DISPOSABLE CORPORA AS INTELLIGENT TOOLS IN TRANSLATION

Krista Varantola
University of Tampere

1. Introduction

Intelligent electronic dictionaries are currently being designed and constructed. What is normally meant by 'intelligence' in this context is that these dictionaries adapt to users' needs and allow user profiling. Intelligent dictionaries will thus take into account the users' level of language skills, the type of work they are doing with the dictionaries and also allow the users to define the type of information they would like to be displayed as an answer to their queries.

The overall objective is thus to improve a well-known search tool by benefiting from the advances of information technology. The purpose is also to make the tool more usable and user-friendly by providing the available information in a form that helps the users locate the information they are looking for without a great deal of irrelevant and redundant data. Alternatively, the users should also have a chance to ask for additional targeted information.

Could we set a similar goal for the representation of corpus information? In other words, could we envisage a way of retrieving relevant corpus data from an unstructured and unspecified mass of text to give users access to the information they are looking for?

I will explore this issue in a relatively narrow context. My users are translators who are either translating a text from their native

language into a foreign language, or producing texts in a foreign language. The native language in this case is Finnish and the foreign language English. The unspecified amount of text comes from the Internet and the tool the translators have at their disposal is a tool box filled with text analysis tools, mainly the WordSmith Tools, including a concordancer, a wordlist and a keyword function.[1]

2. Premises

Why would we resort to corpus information in translation? Are dictionaries not sufficient resources for general lexical information? Why not try to improve dictionaries instead of trying to complicate matters further by introducing a new source of information which presupposes a great deal of technical competence? In an earlier context (Varantola 1998), I have summarized the reasons why translators are frustrated by dictionaries:

- dictionary-makers usually aim at context-free descriptions of word use, whereas dictionary users resort to dictionaries to solve a context-dependent problem.

- translators certainly need equivalents, but they also need reassurance: for this reason translators do not like to find equivalents that they do not recognize

- translators often need information relating to longer stretches of text than a single lexical item

- translators try to find non-dictionary type information in dictionaries because it is not readily and systematically available in other sources

It is obvious that print dictionaries cannot cater to the particular needs of their individual users. In the future, intelligent, electronic

dictionaries may approach this ideal, but I do not think that even they will ever manage to address the very specialized needs of their most demanding and suspicious users. I am referring to the idiosyncratic and sophisticated needs of language professionals who want to be able to make their own decisions on the basis of cumulative evidence they have gathered from different sources during their information retrieval operations.

It has been estimated that translators may spend up to 50 % of their total time performing a particular translation task when trying to find relevant lexical information. We can therefore argue that the more varied tools language professionals have at their disposal, the better their decisions and work will be. This in turn will result in professionals who are much more satisfied with their tools and the ways they can employ these tools.

3. Competence vs. performance in translation

When discussing the qualifications of professional translators, we can make a distinction between translational competence and actual performance. Moreover we can suggest that a mismatch may arise between underlying competence and "surface" performance, because translators have not had access to information that would have been vital in the decision-making process, even though they have been clearly aware of what the problem is. This controversy is often behind the frustration and dissatisfaction that translators feel towards their reference sources, particularly towards dictionaries and terminological glossaries. (For a further discussion about these issues, see e.g. Atkins & Varantola 1997, Varantola 1997 and 2002.)

4. Information needs and corpus information

4.1 Types of translation corpora

Before embarking on a more detailed discussion of how disposable corpora could help in increasing translator satisfaction, a few words are first needed about the types of corpora available and their usability in different translation contexts.

Some ten years ago corpus size was the only thing that seemed to matter, but now when size is no longer an issue and there is an abundance of electronic text available, the "my corpus-is-bigger-than-yours" rhetoric has subsided even in lexicographical contexts. Instead the focus has moved on to a discussion of the adequacy of the corpus for its purpose, onto a "my-corpus-is-smarter-than-yours" rhetoric. Parallel and comparable corpora are the types of corpora that have received the greatest attention in translation contexts. There already exist a number of permanent or semi-permanent collections of bi- or multilingual corpora that have been used in research on the actual translation process, translationese, i.e. the differences between translated and authentic texts, and translator training (See e.g. Baker, 1995, 1996, 1999, Bernardini & Zanettin 2000, Bowker 1999, 2002 Laviosa 1997, META 4/1998, Maia, 1997/2000, Tagnin in this volume, Varantola 1997/2000, 1998 and Varantola 2002).

