be explicitly spelled out. When we move beyond taxonomies in the LKB, we establish many different types of associations which are usefully represented in a conceptual network, and when we move from a “monolingual” to a “multilingual” environment, we also establish associations among different languages. These associations are obtained (for those parts of the languages which can be reduced to a common set of concepts and relations) through the common conceptual network constructed by working on different languages but within the same “research template”, i.e. trying to accomodate in the semantic network:

- the “same” world-knowledge,

- for the “same” purposes (NLP, Text Processing, etc.),
- with the “same” methodology,
- from the “same” type of sources (MRDs),
- into the “same” kind of representation.

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

2.1 Reusability of Bilingual Dictionaries

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

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

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

The overall picture of the bilingual LDB system we have in mind is sketched in Figure 16. Also with regard to bilingual dictionaries, the method we are adopting consists of reusing available data in machine-readable form by analysing and transforming the information already contained in common dictionaries. The procedure of processing the bilingual MRD is rather similar to the one outlined above for monolingual dictionaries (i.e. parsing of the lexical entry, design of a new structure, computational reorganization, etc.). After this preliminary part again comes out the utility of browsing the bilingual

LDB, taking advantage of the structural elements already formalized in the LDB, with the purpose of discovering properties and structures not immediately visible in the printed dictionary, but useful for further exploitation in the computational dictionary.

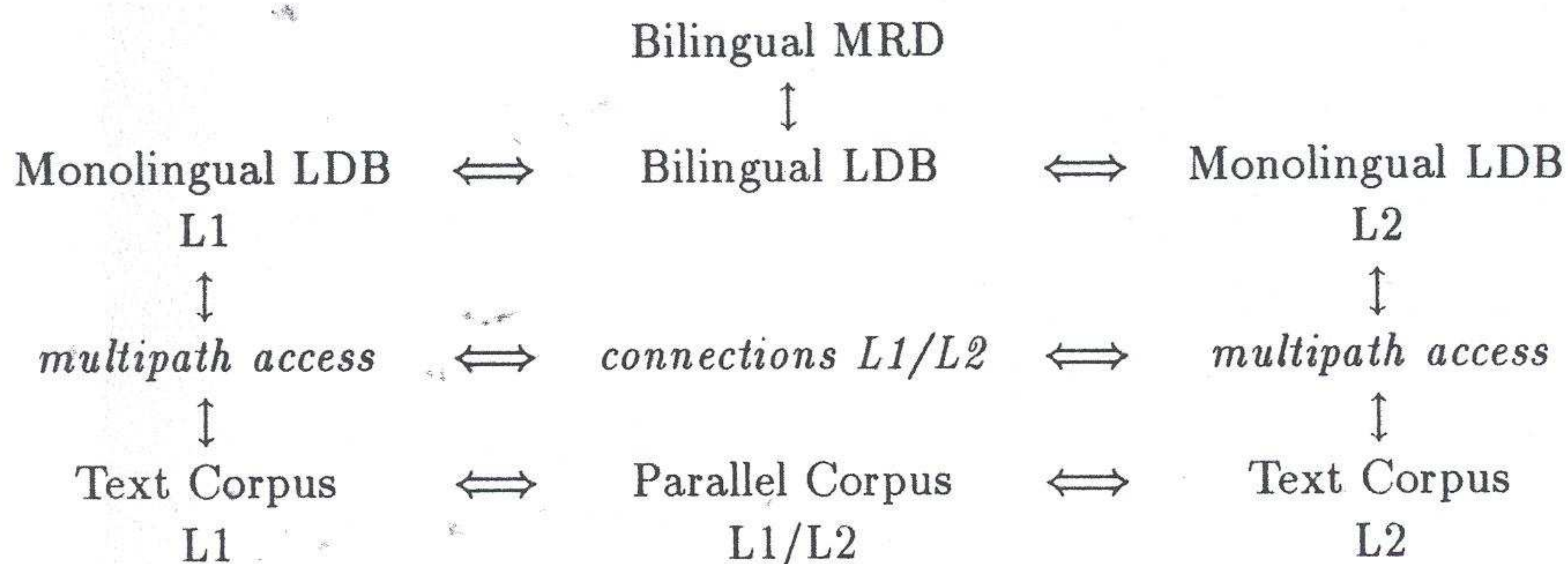


Figure 16: A model of a Bilingual LDB System

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

One of the parts of the bilingual dictionary we are processing that can be partially made explicit in all its different meanings, is the field of the so-called *semantic indicators*. These provide the constraints for selecting one translation equivalent or the other. The problem is that these constraints are of a different nature, being either

