

## **The Internet and the Single Translator**

Janet Carter-Sigglow  
Zentralbibliothek - Sprachendienst  
Forschungszentrum Jülich GmbH  
52425 Jülich  
Germany  
j.carter-sigglow@fz-juelich.de

### **Synopsis**

The Internet is a lifeline for freelancers or translators working in small groups. Especially for scientific translators it represents a link to the larger scientific community. In our special case at Forschungszentrum Jülich (Research Centre Jülich), we are a small translation department within the largest German national research centre. The major priorities of the Research Centre are structure of matter and materials research, information technology, energy technology, environmental precaution research and life sciences, and these fields are reflected in our translation work. In order to satisfy our needs for cutting-edge terminology we have turned to the Internet, which offers an unimaginable wealth of information for translators if used imaginatively and prudently.

### **Introduction**

This paper should be regarded as a complement to the presentation at the Conference itself. Due to the nature of the print medium, it is not meaningful to present page after page from the Internet illustrating the resources available to the translator. I shall therefore outline the motivation for using the Internet and suggest approaches for tapping its riches and not merely give a directory of websites for translators. As the Conference programme says, I would like to propose some strategies for using the Internet economically, rapidly, imaginatively and discerningly. There is a lot of excellent information going for free but language professionals must use the skills of their trade to access the Web and assess the material.

### **Motivation**

As a regular attendee at "Translating and the Computer" conferences, my motivation in submitting a paper was to address the perceived imbalance of the presentations to the disadvantage of practising translators, especially freelancers or those working in small groups.

I always find the conferences stimulating, but I often feel that in our translation service we are working on a different scale since our hallmark is quality not sheer quantity.

A very apt example is provided by the speaker following my presentation at the Conference. Even without having heard the paper, I see a certain degree of overlap with Gregory Grefenstette's "The WWW as a Resource for Example-Based MT Tasks". His first bullet point certainly applies to me "using the WWW as a free linguistic resource". And his final point is at least related to my strategy in using the Internet "87% of ambiguous German compositional compounds are correctly translated using the WWW".

Then, however, our ways part. The magnitude of the Web is of purely practical relevance for me, I am looking for scientific terminology and only rarely for linguistic structures. Also as a practising translator, I have a short-term interest in solving our translation problems and updating our in-house terminology bank. Nevertheless, this proximity illustrates the strength of this conference series. Language professionals from all walks of life have an opportunity to view problems from a different angle. I imagine it is not a coincidence but rather a clever move by the ASLIB organizers that my paper will be followed by one on a related topic but with a completely different perspective.

### **Professional Environment**

Our position in the translation service of Research Centre Jülich can be described by Winston Churchill's famous remark that "Never ... was so much owed by so many to so few". The translation service is staffed by two translators and a foreign language secretary whereas the Research Centre has a total of 4300 employees including 1000 scientists. Our scientists are very diligent and produce 1800 scientific publications per year and present 2700 papers at national and international conferences, a large proportion of which appear in English.

The major fraction of our work consists of translating scientific articles from German to English to be published in learned journals, and in reworking and polishing papers written in English by German colleagues. Our translation service also works with French, but due to the nature of our mission Jülich is largely an English-language environment.

In keeping with the Research Centre's motto "The Future is Our Mission", the breadth of the scientific endeavour at Research Centre Jülich covers:

- life sciences
- environmental precaution research
- energy technology
- information technology
- structure of matter and materials research.

Our texts are drawn from all these fields, which also include areas such as sociology, nuclear safeguards and medicine perhaps not immediately apparent from the bare headlines. To handle all our terminology we use TRADOS MultiTerm '95 Plus, and we have now built up a database of almost 24,000 terms. We need the system to ensure that the terminology is handled correctly in all the different fields, its multilingual capacity is not of paramount interest for us.

### **The Central Library**

A further link to the ASLIB Conference is that for nearly two years now the translation service has been part of the Central Library at Research Centre Jülich. In our new home we have become much more aware of the tasks of information professionals and have discovered many points of contact in our work. The library also provides us with valuable ED facilities and support for our computer tools so that we are in a very comfortable position with access to e-journals and expert in-house assistance with information acquisition.

