Implementing a Bilingual Lexical Database System

Eugenio Picchi, Carol Peters, and Nicoletta Calzolari

Abstract

The current state of progress of a research project for the design and development of a bilingual, Italian-English/English-Italian, lexical database system is presented. The aim is to create an integrated system in which a number of monolingual electronic dictionaries and/or lexical databases can be linked through the medium of a bilingual database. In addition, procedures are being implemented to establish access paths from the dictionary data to archives of texts in machine-readable form and language reference corpora, and vice versa. The system not only provides the standard look-up functions offered by conventional mono- and bilingual dictionaries but the organization of the data on database structures makes it possible to access and exploit the lexical information in many different and new ways. The structuring of the bilingual component is described in some detail and some of the possible applications envisaged for a tool of this type in the fields of pure and applied linguistics, lexicography and language learning are mentioned.

1. Introduction

There is growing awareness that the static structure of the traditional printed dictionary makes it very difficult, and at times impossible, to access much of the wealth of knowledge that it effectively contains. This means that a considerable amount of valuable information concerning the lexical system remains "hidden" and cannot be exploited by the ordinary user. For this reason, increasing attention is being given to the potential of electronic dictionaries and lexical databases which, by offering flexible and dynamic access and storage facilities, make it possible to retrieve and study lexical information in new and interesting ways.

In this context, research is now underway at the "Istituto di Linguistica Computazionale" (ILC), Pisa, into the design and development of a bilingual, Italian-English/English-Italian, lexical database system.[1] The objectives and the initial stages of this project have already been described in detail in (Calzolari and Picchi, 1986) which also discussed how an integrated system of this type should provide new insights into the organization of mono- and bilingual lexical data and into the complex network of relations existing between lexical entries at different linguistic

[1] The project is also supported by the Strategic Project for "Automatic Natural Language Processing" of the Italian National Research Council (CNR).

levels. In the present paper, we will describe the current state of progress of the system and, in particular, the work being done to structure and implement the bilingual component.

An important factor in the development of the system has been the recent rapid advances in information processing and computing technologies as this has meant that it is now possible to store and access large quantities of data rapidly and economically using both mainframe computing systems and powerful personal computers. Indeed, most of our lexical data is now also installed on personal computers with a hard disk storage capacity of up to 120 megabytes, equivalent to 120 million characters. In combination with the PC, we can also use WORM (Write Once Read Multiple) storage devices of up to 200 megabytes.

2. The System Components

2.1 Lexical Data

The system components consist, in the first place, of lexical data actually available at the ILC. This includes the Italian Machine Dictionary (DMI) which is mainly based on the Zingarelli Italian Dictionary, the Garzanti "Nuovo Dizionario Italiano", the Collins Concise Italian/English, English/Italian Dictionary, the Longman Dictionary of Contemporary English (LDOCE) and the Oxford Advanced Learner's Dictionary of Current English (OALD). With the exception of the DMI, the rest of the material was conceded to us by the publishers, for research purposes only, on tapes coded in computer typesetting format. In fact, our philosophy is, where possible, to develop procedures that process and reutilise data which has been prepared in machine-readable form for other scopes. In this way, the prohibitive costs and times which are needed to input dictionary data from scratch can be avoided.

Other important components of the final system will be (i) the extensive archive of texts stored in machine-readable form which has been built up at the ILC over the last twenty years (many of these texts have already been processed and analyzed to varying degrees for particular research purposes), and (ii) the Italian Reference Corpus, which is now being created at the ILC; the Corpus will consist of 20 million words extracted from a collection of texts taken from a wide range of sources to provide a representative sample of contemporaryItalian (Bindi et al, 1989, Saba and Turrini, 1987). Procedures are now being implemented to link these archives to the dictionary data so that both users and application programs can pass easily from dictionary to text and back again. This will mean, for example, that the language learner can pass from a lexical item given in the bilingual database to the equivalent entries in the monolingual components and then, if he so desires, can consult the text archives to see how the item is used in different contexts.

In the future, we hope to have access to similar archives of English texts and to an English Reference Corpus. Work has already begun on procedures to implement and manipulate contrastive corpora.