The terminological controversy between parallel and comparable corpora also seems to have been resolved. Parallel corpora refer to collections of source language texts and their translations. A comparable corpus is a collection of authentic source language texts in more than one language dealing with the same subject matter, style, etc. In other words, these are electronic texts collections of the type translators used to call parallel texts.

The disposable corpora I am focusing on in this article can be seen as an extreme type on the scale of comparable corpora.

Typically, the source text to be translated is merely a text, just material in the source language that forms the background material for the text that is to be produced in the target language, or a text already in the target language that needs to be edited or updated. The bulk of the corpus consists of comparable electronic text material in the target language compiled for the particular translation, editing or text production task.

The most available source for this type of material is naturally the Web, but nothing prevents the compiler from using more permanent sources such as encyclopedias and other CD-ROM publications as source material. Typically the resulting corpus is disposable and does not need to be technically elaborate. Yet we can hold on to the metaphor and say that should the need arise, this disposable material can be recycled and refined to form part of a more permanent collection.

4.2 What is corpus information good for?

I have conducted a number of studies with a group of trainee translators (Varantola 1998, 2000, in print), which seem to indicate that corpora are a very useful source of reference for certain types of linguistic queries. I will list these types below but will not make any claims as to the list being exhaustive. The list should rather be regarded as a way of highlighting some interesting tentative findings that need to be further elaborated and refined in future research.

4.2.1 Translators need equivalents, but they also need reassurance: therefore translators do not like to find equivalents which they do not recognize. A typical example of how corpus information could help translators in these cases is when they wanted to check in relevant corpus material about fishing as to whether ***hatch*** and ***fry*** were adequate terms when discussing fish populations in English. The dictionaries gave these words as equivalents but provided no examples. This did not satisfy the

translators, because they had not come across these terms earlier in fishing contexts. They therefore wanted to make sure from a corpus that these terms were actually adequate and in current use (cf. Atkins & Varantola 1997, Varantola 1997, 2000).

A slightly more complicated case is the one in which the users wanted to find the opposite of a ***fine-meshed*** **net.** Dictionaries typically showed fine-meshed in their usage examples but none gave any examples of ***large-meshed, bigger-meshed, coarse-meshed, loose-meshed*** – all examples of searches the users in the particular experiment tried to find in their monolingual dictionaries. And indeed, this was a problematic case in the particular experiment.

Native speakers also encountered problems when trying to decide what the most adequate expression would be, even though they could naturally exclude some alternatives. The default concept seemed to be ***fine-meshed***, whereas its opposite term was apparently a much fuzzier concept. The BNC was checked and it had only a few examples of ***large-meshed***. Today if we use a search engine on the Web, e.g. Google, to confirm or refute our hunches, we are immediately reassured that ***large-meshed*** is adequate and in current use. The only problem of course is that we first have to come up with this alternative before we can find it. Therefore, in trickier cases, adequate corpus data is indeed the source where we often get our first hunch of what an adequate expression might be in the target language.

Currently, up-to-date electronic dictionaries with free text search provide answers to queries that were hard to solve in the pre-electronic times. At that time, finding the English equivalent for a "reel seat belt" (word-for-word translation from Finnish) was time-consuming and frustrating. This was a new concept at the time which had not yet reached the bilingual dictionaries. Using ***reel*** as the search word did not produce any results in monolingual dictionaries. Now a Google search with a partly correct search word immediately provides you with references and contexts to ***inertia reel seat belts***.

In other words, these days partially correct hunches are enough and there is no need to go to the trouble of compiling a corpus and acquiring corpus evidence, because a simple Web search will be reassuring enough. Moreover, there are still many cases left for which the translator may have no hunch whatsoever. Educated guesses can then arise from corpus evidence.