The Central Library is one of the largest special scientific libraries in Germany and its major task is acquiring, processing and providing literature and information for the Research Centre's scientists.

#### **Holdings**

Monographs	400,000
Reports	280,000
Journal subscriptions	1,300
Newspapers	24
Journals in reading room/open stacks	1,000
Monographs in reading room	25,000
Annual acquisitions (monographs)	4,000
Full-text e-journals	700

## Databases

- In-house catalogue database with 170,000 records
- Database of Research Centre publications
- Approx. 50 CD-ROM databases
- Access to more than 500 external online databases

The Central Library also publishes reports and books by Research Centre staff plus a press cuttings service, so that the library is not a mere consumer but also a provider of information. The Central Library's website at <http://www.kfa-juelich.de/zb/zb.html> is itself an example of a valuable resource for translators looking for state-of-the-art library and information terminology. This two-way relationship is an innovative feature of the Internet meaning that any and everybody can play an active part in the information environment.

## Nature of the Task

I have described our working environment in some detail to illustrate the nature of our task and to indicate what we are looking for on the Internet. As part of a national research centre, we are at the leading edge of science and technology and need up-to-date terminology, mainly in English, which can in many cases be extracted from the Web with a little ingenuity.

On the very simplest level, you can make an educated guess and check it out. For example, it is useless to thumb through dictionaries to discover what term is used for *bioprospecting* in German. Bioprospecting is defined by the World Resources Institute as "the exploration of wild plants and animals for commercially valuable genetic and biochemical resources"<sup>1</sup>. Simply entering *bioprospecting* in a search engine such as Fireball (<http://www.fireball.de>) and selecting German as the query language brings quick results. Since you have a source document you have also established the provenance of the term in scientific usage, not just as a perhaps rather suspect "dictionary term". Of the 11 hits in the list, I selected the shortest site and the most confidence-inspiring address, i.e. apparently an organization not concerned with propagating attitudes, Projekte Ressourcenmanagement at <http://www.biodiv.de/projekte-res.html>. This turned out to provide a valuable collection of biodiversity terminology in German including *Bioprospecting* and the quintessentially German *Bioprospektion*. The Internet provides the term itself and also vouches for its authenticity.

---

<sup>1</sup> World Resources Institute, <http://www.igc.org/wri/wri/wri/biodiv/bp-facts.html>

## **Weakness of Search Engines**

My reliance upon search engines in accessing the Web is immediately apparent, but a word of warning should be given about their coverage of the Internet. In their timely article "Accessibility of information on the web", Steve Lawrence and C. Lee Giles point out that in December 1998 they estimated

"that six major public search engines (AltaVista, Excite, Hotbot, Infoseek, Lycos and Northern Light) collectively covered about 60% of the web. The largest coverage of a single engine was about one-third of the estimated total size of the web."<sup>2</sup>

Given the estimated 800 million pages on the Web, you may not find this particularly disturbing. However, in our field of endeavour we are looking for the latest terminology in rather obscure (at least for the general Internet user) fields. It is therefore salutary to learn from Lawrence and Giles "that indexing of new or modified pages can take several months or longer."<sup>3</sup>

Search engines are commercial undertakings and they can satisfy most users' queries from a relatively small database. This means that precisely the information I am looking for may become more difficult to access as the Internet expands. As Lawrence and Giles say:

For ranking based on popularity, we can see a trend where popular pages become more popular, while new, unlinked pages have an increasingly difficult time becoming visible in search-engine listings. This may delay or even prevent the widespread visibility of new high-quality information.<sup>4</sup>

Nevertheless, much of the scientific information on the Internet is simply not available in traditional databases or is published earlier on the Web than in a printed form so that in spite of the drawbacks the Internet is of paramount importance to a technical translator.