2.2 Lexical Tools

In addition to the system components, a number of tools have been studied and developed which can be used to process and analyze the lexical data in different ways. These include:
— the DBT, a full-text retrieval system which can be used to directly access all occurrences of a word-form, or co-occurrences of more than one form, in any type of text (Picchi, 1983b; Picchi, 1988). Running on a dictionary considered as a text, the DBT can be used to access words not only as headwords but wherever they appear in the dictionary, e.g. in the definitions or in the examples;
— a morphological analyzer and generator which automatically analyzes and generates all possible word-forms for any Italian lemma;
— a procedure for semi-automatic lemmatization which, when run over a text in machine-readable form, associates each word-form with its base lemma. When the form is ambiguous, the procedure either signals all the possible solutions to the user who must select the most appropriate, or the output must be passed through a disambiguation procedure;
— a contrastive multiligual textual database system which is now being developed; this is a full-text retrieval system working in a multilingual environment. The user specifies his "pivot" and "target" languages to retrieve contrastive concordances for parallel texts in different languages.

3. Structuring the Dictionary Components

3.1 The Monolingual Data

The work which has already been done to structure the monolingual Italian dictionary components is described in detail elsewhere (see Calzolari, Ceccotti, 1981; Calzolari, 1984a). The DMI is now structured as a lexical database (LDB) which offers on-line access to approximately 120,000 lemmas and their inflected word-forms (more than 1 million). A number of coded attributes such as morpho-syntactic categories or phonological information and various types of semantic indicators, e.g. usage codes, specialized terminology, can also be used as search keys to query the database (Calzolari, 1983). A similar structure is now being created for the Garzanti dictionary. Much work is being done to develop dictionary definition parsing procedures in order to create not only taxonomical hierarchies but also many other conceptual relations from the semantic data given in the dictionary definitions. This work is described in detail in (Calzolari and Picchi, 1988). We have now begun work on the English monolingual data; the Longman tape, in particular, is already in a highly coded form.

3.2 The Bilingual Data

As has already been stated, our bilingual data is that contained in the Collins Concise English-Italian/Italian-English dictionary which was made available to us on a tape prepared in computer typesetting format.

The drawbacks implicit in the organization of the conventional printed bilingual dictionary are well-known and include:
— the severe restrictions imposed by the traditional alphabetical ordering of the entries and the use of the headwords as the only access keys;
— the lack of space which imposes limits on the amount of information which can be supplied on either side: this means that the user must often turn to the monolingual dictionary, especially a learner's dictionary if available, to find certain detailed information, e.g. definitions, additional grammatical information, other examples of usage;
— the time needed for the user to search a word, which often involves not just a simple look-up from L1 to L2 but looking-up the suggested L2 translation equivalents in their turn to check their in-context validity;
— the fact that lexical items in a bilingual are defined in terms of their translation equivalents and therefore in reference to the lexical system of L2, whereas, for a more complete knowledge of an item, a definition in terms of its own lexical system is often necessary.

Therefore, when structuring the bilingual data we had two main aims:
(i) to provide an autonomous, dynamic tool which would offer new ways to access the information and new possibilities of exploitation, showing how the potential of the printed dictionary can be enhanced by organizing it on a DB structure;
(ii) to organize the data so that it would be possible to design and perform mapping operations between the bilingual and the monolingual data in order to enable both human users and application programs to move easily and rapidly between the two different lexical systems.

3.3 Working on the Bilingual Data

In the printed dictionary, different categories of text, e.g. phonetic, syntactic or semantic, are usually distinguished by different type-faces and reliance is generally made on the user's intelligence and intuition to interpret the precise value of the information he is given. On the other hand, in a machine dictionary or an LDB, each separate information field must be identified explicitly in order to ensure consistency and to provide direct access to the data. A pre-parsing procedure was thus designed to identify the structure of the bilingual lexical entry from the typesetting codes on the tape and to insert markers to separate and tag the different information fields. The next step was to construct a suitably coded structure to represent the bilingual entry in the LDB. In view of our intention to perform mapping operations between the monolingual and bilingual datasets, this representation structure had also to be compatible with the structure which had already been defined for the monolingual lexical entry. These two representation structures are shown in Fig. 1.

