ISSN 1798-4769
Journal of Language Teaching and Research, Vol. 3, No. 5, pp. 1051-1060, September 2012
© 2012 ACADEMY PUBLISHER Manufactured in Finland.
doi:10.4304/jltr.3.5.1051-1060

Different Aspects of Exploiting Corpora in Language Learning

Tayebeh Mosavi Miangah
English Language Department, Payame Noor University, Iran
Email: mosavit@pnu.ac.ir; mosavit@hotmail.com

Abstract—**Corpus-based studies have provided an accurate description of language, and its new potentials for language structure and use have many applications in language learning. This paper tries to define corpora and their different types and discuss the contribution of concordancers to language learning and teaching providing some examples. Then we see how a number of areas of language study have benefited from exploiting corpora. These areas of language that have been examined with the help of language corpora include, but are not limited to: language description, lexicography, morphology, syntax, semantics, cultural studies, computer-assisted language learning, English for specific purposes, and translation.**

Index Terms—**concordancers, corpora, corpus linguistics, language learning**

I. Introduction

We live in a world where information is becoming more freely accessible than ever before, however, this information needs to be processed and translated into knowledge. In order to live, work, and learn, traditional methods of information gathering and storing will no longer be sufficient in the coming centuries. So, developing knowledge construction in teaching and learning as well as providing learners with more autonomous or learner-centered opportunities for learning seems to be of great help for the learners. And this can be achieved by means of exploiting modern technology which is regarded as a very effective element in the learning environment in present and the future.

Corpus-based linguistics has provided an accurate description of language, and its new potentials for language structure and use have many applications in linguistics and language teaching.

In recent years, corpus linguistics has come together with language teaching by recognizing the importance of language corpora as a basis for acquiring facts about the language to be learned and sharing a larger, “chunkier” view of language (Johns, 1991 and McEnery and Wilson, 1997). According to McEnery & Wilson corpus linguistics is a methodology which can be described as a study of natural language on examples of ‘real life’ language use via a *corpus* (McEnery & Wilson, 2001), a body of text that is representative of a particular variety of language.

The accessibility of language corpora provides language learners and teachers with great opportunities in learning a language as well as language analysis with the help of various computer programs in order to reveal many aspects of language use quickly and accurately without any need to manually collect and analyze data. Corpora also provide a wealth of actual rather than made-up examples from different contexts of language use for both teachers and learners. Fortunately, many corpora can be reached freely or at low-cost price (See the appendix.). Working with corpora in language learning has many promising consequences in language descriptions and providing pedagogical materials. Although there have been articles on how teachers with minimal computer resources can make use of corpora (c.f. Johns, 1991a, 1991b; Stevens, 1995; Tribble, 1997a, 2000), few teachers are clear about their nature or their relevance to language teaching. This paper tries to define corpora and their types, discuss their contribution to language learning and teaching providing some examples.

II. Different Types of Corpora

A corpus is a “collection of linguistic data, either written texts or a transcription of recorded speech, which can be used as a starting-point of linguistic description or as a means of verifying hypotheses about a language” (Crystal, 1991, p. 86)

In recent years one can access to a large number of corpora in many different languages via World Wide Web. The Internet is a cumulative source for language data which can be readily available for everyone, everywhere, and for every purpose in a huge size. Of course, the size cannot be considered a determining factor as Fillmore suggest: “I don’t think there can be any corpora, however large, that contain information about all the areas of English lexicon and grammar that I want to explore. . . [but] every corpus I have had the chance to examine, however small, has taught me facts I couldn’t imagine finding out any other way” (Fillmore, 1992, p. 35).

Dealing with size corpora can be divided into *reference* and *monitor* corpora. Reference corpora have a fixed size; that is, they are not expandable (e.g., the British National Corpus), whereas monitor corpora are expandable; that is, texts are continuously being added (e.g., the Bank of English). Corpora can also be divided depending their shapes as the corpora which contain *whole texts* and those which contain *samples* of a specified length. The latter allows a greater variety of texts to be included in a corpus of a given size.

In terms of content, corpora can be either *general*, that is, attempt to reflect a specific language or variety in all its contexts of use (e.g., the American National Corpus), or *specialized*, that is, aim to focus on specific contexts and users (e.g., Michigan Corpus of Academic Spoken English), and they can contain *written* or *spoken* language. Corpora can also represent the different varieties of a single language. For example, the International Corpus of English (ICE) contains one-million-word corpora representative of different varieties of English (British, Indian, Singaporean, etc.). As implied in the previous section, corpora may contain language produced by *native* or *non-native speakers* (usually learners). Finally, corpora can be *monolingual* (i.e., contain samples of only one language), or *multilingual*. Multilingual corpora are of two types: they can contain the *same text-types* in different languages, or they can contain the *same texts translated* into different languages, in which case they are also known as *parallel corpora* (Hunston, 2002; Kennedy, 1998; McEnery & Wilson, 2001; Meyer, 2002). Some corpora are enriched with linguistic information such as part-of-speech annotation, parsing and prosodic transcription. In these cases it is far easier and more specific to retrieve data than in the case of unannotated corpora.

III. Concordances

A *concordancer* is a software for observing a corpus which shows a list of occurrences of a particular word, part of a word or combination of words, in its contexts. The search word is referred to as key word. The most common way of displaying a concordance is by a series of lines with the keyword in context (KWIC-format) in which each word is centered in a fixed-length field (e.g., 80 characters).