- i) synonyms or hyponyms of the entry, or
- ii) contextual indicators such as typical subjects or objects of verbs, typical nouns of which an adjective can be predicated, etc..

It is possible to semi-automatize the process of disambiguation between the different values, after analysing all the different possibilities and designing a typology of what can appear in this field.

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

vivido ... (colori) bright, vivid

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

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

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

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

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

Other purposes of a Bilingual System like the one which appears in Figure 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 the two languages, 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.

2.2 Reusability of Textual Corpora and their Integration into LKBs

We have seen that MRDs are very valuable sources of lexical and also of semantic information, but unfortunately not all what is needed to know about the lexicon is there. There are very important pieces of information which in MRDs are completely missing, or incomplete, or simply are not very good or reliable or easily recoverable. For this type of information, we have to resort to different types of sources (see also [Calzolari, 1989b]).

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

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

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

- aspectual information;
- proper nouns.

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

<i>dividere</i>	NP in NP	268
''	NP	175
''	NP tra NP, NP, ...	80
''	NP con NP	78
		601

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

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

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

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

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

word *libro* is defined in terms of words such as *volume*, *opera*, *scritti*, *stampati*, the same words that appear in the definitions of the above verbs.

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

<i>leggere</i>	187
<i>scrivere</i>	196
<i>pubblicare</i>	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 and Calzolari]).

When analysing a large corpus with millions of words in context, we are in a sense compelled to discover and describe:

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

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

In order to automatize the retrieval of this type of information directly from the corpus we should first be able to tag the corpus for the different POSs. For this task many systems already exist (see e.g. [Hindle, 1988], [Webster and Marcus, 1989]). It should then be possible, even without a complete parser, to apply to the text corpus some pattern-matching procedures (as those we are presently using with dictionary definitions). These pattern-matching procedures should be explicitly geared to the extraction of the type of data we are searching (i.e. prepositional phrases, that-clauses, infinitives, etc.).

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

3 The Lexicographer's Workstation as a Model of Integration of Tools and Data from Different Environments and Expertises

The importance of a collaboration between researchers working in the fields of Computational Linguistics/Natural Language Processing (CL/NLP) and Literary and Linguistic Computing/Text Processing (LLC/TP) is evident when we consider that it is necessary to process large textual corpora in order to achieve better LKBs. The design of these large integrated LKBs can really become the purpose of cooperative projects, where the "typical" data, tools, procedures, knowledge, expertise, results, etc., of the two areas of CL/NLP and LLC/TP "must" work in parallel and cooperate and interact with each other.

In order to achieve at least some of the results outlined so far, we can summarize the needs as follows:

- design and implementation of powerful tools;
- large sets of lexical and textual data;
- very modular systems;
- possibility of sharing resources, data and procedures;
- large cooperation among traditionally different research or industrial communities.

A model of the type of integration we have in mind can be seen in the lexicographer's workstation (LW) we are designing in Pisa (see [Calzolari *et al.*, 1987]). It is conceived as a very modular system, where different types of data and of procedures are integrated. At the level of data the LW contains, or will contain, among other modules:

- a textual data base,
- one or more monolingual lexical databases,
- a thesaurus with taxonomic information,
- bilingual lexical databases,
- a reference corpus

while at the level of procedures, it contains, among others:

- a morphological tool,
- dictionary parsers,
- a hyponym finder,

- an information retrieval system,
- a lemmatization package,
- a pattern-matching procedure for dictionary definitions,
- a redaction tool.

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