As can be seen from the figure, our aim has been to distinguish precisely all the different information supplied in the lexical entry. In this way, we can facilitate the retrieval of "new" information or information which is difficult to access in the printed dictionary. For example, when working on the bilingual data, we paid a lot of attention to the dictionary metalanguage and, in particular, to the information which is given to help the user to select the most appropriate sense of an item and thus the most suitable L2 equivalent. This information, which is grouped together

by Collins under the heading of "Semantic Indicator", is quite varied and the user is given little assistance to recognize exactly what help he is being given; he is expected to be able to distinguish more or less intuitively between subject field labels, style and usage registers, typical collocates, synonyms or near synonyms, superordinates, etc. We developed procedures to disambiguate, automatically as far as possible, the information contained in this field (see Picchi, et al., 1988). The results of these procedures are shown in Fig. 2 in which the printed dictionary entry for the Italian noun **accento** can be compared with the entry for the same data in the bilingual LDB. Once this type of information has been disambiguated as shown in the figure, it can be exploited in various ways. For example, we can retrieve lists of all entries which are associated with a given Field Label, or all entries for which a given word has been tagged as a typical collocate, or all words which have been tagged as near synonyms for a given word, etc.

In a similar way, we worked on the information given in the examples of usage associated to many of the entries. We have isolated a number of example types which can be classified automatically. These include examples with associated style registers or field labels; particular grammatical patterns; prepositional government; etc. Thus, we can produce lists of all verbs which take a particular preposition, or all examples with a particular grammatical pattern; etc. In this way, much previously unaccessible information can be identified and classified and will be available for future studies and applications.

3.3.1 Using the DBT

We have already observed that the printed dictionary contains much "hidden" data, i.e. information which it is difficult to access. In order to capture such information, the parsed bilingual data was next structured in DBT form (see Section 2.1) so that all the occurrences of a word-form can be found wherever they appear in the dictionary, e.g. in definitions, in examples of usage, as translations which are not listed on the other side of the dictionary as headwords. For Italian, we can use the morphological analyzer (see Section 2.1) in combination with the DBT so that a lemma can be searched together with all its associated forms. We intend to implement a morphological analyzer for the English data, too. Using the DBT, much valuable information can be accessed which otherwise could never have been retrieved.

An example of how the DBT can be used to search all the occurrences of the English adverb **aback** throughout the dictionary is given in Figure 3. On the L1 = English side we have just one occurrence of **aback** with no direct translation (⟨Tr⟩NDT) but **to be taken aback** is given as an example of usage with three suggested translations. This is, in fact, all the information which the user is able to access directly in the printed dictionary for **aback**. However, using the DBT, we find that on the L1 = Italian side of the dictionary we have one occurrence of **aback** in **taken aback** which is given as a translation for the headword **frastornato** in its figurative sense and other occurrences, in each case collocated with **taken**, in translations of examples of usage under six, very different, headwords (aperto, bocca, cadere, cascare, interdetto and nuvola). This provides a lot of additional and useful

information to both encoding and decoding users, in particular making it clear that **aback** can only be used safely when it is collocated with **taken**. Such information could not have been accessed without the help of the DBT.

3.3.2 Normalizing the Bilingual Data

The bilingual dictionary is composed of two distinct datasets, each of which serves two types of users, as follows:

1st Dataset	L1 (Italian) = Source Language L2 (English) = Target Language
	Italian user encodes English user decodes
2nd Dataset	L1 (English) = Source Language L2 (Italian) = Target Language
	English user encodes Italian user decodes

These datasets are not symmetric partly because space restrictions impose certain choices on the compiler, e.g. phonetic information is only provided for source language entries and not for target language translations and the encoding user who needs such information is obliged to go to the other side of the dictionary, and partly because the same information is not always useful to the same extent on both sides, e.g. **teacake** is translated as **panino dolce all'uva** but there is no equivalent entry on the Italian side precisely because it is extremely unlikely that an Italian user would need to encode this expression or that an English user would need to decode it — in fact, **panino dolce all'uva** is not really a direct translation but an explanation of a very English food item.

To a certain extent, the DBT system can be used to resolve some of these problems, e.g. an L2 translation which does not appear as an L1 headword and thus cannot be decoded by the L1 user can be immediately retrieved in the DBT. However, we are now studying procedures to normalize the bilingual data. Information which is at present available on just one side (e.g. phonetic or morpho-syntactic information) must be integrated and new access keys will be created to cater for target language translations which do not appear as source language data. In this way, there will no longer be two distinct datasets, and look-up procedures will be independent of source and target languages. For any word searched, all the information on this word, wherever it is stored in the dictionary, will be retrieved.

