

*Centre for English Language Studies*

Postgraduate programmes, Open Distance Learning  
***ESSAY COVER SHEET AND DECLARATION***

**IMPORTANT: Please complete ALL sections of this template and PASTE it into the FIRST PAGE of your essay**

Student ID number	822875
Module Number (1-6)	2
Title of Degree Programme:	MA ODL in Translation Studies
Title of Module:	Corpus Linguistics/ Research Methods
Assessment Task No.	CL/06/06
Submission	First Submission
Date Submitted	May 24, 2007
Name of tutor	Laksmi Holland

## **1. Introduction**

During the process of translation, translators may encounter problems that require special resources and skills in order to be solved. This paper is a discussion of how corpus information may be useful to translators in translating two genuine translation problems. These problems will be discussed through presenting the nature of these problems as well as information and resources needed to solve them. The layout of this assignment starts with a comprehensive literature review about corpus linguistics, corpus-based translation, and translation problems in general. Then the two problems, collocation and ambiguity, will be presented one by one accompanied with a discussion of how corpus data can help in sorting out these problems, and to support the argument some corpus samples will be provided.

## **2. Literature Review**

### **2.1. Corpus Linguistics**

In this electronic age, we live in, computers are considered to be one of the most important needs as well as practical solutions; from there the idea of exploiting electronic corpora has been originated. Corpus is "a collection of texts in an electronic database" (Kennedy, 1998:3). And corpus linguistics is a merge of technology and linguistics, as corpus linguistics is defined as: "the study of language on the basis of text corpora" (Aijmer & Altenberg, 1991:1). Therefore, corpus linguistics has recently become the reliable source of real linguistics data and statistical information about language. Also it offers fast processing, sorting and searching of data, direct access, time and procedures reduction, and sufficient ability to control huge amount of language data (Kennedy, 1998:5). Consequently corpus, when used appropriately, turns to be one of the most important tools of experienced translators (Johansson, 1991:313).

Corpora are established in different ways to serve different functions. Specialized, general, comparable, parallel, pedagogic and historical corpora are different types of corpora (Birmingham CD, 2001: unit 3), and translation-related corpora i.e. comparable and parallel corpora will be presented in the following subsection.

### **2.1.1. Translation and Corpus**

There are many kinds of corpora exploited for translational purposes, but in this paper two kinds will be introduced. First, the mono-lingual comparable corpus is defined as:

Two separate collections of texts in the same language: one corpus consists of original texts in the language in question and the other consists of translations in that language from a given source language or languages ... both corpora should cover a similar domain, variety of language and time span, and be of comparable length (Baker 1995: 234 cited in Olohan, 2003:62).

This kind of corpus is used to conduct either linguistic or translation studies within the same language to draw conclusion about how languages differ when transfer from being the SL into being the TL.

Second, the bi/multi-lingual parallel or comparable corpus consists of texts in two or more languages that are translation of each other. This parallel corpus can be mono-directional in which the translation is only from one language to the other, or bi-directional in which languages involved are ST and TT in the same time. *MultiLingual Concordancer* for Windows, *Wordsmith Tools Text Aligner*, *Paraconc*, and *MultiConcord* are different examples of translational corpora (Zanettin, 1998:12)

### **2.1.2. Translational Corpora: Application and Potentials**

One of the most valuable applications of mono-lingual comparable corpora lies in the theoretical studies that are conducted, using these corpora, to investigate the linguistic features of translated texts in comparison with original texts within one

language, free from the effects of the SL (Zanettin, 1998:1). Also this kind of corpora is useful to track the norms and the principles that control the process and the outcome of translation (Baker, 1993:243). Displaying the TT accompanied with original texts from the same language will help in figuring out the differences and the similarities of the two versions of the language in question. This contrast and comparison can be drawn at different levels: word, sentence and text level. Bowker also adds that the privilege of using corpus is dimensional when saying:

In contrast, translators using an electronic corpus coupled with a corpus analysis tool can go directly to areas of text containing key words. They can then read as much or as little of the discussion as desired — from a sentence, to a couple of paragraphs, to the entire text. Moreover, as mentioned above, they have a larger selection of texts available for consultation, so if one explanation is not particularly helpful, they can quickly move on to the next (1998:4).

In this case, mono-lingual corpus would introduce translators to varieties of fields, styles, registers, and modes to investigate different features of translation, such as acceptability, accuracy, and naturalness.