Examples of such finds are betting terms that could not be found in any reference sources nor predicted in the translator's armchair but were instead found in adequate corpus material:

> team at high odds
> bet is in bettor's favour
> *combination of three matches*

Similarly "**basic medicine**" did not sound like an adequate rendering for a medical concept in Finnish but the disposable corpus highlighted a number of expressions with ***first-line*** and ***first-line*** **antihypertensive therapy** was what the translators were looking for. The relevant texts had been found by using the following key words in the search for texts: **therapy, medication, drug.**

4.2.2 Translators often need information relating to longer stretches of text than a single lexical item. The problem is that relevant passages are hard to find. A case in point are idiomatic paraphrases. The translator may very well realize that the Finnish way of expressing an idea needs to be reformulated when it is expressed in English but it can be very difficult to come up with a satisfactory solution when one is translating into a foreign language. The Finnish text referred to above discusses fish populations and deplored the "***weak age groups***" of a certain fish in a large lake. It was clear to all translator trainees who tried to express this idea in English that "weak age groups" was not an appropriate way, but none of them could think of anything that they were happy with. A comparable corpus text found after this exercise contained the following passage:

This results in <u>reduced populations</u>, and even <u>wholesale losses</u>
of some species, of both invertebrate and vertebrate fauna.

Reduced populations would have been just what the frustrated
test subjects would have needed when they were struggling with the
Finnish way of expressing the idea (Varantola 1997).

4.2.3 Fuzzy searches and serendipitous finds. Translators´
information retrieval is not only linked to problem-solving but also
to the identification of problems. The process of problem-
identification is, however, often subconscious. Translators may not
consciously verbalize the problem or even be aware of it but are
nevertheless able to spot the solution when they come across a
suitable stretch of text in a comparable corpus that highlights the
issue. I call this type of information retrieval "fuzzy searching"
and the results "serendipitous finds". The above example with
reduced populations can be described as a serendipitous find. The
same text where this collocation was found also contained such
expressions as

> *The nature conservation impacts* of *enrichment … rare
> freshwater fish* …Their *limited distribution* …acutely
> *vulnerable to pollution …one of these post-glacial relics*, the
> vendace, was *wiped out* …by sewage discharges. This results
> in *reduced populations*, and even *wholesale losses* …simple
> loss of *habitat diversity* in rivers, streams and lakes …the
> water, *detergent and fertiliser industries.*

> The ecological effects of eutrophication in a pristine *water
> body …sewage discharges and agricultural pollution … fish
> farming* can affect more remote areas severely …*phosphorus
> losses into watercourses* from fertilised forestry plantations
> …and one or two other *waters … eutrophication* – the scientific
> word for *enrichment* … a *remedial action plan* (The source is

> the MicroConcord general text corpus based on *The Independent on Sunday*.)

It is very likely that a translator working on a text dealing with environmental issues in watercourses would benefit from non-definable strings of this type and would be able to apply the information contained in them in the target context.

4.2.4 Text editing. Translators are often asked to edit existing texts. The need for editing may arise because the texts need updating, are stylistically inadequate, unidiomatic or otherwise inappropriate. In one study, two groups of students were asked to improve the quality of two instructional booklets. The booklets had already been published in English and were intended for the general non-Finnish speaking population in Finland. One booklet dealt with hypertension and the other with HIV.

On the basis of corpus evidence, the first group decided on the following types of lexical amendments:

> ***patient*** over ***subject***
> ***irregular heart rhythm*** over ***arrhythmia***
> ***over-the-counter medicine*** over ***non-prescription drug***
> ***elderly*** over ***older people***
> ***drugless treatment*** over ***nondrug treatment***

The second group editing the HIV text decided that such expressions as ***asymptomatic*** or ***seroconversion*** were not general enough to be used without definitions embedded in the text. ***Lymph nodes*** nevertheless appeared in a number of corpus texts and seemed thus also to be common in general texts. ***Lymph nodes*** were, however, sometimes also referred to as (***swollen) glands*** in the corpus. The group decided to use both of these expressions in their own text but at the same time make sure that it was explicitly indicated in the text that the expressions

were used synonymously. The group also decided to watch out for synonym use in general in order to avoid comprehension problems. They noticed that synonymous terms can cause a great deal of confusion if their internal relations are not explicitly spelled out in the text.

Frequency of occurrence was the main criterion used when deciding on the adequacy of the terms. This criterion gave the editors the justification they were looking for. If the more technical terms were rare in the corpus, which consisted of a number of authentic English texts intended for the general public, the students felt that they should opt for the more general expressions and that their choices were well justified. Corpus evidence on the frequency of usage therefore helped the student editors to adjust the level of the approach in the editing process and it also provided the editors with the necessary self-confidence when they wanted to make more radical changes.