According to the *Collins Cobuild English Dictionary*, a “concordance” is:

“An alphabetical list of the words in a book or a set of books which also says where each word can be found and often how it is used”

Language teachers can make use of concordancers to help their students understand word associations. A concordance demonstrates authentic patterns of a language more explicitly and accurately in terms of language samples in their usual context. In fact, the learners can get a much better idea of the use of the word looking into the context than they would achieve by merely looking it up in the dictionary. A concordancers are particularly helpful when used in conjunction with a thesaurus. One cannot find information about unknown words and their authentic usage in the context of the language, but with the help of a concordancer s/he can

Tim Johns was one of the first language teachers to make use of concordancers in the languages classroom. Back in the early 1980s he was making use of the concordancing packages available at the University of Birmingham, and he wrote a micro-concordance program that ran on one of the first popular microcomputers, the Sinclair ZX81 - and it worked: see Higgins & Johns (1984:88-93). Johns later developed the concept of Data Driven Learning (DDL) and wrote one of the first commercially available classroom concordancers, which was published by Oxford University Press: MicroConcord (Lamy, Klarskov Mortensen, and Davies, 2005). The advantage of the DDL approach is that, in a classroom situation, it enables the teacher to play a less active role whilst at the same time exposes the student to authentic texts like those found in a monolingual corpus. What distinguishes the DDL approach is the attempt to cut out the middleman as much as possible and give direct access to the data so that the learner can take part in building his or her own profiles of meanings and uses. The assumption that underlies this approach is that effective language learning is itself a form of linguistic research, and that the concordance printout offers a unique resource for the stimulation of inductive learning strategies -- in particular, the strategies of perceiving similarities and differences and of hypothesis formation and testing. (Johns, 1991b, p. 30)

Examples of Using Concordances in Learning and Teaching

Examples from the corpus provide the learners with the kinds of sentences that they will encounter when using the language in real life situations. By searching for key words in context, the learners can extract the rules of grammar or usage and lexical features and based on their observation of patterns in authentic language they can question some of the rules. In the case of vocabulary they can be critical of dictionary entries. They can also compare texts produced by native and non-native speakers of a language in terms of appropriate position of certain lexical items in the context by means of a concordancer. A concordance with some deleted keywords can be exposed to the learners and they are asked to fill the gaps based on their lexical and grammatical knowledge. This is useful for learners to assess whether or not they have fully grasped a certain item of vocabulary or structure. Furthermore, the learners can use a concordance to work with multi-meaning and multi-usage words in that they are given some concordances of a single word and try to group them according to their usage.

Teachers also can benefit from the concordances in many ways. Using the examples taken from a variety of corpora, a teacher can make appropriate exercises which can be used in the classroom. These exercised can be in the form of cloze tests, typical collocations, a point of grammar, or proper vocabulary usage. Teachers also make use of concordancers to criticize the contents of textbooks in terms of vocabulary and grammar. Many traditional grammar books, textbooks and dictionaries contain only invented examples, and reflect their authors’ insights. McEnery & Wilson list four separate studies of ESL textbooks that have shown that teaching materials not based on authentic data can be positively misleading to students (McEnery & Wilson 1996). Concordances can also be used as test material providing assessment items. With the help of a concordancer the teachers can also develop strategies for inferring the meaning of unknown lexis in the text.

Some teachers tend to use concordances *inductively*. They find the concordances as examples of a language structure already taught. Some others tend to use concordances *deductively* in that they present the concordances as data for the learners to analyze.

IV. The Role of Corpora in Language Learning

Corpus use contributes to language teaching in a number of ways (Aston, 2000; Leech, 1997; Nesselhauf, 2004). The insights derived from native-speaker corpora contribute to a more accurate language description, which then feeds into the compilation of pedagogical grammars and dictionaries (Hunston & Francis, 1998, 1999; Kennedy, 1992; Meyer, 1991; Owen, 1993). In the following sections we see how a number of areas of language study have benefited from exploiting corpora.

A. Language Description

The first important contribution of corpus-based research to language teaching seems to be the accurate descriptions of a given language. The use of L1 corpora in linguistic research has provided the most convincing evidence of discrepancies between actual use and traditional, introspection-based views on language (Sinclair, 1997, pp. 32-34).

Corpus-based research has revealed the inadequacy of many of the rules that still dominate ELT materials. For example, in a study of a random sample of 710 *if*-conditionals from the written section of the BNC, the conditional sentences were examined against the information about form, time orientation and attitude to likelihood given within the currently favored framework of five types (zero, first, second, third and mixed). The rules presented in fifteen recent intermediate-to-advanced course books, taken collectively, accounted for only 44% of the sentences (Gabrielatos, 2003b).

B. Lexicography

Corpus linguistics can have a considerable contribution to the field of lexicography through extracting and studying empirical data from corpora. This includes improving, updating and classifying words or expressions as well as gathering additional information such as typical collocations, sub-categorizations, requirements, or definitions.

The importance and benefits of corpus studies in dictionary compiling can be observed in the considerable number of dictionary publishers investing in corpus technology and it is easy to find many corpus-based monolingual as well as bilingual dictionaries.

