



Michael Wilkinson was born and brought up in Newcastle upon Tyne in the northeast of England. He attended Cambridge University, and, after graduating with a degree in Economics, subsequently attended Coventry College of Education, where he obtained a Post-Graduate Certificate in Education. In 1975, after having taught for one year in England and one year in Belgium, he took up a teaching post in eastern Finland. Since 1981 he has been a lecturer at Savonlinna School of Translation Studies, a department of the University of Joensuu. Nowadays he mainly teaches courses in translation from Finnish to English, oral expression and liaison interpreting. His wife is a professional translator, working mainly from Finnish into English.

Mr. Wilkinson can be reached at Michael.Wilkinson@joensuu.fi.

Front Page

Select one of the previous 33 issues.
Select an issue:

[Index 1997-2005](#)

[TJ Interactive: Translation Journal Blog](#)

Translator Profiles

[Translators and Translations: Paintings and Shades in Their Frames](#)

by Regina Alfarano, Ph.D.

The Profession

[The Bottom Line](#)
by Fire Ant & Worker Bee



Discovering Translation Equivalents in a Tourism Corpus

by Means of Fuzzy Searching

by Michael Wilkinson

Corpora and corpus analysis tools

In [Wilkinson \(2005\)](#)—the July 2005 issue of *Translation Journal*—I showed some of the ways in which a monolingual target-language corpus can be a useful performance-enhancing resource in translating and described how students at the Savonlinna School of Translation Studies are able to exploit a 670,000-word corpus of English-language tourist brochures using the corpus analysis program *WordSmith Tools* (Scott, 2004) in order to improve the quality of their translations.

The strategies described for finding potential translation equivalents focused mainly on targeted searches where the translator has some idea of what he or she is looking for—for example obtaining information about collocates; choosing between terms suggested by other translation aids such as dictionaries or the Internet; confirming or rejecting intuitive decisions; and extracting multi-word chunks that help the translator to produce natural-sounding text. However, in many cases, it is by no means obvious how to carry out an effective search, and frequent complaints from first-time users of corpus analysis tools on my translation courses are: "I don't know how to find what I'm looking for" and even "I don't know what to look for."

Fuzzy searching

As they gain experience in searching corpora with corpus analysis tools, translators gradually learn how to implement creative searches that increase their chances of finding potential translation equivalents. The guiding hand of an experienced corpus user can also speed up this learning process.

Examples of searching for unknown terms and phrases in a monolingual corpus are given in [Bowker & Pearson \(2002, pp 200-202\)](#), where it is shown how creative techniques can provide possible equivalents for French source-text terms such as *virus dans la nature* (viruses in the wild), *les virus furtifs et semi-furtifs* (stealth viruses and semi-stealth viruses) and *réseau poste à poste* (peer-to-peer network).

[Varantola \(2002, p 180\)](#) has also pointed out that search strategies must sometimes be elaborate. In a workshop experiment, in which her students exploited relatively small self-compiled corpora, some groups employed "sophisticated, indirect deduction chains when searching for corpus information" ([Varantola, 2003, p 66](#)).

Below I shall provide two examples to illustrate how my students have been able to find translation equivalents through creative searching with the Tourism Corpus when translating Finnish texts into English. The search strategies described may seem obvious to experienced users of corpus analysis tools, but are not always apparent to novices translating into a foreign language.

The following examples attempt to illustrate the thought processes of two "typical" novice translators trying to find suitable translation candidates with the help of the Tourism Corpus and other aids. Their thought processes are shown in boxes with a bluish background.

Independent travellers don't need guides

Finnish source text	<i>Järvi-Suomen komeaan luontoon tutustut vesiltä tai maasta käsin opastetuilla tai omatoimisilla retkillä.</i>
---------------------	--

TJ Cartoon

- [Great Moments in Languages: Twelve-step Program to Recover from Translationese](#)
by Ted Crump

Translators Around the World

- [Translation Accreditation Boards/Institutions in Malaysia](#)
by Dr. Kulwinder Kaur d/o Gurdial Singh

Translators and Computers

- [La traduction automatique par opposition à la théorie interprétative — analyse d'un corpus de productions réelles](#)
Chidi Nnamdi Igwe

Interpretation

- [Strategies for New Interpreters: Interpreting in the Indonesian Environment](#)
by Izak Morin

German

- [Picturesque German —German Idioms and Their Origins](#)
by Igor Maslennikov

Translator Education

- [Training of Interpreters: Some Suggestions on Sign Translation Teaching](#)
by Elif Ersozlu, Ph.D.