On the other hand, bi/multi-lingual parallel-comparable/corpora can be useful in different ways. They are absolutely resourceful for compiling terminology, teaching/learning translation and languages, and exploiting translation memories which are used as retrievable translation reference (Zanettin, 1998:2). Therefore, Olohan demonstrates the importance of this kind of corpora when says: "the use of electronic collections of texts, compiled according to specific research criteria and questions, to analyze occurrences of linguistic patterns or features is a relatively new approach to translation research, but one which has rapidly gained in popularity." (2003:59). To sum up, "the use of corpora in the discipline of translation studies is enjoying increasing popularity", because corpora are applicable to reduce time consuming and improve data accessing, and concordance lines provide better

visualized view of words to draw language patterns (Bowker, 1998:3-5). All these application of corpora considerably provide translators with great help to deal with translation problems.

## **2.2. Translation Problems: Theoretical Consideration**

In translation, roles are manipulated, whereas the translator becomes the reader of the source text ST and the creator of the target text TT; this manipulation would create translation problems. Moreover, translating from one language to another is not a simple direct action that is accomplished in one step, but it is a multi-dimensional work that takes place in many levels. In other words, the translator has to do his/her job through passing different stages starting at reading the ST and ending at producing the final version of the TT. Meanwhile every stage has its difficulties and holds its own problems. The translation problem can be defined as:

A verbal or non-verbal segment that can be present either in a text segment (micro level) or in the text as a whole (macro level) and that compels the student/translator to make a conscious decision to apply a motivated translation strategy, procedure and solution from amongst a range of options (González Davies, 2005:164).

At the early stage of understanding the ST, for example, the translator may face semantic or syntactic ambiguity which needs deep understanding of the text to come up with an appropriate translation.

One of the most prominent problems that occur at the stage of producing the TT is the non-equivalence. The problem of non-equivalency has enlarged because the need of translation is getting more and more demandable in this communication era in which new doors are opened for information exchange and new areas were created, leading to an urge for constructive translation industry (Torrejon and Perz, 2004). This non-equivalency probably is the base of some major translation problems, i.e. lack of grammatical, lexical, and cultural equivalences.

### **2.2.1. Solutions: Skills and Information**

It is generally assumed that the three most important criteria required to produce a high-quality translation are: an understanding of the subject field, an excellent command of the target language, and a good knowledge of the source language (Bowker, 1998:1). Managing translation difficulties and solving problems demands specific skills and requires awareness of some information, consequently translation strategies are the outcome of integrating skills with information. Abdellah lists some macro-skills, to include: "reading comprehension, researching, analytical, and composing skills. These macro-skills include many sub- or micro-skills that need to be mastered." (2002:135,136). One of these micro-skills is the ability to analyze the complex meaning of the ST as well as figuring out the exact intended meaning of an ambiguous structure. Corpus information would facilitate analyzing the ST structure by offering different analytical utensils, such as tagging and parsing.

On the other hand, information is the backbone on which the translator would depend on to attain success. This needed information includes linguistic, non-linguistic, and cultural knowledge that all merge to improve the quality of the TT. For one thing, wide awareness of lexical meaning and how words behave within sentences would improve translators approach in dealing with difficulties, because "a focus on lexis is thus encouraged, but along with this comes a realization of how context-dependent the meaning is" (King, 2003:163). For example, dictionaries and thesauruses may help in finding several possible meanings of one word but will not be enough to figure out the exact meaning. Therefore, more exposure to resources of real language in context will enhance the translator's ability to get hold of the exact meaning. Moreover, reading the text while keeping in mind the cultural aspects of the ST will ease translators' mission. In other words, the more translators are exposed to

quantitative and qualitative linguistic data like, concordance lines in corpus, the more sufficient the translation product will be.

### **2.2.2. Technology: Potential Source for Solution**

To overcome the problems, mentioned above, translators need to have an access to material resources, like text in certain fields or genres, supportive tools, like dictionaries and real data of real language in context. The best provider of the above resources is technological software and computers. Information technology plays a pivotal role in modern translators' professional life, since electronic/online dictionaries and database, translation memories, online translation, and translation machine have enriched translators' reliable resources (Laufer, 2004). Corpora either translational or non-translational are the best resource for information that the translators need. While machine translation or online translation may directly feed the comparable/parallel corpora with material, translation memories can be used as corpora that contain ready-made translation of texts, among which solutions for translation problems may emerge. And technology is the best tool to put all the above resources into handy access. Consequently rich corpora are the best raw material for compiling dictionaries which are considered the translators' right hands. Therefore, corpora are practically beneficial either directly as database of natural language or indirectly as banks of resourceful applications.

According to Arnold, translating problems are of three kinds: ambiguity, structural and lexical mismatches, and multi-word expressions (2001:105). So ambiguity and collocation as 'multi-word expression' will be discussed as examples of genuine translation problems.