For text editing, the issue is clearly adequacy and rarely correctness and it can be claimed that the changes made are often based on subjective decisions. It is arguable whether the editorial changes were always relevant and necessary, but with recourse to corpus data, the editors could at least be sure that their opinions and decisions were justified and corroborated by other writers. In brief, corpus evidence sensitizes its users to the appropriate style. The corpus highlights the terminological conventions, it gives writers additional peripheral vocabulary and suggests idiomatic ways of expressing ideas. An additional point is that corpus evidence is not obtrusive and categorical, but leaves the decision-making clearly to the translator or to the editor.

Corpus evidence also needs to be thoroughly processed before it is used. Even the search strategies must sometimes be elaborate. If no adequate search string or term springs to mind, the corpus compilers need to think of indirect ways of finding what they are looking for. In the above case, for example, the students were sure that ***mucosa*** or ***mucose membrane*** were far too technical in the

target context and that they had to find better ways of expressing the contextual meaning of these terms. By concordancing with ***transmit***, they noticed that ***lining*** and ***tissue*** were the expressions that occurred most frequently in their corpus of texts intended for the general public, instead of the more exact technical terms.

4.3 The Web and other corpus sources

Earlier on I claimed that corpora will help translators in their information retrieval and in obtaining adequate and satisfactory information for their queries. One issue I deliberately omitted from the discussion was the time needed for the compilation of the disposable corpus and yet, time is certainly a critical dimension that cannot be overlooked. If it takes a long time to compile a well-targeted disposable corpus, why not make it part of a permanent corpus instead?

I am sure that, in the future, we will be able to increase the speed at which we can compile a disposable corpus. The critical aspects are the search strategies and the tools we have at our disposal. We will come a long way if we develop the methodology of searching for adequate texts. We can do this by designing strategies for creating key search strings, by determining the criteria for checking the reliability of potential corpus texts or the criteria used for text profiling, etc. We can then use these criteria in search engines when we are looking for texts in dynamic resources such as the Web. In addition, we can combine intelligent search strategies with dedicated, ready-to-use tools in a translator's work station, but the issue of permanency still remains.

Disposable corpora can naturally be recycled and refined if the user knows that they will also be relevant for future assignments. Disposable corpora can also form clearly defined subsets in dynamic corpus repositories intended as collective resources. Ahmad et al. suggested in 1994 that *virtual* corpora for individual needs could be downloaded from standard electronic sources (Ahmad et al. 1994). Kilgarriff (in print) has argued that permanent collections such as

the BNC should in the future be replaced by a more dynamic concept of a regularly monitored and balanced Web corpus, which would consist of replaceable URL addresses of the selected Web texts.

Yet, I am sure that there are always text production assignments that will be better served by targeted corpus collections than by permanent corpora. We are all familiar with the frustrating fact about language in use that it is very evasive. Even large collections of over 100 million words do not necessarily contain the information we need. I therefore think that the need for disposable corpora remains and that the best resource for these is the Web precisely because of its dynamic character.

Obviously, small (a few million words) general corpora also have their uses. They are a very good source of reference for grammatical and syntactic information. It is certainly worth having them as on-line sources in stand-alone computers, because it is much faster to find information in small on-line corpora about grammatical collocations and syntactic patterns than in dictionaries or on the Web.

It is likely that in the future language professionals will be much more aware of what resource is the best and fastest source for any particular type of lexical information need. If they also have the various sources at their fingertips and can effortlessly and quickly resort to any one of them, language professionals will also learn to use the sources intelligently, be more satisfied with the results and make information retrieval much more cost effective.

5. Tools and their usability

WordSmith Tools was the tool that was used in the experiments and all the different functions in this software – the wordlist tool, the concordancer and the keyword tool have proved useful in analyzing retrieved corpus data. In addition, they have been used in profiling the source text data and in highlighting potential new search

words for targeted text retrieval from the Web (e.g. the keyword tool, cf. Varantola 2002).

Yet, problems do arise. For instance, the learning threshold for a text analysis tool is much higher than that for dictionaries. Even if the WordSmith Tools functions are a kind of snap-on tools in that particular toolbox, much could be done to make them easier to use. What I am looking for is a special-purpose set of analysis tools for translation that has a modular structure and can be easily complemented with other textual snap-on tools. In practice this would mean improving the consistency of the procedures, smoother management of multiple windows, more comprehensive user instructions, compatibility with other software and automatic text formatting.