- [The Contact Between Text, Mind, and One's Own Word in a Translation Workshop](#)
by Leandro Wolfson

- [A Competent Translator And Effective Knowledge Transfer](#)
by Dr. Kulwinder Kaur a/p Gurdial Singh

Literary Translation

- [L'Epreuve de l'autre dans la traduction espagnole de Vivre me tue](#)
Dr. Nadia Duchêne

Translators' Tools

- [Translators' Emporium](#)
- [Discovering Translation Equivalents in a Tourism Corpus by Means of Fuzzy Searching](#)
by Michael Wilkinson

Initial translation	<i>You can admire the splendid scenery of the Finnish Lakeland by boat or overland on either guided or independent trips.</i>
---------------------	--

Imagined thought processes of Novice Translator A

I wonder if I can use *independent* like this? At least my bi-lingual electronic dictionary gives this as the only equivalent for *omatoiminen*. Perhaps I'll check it out in the corpus.

(See Figure 1).

Well there are several references to *independent tour itineraries and packages*, and in line 5 *independent tours* are contrasted with *guided tours*. And a couple of references to *the independent traveller*.

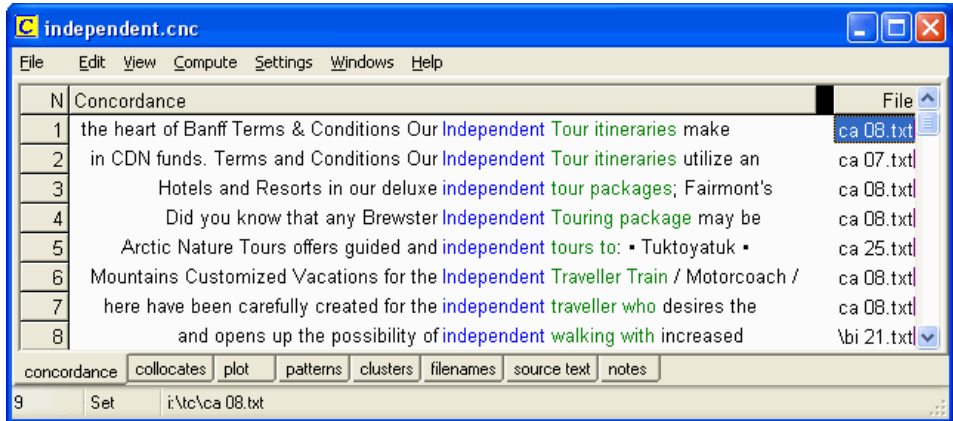


Figure 1: Edited display of the concordance lines generated for the search word *independent*, sorted alphabetically to the right

(In the above screenshot, as in most of those that follow, the display has been heavily edited, mainly to reduce multiple occurrences of the same collocation pattern. However it should be noted that in practice multiple occurrences of the search pattern, or of the search pattern with a specific collocate, is what catches the translator's attention, and reveals the most common way of expressing a term or phrase.)

Maybe I should also try a search for *independently*.

(See Figure 2).

Yes this seems to be possible too. But I would have expected to get more hits for these searches.

Some lines include the phrase *without an escort*—I wonder if I should follow that up... *with an escort? escorted? unescorted?* Maybe I'll check those out later.