### **3. Collocation**

Collocation is the term that has been used as "a phenomenon in language whereby a lexical item tends to keep company with other words. It is a lexical relation of co-occurrence that binds words together with varying degrees of occurrence strength" (Bahumaid, 2006:133). The most common collocation types, Newmark states, are: adjective-noun, e.g. (tasty food), noun-noun, e.g. (arm-chair), and verb-noun, e.g. (pay a visit) (1988:212,213). Another classification of collocation is dividing them into: lexical collocation, between two or more lexical item, e.g. (fair lady), and grammatical collocation which is between one lexical item and a grammatical item mostly preposition, e.g. (go through) (Carter, 1998:51)

#### **3.1. Collocation as Translation Problem**

Collocation can be problematic either at the stage of ST comprehension or at the stage of TT production. Collocation as a comprehensive problem takes many phases; Bahumaid refers to collocation comprehensive problem as the intralingual problematic aspect of collocation. For him this problematic feature may come out of the collocation acceptability which purely depends on native speakers' subjective intuition and the changeability of some elements of collocations. Meanwhile new collocations are added to the lexicon of languages which may make translating them more challenging. On the other hand, he calls the productive problematic phase of collocation 'interlingual' (2006:135,136), which occurs when trying to translate one of these collocations into another language. To illustrate, every language has its range of collocation some of which are language-specific or cultural-bound and the collocability of lexemes differs from one language to another. An example of cultural-bound collocations is the collocation of 'bread' and 'butter' in English which does not exist in Arabic (Baker, 1992:49). To explain that every language has its system of

collocations, Baker used the verb 'deliver' to illustrate how the verb collocates with a variety of nouns, and every collocation has different verb equivalent in Arabic; for instance, 'deliver a letter' is translated as 'yusallimu risaalatan': 'hand a letter' while 'deliver a speech' means 'yulqii khutbatan': 'throw a speech' in Arabic (1992:48). If a translator would check a bilingual dictionary like *Al-Mawrid* to find the meaning of 'deliver', s/he would find six different verbs<sup>1</sup> with no explanation of the different words that may collocate with the verb. This kind of variation would cause a problem in translating especially when collocations of the same verb occur within the same text.

Another difficulty of translating collocation occurs when the meaning of collocation has nothing to do with the meanings of its constituents, for example, 'run a car' does not mean 'drive a car fast' (Baker, 1992:53). Newmark stresses the difficulty of translating collocations when he points out that, "the translator will be 'caught' every time, not by his grammar, ... not by his vocabulary, ... but his unacceptable or improbable collocations" (1988: 180). Also "The problem of searching for 'acceptable' collocations is aggravated by the lack of adequate bilingual resources on collocation (dictionaries, thesauruses, etc.) to which the translator may look" (Bahumaid, 2006:138).

To overcome collocation as a problem, translators need to look after comprehensive as well as productive difficulty of collocations. At comprehensive stage of ST, using mono-lingual collocation dictionaries in SL and comprehensive exposure to more texts within the same field would cultivate translator's capability to digest the ST collocation. There is a number of collocation dictionaries available, among them: *the BBI Combinatory Dictionary of English* (1997), *LTP Dictionary of*

---

<sup>1</sup> Al-Mawrid English/Arabic Dictionary 2000.

*Selected Collocations* (1991), *A Dictionary of Collocations* (1994) and *Oxford Collocations Dictionary* (2002) (Bahumaid, 2006:137). Another provider of translation aid is corpus data, "further, online corpora (e.g. *The Bank of English*) with their concordances provide an enormous amount of information on permissible combinations for the translator" (*ibid*). In addition, an attempt to understand the collocation within its culture would solve half of the problem because a good number of collocations are culturally bound. Meanwhile, bilingual collocation dictionaries (if available) as well as good knowledge of TL collocation system and avoidance of literal translation would improve the quality of translation. Similarly parallel/comparable corpora would reduce translators' effort and time by providing direct equivalences within their contexts.

### **3.2. Collocation and Corpus**

Translating collocation, as mentioned above is one of major problems for translators, so s/he is in need for tools to assist the process of translation. One of the most important support providers is corpus. Corpus information, either mono-lingual or bi/multi-lingual ones, can help in many ways. First the use of concordance lines of mono-lingual corpus is beneficial in different aspects, and concordance "provide a simple way of placing each word back in its original context, so that the details of its use and behavior can be properly examined" (Barnbrook, 1996:65); concordance line, thus, would display collocate words within context (Bahumaid, 2006:137), which reveal different meanings within different frameworks (Renouf & Sinclair, 1991:128). Collocations are a merge of grammar and lexical words, so "the concordancing feature of a corpus analysis tool, however, can quickly bring together all occurrences of a given pattern." (Bowker, 1998:5).