Now, availability of different corpora has eased the task of lexicographers to a high degree in that they can access to all the examples of the usage of a word or phrase out of corpora in a few seconds. Consequently, compiling and revising dictionaries would be more precise and faster.

In order to show the benefits of corpus studies in lexicography we mention the work cited by Atkins and Levin (1995). They studied verbs in the semantic class of *shake* and quoted definitions from three dictionaries. Two of the entries which Atkins and Levin discuss are *quake* and *quiver*. Both the Longman and COBUILD dictionaries list these verbs as being intransitive, while the Oxford dictionary lists *quake* as being intransitive, but lists *quiver* as being transitive. Looking at the occurrences of these verbs in a corpus of 50,000,000 words, Atkins and Levin were able to discover examples of both *quiver* and *quake* in transitive constructions. E.g. *It quaked her bowels* and *quivering its wings*. In other words, the dictionaries had got it wrong - both verbs could be transitive as well as intransitive. This small example demonstrates how a sufficiently large and representative corpus can either supplement or refute the lexicographer's intuitions and provide information which will in future result in more accurate dictionary entries (McEnery and Wilson, 1996).

It is in dictionary building that the concept of an open-ended monitor corpus has its greatest role since it enables lexicographers to keep on top of new words entering the language, or existing words changing their meanings, or the balance of their use according to genre, formality, and so on. The ability to call up word combinations rather than individual words, and the existence of mutual information tools which establish relationships between co-occurring words mean that we can treat phrases and collocations more systematically than was previously possible (McEnery and Wilson, 1996).

C. Morphology

In the realm of morphology we can also see the effect of corpus-based works. Although it may seem that corpus studies cannot have any great advantage in word structure which is the core concept of morphology, researchers in morphology can exploit corpora through studying the frequencies of different morphological variations (such as got/gotten) and the productivity of different morphemes.

In a study carried out by the author it has shown that some translation problems in the morphological level can be easily resolved with the help of statistical data gained from a corpus. The main concern of the study is automatic morphological analysis for machine translation from English into Persian. Using a stem dictionary, a previously compiled table and an untagged monolingual corpus of Persian language, the algorithm can select the most plausible and appropriate Persian equivalent for English words with suffixes and/or prefix. The algorithm has been tested for a set of non-simple English words (word-forms) and the results were encouraging in respect to their Persian translation (Mosavi Miangah. T. 2007b).

In another experiment Opdahl used the LOB and Brown corpora to differentiate between adverbs with and without –ly suffix (like low/lowly) and their frequencies in term of usage in American and British English (Opdahl, 1991).

D. Syntax

In fact, syntax (syntactic studies) can be considered the most interesting and challenging branch of language research which makes the highest use of corpora. The majority of studies in this discipline use the quantitative data analyses to provide information about relative frequencies of certain syntactic structures. The proper position of certain adverbs can be recognized using a corpus. The adverb *therefore*, for instance, may be at the beginning of a sentence or at the middle or probably after a comma, that is, at the beginning of a subordinate clause. The learners do a search in concordancer based on the word *therefore*, and then make conclusions out of the search results for the most probable and natural position of *therefore*.

Bernhard Kettemann (1996) looked at *if*-clauses, Reported Speech, the contrast between Present Perfect and Past Tense and some examples of possible contrasts between *since* and *for* using the corpus.

Some researchers tried to study the relative clauses in English using the quantitative information provided by the LOB and Kolhapur corpora (Schmied, J. 1993), and some others used corpora to analyze noun phrase structure for correct determination of head nouns as well as modifying nouns or adjectives of a noun phrase for indexing to improve retrieval performance (Mosavi Miangah, T. 2007a). In fact, the large volume of works in corpus studies in the field of grammar (especially syntax) indicates the effectiveness of corpus-based methods in teaching and learning grammatical structures.

The study of prepositions using corpora is perhaps the most interesting one. The use of preposition in different languages may vary a lot. Consider, for instance, the use of *at/in* as a preposition of place in English in collocation with different nouns following them.

TABLE 1.
FREQUENCIES RELATED TO TWO PREPOSITIONS OF PLACE PRECEDING NOUNS *UNIVERSITY*, *HOTEL* AND *OFFICE*

preposition	noun	frequency
at / in	university	1113 / 330
at / in	hotel	344 / 368
at / in	office	233 / 698

The information demonstrated in Table 1. has been gained from the British National Corpus (BNC) and shows that some nouns tend to come with certain prepositions rather than with some other ones. So, we see how corpora can meet the needs of grammar teaching in certain fields.

Corpora also has a considerable role in finding word's collocations especially in selecting the most appropriate adjective out of several synonymous adjectives for certain nouns. In Table 2 consider which of the following five adjectives collocate with the nouns in the left column.

The data shown in Table 2 which is again extracted from the BNC helps the learners to predict the appropriate adjective for each of the given nouns, though we know that some of these adjectives may replace some other ones depending on formal and informal situations.