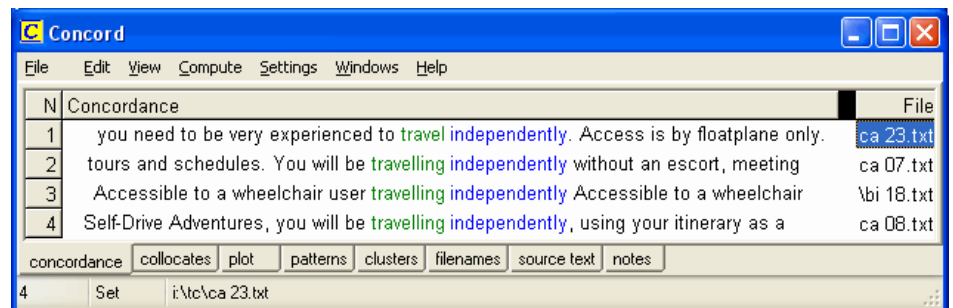


Figure 2: Edited display of the concordance lines generated for the search word *independently*

Omatoiminen is being used in the source text as an alternative to *opastettu* (= guided). Maybe the corpus could help me here. I'll try a search for *guided and / guided or* in order to see what they tend to be paired up with.

CAT Tools and Productivity:
Tracking Words and Hours
by Fotini Vallianatou

Caught in the Web

Web Surfing for Fun and Profit
by Cathy Flick, Ph.D.

Translators' On-Line Resources
by Gabe Bokor

Translators' Best Websites
by Gabe Bokor

Translators' Events

Call for Papers and Editorial Policies

(See Figure 3).

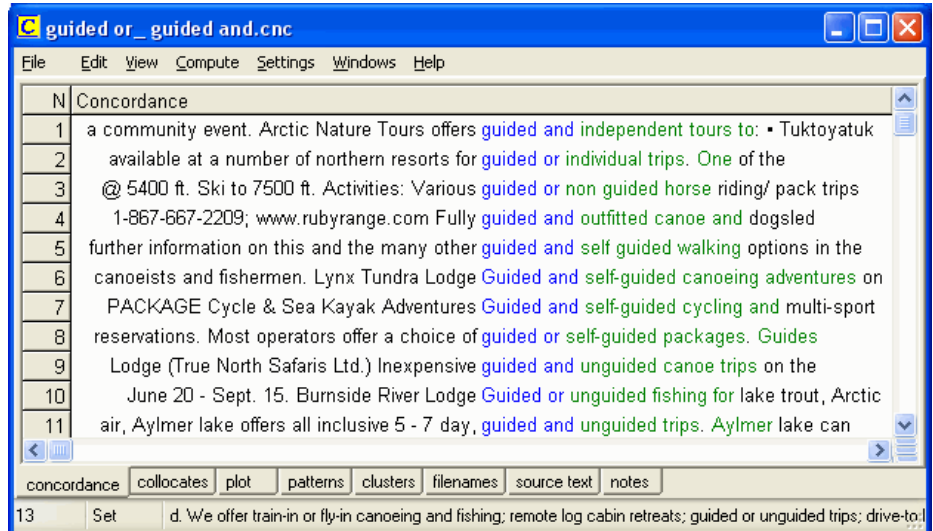


Figure 3: Edited display of the concordance lines generated for the search pattern *guided and/guided or*, sorted alphabetically to the right

Well this is revealing. Lots of lines with *self-guided* being used in contrast to *guided*, and also quite a few incidences of *unguided*. Also a line with *independent*, so this does seem possible, but not so common as the other alternatives. And one incidence of *individual trips*.

Perhaps I'll try a separate search for *self-guided / unguided*.

(See Figure 4).



Figure 4: Edited display of the concordance lines generated for the search pattern *selfguided/self-guided/self guided/unguided*, sorted alphabetically to the right

Okay, 56 hits—but only 5 are for *unguided*, and they are all in Canadian texts. *Self-guided* is sometimes written as two words and sometimes as one, but in 40 cases it is hyphenated. There are piles of hits for *self-guided tour* and *self-guided tours*. Maybe I'll use that in my translation for now.

Finnish source text	<i>Golfin, ratsastuksen, maastopyöräilyn ja tenniksen ohella tarjolla on veneilyä, kalastusta, patikointia, melontaa sekä mönkijäsafareita.</i>
Initial translation	<i>In addition to golf, horse-riding, mountain-biking and tennis, we provide opportunities for boating, fishing, hiking, canoeing and ???safaris.</i>

Imagined thought processes of Novice Translator B

What on earth is a *mönkijä* in English? Can't find it in my online bilingual dictionary or in any of my printed dictionaries and glossaries. It's used here as a compound noun with *safaris*. Let's see what words collocate with *safari(s)* in the Tourism Corpus.