We can envisage a special-purpose version that would make it possible to automate operations that are frequently needed in translational text analysis. It would help the users to organize and systematize the disposable corpus data, automatically save the retrieved text material in a form that can be easily processed further, include standard word lists from different special fields for the keyword function, provide compatible lemmatizers for different languages, etc. The toolbox could also be complemented with a tool like WordSketches (Cf. Kilgarriff at http://www.itri.bton.ac.uk/ ~ ~ Adam.Kilgarriff) that would let the users retrieve a collocational profile of the lexical items they are interested in. Ideally, a translator's electronic toolbox would consist of tools that are interactive and adapt to the advanced users' needs.

The major problem with advanced software is naturally the users' limited willingness to learn to use them. Software literacy should not be taken for granted. Impatient translators have to be sure that learning to use the software is worth their while, cost-effective and does things that they can benefit from. The problem with most language technology applications is that software developers and software users still find it hard to communicate with each other.

Some of the operations experienced translators would like their software to do are difficult to explain because they result from complicated and sometimes highly intuitive decision-making procedures. Likewise, software producers find it hard to explain to the translators what lies behind their designs and user interfaces. Collaborative design environments in which different professionals have equal rights are therefore a necessary prerequisite when language technology products are designed for translation. Obviously, in addition to text analysis tools, this wish applies equally well to electronic dictionaries, translation memories and machine translation.

6. Corpora in translator training

Modern translational competence also includes an understanding of how corpus information can be used in translation and text production. In practical terms, this means that translator trainees need to be taught to understand the various uses of corpora, how to compile them and apply them in an intelligent way. Translator training should thus include courses in the compilation and use of corpus information. Broadly speaking, instruction in corpus linguistics could be divided into courses in corpus use, corpus compilation and corpus analysis software. (Cf. also Varantola 2002)

Corpus application would deal with the different qualities and applications of permanent, semi-permanent and disposable corpus sources, deductive corpus analysis and corpus manipulation skills, as well as the use of corpus evidence for translational decisions.

Corpus compilation would describe the methodology of do-it-yourself corpus building and include such elements as corpus design and design criteria, corpus sources, text search strategies and search word selection, source criticism and assessment of the reliability of the retrieved data, as well as assessment of the relevance and adequacy of the corpus.

Corpus analysis tools should be taught on hands-on courses where different types of software would be used to dissect corpus data in different ways in order to retrieve relevant information for different types of linguistic queries. These courses could also be used to promote the development and usability of dedicated software for translators.

7. Concluding remarks

Can disposable corpora be described as intelligent tools in translation? The answer naturally depends on how we define intelligent in this context. As stated above, in the context of intelligent dictionaries, "intelligent" usually means that the dictionary can adapt to the needs of the users, i.e. allow user profiling and give selective answers from the dictionary database thus highlighting the information categories that interest the user.

Disposable do-it-yourself corpora are highly adaptable because the user decides what is included in them. They are also adaptable in the sense that the user defines what type of information he or she wants to extract from the corpus. Users should obviously also decide how they wish to apply the information they have retrieved. As always, the formulation of the particular information need, the problem, is left to the user, but unlike most sources, disposable corpora and corpora in general provide different types of answers depending on how the user wants to structure and display the data. We could thus claim that, providing the corpus is adequate and the analysis tools are versatile and flexible, the intelligence of the information is dependent on the user´s skills. Disposable corpora are interactive and demanding tools and it is up to their users to recognize whether the corpus adequately answers the query, whether the answer is preliminary, definitive, exhaustive or just absent in the available data. The users can refine the information further. If necessary, they can complement a corpus according to insights gained during the preliminary analysis and find answers to

questions they did not think of asking. In comparison with permanent reference sources, a disposable Web corpus is dynamic and up-to-date. If the user has a hunch, it is relatively easy and quick to follow it up on the Web. A Web-based corpus is also likely to have the latest terminology, which might be time-consuming to locate elsewhere. The downside is, of course, that the information is not vetted and there is no guarantee of its quality. The user is thus alone responsible for applying the information further.

A fair conclusion might thus be that disposable corpora are a valuable addition to the continuum of reference sources. At one end we have dictionaries that are systematic and highly structured sources and, at the other, we have collections of running text that are non-systematic, even random. Together all these sources can help the language professionals in their lexical knowledge management (Carliner 1999:85). The provision is, however, that the sources have to be easy to access, user-friendly and compatible with each other and that the user has to be competent enough to use such advanced tools.

