Abstract
The main aim of this paper is to describe a project aimed at providing Italian learners of English with a collection of English multiword units (MWUs) which they may learn with relative ease. The MWUs in question are those which have similar forms and meanings in Italian and English. Over 2,400 pairs of such items have been inserted into a database. Many different types of MWU are involved since the cross-language phenomenon of phraseological similarity cuts across traditional categories. As regards the language pair Italian-English, it applies, for example, to both idiomatic and non-idiomatic phrases, and to both sentence-level and sub-sentence-level items. The items in question are not usually a problem for the learner from the point of view of receptive skills. The pedagogical usefulness of presenting learners with a collection of such phrases lies, above all, in the fact that they will readily become available for production skills, assuming that any eventual differences in usage have been highlighted.

The paper is divided into 4 sections. The first consists of a description of the phraseological phenomenon in question. The second explains the pedagogical usefulness of the project. The third briefly describes aspects of data collection and management. And lastly, an outline is given of expected lexicographical presentation.

1 Introduction
This paper outlines the salient features of a project whose main objective is to help Italian mother-tongue learners of English increase the repertoire of multiword units at their disposal when engaged in production skills. The culmination of the project, now in its final stages, will be the presentation in book form of approximately 1500 English multiword units (MWUs). The selection of MWUs is not based primarily on the presumed usefulness to the learner, but on the ease with which it should be possible for students to learn the items.

The English MWUs to be presented are definable only within the context of interlingual comparison and are perhaps best described with reference to a 'phenomenon' rather than in terms of any particular phrasal type or types. The phenomenon in question is that of the existence of MWUs of similar form and meaning in a given pair of languages, in this case English and Italian. Phrases of this sort have been studied mostly with reference to those items which are loan translations, or 'calques'. Thus, for example, the Italian cortina di ferro derives from the English iron curtain (see Klajn [1972: 133] and Rando [1987: 37]). Parallel MWUs have also been described in terms of synchronic contrastive description, whether of a semantic, functional or formal nature, though analysis is usually limited to a particular phrasal type, for example 'idiomatic expressions' or 'proverbs'. Some such description is to be found within Cragie [1984], Dobrovolskij [2000], Jernej [1982], and Negreanu [1975].
2 ‘Phrasal friends’ and foreign language learning

As well as being of interest both from a historical point of view and from the viewpoint of synchronic contrastive description, the existence of comparable MWUs in a given language pair is also of potential interest within the applied field of foreign language learning. It is my belief that ‘phrasal friends’, as they might be termed, can provide an excellent source of input for the foreign language learner, all the more so within a language pair such as Italian and English which have a large number of comparable MWUs. The database of phrasal friends which I have set up contains in excess of 2,400 pairs of items, and very few of these belong to the probably larger set of specialist noun compounds pertaining to scientific and technological sub-disciplines.

2.1 The relevance of ‘phrasal friends’ to foreign language learning

The basic concept behind the presentation of phrasal friends is that if students are familiar with the L2 lexical items $A$ and $B$, then they will very quickly learn the phrasal item $A+B$ if a comparable MWU exists in their mother tongue, and thereafter have the phrase available for production purposes. Thus, for example, the Italian learner who knows the words Christmas and tree will very readily learn the MWU Christmas tree since its lexical components are mirrored in the Italian phrase albero di Natale. There will sometimes be cases where the individual learner is not familiar with both or all items in a given MWU. Such items, however, are still useful since familiarity with the L1 MWU and the latter's strong resemblance to the L2 MWU should facilitate learning of both the L2 MWU and the unknown single word within it. A potential example of this is the phrasal pair un piede nella fossa / one foot in the grave, where the Italian learner of English might not be familiar with the word grave.

Phrasal friends have so far received little treatment in the language teaching literature, at least as regards English teaching. This is probably for two main reasons. Firstly, much of the literature concerned with EFL teaching discusses L1-independent issues. In this respect, it is worth quoting the ELT author Michael Swan who, while discussing the “British version of the Communicative Approach”, comments that “students might as well not have mother tongues. Meanings, uses, and communication skills are treated as if they have to be learnt from scratch ..... This is a peculiar state of affairs. It is a matter of common experience that the mother tongue plays an important part in learning a foreign language” [Swan 1985: 85]. For some discussion of the relevance of contrastivity to second language acquisition, see Sajavaara [1996].