To elaborate, there are three features available in corpora, such as *Bank of English*, which integrate together to help the translator in dealing with collocation. These features are: collocation views, concordance lines, and frequency lists. The collocation view of the verb 'pay' shows the lexical words collocate with the verb, such as 'attention, money, and rise', and the grammatical words, like 'to, for, and, have', (please refer to Figure1 in Appendix 1), while the concordance lines view the verb behaving within the context to figure out the meaning (please refer to Figure 2 in Appendix 1). These corpus features would help translators to build up a good background about all words that may collocate with verb 'pay' in order to be able to pick up the best collocation especially when translating into English. Also concordance would present every collocation of 'pay' within its context to smooth the task of finding out the meaning of every collocation. Another facility of concordance is sorting, which is the ability to alter the order of concordance lines according to the appearance of the keyword in the corpus or a specified number of words to the right or the left of the keyword (Barnbrook, 1996:73). If we sort the first five lines of figure 2 in Appendix 1 alphabetically according to the first word to the right of 'pay', the following two lines will be the first:

right to have people pay close attention to  
It's a small amount to pay for a room that makes

This sorting would help the translator to find the collocation under investigation quickly within its context.

In addition corpora help in calculating the frequency score of words collocate with the word in search. For example, 'to' has the highest frequency score of grammatical words that collocate with the verb 'pay', while 'rise' has the highest score of lexical words. This frequency calculation may help the translators particularly when translating into his/her non-native language because frequency calculation

would help him/her to choose the best collocation and the best pattern. Hence all the above tools would help to reach the exact meaning of the collocation. In addition, Carter emphasizes the potential of corpus data to help in recognizing the meaning of collocation words by spotting the grammatical pattern when states:

Francis (1994) points out that the two verbs *find* and *make* occur in 98 per cent of cases in the extraposed structure with *it* in clauses such as: *I find it amusing that he never replies to my faxes, can you make it more exciting? And I owed it to you that I passed the final exam* (1998:63).

Meanwhile, a mono-lingual corpus would be indirectly resourceful, using collocation lists and concordance lines as examples, to compile electronic comprehensive collocation dictionaries, such as *Collins Cobuild English Words in Use* (1997), which includes over 100,000 different collocations of different words (Carter, 1998:171). "The corpus can be used extensively, as the main source of information for the lexicographer..... The result is a dictionary *Collins COBUILD English Dictionary* that prioritizes the phraseology of a word." (Birmingham CD, 2001:Unit 3). For example, the above dictionary lists all the possible collocation of 'sharp' provided by illustrative examples that show how the adjective used within texts differently (*ibid*). Therefore, such a corpus-based dictionary will be absolutely helpful in translating collocation because it promotes the chance of understanding the ST collocations for better and proper translation.

Second bi-lingual parallel/comparable corpus will be a great aid by providing translators with previous translation of specific collocations in different fields which means a better chance to produce better translations. Zanettin stresses this crucial application as he says: "comparable corpora offer no single translationally established correspondence, but present a repertoire of recurring collocational or structural features which provide a basis for establishing equivalence between stretches of texts in two languages." (1998:12). He supports this advantage by displaying some

concordance lines, from English-Italian comparable corpus<sup>2</sup> about some collocations of 'prices', 'prezzi' (Please refer to Appendix 2) and concludes that nominal groups such as '*l'aumento dei prezzi*' or '*la discesa dei prezzi*' are more common in Italian, while verbal groups such as '*prices soared*' or '*prices fell*' are more common in English (1998:11).

The *Collins Cobuild English Language Dictionary*, mentioned above, is a monolingual dictionary that is crucial for collocation translation; however a bi-lingual collocation dictionary will be more beneficial because it will display SL collocation accompanied with the TL equivalences. Bi-lingual parallel corpora will be a good resource to exploit such bi-lingual dictionaries.

#### **4. Ambiguity**

Ambiguity "is the property of words, terms, notations and concepts (within a particular context) as being undefined, undefinable, or without an obvious definition and thus having an unclear meaning", this is the definition of ambiguity in its broad concept, which differs from vagueness, the unclear boundaries of words meanings (Wikipedia).

Ambiguity is in two kinds: lexical (semantic) ambiguity, when the word has more than one meaning and structural (syntactic) ambiguity, which occurs when a phrase or a sentence has more than one meaning (Arnold, 2001:105). Polysemy (one word with two or more related meanings) and homonymy (one word with two or more distinct meanings) are two different aspects of lexical ambiguity, for example 'bank' as a river side or a financial institution are two homonyms of 'bank', while 'bank' as a financial institution and to rely upon are two different polysemies of the word (Wikipedia).

---

<sup>2</sup> comparable corpus of foreign news articles about France (from *The Independent*, *the Daily Telegraph*, and *Il Sole 24 Ore* of 1992, of about 152,000 in English and 102,000 words in Italian) (Zanittin, 1998:12, 13)

On the contrary, structural (syntactic) ambiguity takes place when a phrase or a sentence can be understood in more than one way; for example 'he ate the cookies on the couch' can have two underlined interpretations: 1) the man ate the cookies which were on the couch, or 2) he ate the cookies while he was on the couch (ibid).