Note

1. This article is based on on-going research dealing with the use and usability of dictionaries and corpora in translation and as such attempts to summarize different aspects of this research.

References

Ahmad, K. , Holmes-Higgin, P. & Sibte Abidi, R. 1994. A description of texts in a corpus: "virtual" and "real" corpora". In W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg and P. Vossen. Eds. *EURALEX 1994 Proceedings*. Amsterdam: Vrije Universiteit. 390-402.

Atkins B.T.S and K. Varantola. 1997. Monitoring Dictionary Use. *IJL 10/ 1(International Journal of Lexicography).* 1-45.

Baker, M.1995. Corpora in Translation Studies: An Overview and Some Suggestions for Future Research', *Target,* 7(2): 223-243.

Baker, M.1996. Corpus-based Translation Studies: The Challenges that Lie Ahead. In Harold Somers (ed.) *Terminology, LSP and Translation: Studies in Language Engineering, in Honour of Juan C. Sager,* Amsterdam and Philadelphia: John Benjamins, 175-186.

Baker, M. 1999. The Role of Corpora in Investigating the Linguistic Behaviour of Professional Translators, *International Journal of Corpus Linguistics* 4(2): 281-298.

Bernardini, S. and F. Zanettin (eds.) 2000. I corpora nella didattica della traduzzione. Biblioteca della Scuola Superiore di Lingue Moderne per Interpreti e Traduttori. Proceedings of the Workshop "Corpus use and learning to translate". Forli/ Bertinoro, 14-15.11. 1997.

Bowker, L.1999. Exploring the Potential of Corpora for Raising Language Awareness in Student Translators, *Language Awareness* 8(3/4): 160-173.

Bowker, L. 2002. Towards a Collaborative Approach to Corpus Building in the Translation Classroom. To appear in Baer, Brian. B. (ed.) 2002 *Beyond the Ivory Tower: Rethinking Translation Pedagogy.* John Benjamins.

Carliner, S. 1999. Knowledge Management, Intellectual Capital, and Technical Communication. *In Communication Jazz: Improvising The New International*

Communication Culture. Proceedings 1999 IEEE International Professional Communication Conference. New Orleans, September 1999. 85-91.

Friedbichler, I. and M. Friedbichler. (1997). 2000. The potential of domain-specific target-language corpora for the translator's workbench. In Bernardini, S. and F. Zanettin (eds.) Also available under: http://www.sslmit.unibo.it/cultpaps.

Kalliokuusi, V. and K. Varantola. 2000. User-Sensitive Lexical Databases. A case of lexical knowledge management. In Ulrich Heid, Stefan Evert, Egbert Lehmann, Christian Rohrer (eds.). *Proceedings of the Ninth Euralex International Congress, EURALEX 2000.* Stuttgart. 393-401.

Kilgarriff, A. Comparing Corpora. To appear in *International Journal of Corpus Linguistics.*

Laviosa, Sara. 1997. How Comparable Can Comparable Corpora Be?. *Target* 9(2): 289-319.

Maia, B. (1997). 2000. Making corpora: A learning process. In Bernardini, S. and F. Zanettin (eds.) Also available under: http://www.sslmit.unibo.it/cultpaps.

META XLIII. 4/1998. Special issue "The Corpus-Based Approach".

Sinclair, J. (ed.). 1987. *Looking up: an account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary.* London: Collins.

Varantola, K. 1998. Translators and their Use of Dictionaries. User needs and user habits. In Atkins, B.T.S. Ed. *Using Dictionaries. Studies of Dictionary Use by Language learners and Translators.* Tübingen: Niemeyer, 179-192.

Varantola, K. (1997). 2000. Translators, dictionaries and text corpora. In Bernardini, S. and F. Zanettin (eds.) 2000, 117-133. Also available under: http://www.sslmit.unibo.it/cultpaps.

Williams, I.A. 1996. A Translator's Reference Needs. Dictionaries or Parallel Texts. Target 8:2.277-299.

Electronic references

BNC website. http://info.ox.ac.uk/bnc

COBUILD website. http://titania.cobuild.collins.co.uk

Killgarriff, Adam homepage. http://www.itri.bton.ac.uk/ ~ ~ Adam.Kilgarriff

WordSmith Tools website. http://www.liv.ac.uk/~ ms2928/index.htm