A second reason for phrasal friends receiving very little attention in the literature is that, where contrastive approaches are given importance, the items I am discussing are probably judged to constitute a somewhat banal phenomenon. They present little difficulty as regards comprehension, and some of the phrases are also relatively easy to learn from the point of view of production. Their consequent lack of importance may be seen as part of a more general state of affairs whereby that which is similar in the L2 tends to be ignored, while that which is different is emphasized. It is worth pointing out in this respect that the affinity between phrasal friends is important not only in relation to lexical acquisition but also from a psychological point of view. Phrasal friends help the foreign language to lose some of its ‘foreignness’, and thus appear more welcoming to the language learner.
Although phrasal friends have not received a great deal of attention in the literature, there are a few studies which may be mentioned, and there may well be others with which I am not familiar, since much contrastive literature is published within specific L1 countries. One author who has written a number of studies dealing with the phenomenon is Suzanne Irujo, [1984, 1986a, 1986b, 1993], who discusses comparable Spanish and English idiomatic expressions in the context of language learning. Cornell [1999: 9-10] also mentions the phenomenon in relation to idiomatic expressions. Furthermore, it is worth noting that quite a number of authors discuss a very similar phenomenon within the context of second language learning, that of single-word cognates. See, for example, Granger [1993] and Laufer [1990].

The presentation of phrasal pairs is also to be seen in relation to the importance now given to the role of phraseology in both L1 language use and L2 acquisition. Any approach is useful which can help foster an awareness on the part of the language learner of the importance of MWUs and of the various forms they can take. A user-friendly approach, such as the one suggested, is perhaps even more important. For some discussion of the relevance of phraseology to second language learning, together with further bibliographical references, see Howarth [1998], Nattinger & DeCarrico [1992], and Weinert [1995].

2.2 The types of MWU involved

Intuition suggests that for the L2 phrase to be learnt quickly, through assimilation with the comparable L1 phrase, the MWU in question should consist of at least two content words. Apart from this constraint, any type of MWU is of potential interest. In my own database, some of the items are semantically opaque (e.g. rompere il ghiaccio / break the ice) while others are relatively transparent (e.g. crimine di guerra / war crime). Some consist of one or more cognate lexical items (e.g. intelligenza artificiale / artificial intelligence), while others consist of words of completely different etymology (e.g. caccia alle streghe / witch-hunt). The latter example (witch-hunt) also exemplifies the fact that I have included as MWUs those items which may be written as one single orthographic word, this since their composite nature can be taken advantage of just as well as if they were written as a succession of single words (compare comments by Cermák [2001: 8]). Some of the items on the database are represented as single words in both languages, e.g. madreperla / mother-of-pearl.

A further point regarding the heterogeneity of the MWU pairs relates to ‘syntactic length’. Whereas most items on the database usually function as sentence constituents, as in the various examples already given, there are also a considerable number of sentence-length items (e.g. Tutte le strade portano a Roma / All roads lead to Rome), as well as a small number of text units which stretch beyond a single sentence. A final example I will give of the different types of MWU involved is that of proper names (as opposed to ‘common’ units of language), for example the names of places, e.g. il Ponte dei Sospiri / the Bridge of Sighs. The range of phrasal types included in the planned volume thus contrasts with most pedagogical works dealing with phraseology, since, most commonly, only one type of MWU is dealt with, and that one type is usually ‘idiomatic expressions’.

Finally, it should be pointed out that not all pairs of items on the database will be included in the volume in preparation. Some items have been omitted because they are too specialized in nature (e.g. teoria dei giochi / games theory). Other items have been excluded because they
are too easy for the intended readership (e.g. “Buona sera” / “Good evening”). I should state at this point that the work being compiled is aimed at learners of an intermediate level of English or above. This is an obvious requisite since a certain degree of familiarity with English vocabulary and syntax will be necessary in order to take advantage of the main pedagogic principle underlying the book. With regard to the omission of items judged to be too easy, it is interesting to note that there were, in fact, very few phrases involved. Perhaps this can be seen as a reflection of the fact that the average second language learner has an underdeveloped knowledge and awareness of phraseology.

2.3 Equivalence, comparability and choice

As has often been pointed out, notably in the context of bilingual lexicography, there are relatively few pairs of lexical items which may be deemed to be absolutely equivalent in a given language pair. This fact, however, does not prejudice the work in hand, and this for a number of reasons. Firstly, a considerable number of the MWU pairs examined do, in fact, appear to be equivalent to a very high degree. Secondly, where it is possible to establish precise differences between a given pair of items, such differences will be explained or illustrated in the planned volume, assuming this may be done with relative clarity and brevity. Thirdly, it is unrealistic and counterproductive in the context of second language learning to give too much importance to the notion of absolute equivalence. If one did, L1-L2 vocabulary lists of any sort—whether lists appearing in publications or lists drawn up by students in the course of their studies—would be very short indeed, either that or full of annotations.