## **Cheap and Cheerful**

In view of the patchy way in which search engines cover the Internet, you could try using more than one, which would, of course, be time-consuming and thus expensive. Lawrence and Giles recommend "using metasearch engines such as MetaCrawler, which combine the results of multiple engines." In my experience, however, this takes too long since a translator's strategy is different from that of scientists looking for literature to read at their leisure. I am

---

<sup>2</sup> Lawrence, Steve and Giles, C.Lee, "Accessibility of information on the web", *Nature*, Vol. 400, 8 July 1999, pp. 107-109

<sup>3</sup> Lawrence and Giles, p. 109

<sup>4</sup> Lawrence and Giles, p. 109

<sup>5</sup> Lawrence and Giles, p. 108

looking for a quick fix to a terminology problem and it is quicker and just as productive to use a simple search with a search engine

I searched for the term "hornwork" from mediaeval military terminology, to which I shall return later, and with MetaCrawler I was offered 31 hits, of which 19 referred to "horn works", i.e. music for the horn. With a simple search using AltaVista, I had 63 hits and the machine asked me if I really meant "horn works", so I knew what to expect. The MetaCrawler also took much longer since it had to search several search machines. For the purposes of terminology searches where you don't intend to read the documents found I see no advantage to a metasearch engine. With obscure terminology it is probably more economical from the time aspect to refine your search using an advanced search (as AltaVista calls it) rather than using several overlapping search engines.

Especially as a freelancer, you may feel that time is money on two fronts. Firstly, you could be translating instead of surfing the Web and, secondly, you must keep the phone bills in mind. I announced that I would be discussing ways of using the Internet economically and all the information sources are free in the sense that you don't have to pay to access them, but BT or Telekom are always looking over your shoulder.

Aptly enough a piece in ASLIB's *Managing Information* of July/August 1999 by Ian Watson concentrated our minds on telephone charges. At home I have to pay my Internet Service Provider (Telekom) for access to the Internet. The current rates are DM 8.00 per month, which includes 2 hours online. Every further online minute costs 6 pfennigs plus a charge of 6 pfennigs to establish the Internet connection, as well, of course, as local telephone charges. As Watson points out many providers now offer free access to the Internet:

But you do pay the local charge for the time you are online and this remains a significant deterrent to effective use of the Web. If you are paying by the minute there is little incentive to browse, to window shop, to find out how it works. Maybe a penny a minute at off-peak times is not really that much – only 60p an hour – and maybe 4p per minute peak rate is still reasonable.<sup>6</sup>

I take Watson's point that free local calls as in North America or a flat-rate access fee would encourage use of the Internet and I must confess that I enjoy the luxury of being permanently online at work. Nevertheless, I still feel the fees are reasonable for a translator searching for specialist terminology. An hour online is a long time and I try to extract as much information

---

<sup>6</sup> Watson, Ian. "The Information Society", *Managing Information*, July/August 1999, p.56

as I can from the title of the document displayed by the search machine and I check the size of the site before I click on the link.

For example, I searched for the abbreviation FMCT (Fissile Material Cutoff Treaty) using Fireball and discovered it in the first hit: <http://www.isis-online.org/publications/fmct/index.html>. As an added bonus, the page turned out to include a very informative discussion of the treaty itself with useful links. That is the beauty of Internet searches.

### **Dictionaries and Glossaries**

As is apparent, support from outside sources of information is invaluable for our range of topics. Bilingual or multilingual dictionaries are not hard to find on the Internet, but in general they are not specialized enough for our work, let alone reliable enough. Sometimes quick and dirty suggestions can be found in sites such as the LEO English/German Dictionary <http://dict.leo.org/dict/> with more than 180,000 entries. Visitors are invited to add terminology and although exhorted to use the "Add Vocabulary" function wisely there is no guarantee that they do so.

An example of the quality of the LEO dictionary can be given by entering the German "unerschöpflich" which turns up "exhaustless, inexhaustible, inexhaustibly (for unerschöpfliche (sic)), inexhaustable, unexhaustible". Most of these suggestions are, quite simply, wrong. Another problem is that although LEO is very user-friendly in that you only have to type in the German or English expression and you will get a corresponding term in the other language, the system cannot distinguish between the two languages. I entered what I intended to be the German "Serpentine" expecting to get something like "hairpin bend" and ended up with the following: for the English "serpentine" I was offered "gewunden, Serpentina, schlangenartig, sich schlängelnd". This is a cautionary tale of the quality of information available on the Internet and language professionals must therefore not be afraid to exercise their own judgement.