TABLE 2.
FREQUENCIES RELATED TO FIVE SYNONYMOUS ADJECTIVES FOR DIFFERENT NOUNS

	great	**grand**	**large**	**big**	**high**
asset	34	0	0	6	1
brother	2	0	0	131	0
bang	1	0	3	365	0
difficulty	260	0	0	2	0
degree	40	0	82	0	478
effect	82	1	15	10	0
event	31	4	2	49	0
extent	141	0	342	0	0
fun	238	1	0	6	o
house	95	11	109	250	20
idea	98	4	0	30	0
impact	24	0	6	32	17
majority	389	0	116	5	0
opera	5	67	0	3	0
picture	11	1	24	43	0
pleasure	187	0	0	0	0
population	2	0	54	3	31
Problem	66	0	10	125	0
Quantity	25	0	97	0	2
Question	7	1	9	72	0
money	0	0	0	99	1
price	15	0	8	9	182
quality	12	0	1	0	864
Scale	7	102	500	8	1
Success	370	0	1	40	27
Tour	1	54	1	8	0
measure	20	0	136	1	3

Another application of such a table is categorizing the nouns collocating with each of the adjectives. The learners may be asked: Is there any relationship between the type of noun associated with each of these adjectives? Actually they can find a possible answer to this question in terms of abstractness and concreteness of the nouns collocating with each adjective. However, it should be said that some collocations of the mentioned adjectives with nouns are not the only strict solutions, that is, an absolute yes or no case, but they can order in degrees of strong and weak collocations.

E. Semantics

The concept of collocation to which we referred in the previous part is referred to here again, but this time in terms of semantic association of collocates with each other.

Semantic association (Hoey, 2003) or semantic prosody (Louw, 1993) is a concept that has served to deepen our knowledge of the relationship between a word and its collocates. Through computer-based corpus analysis, first Sinclair (1991), and then in more detail, Louw (1993), discovered that collocates themselves can have a semantic patterning that is not random. Just as the word *blonde* collocates typically with the word *hair*, certain words can collocate with groups of either distinctly positive or negative words (Stubbs, 1995) or with semantic sets of meaning.

The notion of semantic prosody was taken up and expanded by Stubbs (1995) who suggested that as well as collocating with purely positive or negative semantic groupings of words, words can also collocate with semantic sets: Semantic prosodies may be of a very general kind: such as the shared semantic feature “unpleasant”. Alternatively, one may be able to predict that a node will most likely co-occur with collocates from a restricted lexical set: for example, from the semantic field of “care” (Stubbs, 1995, p. 249).

Mike Nelson studied the two words *global* and *international* which frequently used with the names of companies in the business environment. Based of his reports despite the similarities, there were clear differences. *Global* collocated more with “business activities” than *international*, and *international* more with “companies and institutions” than *global*. Both these words were in contrast to local. The words collocating with local included a large number of non-business-related words and this can be seen most clearly in the “companies/institutions” category. All three words, *global*, *international* and *local* shared this semantic prosody with “companies/institutions”, but the institutions collocating with local were noticeably more often of a distinctly non-business nature (Nelson, 2005).

Mindt considers two important roles of the corpus in semantics: providing objective criteria for assigning meaning to linguistic items, and establishing more firmly the notions of fuzzy categories and gradience (Mindt, D. 1991).

Another application of corpora in semantics seems to be in dealing with false friends. False friends or unfaithful friends are two words in different languages that appear to be the same but have very different meanings. The word *machine* can be considered as a false friend in English and Persian. Although Persian has borrowed this word from English, its range of usage in Persian is very different from that of English. So, we can use the corpus to sensitise learners to the range of semantic differences between the English and Persian false friend *machine*. The word *data* can also be regarded as a false friend in English and Persian. In Persian it may be translated as singular (dade) or as plural (dadeha), based on their grammatical differences in a sentence in which they occur. So the corpus can be used to work out what the difference is between the singular and plural uses of the Persian word.

F. Cultural Studies

It is only recently that the role of corpora in cultural studied has been revealed. Analyzing corpora in order to find the cultural elements which reflect habits of people of the culture specific to those corpora can be of great help. If the word *steak*, for instance, is compared in the two corpora of English and Persian in terms of frequency it will reveal that the usage of this kind of food in English corpus is much higher than in Persian one. In the same way, the frequency of occurrence of *coffee* comparing to *tea* in English corpus is very high. While this case is reversed in Persian corpus. That is, Persian-speaker people tend to drink tea rather than coffee in most occasions, hence the higher frequency of the word *tea* in Persian corpus comparing to *coffee*. The results gained from the two corpora indicate each culture’s interest in each of these drinks.

Leach and Fallon compared the two corpora of American and British English in terms of frequency to check up on the senses in which words were being used. Difference between the frequencies of some concepts in the two corpora revealed findings which were suggestive not only from the linguistic point of view but also from the cultural point of view. Travel words, for instance, were more frequent in American English than in British English and this is naturally due to the larger size of the United States (Leech, G. and Fallon, R. 1992).

G. Computer-assisted Language Learning (CALL)

For more than a decade, corpus and concordance have been regularly described as one of the most promising ideas in computer-assisted language learning (Leech & Candlin, 1986; Johns, 1986; Johns & King, 1991; Hanson-Smith, 1993).

Recent work at Lancaster University has looked at the role of corpus-based computer software for teaching undergraduates the rudiments of grammatical analysis (McEnery and Wilson 1993). This software - Cytor - reads in an annotated corpus (either part-of-speech tagged or parsed) one sentence at a time, hides the annotation and asks the student to annotate the sentence him- or herself. Students can call up help in the form of the list of tag mnemonics, with examples or in the form of a frequency lexicon entry for a word giving the possible parts of speech with their frequencies. Students can also call up a concordance of similar examples. Students are given four chances to get an annotation right. The program keeps a record of the number of guesses made on each item and how many were correctly annotated by the student (McEnery and Wilson, 1996).