With regard to the lack of absolute equivalence, it should be remembered that if a learner fails to use a given item in a completely appropriate way, this fact should neither surprise nor shock us. The process of lexical acquisition—in the sense of both knowing how to use a given item and actually using that knowledge correctly when engaged in language production—can be a long one, especially in the case of complicated items. ‘Knowing’ a given multiword unit may mean being familiar with a wide range of syntagmatic and paradigmatic constraints on its usage. Almost inevitably, interlanguage will contain traces of the ‘foreign accent’ in all component parts of the L2, and that includes lexis and phraseology.

Another aspect of the comparability of phrasal friends is the presence or absence of paradigmatic choice. That is, the two items in a given pair may not be in a one-to-one relationship with each other. Considering the Italian item (the L1) as our starting-point, if the pair of items is in a one-to-'more-than-one' relationship, indications will sometimes be given of alternative ways of expressing the Italian MWU in English, together with indications of any eventual circumstances which may result in one alternative being preferred. This type of information will be given when it appears important to do so, and when it is possible to do so with relative clarity and brevity.

3 Data collection, management and analysis

The majority of phrases were found by reading through Italian lexicographical works and noting those phraseological items that had similar form and meaning to English phrases. Some of the main works consulted were [Hall et al. 1981], [Lurati 1990], [Pittâno 1992],
A number of additional phrases were noted down in the course of everyday contact with the language. Where doubts existed regarding the comparability of meaning and usage, on the English side of the equation corpus-based dictionaries were consulted, and on the Italian side, native speakers were consulted, as well as a corpus of contemporary Italian.

All pairs of items found were recorded on a database, together with a substantial amount of additional information. Some of this was relevant to the present project, while other data is being used for other purposes. It should be pointed out in this respect that the 'phrasal friends' being discussed in this paper form part of a larger collection of lexical and phraseological items which are being studied for a number of different purposes.

The database fields of most relevance to the current project relate, on the one hand, to the MWUs considered as items in the lexicon, and, on the other hand, to the same items functioning as contextualized units of meaning. With regard to the former, some of the more important types of information included in the database relate to: (i) meaning, (ii) degree of semantic transparency (in relation to component parts of the MWUs), (iii) part of speech, (iv) syntactic length (whether single word, phrase, clause or sentence-length), (v) type of sentence-length item (proverb, quotation, routine formula, etc). 'Meaning' is recorded with reference to about 340 semantic categories listed in a separate part of the database. These are further grouped into semantic areas of a more general nature. An example of the latter is "War & Peace", while an example of the former is "military weapons".

Fields relating to 'contextualized units of meaning' include the following information: (i) the number of Italian corpus examples, (ii) corpus extracts containing the Italian MWUs in question (extracts are usually about 80 words in length), (iii) probable English translations for the Italian items on the basis of at least some of the corpus extracts (where a sufficient number of tokens of a given item were found in the corpus).

Finally, it should be pointed out that the database is being used to store and manage data, but not to automatically generate the work in preparation, as will be apparent from the next section, in which I outline the structure of the work and some aspects of its contents.

4 Lexicographical presentation

4.1 Overall structure

The volume in preparation will bear characteristics of a thesaurus, a dictionary and a 'normal' book. The most evident of these three will, perhaps, be that of the thesaurus, since the phrases presented will be grouped into a number of thematic sections, and most of these will be semantic in nature. There will, for example, be a section dealing with phrases from the semantic field of 'Religion'. Some sections, however, will be organized around other linguistic principles. For example, there will be one devoted to 'Proverbial sayings and Quotations' and a short section devoted to 'phraseological false friends' (pairs of items such as servizio civile and civil service). The dictionary aspect of the work lies in key-word indexes for the two languages which may be consulted to see whether a specific word or phrase is present.
The 'normal book' aspect of the work lies in the fact that the various thematic sections will consist partly of expository text. Within that text, there will be a number of shorter or longer lists of items, either within the flow of the text or as separate boxes, according to the length of the lists. Presentation, then, will not be as schematically regular as it is in most lexicographical works. This, however, is a problem for the compiler, not for the reader. Different phrasal pairs, or types of phrasal pair, require different types of treatment, and to homogenize presentation would be to do a disservice to the language learner.

A final point to make regarding macro-structure is that the various thematic sections will be preceded by an introductory chapter in which the various interlingual phenomena are described and exemplified.iii

4.2 Aspects of the presentation of individual phrasal pairs

A general distinction will be made in presentation between items which are more or less equivalent and in a one-to-one relationship, and those which are not. Some phrasal pairs will, in fact, be quite straightforward to present. It is worth pointing out in this respect that over half the English MWUs to be included in the volume are noun phrases, and that, at least from a syntactic point of view, most of these should present little difficulty for the language user. This applies not only to NPs with physical referents, e.g. pesce spada / sword fish, but also to many NPs which are more abstract in nature, e.g. conferenza stampa / press conference. It even applies to some semantically opaque NP idiom pairs, e.g. scheletro nell'armadio / skeleton in the cupboard.