(See Figure 5).

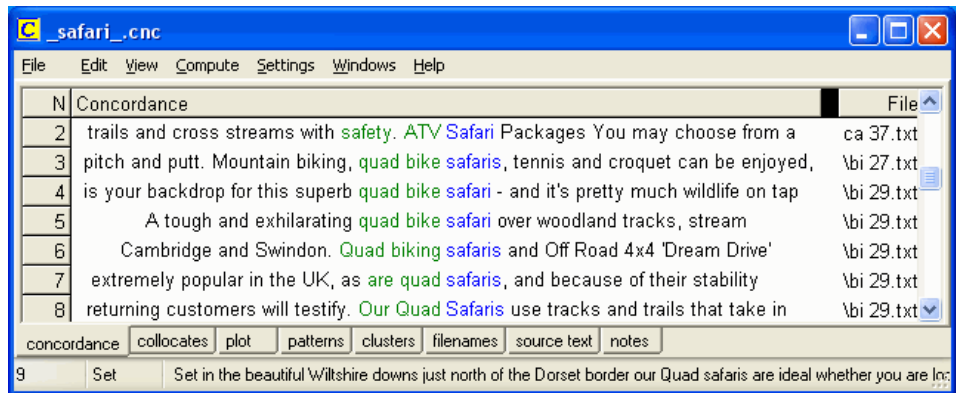


Figure 5: Edited display of the concordance lines generated for the search pattern *safari?*, sorted alphabetically to the left

Over 100 hits. Quite a few *photo safaris* and *wildlife safaris*. And quite a lot of *quad bike* and *quad safaris*. Could that be what I'm looking for? This only occurs in British brochures though. And what are those *ATV safari(s)*?

I'll try out a search for *quad*.

Okay—lots of hits for *quad bikes* and *quad biking*. But again, only in British texts. A lot of hits for *quad* on Canadian sites, but mainly as an adjective preceding *chairs* and *chairlifts* at ski resorts.

Let's try out a search for *ATV*.

(See Figure 6).