H. English for Specific Purposes

McEnery and Wilson identify ESP as a particular domain-specific area of language teaching and learning, where “corpora can be used to provide many kinds of domain-specific material for language learning, including quantitative accounts of vocabulary and usage which address the specific needs of students in a particular domain more directly than those taken from more general language corpora” (McEnery and Wilson 2001, p. 121). In professional domains, various corpora are being built. Most of them are of finite size, with the exception of so-called *monitor* corpora – open-ended collections of texts, to which new texts are being constantly added until the corpora “will get too large for any practicable handling, and will be effectively discarded” (Sinclair, 1991, p. 25).

According to Dlaska ESP teaching need not be “dire and difficult pedagogical ground”, forcing language teachers to surrender their expertise in favor of teaching unfamiliar subjects, but on the contrary, it needs to “address, and eventually bridge, the discrepancy between general language ability and specialized language ability … since the two areas are not in opposition but complement each other” (Dlaska, 1999, p. 403).

Olga Mudraya tries to show how the integration of the lexical approach with a corpus-based methodology in teaching English for Specific Purposes (ESP), especially Engineering English, can improve the way ESP is taught. Her particular point has been to demonstrate how a technical student can benefit from the data-driven lexical approach. In her paper, Mudraya has argued for the integration of the lexical approach with a data-driven corpus-based methodology in ESP teaching, as she believes that the use of language corpora in the classroom can improve students’ knowledge of the language and their ability to use it effectively. This leads her to the conclusion that corpora can also improve the way ESP teaching is approached. It can inform teaching and learning, producing students who know what it means to use a corpus, who know how to extract material from it, and who, consequently, can learn a great deal about language via a corpus (Mudraya, O. 2005).

In another study Curado Fuentes outlines a way of dealing with vocabulary in English for Academic Purposes (EAP) instruction in the light of insights provided by empirical observation. Focusing mainly on collocation in the context of English for Specific Purposes (ESP), and, more precisely, within English for Information Science and Technology, He showed how the results of the contrastive study of lexical items in small specific corpora can become the basis for teaching / learning ESP at the tertiary level. In the process of his study, an account was given of the functions of academic and technical lexis, aspects of keywords and word frequency were defined, and the value of corpus-derived collocation information was demonstrated for the specific textual environment. (Curado Fuentes, A. 2001).

The Guangzhou Petroleum English Corpus containing about 411,000 words of petrochemical domain is an example of domain-specific corpora which provide the learners with many kinds of materials including quantitative information about the vocabulary as well as usage in a particular domain.

The largest current professional corpus is to be the Corpus of Professional English (http://www.perc21.org/cpe_project/index.html). It is being developed in collaboration between the Professional English Research Consortium (PERC), Japan, and Lancaster University, UK. When finished, it will consist of a 100-million-word database of English used by professionals in science, engineering, technology and other fields.

I. Translation

In recent years many researchers and trainers in the field of translation studies have tried to integrate the analysis of corpora into translator education. There have been so many attempts to trace links between corpus linguistics and translation practice.

Today Translators (professional as well as student translators) are able to investigate and manipulate the information contained in corpora using corpus analysis tools in different ways. All machine readable corpora can help the translator to count the occurrence frequency of the search (query) word or phrase in the corpus, then clicking on the word in the display screen, all concordance lines in which the search word or phrase appear will be shown. In addition, most corpus analysis packages comprise a "concordancer", which enables the translator to find all the occurrences of a search word, or search phrase, with the possibility of sorting the displaying data in the screen together with a span of co-text to the left and right. Also the part of speech of the search word can be specified first, and then the search is done, and so many other possibilities.

There are many ways in which the translators are able to exploit the corpora in order to improve the quality of their translations. We mention here only two ways. The first one goes with the collocations. Referring to a monolingual target corpus for finding information about collocates, especially adjectives that collocate with nouns has been proved to be very useful. For example, when looking for translation equivalents of the adjective *hameh ja:nebeh* in the noun phrase *hamlehe hameh ja:nebeh,* Persian-English dictionaries suggest, for example, “multilateral” as the equivalent for *hemeh ja:nrbeh*. Hence, translating the phrase *hamlehe hameh ja:nebeh* as “multilateral attack”, “multilateral strike” or “multilateral rush”. However, of the 304 lines generated by BNC for the search word “multilateral”, none of the above nouns appear immediately to the right of the search word, while there are 11 occurrences of the adjective “massive” in front of the noun “attack” and 1 occurrence of the adjective “massive” in front of the noun “strike” with no phrase as “massive rush”, “multilateral attack”, “multilateral strike” or “multilateral rush”.