The data will then be organized into an interactive work-station where the user can request information on lexical items on an as-needed basis, restricting and expanding his query as he desires to specify his own particular search and display profiles.

3.3.3 Mapping Between Monolingual and Bilingual Data

The problem which we are now tackling is to link the bilingual data to the two sets of monolingual data. This is a complex task as in most cases we are not faced with a simple **1:1** mapping between lexical items but generally with **1**:many or many:many relationships.

An idea of a **1**:many sense mapping and of how important information is often missing in the bilingual entry is given by the word **nipote** which, in Italian, is used indifferently for grandchild/son/daughter and nephew/niece. In the bilingual dictionary entry, the Italian encoding user is not given too much information.

```
Entry  = nipote
Pron   = |     |
POS    = sm/f
         SI context    = di nonni
         Trans         = grandson/daughter, grandchild
         SI context    = di zii
         Trans         = nephew/niece
```

Bilingual DB entry for NIPOTE

He has to go to the other side (L1 = English) to be sure, for instance, that **nephew** is the translation of **nipote** (di zii) **m**, and **niece** of **nipote** (di zii) **f**. He also has to go to the English side to disambiguate grandchild/son/daughter and to discover the irregular plural of grandchild which is not even hinted at in the L1 = Italian entry. The English entry for grandson and granddaughter also gives the Italian diminutives **nipotino** and **nipotina**. These words do not appear as headwords on the Italian side and it is necessary to consult an Italian monolingual dictionary to find them effectively indicated as diminutives. The monolingual also gives a third meaning for the word **nipote** when used in the plural (nipoti = discendenti). The bilingual ignores this use and gives no assistance to the English user who has to decode **nipoti** in this sense.

However, in our opinion, at times it is the bilingual sense division, made on the basis of differences in use between the two languages which are put into relationship, which throws extra light on the monolingual entry. To give a very simple example, in the bilingual dataset, the English noun **hair** is given three sense divisions and three different translations in Italian, i.e. **hair** collective: of person = **capelli** mpl (the examples make it clear that **capelli** is used for hair on the human head); on body = **peli** mpl; of animal = **pelo**, whereas the definition in an English monolingual is "...all the thread-like growths on the skin of animals, esp. on the human head". It is clear from this that what in English is regarded as essentially the same phenomenon is considered differently in Italian.

We feel that automatic procedures can be developed to deal with **1:1** sense mapping and perhaps also in cases of **1**:many relationships of the nipote: grandchild/son/daughter, nephew and niece type. In cases of complex sense division, we hope that the semantic constraints used to define and discriminate between the different translations of a lexical item in the bilingual should be of help to us in the design of procedures to achieve at least a partially automatic sense mapping with

the monolingual. Fig. 4 gives an idea of how the semantic indicators of the bilingual entry for **dope** can be used to help to link the sense divisions and corresponding translations in the bilingual to those of monolingual entries. However, we are sure that in such cases considerable manual intervention will be necessary to create all the data links.

Once the mapping operations have been completed, we will have created a first prototype of a bilingual lexical database system, which will be implemented both on mainframe and personal computers and will be accessible on-line for interactive use.

4. Applications

We feel that a bilingual system of the type we are implementing has considerable potential in a wide number of fields. Two applications which particularly interest us are the creation of translator and lexicographer workstations in which, even though in different ways, the user can have on-line access not only to bilingual and monolingual dictionary data but also to on-line reference corpora and to textual archives in the two languages, so that he can call up a large number of in-context examples of usage for any lexical item. Of course, a system of this type should not only be useful for human translating or dictionary compiling but for other tasks such as language learning, machine(-aided) translation systems and contrastive linguistics.

References

Cited Dictionaries

OXFORD ADVANCED LEARNER'S DICTIONARY OF CURRENT ENGLISH (OALD). 1974[3], 1980. A.S. Hornby (ed.). Oxford: Oxford U.P.

COLLINS CONCISE ITALIAN DICTIONARY: Italian-English/English-Italian. 1985. London and Glasgow: Collins/Firenze: Giunti.

IL NUOVO DIZIONARIO ITALIANO GARZANTI. 1984. Milano: Garzanti.