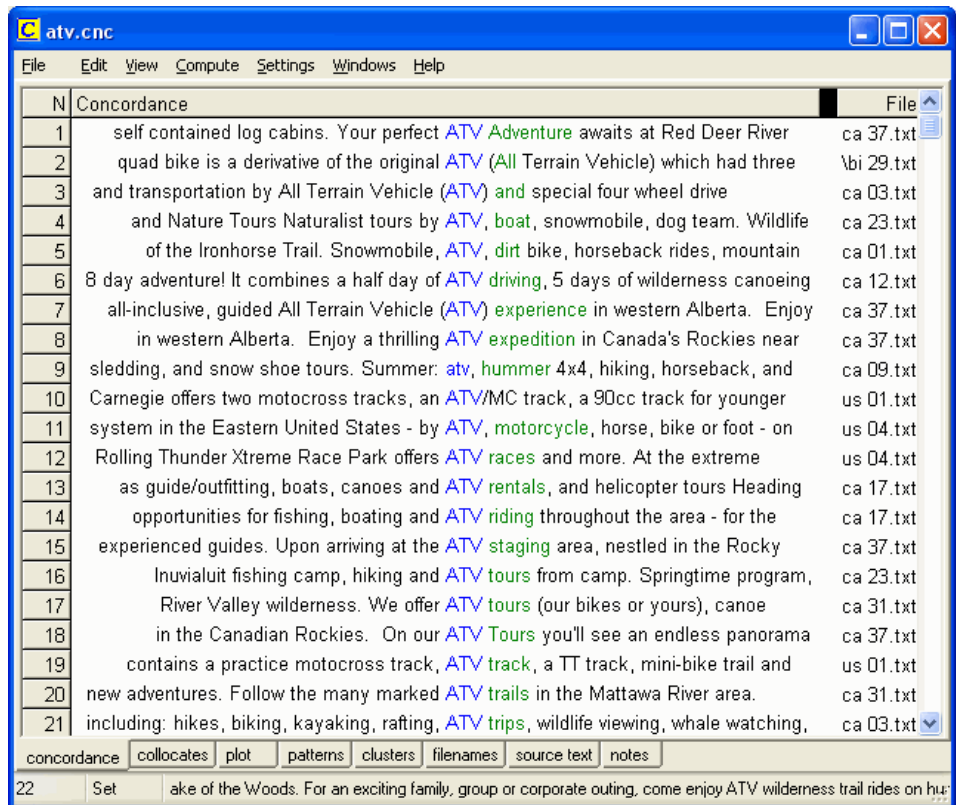


Fig. 6: Edited display of the concordance lines generated for the search word *ATV*, sorted alphabetically to the right

Over 60 hits. Only one line from a British brochure. But widely used in both Canadian and US brochures. And I can see that it's an abbreviation for *All Terrain Vehicle*.

Perhaps I'll check out *ATV* and *quad bike* in an online encyclopaedia.

Well [Wikipedia](#) indicates that *ATV* is a generic term used to describe a range of small open vehicles designed for off-road use, and that the 4-wheeled version is often called a *quad bike*.

If I search for *quad bike* on the Internet, I get hits from, for example, Australian and New Zealand sites, as well as UK sites—so this isn't a purely British term. I also get hits if I restrict my searches to "site:.ca" and "site:.us", but not as many as I'd expect. And since *quad bike* doesn't appear in any of the North American brochures in the Tourism Corpus, maybe I should avoid using it in my translation, since the brochure I'm translating is aimed at an international audience, and North Americans may be unfamiliar with this term.

On the other hand, if I search for *All Terrain Vehicle* on the Internet and restrict my search to "site:.uk", I get over 10,000 hits—so this seems to be a reasonably well-known term on both sides of the Atlantic.

So maybe I'll play safe and go for *ATV safaris* in my translation. At least the vehicles shown in the picture in Wikipedia look just like those in the picture in the brochure I'm translating.

(See Figure 7).



Fig 7. Take an ATV safari with Tahko Safarit

Advanced searching with context words

Although the above depictions of thought processes are imagined, they are based on discussions with and feedback from student groups about the search strategies they have employed. If these thought processes were portrayed more faithfully (e.g. if they were gathered using a think-aloud method), they would no doubt be more untidy, with more occurrences of frustrating unproductive searches plus a liberal sprinkling of expletives.

WordSmith Tools also has an *Advanced Search* feature that facilitates concordancing with contextually-relevant search words. This works in a way similar to the proximity operators used by search engines—you can restrict a concordance search by specifying a context word or context words which either must (or must not) be present within a certain number of words of your search word. Initially this feature tended to cause the program to "freeze", but the fault seems to have been corrected now, thus making the range of fuzzy search strategies available to users of the *WordSmith* concordancer even wider. I look forward to seeing how my students exploit this feature during the forthcoming academic year.

References

Bowker, Lynne & Jennifer Pearson (2002). *Working with Specialized Language: a practical guide to using corpora*. London: Routledge.

Scott, Mike (2004). *Oxford WordSmith Tools* version 4, Oxford University Press.

Varantola, Krista (2002). "Disposable corpora as intelligent tools in translation", in: Tagnin, S. E. O. (Org.). *Cadernos de Tradução: Corpora e Tradução*. Florianópolis: NUT, 2002, v. 1, n. 9, p. 171-189. Viewable online at: <http://www.cadernos.ufsc.br/online/9/krista.htm>

Varantola, Krista (2003). "Translators and Disposable Corpora", in Federico Zanettin, Silvia Bernardini and Dominic Stewart (eds.) *Corpora in Translator Education*. Manchester: St Jerome, pp 55-70.

Wilkinson, Michael (2005). "Using a Specialized Corpus to Improve Translation Quality", in *Translation Journal*, Volume 9, No 3. Viewable online at: <http://accurapid.com/journal/33corpus.htm>

Acknowledgements

Thanks to Mike Scott and Oxford University Press for permission to use screenshots from *Wordsmith Tools*, and to Mikko Oinonen of Tahko Safarit Oy (<http://www.tahkosafarit.fi/tahkosafarit/main.php>) for permission to use the photo in Figure 7.