The second way in which the corpora can help the translators is verifying or rejecting decisions taken based on other tools. Corpora can be used in finding the most suitable equivalent of certain terms in target language for which other translation tools, mainly dictionaries, suggest unusual or unsuitable translations. That is, if we are not sure about the suggested translation(s) of a certain word in the dictionaries, or the given translation(s) are not desirable, referring to corpora can be of great help in verifying or rejecting the suggested translation(s). If we consider the words *dusti* and *ref:ghat* in Persian-English dictionaries we find the word “fellowship” as the first equivalent for them. However, when we refer to the *British National Corpus* (BNC), we can hardly find “fellowship” with these meanings based on the analysis of surrounding words, i.e. the context to which the word “fellowship” belongs. Instead the majority of the word occurrence mean “membership” not “friendship”. So, based on information gained from a naturally occurred corpus, a translator (especially student-translator) can decide on which of the target alternative translations of a certain word in his or her native language to use in order to produce a more natural and sound target text (Mosavi Miangah, T. (2006).

In an experiment the co-occurrences of the multiple-meaning words in a monolingual corpus of the target language, namely Persian, were considered. By calculating the frequencies of these words in the corpus, the most probable sense for these multiple-meaning words can be selected. The method has been tested for a selected set of English texts containing multiple-meaning words with respect to Persian language and the results are encouraging (Mosavi Miangah, and Delavar Khalafi, 2005).

Some other experiments have also reported the uses of bilingual corpora and monolingual corpora (Zanettin 1998, 2001; Bowker 1998, 2000; Pearson 1998; Gavioli and Zanettin 2000) as sources to compile term banks, and as aids during the translation task.

Other examples of areas that have been examined with the help of language corpora are the use of lexical chunks (De Cock et al., 1998), collocations (Nesselhauf, 2005), complement clauses (Biber & Reppen, 1998), the progressive and questions (Virtanen, 1997, 1998), overstatement (Lorenz, 1998), connectors (Altenberg & Tapper, 1998), speech-like elements in writing (Granger & Rayson, 1998), and epistemic modality (McEnery & Kifle, 2002).

V. CONCLUSION

Corpora are frequently used in many academic and applied linguistic fields such as knowing facts about language; language for specific purposes (e.g. use newspaper corpora, corpora of scientific texts); preparing vocabulary lists based on high-frequency lexical items; preparing cloze tests; answering ad hoc learner questions ('What's the difference between *few* and *a few*?'); etc. Both teachers and learners can benefit from exploiting the corpora in teaching and learning tasks.

Using corpora in different areas of language study does not mean that the traditional methods of language learning are no longer applicable or appropriate in the present language learning environment. Instead, it can enrich and enhance the existing teaching approaches, that is, a welcome addition to them.

APPENDIX . FREE/AFFORDABLE CORPORA AND CORPUS TOOLS

1- British National Corpus Sampler (1 million words or written and 1 million words of spoken English): http://www.natcorp.ox.ac.uk/getting/sampler.html. Also, free, but restricted, access to the full BNC: http://sara.natcorp.ox.ac.uk/lookup.html

2- Collins Wordbanks Online English corpus (concordance and collocation samplers): http://www.collins.co.uk/Corpus/CorpusSearch.aspx

3- The Complete Lexical Tutor: http://132.208.224.131

4- Michigan Corpus of Academic Spoken English (MICASE): http://www.hti.umich.edu/m/micase

5- Variation in English Words and Phrases (Mark Davies, Brigham Young University). Interface to the full British National Corpus (100 million words): http://view.byu.edu/

6- Web Concordancer (works with a variety of corpora): http://www.edict.com.hk/concordance/

7- WebCorp: The Web As Corpus (University of Liverpool): http://www.webcorp.org.uk/

8- WordNet: A Lexical Database for the English Language (Princeton University): http://www.cogsci.princeton.edu/~wn

9- WordSmith Tools: http://www1.oup.co.uk/elt/catalogue/Multimedia/WordSmithTools3.0

REFERENCES

[1] Altenberg, B. & Tapper, M. (1998). The use of adverbial connectors in advanced Swedish learners' written English. In: S. Granger (Ed.), *Learner English on computer* (80-93). London: Longman.

[2] Aston, G. (1997). Enriching the learning environment: Corpora in ELT. In: A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora* (51-64). New York: Addison Wesley Longman.

[3] Atkins and Levin (1995). Building on a corpus: a linguistic and lexicographical look at some near-synonyms. *International Journal of Lexicography* 8 (2), 85-114.

[4] Biber, D. & Reppen, R. (1998). Comparing native and learner perspectives on English grammar: A study of complement clauses. In: S. Granger (Ed.), *Learner English on computer* (145-158). London: Longman.

[5] Bowker, L. (1998). Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study. *META*, 43 (4), 631-651.

[6] Bowker, L. (2000). Towards a methodology for exploiting specialized target language corpora as translation resources. *International Journal of Corpus Linguistics*, 5 (1), 17-52.

[7] Collins Cobuild English Language Dictionary. (1987). ed. John Sinclair, London and Glasgow: Collins.

[8] Crystal, D. (1991). A Dictionary of Linguistics and Phonetics. (third ed.), Blackwell, London.

[9] Curado Fuentes, A. (2001). Lexical Behavior in Academic and Technical Corpora: Implications for ESP Development . *Language Learning Technology*, 5 (3), 106-129.

[10] De Cock, S., Granger, S., Leech, G. & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In: S. Granger (Ed.), *Learner English on computer* (67-79). London: Longman.

[11] Dlaska, A. (1999). Suggestions for a subject-specific approach in teaching foreign languages to engineering and science students. *System* 27 (3), 401–417.