#### **4.1. Ambiguity as a Translation Problem**

Generally speaking "English word *string* has 17 different meanings according to WordNet. To make everything even more difficult, these 17 meanings can be (in this case) either nouns or verbs. Even if the word *string* was found in a dictionary it would have 17 senses." (Oikarinen, 2007:3). This statistical notion indicates higher probability of facing lexical ambiguity within a text and proves that ambiguity is one of common translation problems that translators may face. To translate a lexical ambiguous word, such as 'fair', or a syntactic ambiguous phrase like 'flying planes can be dangerous', translators need to look carefully at the context in order to draw the exact meaning of the word or the phrase.

The translators have two options: either to add more contextual information to disambiguate the text or preserve some degree of the ambiguity (Falk, 2007:15-17). In other words, disambiguation can be done by paraphrasing the sentence in order to bring the intended meaning to the surface. Otherwise, maintaining the ambiguity is applicable because, according to Prahl & Petzolt's study (1997, cited in Falk, 2007), titled "Standard Assumptions", on how to translate an ambiguous text, experienced translators tend to preserve some of the ambiguity and keep some of the ST implicit elements which can be assumed within the text, especially when there is a lack of contextual and linguistic information, which are vital to disambiguate.

## **4.2. Ambiguity and Corpus**

Ambiguity is a sort of problem that needs acquiring more knowledge about words behavior within texts, and corpus offers this knowledge in different aspects. For one thing, using collocation analysis provided by corpus data would show that the occurrence of a specific meaning of an ambiguous word or phrase depends on a specific framework (Barnbrook, 1996:103). Hunston and Francis (1998) proved in their corpus-based study, titled "Verbs Observed: a Corpus-driven Pedagogic Grammar" that patterns and framework are strongly associated with meaning, which led them to conclude that meaning can be derived from the grammatical pattern. For example, if the pattern 'V+by+-ing' contains an ambiguous word like 'set' the meaning of the phrase will be either 'start/finish' or 'respond to' (Carter, 1998:64); however, this kind of conclusion will not be easily drawn without accessing corpus data. Also such corpus-based assumptions about meanings of grammatical patterns would help translators to draw the meanings of ambiguous words from the pattern or the framework rather than the lexical constituents themselves.

Furthermore, semantic and syntactic tagging, which is one of the common techniques offered by corpora, would solve the lexical as well as structural ambiguity (Barnbrook, 1996:103). To illustrate, 'I do like you' is structurally and semantically ambiguous. It has two underlying meanings: 1) I love you and 2) I do as what you do. The following concordance lines are cited from the Bank of English<sup>3</sup>:

- 1- Steely Dan record but I do like the band I think they're
- 2- talk to you John and I do like your programme. <M01> Ah
- 3- to the people is to do like they did in Hull but can't

---

<sup>3</sup> Bank of English, search keyword: do+like in academic books sub corpus. These lines are randomly selected.

- 4- is if every day I could do like a minute. <F04> That'd be
- 5- have anything else to do like I do like to make people
- 6- Education Centre. They do like all aspects of education

In lines 1, and 2, 'like' has been used as a verb while 'like' in the rest is a preposition. Tagging the words with part of speech, within the text, would demolish the vagueness of the word and eventually translating it will not be a hard task. However which part of speech the word 'like' would hold is absolutely context-dependent.

Another way of using corpora in disambiguating a text or a word is to use parallel corpus to check how an ambiguous word such as 'fair' has been translated from one language to another. In *Al-Mawrid* (2000), the English/Arabic dictionary, for example, 'fair' has seventeen different meanings, which makes it very hard to spot one meaning. As mentioned previously, using concordance lines from a mono-lingual corpus (refer to Appendix 3) might be helpful, but using an English/Arabic parallel corpus would be more helpful, because it provides the translator with different possible translations, which would show how translators tend to disambiguate the word. King illustrates this advantage of parallel corpus when suggests that: "a number of translations of a single ST will show considerable variation in the choices made by the translators, and that good cases may be makeable for various of the choices" (2003:163). Moreover, Rogati (2003:2), in his study about Arabic stemming, mentioned that Diab and Rensik (2002), using a parallel corpus, proved that translators tend to disambiguate lexical ambiguous words by providing different equivalences.

After all, parallel corpora can be used as translation banks to which translators can go back to find ready-made solutions while mono-lingual corpora can be used as an illustrative analytical tool to elucidate ambiguity.

## 5. Conclusion

Translating some genuine translation problems such as collocation and ambiguity is a challenging job that requires wide knowledge of languages involved, rich translation resources, and cultivated translation skills. Among the most useful translation recourses are corpus-related resources either database corpora or corpus-based dictionaries. These corpus resources can aid the translation process in different ways: 1) displaying the problematic word or phrase within its context by using concordance lines to show how these words behave in different sentences, 2) providing statistical quantitative information about these problematic words, by listing all possible meanings which may assist translators to draw conclusions about alternative meanings, 3) suggesting different possible translations through surfing parallel corpora, which offer real translations of the phrase or the word in search.