LONGMAN DICTIONARY OF CONTEMPORARY ENGLISH (LDOCE). 1978. Paul Procter et al. (eds.). Harlow and London: Longman.

VOCABOLARIO DELLA LINGUA ITALIANA. 1970. N. Zingarelli (ed.). Bologna: Zanichelli.

Other Literature

Bindi, R., M. Monachini, and P. Orsolini. 1989. Corpus of Contemporary Italian Language. Technical report, ILC—TLN, Pisa, 1989.

Byrd, R.J., N. Calzolari, M.S. Chodorow, J.L. Klavans, M.S. Neff, and O.A. Rizk. 1987. *Tools and Methods for Computational Lexicology*. Special Issue of Computational Linguistics on the Lexicon, 1987.

Calzolari, N., M.L. Ceccotti. 1981. 'Organizing a Large Scale Lexical Database' in *Actes du Congres International Informatique et Sciences Humaines*. Liege: L.A.S.L.A., November 1981. 275—288.

Calzolari, N. 1983. 'Semantic links and the dictionary' in S.K. Burton and D.D. Short (eds.). *Proceedings of the Sixth International Conference on Computers and the Humanities*. Rockville (Maryland): Computer Science Press, 47—50.

Calzolari, N. 1984a. 'Detecting patterns in a lexical data base' in *Proceedings of* COLING84. Stanford University (Calif.), Association for Computational Linguistics. 170—173.

Calzolari, N. 1984b. 'Machine-readable dictionaries, lexical data bases and the lexical system' in *Proceedings of* COLING84. Stanford University (Calif.), Association for Computational Linguistics. 460.

Calzolari, N. and E. Picchi. 1986. 'A Project for a Bilingual Lexical Database System' in *Proceedings of the Conference "Advances in Lexicology" of the UW Centre for the New Oxford English Dictionary*. Waterloo, Canada, Nov. 1986. 79—92.

Calzolari, N. and E. Picchi. 1988. 'Acquisition of Semantic Information from an On-Line Dictionary' in D. Vargha (ed.). COLING *Budapest Proceedings*, Budapest: J. von Neumann Society for Computing Sciences. 87—92.

Gruppo di Pisa. 1979. 'Il Dizionario di Macchina dell'Italiano" in D. Gambarara, F. Lo Piparo, G. Ruggiero (eds.). *Linguaggi e Formalizzazioni*. Roma: Bulzoni. 683—707.

Nagao, M. 1986. 'Dictionaries for Machine Translation', presented at the Workshop *'On Automating the Lexicon'*, Grosseto, May 1986.

Neff M.S., R.J. Byrd, and O.A. Rizk. 1987. *Creating and Querying Lexical Data Bases*. IBM, Research Report, RC 13347 (# 59792).

Picchi, E. 1983a. 'Problemi di documentazione linguistica. Archivio di testi e nuove tecnologie' in *Studi di Lessicografia Italiana* 5: 243—252.

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

Picchi, E. 1988. *D.B.T.: A Textual Data Base System: Methods and Tools for Searching Full-Text Data*. ILC Report.

Picchi, E., C. Peters, and N. Calzolari. 1988. 'A Tool for the Second Language Learner: Organizing Bilingual Dictionary Data in an Interactive Workstation', Proceedings of the ALLC Conference, Jerusalem (Israel), June 1988.

Saba A. and G. Turrini 1987. *Progetto per la Costruzione di un Corpus di Riferimento dell'Italiano Contemporaneo*. ILC—TLN, 1987—01.

GRAPHICAL WORD =

ENTRY =

HOMOGRAPH = n
PRONUNCIATION =
PARADIGM LABEL =

POS =
SYNTACTIC CODES =
USAGE LABEL =
POINTERS to base-lemma and derivatives =
POINTERS to graphical variants

SENSE = n
FIELD LABEL =
SYNTACTIC CODES =
DEFINITIONS =
POINTERS TO SYNONYMS =
POINTERS TO ANTONYMS =
POINTERS TO HYPONYMS, HYPERONYMS =
POINTERS to other entries =
SEMANTIC (inherent) FEATURES =