A considerable number of NPs were also judged to be in a one-to-one relationship with their English counterparts. This is the case, for example, of many items which denote aspects of the natural world, for example polo nord / north pole. Interestingly, some idiomatic phrases of a much more abstract nature were also invariably best translated—in the extracts examined—by their 'phrasal friend'. This was the case, for example, of the phrase prendere il toro per le corna (to take the bull by the horns).

Elsewhere, different features of the English MWU, or of its relationship with the Italian item, need to be commented on. The following is a list of some of the types of information which will be included whenever thought useful:

(i) contextualized examples will be given for many sub-sentence verbal phrases. This will partly be with reference to item-specific aspects of usage, but also as a way of re-inforcing typical behaviour of English verbal phrases as opposed to their Italian counterparts, for example the use of the definite article instead of possessive adjective (e.g. prendere il coraggio a due mani / to take your courage in both hands).

(ii) there will sometimes be the opportunity for vocabulary expansion. For example, along with the English word honeymoon (luna di miele) will be presented the item honeymooner, for which their is no formal correspondent in Italian (in this particular example reference will also be made to the Italian phrase viaggio di nozze). Sometimes, the 'new' word presented will have the same form but constitute a different part of speech. For example, the noun
phrase *lista nera* / *blacklist* has a verbal counterpart in English (*to blacklist*) but in Italian it is necessary to introduce a collocating verb.

(iii) it will sometimes be necessary to explain individual items within the English MWU. Examples are the unusual word *wailing* found in the phrase the *Wailing Wall* (*il Muro del Pianto*) and the word *gauntlet* found in the phrase *to throw down the gauntlet* (*gettare il guanto*). This 'necessity' is, of course, also a virtue since it offers a further opportunity for vocabulary expansion.

(iv) historical and etymological information will be included where it is felt that it might be of interest to the reader. Thus, for example, the origin of the phrase *to be in seventh heaven* (*essere al settimo cielo*) will be provided. This historical dimension may also help bind the phrase in the memory. With regard to the relevance of etymology (in a broad sense) to foreign language learning, see Ilson [1983].

(v) where there are clearly different ways of expressing a concept in English, these will sometimes be listed as possible alternatives. For example, the concept expressed by the already cited phrase *to be in seventh heaven* may also be expressed by such phrases as *to be over the moon*, *to be on cloud nine*, and *to be absolutely delighted*. Indications will be given of such phrases, and, where thought necessary, examples of usage.

(vi) occasionally, distinction will be made between British and American English. An example is the already mentioned phrase *skeleton in the cupboard*, where *closet* would tend to be used in American English instead of *cupboard*.

5 Concluding remarks

By way of conclusion, I would like to summarise the additional things the book has to offer, in addition, that is, to its primary purpose of providing the reader with a learnable 'vocabulary list'. Firstly, it will help the learner to appreciate the ubiquity and importance of multiword units in English. Secondly, it will provide a bridge between the two languages and cultures, thus rendering the foreign language 'less foreign'. Thirdly, since the collection of phrasal friends is very complete, the volume implicitly distinguishes between those Italian MWUs which have a corresponding phrase in English and those which do not. Fourthly, the different ways in which phrasal items will be presented will re-inforce the fact that some L2 items exist in a one-to-one relationship with L1 items but that others do not. And finally, the work should also prove a useful starting point for researchers interested in studying particular aspects of subsets of the phrasal pairs presented.

Acknowledgements

I would like to thank Prof. Antonio Zampolli of the Institute of Computational Linguistics in Pisa for permission to consult the Italian corpus mentioned in this study, as well as Dr. Laura Cignoni of the same Institute for help given at various stages of the project.

References

REPORTS ON LEXICOGRAPHICAL AND LEXICOLOGICAL PROJECTS


Endnotes

i Loan translation is defined by Trask [2000] in the following way: “A word or phrase constructed by taking a word or phrase in another language as a model and translating it morpheme by morpheme…” Some works dealing with, or including discussion of, loan translations are the following: [Carstensen 1988], [Delmay 1990: passim], [Gusman 1983: 3-35, 59-70], [Haugen 1950], [Nicolas 1994], [Weinreich 1968, 50-53], and, with specific reference to the influence of English on Italian, [Dardano 1986], [Klajn 1972: 108-154], [Pulcini 1994] and [Ross 1995].

ii The corpus in question is the Italian Reference Corpus located at the Italian Research Council's Institute of Computational Linguistics in Pisa.

iii Some other Italian-English publications of a phraseological and/or thesaural nature are [Cambridge Word Routes], [Hall et al. 1981], [Harvey & Ravano 1999], and [Take My Word].