To conclude, exploiting more parallel/comparable corpora either general or field-specific is more demandable these days because the world is turning to be a global village in which exchanging information is part of everyday life, and one of the couriers of this information circulation is translation. On the other hand, one of the most exhausting tasks, I faced, was finding an English-Arabic parallel corpus to gain some illustrative examples from. Few English-Arabic corpora are available on line, but unfortunately they are not free and long time would be needed to subscribe to them. So exploiting such corpus will be absolutely helpful for future studies. Also the lack of bi-lingual collocation dictionaries was one of the most significant shortages,

though parallel/comparable corpora would provide rich resources to compile these specialized dictionaries.

## Bibliography

- Abdellah, A. July 2002. "What Every Novice Translator Should Know". In *Translation Journal*, vol: 6, no: 3.
- Aijmer, K and Altenberg, B (Eds). 1991. *English Corpus Linguistics*. Longman London and New York.
- Al-Mawrid 2000. *A Modern English-Arabic Dictionary*. Dar El-Ilm Lil-Malayen, Beirut, Lebanon.
- Arnold, D, et al. 1994. *Machine translation: An Introductory Guide*. Electronic PostScript/ PDF book, Retrieved on 14 April 2007-05-13 from <http://www.essex.ac.uk/linguistics/clmt/MTbook/>
- Bahumaid, S. 2007. "Collocation in English-Arabic Translation". In *Babel*, vol: 52, issue: 2.
- Baker, M. 1992. *In Other Words: A Coursebook on Translation*. Routledge, London and New York.
- Baker, M, et al (Eds) 1993. "Corpus Linguistics and Translation Studies: Implications and Applications". In *Text and Technology*, Philadelphia/Amsterdam, John Benjamins.
- Barnbrook, G. 1996. *Language and Computers: A Practical Introduction to the Computer Analysis of Language*. Edinburgh University Press.
- Bowker, L. 1998. "Using Specialized Monolingual Native-Language Corpora as a Translation Resource: a Pilot Study". In *Meta*, Vol: XLIII, Issue: 4.
- Carter, R. 1998. *Vocabulary: Applied Linguistics Perspectives*. Routledge Taylor and Francis Group, London and New York.
- Falk, C, et al. 2007. "Translation Error Types". PDF Presentation, Retrieved on 6 May 2007 from

[http://coral.lili.uni-bielefeld.de/~ttrippel/cat/Translation\\_Error\\_Types-1.pdf](http://coral.lili.uni-bielefeld.de/~ttrippel/cat/Translation_Error_Types-1.pdf)

Gonzalez Davis, M. 2005. "A Problem-Solving and Student-Centered Approach to the Translation of Cultural References". In *Meta*, Vol: L, Issue: 1.

Hundston, S and Francis, G. 1998. "Verbs Observed: A Corpus-driven Pedagogic Grammar". In: *Applied Linguistics*, Vol: 19, (1), pp. 45-72. Oxford university press.

Johansson, S. 1991. "Times Change, and So Do Corpora". In Aijmer, K and Altenberg, B (Eds). 1991. *English Corpus Linguistics*. Longman ondon and New York.

Kennedy, G. 1998. *An Introduction to corpus Linguistics*. Longman, London and New York.

King, P. 2003. "Parallel Concordancing and Its Applications". In *Corpus-based Approaches to Contrastive Linguistics and translation Studies*. Amsterdam-New York, NY.

Laufer, S. 2004. "Human Interaction with Computer-assisted Technology: A Study of Translator Experience". In *International Conference on Translating with Computer-Assisted Technology: Changes in Research, Teaching, Evaluation and Practice*, retrieved on 19 April 2007 from

[http://w3.uniroma1.it/diplingue/convegno2004/abstracts.html#lauffer\\_relaz](http://w3.uniroma1.it/diplingue/convegno2004/abstracts.html#lauffer_relaz):

Newmark, P. 1988. *A Textbook of Translation*. New York: Prentice Hall International.

Oikarinen, L. 2007. "Dictionary Based Translation in Cross-Language Information Retrieval System". Essay in Language Technology and Applications, retrieved on 6 May 2007 from

[www.cs.helsinki.fi/u/linden/teaching/Ita07/drafts/LTA07\\_Liisa.pdf](http://www.cs.helsinki.fi/u/linden/teaching/Ita07/drafts/LTA07_Liisa.pdf)

Olohan, M. 2003. "How Frequent Are the Contractions?". In *Target*, Vol: 15, Issue: 1.