EXAMPLE =
FIGURATIVE, RARE, ... =
DEFINITIONS OF PARTICULAR USAGE =
IDIOMS =
CITATIONS =
PROVERBS =

Representation Structure for a Monolingual Entry

GRAPHICAL WORD =
SOURCE LANGUAGE =
ENTRY =
HOMOGRAPH = n
PRONUNCIATION =
SPELL.DIVISION =
POS =
POS.SUBCATEGORY =
OTHER GRAM.INF =
(IRREG.)MORPHOLOGY =
AUX =
CROSS-REFERENCE =
SENSE = n
FIELD LABEL =
USAGE =
STYLE =
SYN/HYPERNYM =
CONTEXT/COLLOCATION =
TRANSLATION =
TRANS.GRAM.INF =
TRANS.SI =
EXAMPLE =
EXAMPLE OF USAGE =
GRAMMATICAL PATTERN =
PREPOSITIONAL GOVERNMENT =
ESCLAMATIVE =
PHRASEOLOGY =
EXAMPLE.SI =
EXAMPLE TRANSLATION =

Representation Structure for a Bilingual Entry

Figure 1

accento |aʧ'ʧɛnto| *sm* **(a)** *(pronuncia)* **accent; parla con un ~ straniero he speaks with a foreign accent. (b)** *(Fonetica)* **accent, stress;** *(fig)* **stress, emphasis; mettere l'~ su qc to stress sth. (c)** *(segno grafico)* **accent; ~ grave/acuto/circonflesso grave/acute/circumflex accent. (d)** *(inflessione)* **tone (of voice); un breve ~ di tristezza a slight note of sadness.**

Printed bilingual dictionary entry for "accento"

```
Entry = accento
  SL = Italian
  Pron = |
    POS = sm
      Sense = a
        SI syn = pronuncia
      Trans = accent
          Ex = parla con un ** straniero
          ExTrans = he speaks with a foreign accent
      Sense = b
        SI FL = Fonetica
      Trans = accent, stress
        SI Styl = fig
      Trans = stress, emphasis
          Ex = mettere l'** su qc
          ExTr = to stress sth
      Sense = c
        SI syn = segno grafico
      Trans = accent
          Ex = ** grave/acute/circonflesso
          ExTr = grave/acute/circumflex accent
      Sense = d
        SI syn = inflessione
      Trans = tone (of voice)
          Ex = un breve ** di tristezza
          ExTr = a slight note of sadness
```

Bilingual LDB entry for "accento" with Semantic Indicator disambiguation

Figure 2

D.B.T. (E. Picchi) Collins Bilingual English-Italian Parola: ⟨Entry⟩ **ABACK** Frequenza: 1

1) **ABACK** ⟨POS⟩adv ⟨Tr⟩NDT ⟨Ex⟩to be taken ** ⟨ExTr⟩essere colto(a) ⟨or⟩ preso(a) alla sprovvista, rimanere sconcertato.

D.B.T. (E. Picchi) Collins Bilingue Italiano-Inglese Parola: ⟨Tr⟩ **ABACK** Frequenza: 1

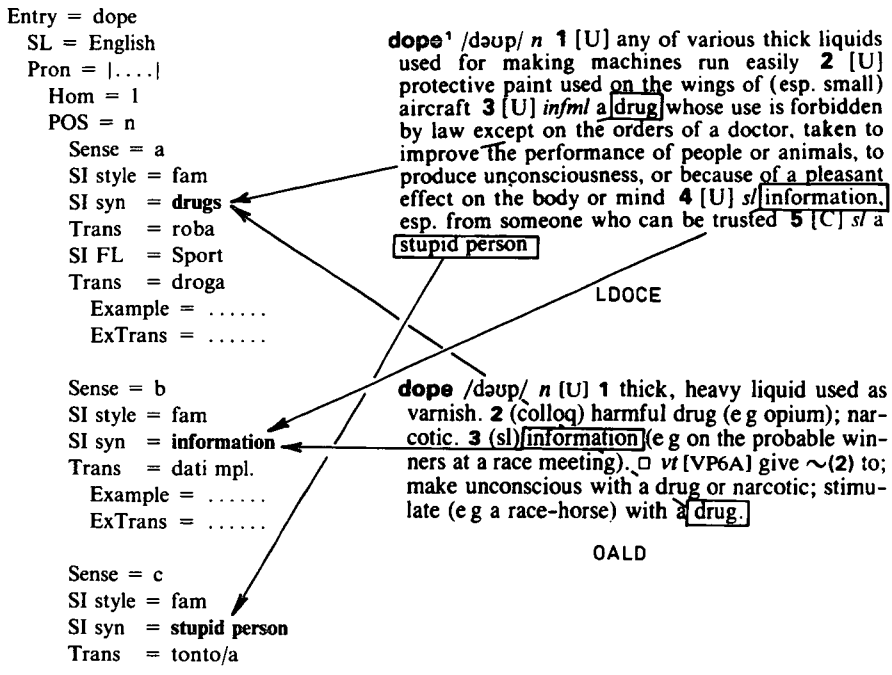
1) FRASTORNATO ⟨POS⟩ag ⟨Tr⟩deafened; ⟨SI styl⟩fig ⟨Tr⟩taken aback