There is, however, nothing that a translator enjoys more than a collection of dictionaries. Such addresses are passed from hand to hand. I was very impressed by the selection at <http://people.a2000nl/mderaaij/irt/index.html>. I tried to test "Glossary of Agricultural and Food Policy Terms" and "Feudal Dictionary" but I was unable to contact them. If time is

money, I found trying to use conventional dictionaries on the Internet rather frustrating. Naturally, you will have your own tried and tested bookmarks but in my experience it is quicker to find scientific terminology via a search engine, in spite of the drawbacks, rather than in predefined dictionaries.

### **Open Source Information**

I find it useful to explore the Internet for what the Secret Service likes to call "open source information". In his paper "INTERNET - An Unlimited Open Source?", Ulrich Kotte quotes a definition of this material referring to "publicly available information appearing in printed or electronic form (i.e. any member of the public could lawfully obtain the information by request or observation)".<sup>7</sup>

If intelligence organizations can obtain 80% of their requested information from open sources then so can the translator. For example, German ministries and government agencies offer rich pickings since their home pages often have an overview in English and sometimes in French. For instance, you will discover that the Bundesamt für Seeschifffahrt und Hydrographie at <http://ws2212.bsh.de> wishes to be known as the Federal Maritime and Hydrographic Agency of Germany. The quality of the language itself can be patchy as illustrated by the "Informations" on the home page of a German ministry whose name I shall not disclose.

The home page of the German government itself <http://www.bundesregierung.de> is very illuminating because it provides versions in English, French and Spanish. The English texts are exact translations of the German and one assumes that this is the preferred terminology for rendering German institutions in English. An instructive example is the term "Spiegelreferate", which they define as follows:

"the Federal Chancellery is organized in divisions that attend to and support matters dealt with by the ministries. These divisions are known as "duplicate divisions" because they reflect the structure found in the respective ministries."<sup>8</sup>

I, for one, would never have hit upon this term, or even have understood it in a text. It should be noted that both "Spiegelreferate" and "duplicate divisions" are always used in inverted commas so that they cannot simply be transferred to other contexts.

---

<sup>7</sup> Kotte, Ulrich, "INTERNET- An Unlimited Open Source?" ESARDA, 21<sup>st</sup> Symposium, Seville, 4-6 May 1999

<sup>8</sup> <http://www.bundeskanzler.de/kanzlerenglisch/02/0201/index.html>



Sometimes the organizations philosophize about the nature of translation. The Deutsche Forschungsgemeinschaft muses on whether its name should be translated:

DFG is internationally known as DFG, Deutsche Forschungsgemeinschaft. There are, of course, translations such as "German Research Council/Society/Foundation", but we feel that "Deutsche Forschungsgemeinschaft" is a proper name, and proper names habitually resist translation.<sup>9</sup>

Ultimately, translators must use their professional judgement on this issue, the information is there on the Internet but nobody can dictate how you should use it.

### **Misusing the Internet**

I am not above misusing, or shall we say misappropriating, sources on the Internet. For example, checking the spelling of chemical compounds can be quite hair-raising, but help is at hand with the ChemFinder at <http://www.chemfinder.com>, which is really intended to provide information on the structure and application of chemicals. You will be reassured, for instance, that *polybutadiene* is spelt correctly with an *e* at the end, in contrast to the German spelling. The ChemFinder compilers are well aware of the linguistic pitfalls involved and include an amusing Typos Quiz at <http://www.chemfinder.com/typos.shtm> for the "nomenclaturist extraordinaire". For example, is chloro(4-chlorophenyl)phenylmethane spelt correctly?

The extremely opulent home page of the Russian Bryansk Regional Authority, in royal blue with the imperial eagle, <http://www.admin.debryansk.ru>, helped me to check the correct English spelling of place names transliterated from Cyrillic when I was revising a text on radioactivity measurements in the Chernobyl area. These names are, of course, transliterated differently into German, e.g. Brjansk, Tschernobyl, and it can be difficult to persuade your clients of the need to change the spelling unless they can see for themselves. Altogether, it can be said that entries on the Internet are very effective in convincing clients of the authenticity of your terminology.