Open Distance Learning Postgraduate Programmes CD ROM Course Material.

University of Birmingham.

Renouf, A and Sinclair, J. 1991. "Collocation frameworks in English". In Aijmer, K and Altenberg, B (Eds). 1991. *English Corpus Linguistics*. Longman London and New York

Rogati, M, et al. 2003. "Unsupervised Learning of Arabic Stemming Using a Parallel Corpus". Proceeding of the 41<sup>st</sup> annual Meeting of the Association for Computational Linguistics, July 2003, pp. 391-398, retrieved on 6 May 2007 from <http://citeseer.ist.psu.edu/rogati03unsupervised.html>

Torrejin, E and Prez, C. 2004. "The translator in a Controlled translation environment: Key Elements for Understanding Technology". In *International Conference on Translating with Computer-Assisted Technology: Changes in Research, Teaching, Evaluation and Practice*, retrieved on 19 April 2007 from:

[http://w3.uniroma1.it/diplingue/convegno2004/abstracts.html#lauffer\\_relaz](http://w3.uniroma1.it/diplingue/convegno2004/abstracts.html#lauffer_relaz)

Zanenttin, k. 1998. "Bilingual Comparable Corpora and the Training of Translators". In *Meta*, Vol: XLII, Issue: 4.

University of Birmingham. *Bank of English*. Retrieved on 8 May 2007.

Wikipedia, the Free Encyclopedia. [http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page), retrieved on 3 May 2007.

## Appendix 1

**Figure 1**

Search word: pay. Sort order: two words to both sides. Retrieved on 8 May 2007

to	397252	2140	38.204348
for	100986	935	27.479608
have	109742	516	18.183562
rise	516	114	10.631742
money	11712	118	9.851344
tax	1457	96	9.658459
t	210204	380	9.377843
rent	649	85	9.153508
you	475890	680	8.956924
got	69009	186	8.891414
attention	758	68	8.159980
them	47377	138	7.963988
ll	35586	120	7.906996
more	32360	115	7.893011
off	16413	88	7.739504
can	70638	165	7.686474
had	51682	139	7.677565
their	22354	96	7.657690
back	22651	95	7.566705
bills	404	51	7.088359
your	46058	121	7.072094
afford	1186	52	7.056815
will	19436	82	7.041899
should	13381	71	6.936418
rises	137	48	6.909653

**Figure 2**

Search word: pay in British spoken English sub-corpus. Retrieved on 8 May 2007.

promise the greatest pay-offs for the economy  
right to have people pay close attention to  
without having to pay out a fortune. She  
It's a small amount to pay for a room that makes  
bit pathetic to have to pay to get aroused." <p>  
led President Reagan to pay lip service to AIDS,  
the cost of maternity pay above the legal  
company is helping to pay for it. He tells me  
She certainly made me pay, made me feel like  
are prepared to pay the price. This is a  
work is major, it will pay to hire a chartered  
Romeo and Juliet who pay for love with their  
the Close Season, yet I pay rates for 12 months,"  
is going to have to pay to see the Milk Race  
biggest stars gather to pay their respects, but  
collected money to pay the damages awarded  
of the bride has to pay. Maybe it was a  
artist, she agreed to pay for his passage to  
supermarkets should pay the same, or higher,  
actress Kim Basinger to pay \$8.9 million (later  
12's Youth Orchestra to pay 'Happy Birthday'  
signed Deano did they pay &hellip; <p> Was it  
from various teams to pay for the damage. <p>  
<p> WISEST WORDS <p> T: Pay attention to your  
debited. But if you do pay charges, you never  
idea of being able to pay the premiums each  
laws ensuring equal pay and rights. It would  
Herts wdl 1DN, or pay by credit card by  
the facility to pay over 50 weeks but  
in love-now stars to pay golden dividends. On  
the classroom teachers' pay spine to at least

unwilling to pay for anyone to learn  
 been suspended on full pay pending the findings  
 London. Our commuters pay by far the highest  
 balls and brains to pay." However, we  
 know your patience will pay off. Timing is  
 demanded that Combined pay him huge sums of  
 moul, while the AZ/DWV pay £12,000 a year to the  
 Children under 15 pay £4 (two fish).  
 after neglecting to pay the bill for last  
 extension of bskyb's pay channels. It quickly  
 is ordered to pay Pele £2million in  
 and that he would pay money for even half  
 there's a price to pay if they step outside  
 them to one side and pay lip service to your  
 killed. That's what we pay for, isn't it? So  
 s inability to pay a 300 per cent

## Appendix 2

These data are cited from Zanittin, F. 1998. "Bilingual Comparable Corpora and the Training of Translators". In *Meta*, Vol: XLIII, Issue: 4.