D.B.T. (E. Picchi) Collins Bilingue Italiano-Inglese Parola: ⟨ExTr⟩ **ABACK** Frequenza: 6

1) APERTO ..⟨Ex⟩a cuore ** ⟨Styl⟩fig ⟨Extr⟩frankly, sincerely ⟨Ex⟩a bocca **a ⟨ExTr⟩open-mouthed; ⟨Ex⟩rimanere a bocca **a ⟨Styl⟩fig ⟨ExTr⟩to be taken **aback;** ⟨Ex⟩all'aria **a ⟨ExTr⟩in the open air; ⟨Ex⟩all'** ⟨POS⟩av ⟨ExTr⟩outdoors; ⟨***⟩ ⟨ExTr⟩open-air ⟨***⟩ ⟨SI⟩giochi ⟨SI⟩vacanze ⟨ExTr⟩ outdoor..
2) BOCCA ..⟨ExTr⟩orally; ⟨Ex⟩rimanere a ** asciutta ⟨ExTr⟩to have nothing to eat ⟨Styl⟩fig ⟨ExTr⟩to be disappointed ⟨Ex⟩rimanere a ** aperta ⟨Styl⟩ fig ⟨ExTr⟩to be taken **aback;** ⟨Ex⟩non ha aperto ** ⟨Si⟩parlare ⟨ExTr⟩ he didn't open his mouth; ⟨Ex⟩vuoi chiudere la **? ⟨SI⟩star zitto ⟨ExTr⟩will you shut up?; ⟨Ex⟩essere sulla..
3) CADERE ..to knock over o down; ⟨Ex⟩fa sempre ** tutto dall'alto ⟨ExTr⟩ he does everything as if it were a great favour; ⟨ExTr⟩** dalle nuvole ⟨ExTr⟩to be taken **aback** ⟨Ex⟩la conversazione cadde ⟨ExTr⟩the conversation died ⟨Ex⟩la conversazione cadde su Garibaldi ⟨ExTr⟩the conversation came round to Garibaldi; ⟨Ex⟩questi pantaloni cadono bene ⟨ExTr⟩these trousers hang..
4) CASCARE ..** dalla fame ⟨ExTr⟩to be faint with hunger ⟨Ex⟩** dal sonno ⟨ExTr⟩to be falling asleep on one's feet; ⟨Ex⟩** dalle nuvole ⟨Styl⟩fig ⟨ExTr⟩to be taken **aback;** ⟨Ex⟩** bene/male ⟨Styl⟩fig ⟨ExTr⟩to land lucky/unlucky ⟨Ex⟩gli ho detto che tu eri partito e lui c'e' cascato ⟨ExTr⟩I told him you..
5) INTERDETTO ⟨Hom⟩1 ⟨POS⟩pp di interdire ⟨Hom⟩2 ⟨POS⟩ag ⟨SI⟩sconcertato ⟨Ex⟩rimanere ** ⟨ExTr⟩to be taken **aback;** ⟨Ex⟩lascia|re qn ** ⟨ExTr⟩to disconcert ⟨Hom⟩3 ⟨POS⟩sm ⟨FL⟩Rel ⟨FL⟩Dir|⟨Tr⟩interdict.
6) NUVOLA sf ⟨Tr⟩ cloud ⟨Ex⟩avere la testa fra le **e ⟨ExTr⟩to have one's head in the clouds; ⟨Ex⟩cascare dalle **e ⟨ExTr⟩to be astounded, be taken **aback.**

DBT used to search all occurrences of "aback"

Figure 3



Mapping between bilingual and monolingual entries

Figure 4