[12] Fillmore, C.J., (1992). Corpus linguistics or Computer-aided armchair linguistics In: Svartvik, J. (Ed.), *Directions in Corpus Linguistics*. Mouton de Gruyter, Berlin, 35–60.

[13] Firth, J. R. (1957). A synopsis of linguistic theory. 1930–55 In: F. R. Palmer (Ed.), (1968) *Selected Papers of J.R. Firth 1952–59* (168–205). London/Harlow: Longmans.

[14] Gavioli, L. and Zanettin, F. (2000). I corpora bilingui nell'apprendimento della traduzione. Riflessioni su un'esperienza pedagogica. In: Silvia Bernardini and Federico Zanettin (eds.) *I corpora nella didattica della traduzione.* Bologna: CLUEB, 61-80.

[15] Gabrielatos, C. (2003b). Conditionals: ELT typology and corpus evidence. *36th Annual Meeting of the British Association for Applied Linguistics (BAAL)*, University of Leeds, UK, 4-6 September 2003.

[16] Granger, S. & Rayson, P. (1998). Automatic profiling of learner texts. In: S. Granger (Ed.), *Learner English on computer* (119-131). London: Longman.

[17] Higgins J. & Johns T. (1984). Computers in language learning, London: Collins.

[18] Hoey, M. (2003). Lexical priming and the qualities of text. Retrieved. 17.5.2004, from: http:// www.monabaker.com/tsresources/LexicalPrimingandthePropertiesofText.htm.

[19] Hunston, S. (2002). Corpora in applied linguistics, Cambridge: Cambridge University Press.

[20] Hanson-Smith, E. (1993). Dancing with concordances. *CÆLL Journal*, 4 (2), 40p.

[21] Hunston, S. & Francis, G. (1998). Verbs observed: A corpus-driven pedagogic grammar. *Applied Linguistic,* 19(1), 45-2.

[22] Hunston, S. & Francis, G. (1999). Pattern grammar. Amsterdam: John Benjamins.

[23] Johns, T. (1986). Micro-concord: A language learner's research tool. *System, 14* (2), 151-162.

[24] Johns, T. (1991a). Should you be persuaded: Two examples of data driven learning. In: T. Johns & P. King (Eds.), *Classroom concordancing.* ELR Journal 4 (1-16). Birmingham: University of Birmingham.

[25] Johns, T. (1991b). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. In: T. Johns & P. King, P. (Eds.), *Classroom concordancing*. *ELR Journal*, 4 (27-45). Birmingham: University of Birmingham.

[26] Johns, T. & King, P. (eds.) (1991). Classroom concordancing. *English Language Research Journal, 4.* University of Birmingham: Centre for English Language Studies.

[27] Kennedy, G. (1998). An introduction to corpus linguistics, London: Longman.

[28] Kennedy, G. (1992). Preferred ways of putting things with implications for language teaching. In: Svartvik, J. (Ed.), *Directions in corpus linguistics.* Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991 (335-378). Berlin: Mouton de Gruyter.

[29] Kettemann, B. (1996). Concordancing in English Language Teaching. From: http://www-gewi.kfunigraz.ac.at/ed/project/concord1.html, Retrieved: 18 November, 2008

[30] Lamy, M-N, Klarskov Mortensen, H. J. and Davies, G. (2005). Using concordance programs in the Modern Foreign Languages classroom. From: http://www.hull.ac.uk/ict4lt/en/en_mod3-4.htm, Retrieved: 7 October 2008

[31] Leech, G. (1997). Teaching and language corpora: A convergence. In: A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora* (1-23). New York: Addison Wesley Longman

[32] Leech, G. & Candlin, C. N. (1986). Computers in English language teaching and research. London: Longman.

[33] Leech , G. and Fallon, R. (1992). Computer corpora – what do they tell us about culture? *ICAME Journal*, 16, 29-50.

[34] Lorenz, G. (1998). Overstatement in advanced learners' writing: Stylistic aspects of adjective intensification. In S. Granger (Ed.), *Learner English on computer* (53-66). London: Longman.

[35] Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In: M. Baker, G. Francis & E. Tognini-Bonelli (Eds.), *Text and Technology. In Honor of John Sinclair* (157–176). Philadelphia/Amsterdam: John Benjamins.

[36] McCarthy, M. (1990). Vocabulary. Oxford: Oxford University Press.

[37] McEnery and Wilson, (1993). The role of corpora in computer-assisted language learning. *Computer Assisted Language Learning (CALL),* 6 (3), 233-48.

[38] McEnery T. & Wilson A. (1996). Corpus linguistics. Edinburgh: Edinburgh University Press.

[39] McEnery, T., Wilson, A., & Baker, P. (1997). Teaching grammar again after twenty years: Corpus-based help for teaching grammar. *ReCALL, 9* (2), 8-16.

[40] McEnery, T. & Wilson, A. (2001, 2nd ed.) Corpus linguistics. Edinburgh University Press.

[41] McEnery, T. & Kifle N.A. (2002). Epistemic modality in argumentative essays of second-language writers. In: J. Flowerdew (Ed.), *Academic discourse* (182-195). Harlow: Longman.

[42] McEnery, T., Xiao, Z. & Tono, Y. (2005, in press). *Corpus based language studies*. London: Routledge.