### Figure 1

Search word: PREZZI. Sort order: 2<sup>nd</sup> word to left, then 1<sup>st</sup> word to left

recente, sensibile abbassamento dei **prezzi** sui mercati tedeschi ha reso meno  
 tto Saddam", che si e' abbattuto sui **prezzi** petroliferi dall' inizio di  
 e di flessione si e' avuto anche sui **prezzi** del provolone, mentre il burro ha  
 riforme. Gli stock aumentano, alcuni **prezzi** scendono e inevitabilmente  
 di attendersi ulteriori aumenti dei **prezzi**". Secondo la fonte, l' incertezza  
 lettriche. Concedendo l' aumento dei **prezzi** chiesto dall' Enel, il Governo  
 erso il terzo round e un aumento dei **prezzi** petroliferi in una condizione  
 ella degli interessi; l' aumento dei **prezzi** si scarichera' su BoT e Cct e  
 usa prima del recupero autunnale dei **prezzi**. Il movimento rialzista non era  
 l mercato comunitario, la caduta dei **prezzi** e la necessita' per la  
 in malinconici ammassi; il calo dei **prezzi** internazionali del frumento fa  
 a scorta di manovra ogni volta che i **prezzi** scendevano sotto un livello  
 na allo Sme il tasso di crescita dei **prezzi** in Gran Bretagna dovrà scendere  
 a del mercato del Golfo e crollo dei **prezzi**, anche sull' onda della caduta  
 nsieme, hanno generato un crollo dei **prezzi** e reso necessarie delle misure d'  
 he da carne). Domanda molto debole e **prezzi** nettamente cedenti, per contro,  
 Medio Oriente ha mantenuto elevati i **prezzi** delle scadenze lontane e non ha

### Figure 2

Search word: PRICES. Sort order: 1<sup>st</sup> right, 2<sup>nd</sup> right

tions against the steep rise in food **prices**. A court in Marrakesh sentenced  
 tonnes might be delivered, squeezing **prices** acutely. In that event, smaller  
 day, but said no relief from soaring **prices** and food queues could be expected  
 will step in themselves to maintain **prices**, as they did in the wake of the  
 eir nadir because of the hike in oil **prices**. At the same time inflation and  
 ven in the 1973/74 collapse, nominal **prices** did not fall. The 'wealth effect"  
 s. Properties will stick rather than **prices** fall." Barclays Bank said that t  
 p 3.9 points at 1,775.6. Bombay: **Prices** fell sharply for the second day  
 are traded publicly. Last month junk **prices** fell sharply on news that Campeau  
 market drifted in sympathy. Bombay: **Prices** fell on heavy speculative selling  
 poll discloses a decline in selling **prices** for the second successive quarter  
 weekend sales in New York and Paris **prices** for post-war French artists soared  
 ominated a moderate session in which **prices** generally fell back from early ad  
 delayed to the end of 1991." House **prices** have already begun to fall across  
 ial said. Jaguar, which dropped its **prices** in North America last year, now  
 yield the best opportunities, as the **prices** of companies fall. Jon Moulton,

st's cocoa crop indirectly pushed up **prices** on the London Futures and Option  
ors. Page 22 World Markets New York: **Prices** plunged and by the

## Appendix 3

Search word: fair in British spoken English sub-corpus. Retrieved on 19 May 2007

team cricket at the time so to be fair to his development <ZF1> we  
answer this if you don't want to it's fair enough. I mean if you say I  
to go on strike as well. Which is not fair to the. <M01> Well it's very  
Okay MX <ZF1> that's er <ZF0> that's fair enough that's your comment  
leaders and said Well it's just not fair erm I've had to go home  
<ZF0> you decide you don't want to go fair enough <ZF1> then <ZF0> then  
it's going? <M0X> Er yes I think it's fair enough to say in between  
damaged. <M01> Sometimes to be fair though I mean it's even  
t erm <ZG1> that hygiene is <ZGY> fair enough right you dunno where  
Home. <M01> Well that sounds like a fair ambition to me MX. <M02>  
John. Good evening. <M01> How is the fair city of <PN1> Chadsley  
accident. <M02> Yes well that's fair enough you've got a job to  
<ZZ0> <ZZ1> <!--music--> <ZZ0> Fair Blows The Wind For France.  
all the time because it just isn't fair basically <F01> No. <F02>  
respect in the sport. And an ethos of fair play. <M01> Yes. You quote  
I was I must admit I was up town a fair while because I had to get  
is that I <ZGY> <M01> No. No. That's fair enough. <M02> Erm I still  
<ZG1> bared <ZG0> care loud nurse fair look place luck fur plaice  
unaffected. <M01> <ZGY> I mean <ZGY> fair <F01> Mm. Oh I think a  
<!--of--> <!--muezzin--> <ZZ0> To be fair to the guidebooks  
is very very far from being fair to most nurses. <M01> And