### **Glossaries and Lists**

However, sometimes even I need to use a conventional resource. Single-language glossaries are especially valuable when the subject matter does not fall within the translator's core competence, where it is simply not feasible to buy the specialist literature. As I mentioned above, mediaeval military architecture was a case in point for us. By entering "hornwork" in

---

<sup>9</sup> <http://dfg-bonn.de/english/name/html>

AltaVista I was able to find a glossary at <http://digitalhistory.org/glossary.html>. For those of you not *au fait* with the subject, a hornwork is lucidly defined as: "An outwork consisting of demi-bastions joined by a curtain." The German expression is "Hornwerk" and my approach was simply to guess what it might be in English.

You may wonder how a scientific and technical translator got caught up in mediaeval architecture. Well, we were working on a brochure for the town of Jülich, which, you may be interested to know, is home to the most significant Renaissance citadel north of the Alps. We therefore not only use Internet resources but also create them as you will appreciate if you care to take a trip around Virtual Jülich in English or German at <http://www.juelich.de/virtuell/>

### **Learned Journals**

For information professionals, the figures quoted by Ulrich Kotte will come as no surprise stating that of 5593 known online publications in March 1999, 142 concern scientific topics, or that the number of online newspapers reached 4925 in September 1998. Indeed, he says:

The number of scientific and technical journals that appear only in an electronic form has been mentioned earlier (around 1994) as already exceeding 600.<sup>10</sup>

Librarians know only too well the cost of making online journals available to their users, but the single translator without the enviable library support we enjoy can glean a good deal of information free of charge. As an example, access to the home page of the respected weekly science journal *Nature* is free of charge at <http://www.nature.com/>. You can read abstracts of the week's special topics and look at the contents. This is often sufficient to help with terminology. For example, *Nature* offers a daily *Science Update*, where on 21 September 1999 you could read full-text articles about genetically modified crops, forensic entomology, cocaine addiction, and tuberculosis to name but a selection. These popular science articles are very useful for a translator and the terminology can be extracted fairly readily even by a layperson. You can even get one sample copy to browse through without subscribing. Of course, *Nature* may be scientific but it is not primarily philanthropic and they try to persuade you to subscribe at every turn. However, this problem can be solved with a click of the mouse.

---

<sup>10</sup> Kotte 1999

## Collaboration and Interaction

Thus far we have mainly considered the resources of the Internet to be consumed by the translator, but the innovative aspect of the Internet is its interactive nature. As Kotte puts it in his paper "Internet – the New Pathway for Information",

Unlike the development of other information and communication services, the Internet is essentially user-driven, i.e. the content is generated by many active participants within the network.<sup>11</sup>

I have indicated that the Central Library at the Research Centre and also the town of Jülich participate actively in providing information on the Internet, but the single translator does not need a home page in order to join in. You can do so simply by making use of the e-mail addresses so liberally supplied on the Internet. An e-mail message is not so intrusive as a phone call and most people are willing to help. For example, after wading through pages of open-source material on Escom, the South African electricity utility, I was still none the wiser about what the name actually stood for, so I reached for my mouse and posted an e-mail with the following result:

Dear Janet

ESKOM is not an abbreviation. It is our organisation's name. It is not an English name, neither an Afrikaans one. It is just Eskom.

Previously it was Escom (English) and Evkom (Afrikaans). That was an abbreviation for the "Electricity Supply Commission" and the "Elektrisiteitsvoorsieningskommissie". They changed the name to Eskom in 1987 after the dissolution of the commission.

Hope the information will help you!

Regards from South Africa!<sup>12</sup>

Kotte reminds us that

Literally anyone can participate everywhere, anybody can share globally available knowledge, state their position and opinion, and contribute, modify and evaluate information either by a presence in the common mainstream or segregated niches.<sup>13</sup>

This applies to translators working in specialized fields or with minority languages. The Internet is admittedly currently dominated by English with an estimated 60% US-based hosts (identified by IP address) in January 1999, out of a total of 43,230,000 in an Internet Domain Survey by Network Wizards. But this proportion is set to decline as more countries become involved and by July 1999 the number of hosts had already increased to 56,218,000<sup>14</sup>.