[43] Meyer, C. F. (1991). A corpus-based study of apposition in English. In: K. Aijmer & B. Altenberg (Eds.), *English corpus linguistics. Studies in honor of Jan Svartvik* (166-181), London: Longman.

[44] Meyer, C. F. (2002). English corpus linguistics: An introduction. Cambridge: Cambridge University Press.

[45] Mindt, D. (1991). Syntactic evidence for semantic distinctions in English. In: Aijmer and Altenberg (eds.) *English Corpus Linguistics: Studies in Honour of Jan Svartvik,* London, Longman, 182-96.

[46] Mosavi Miangah, T. and Delavar Khalafi, A. (2005). Word sense disambiguation using target language corpus in a machine translation system. *Literary and Linguistic Computing,* 20(2), 237-249.

[47] Mosavi Miangah, T. (2006). Applications of corpora in translation. *Translation Studies*, 12, pp: 43-56.

[48] Mosavi Miangah, T. (2007a). A corpus-based approach for noun phrase parsing. In: *5th International Conference on Natural Language Processing, ICON-2007,* pp. 153-158, Hyderabad, India.

[49] Mosavi Miangah, T. (2007b). An unsupervised method for solving translation ambiguities at the morphological level. *Proceedings of the Conference Language and Technology (CLT07),* PP. 125-129, University of Peshawar, Pakistan.

[50] Mudraya, O. (2005) Engineering English: A lexical frequency instructional model. *English for Specific Purposes (*Available online 20 June 2005).

[51] Nelson, M. (2005). Semantic associations in Business English: A corpus-based analysis. *English for Specific Purposes*, 25, PP. 217-234.

[52] Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In: J. McH. Sinclair (Ed.), *How to use corpora in language teaching* (125-152). Amsterdam: Benjamins.

[53] Nesselhauf, N. (2005). Collocations in a learner corpus. Amsterdam: John Benjamins.

[54] Opdahl, L. (1991). *–ly* is adverbial suffix: corpus and elicited material compared. *ICAME Journal*, 15, 19-35.

[55] Owen, C. (1993). Corpus-based grammar and the Heineken effect: Lexico-grammatical description for language learners. *Applied Linguistics,* 14(2), 167-187.

[56] Pearson, J. (1998). Terms in context. Amsterdam & Philadelphia: Johns Benjamins.

[57] Schmied, J. (1993). Qualitative and Quantitative research approaches to English relative constructions. In: Souter and Atwell 1993, 85-96.

[58] Sinclair, J. M. (1997). Corpus evidence in language description. In: A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora* (27-39). New York: Addison Wesley Longman.

[59] Sinclair, J. (1991). Corpus, concordance, collocation. Oxford University Press, Oxford.

[60] Stevens, V. (1995). Concordancing with language learners: Why? When? What? *CALL Journal,* 6 (2), 2-10. Available online, http://www.ruf.rice.edu/~barlow/stevens.html.

[61] Stubbs, M. (1995). Corpus evidence for norms of lexical collocation. In: G. Cook & B. Seidlhofer (Eds.), *Principle and Practice in Applied Linguistics*. Studies in Honor of H.G. Widdowson (245–256).

[62] Tribble, C. (1997a). Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching. In: J. Melia. & B. Lewandowska-Tomaszczyk (Eds.), *PALC '97 Proceedings, Practical Applications in Language Corpora* (106-117). Lodz: Lodz University Press. Available online, http://www.ctribble.co.uk/text/Palc.htm.

[63] Tribble, C. (2000). Practical uses for language corpora in ELT. In: P. Brett & G. Motteram (Eds.), *A Special interest in computers* (31-42). Whitstable, Kent: IATEFL.

[64] Virtanen, T. (1997). The Progressive in NS and NNS student compositions: Evidence from the International Corpus of Learner English. In: M. Ljung (Ed.), *Corpus-based studies in English.* Papers from the Seventeenth International conference on English Language Research on computerized Corpora (ICAME 17) (299-309). Amsterdam: Rodopi.

[65] Virtanen, T. (1998). Direct questions in argumentative student writing. In: S. Granger (Ed.), *Learner English on computer* (94-106). London: Longman.

[66] Zanettin, F. (1998) Bilingual Comparable Corpora and the Training of Translators. *ETA*, 43 (4), 616-630.

[67] Zanettin, F. (2001). Swimming in words: corpora, language learning and translation. In Guy Aston (ed.) *Learning with corpora,* Houstox, TX: Athelstan, 177-197.

Tayebeh Mosavi Miangah, an academic staff in Payame Noor University of Yazd, is an assistant professor of “Applied and Mathematical Linguistics”. She was previously the academic staff of Shahrekord University. Her research interests are as follows:

Computational Linguistics: Machine Translation; Corpus-based Approaches in Language Studies; Word Sense Disambiguation; Part of Speech Tagging; Automatic Morphological Analysis.

Dr. Mosavi Miangah has recently published scientific papers in national as well as international journals like “Translation Studies”, “Literary and Linguistic Computing”, Journal of Quantitative Linguistics”, “International Journal of Translation”, “Meta” and in a number of international conferences throughout the world. Among her recent endeavors are Compiling and working on monolingual and parallel bilingual corpora in language analysis.