---

<sup>11</sup> Kotte, Ulrich, "Internet - the New Pathway for Information", ESARDA Bulletin No. 29, December 1998, pp. 10-13

<sup>12</sup> Personal communication from Hanlie Smit, 15 July 1999

<sup>13</sup> Kotte 1999

<sup>14</sup> <http://www.isc.org/dsview.cgi?domainsurvey/index.html>

Especially translators working on their own should therefore develop an "Internet lifestyle", which may include participating in news groups and mailing lists.

### **Privacy**

It may brighten up a lonely translator's day to get a friendly e-mail, but I take it none of us are looking for pen pals on the Internet. I would therefore mention the importance of ensuring privacy. The e-mail message I quoted is from an individual at his place of work, which also applies to my query. I would urge against putting a signature (i.e. address, telephone or fax number) on your e-mail if you are working from home since as Kotte (1998) remarks:

Internet users should be able to specify their privacy preferences for information exchange with the variety of Web sites, and to control the use of their personal information profile (e.g. "cookies").<sup>15</sup>

In the example I mentioned above, *Nature* encourages visitors, even those who don't want to subscribe, to fill in a questionnaire and become "registered users". I am convinced that *Nature* quite innocently wishes to establish a user profile, but before setting out to explore the Internet you should seriously consider whether you wish your personal data to be revealed. In any case, you can read the pages without registering. Do not give up your privacy too readily.

### **Conclusions**

Used imaginatively, the Internet is a cheap and easy tool for translators if they are prepared to use the information critically and apply their own judgement as language experts. By this I mean, for instance, noting the country of origin to make a rough check on the authenticity of the expression and avoid the problem of loan translations. It is also prudent to ascertain when the document was last updated to make sure that the terminology really is state of the art. We feel no shame in using the Internet, it is not a second-class source. Its use has been sanctioned by the highest authority:

The publisher of the *Oxford English Dictionary (OED)* is issuing a broad appeal to residents of English-speaking countries for documentary evidence of new words. In the past 18 months, the *OED* has appealed for evidence on words such as space junk, termiting, tipitiwitchet, asteroid field and alkalimetrically.<sup>16</sup>

Once again this emphasizes the interactive nature of the Internet and you are invited to join in by submitting evidence at <http://www.oed.com>

---

<sup>15</sup> Kotte 1998

<sup>16</sup> *Nature*, Vol. 400, 5 August 1999, p.497

At the moment translators have to fall back upon their own skills in extracting information from the Internet, but, in the sense of this ASLIB Conference, if we look to the future automated assistance may soon be at hand. As Kotte (1999) puts it:

In the coming years, the contributions of "human users" may be complemented further by intelligent "information appliances" as participants in the Internet, as well as "machine users" such as household devices and computer desks.<sup>17</sup>

This is, I feel, the ultimate mission of the "Translating and the Computer 21" conference – to bring together automated aids and human skills.

## **References**

Deutsche Bundesregierung, <http://bundestkanzler.de/kanzlerenglisch/02/0201/index.html>

Deutsche Forschungsgemeinschaft, <http://dfg-bonn.de/english/name/html>

Glossary of Eighteenth-Century Military Terms, <http://digitalhistory.org/glossary.html>

Kotte, Ulrich, "Internet - the New Pathway for Information", ESARDA Bulletin No. 29, December 1998, pp. 10-13

Kotte, Ulrich, "INTERNET - An Unlimited Open Source?" ESARDA, 21<sup>st</sup> Symposium, Seville, 4-6 May 1999

Lawrence, Steve and Giles, C. Lee, "Accessibility of information on the web, *Nature*, Vol. 400, 8 July 1999, pp. 107-109

Network Wizards, <http://www.isc.org/dsview.cgi?domainsurvey/index.html>

News in Brief, *Nature*, Vol. 400, 5 August 1999, p.497

Watson, Ian, "The Information Society", *Managing Information*, July/August, 1999, p.56

World Resources Institute, <http://www.igc.org/wri/wri/wri/biodiv/bp-facts.html>

---

<sup>17</sup> Kotte 1999