JoSTrans
The Journal of Specialised Translation

www.jostrans.org · ISSN: 1740-367X

Wilkinson, M. (2007). Corpora, Serendipity & Advanced Search Techniques. *The Journal of Specialised Translation, 7*, 108-122. https://doi.org/10.26034/cm.jostrans.2007.719

Corpora, Serendipity & Advanced Search Techniques

Michael Wilkinson, University of Joensuu, Savonlinna Campus, Finland

ABSTRACT

Exploring corpora with concordancers can help translators to improve the quality of their translations by, for example, providing them with information about collocates; by helping them to choose between terms; or by enabling them to confirm intuitive decisions. But corpora also allow unpredictable, incidental learning: the user may notice unfamiliar uses in a concordance and follow them up by exploratory browsing. The article discusses the potential of corpora to throw up previously unknown information that may be relevant to a translation assignment, and illustrates how advanced search strategies can increase the likelihood of "accidentally" finding relevant information.

KEYWORDS

translation quality; corpora; corpus analysis tool; concordancer; key-word-in-context; serendipity; incidental learning; advanced searching

Introduction

> *As we know, there are known knowns; there are things we know we know.*
>
> *We also know there are known unknowns; that is to say we know there are some things we do not know.*
>
> *But there are also unknown unknowns – the ones we don't know we don't know.*

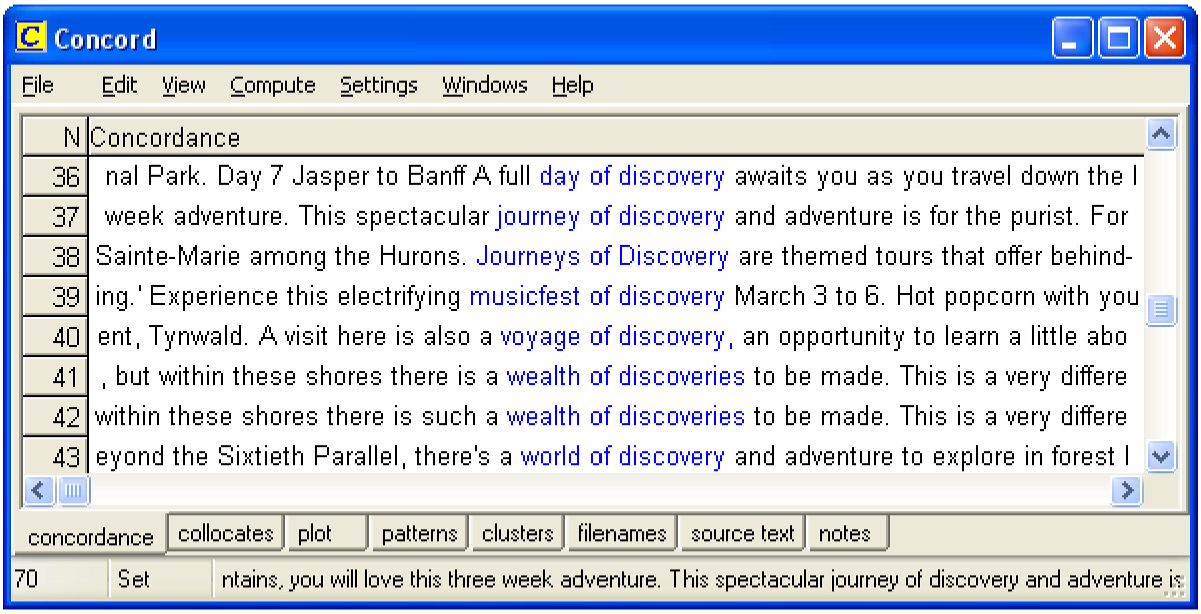
– Thus spake Donald Rumsfeld, United States Secretary of Defence, in 2002 referring to the situation in Iraq, though he might well have been talking about searching corpora for translation candidates.

One way the translator can find out more about the "known unknowns" and the "unknown unknowns" is by exploratory browsing through relevant material. In this respect, a number of researchers in the fields of language learning and translating have drawn attention to the potential that electronic corpora, when used in conjunction with corpus analysis tools, provide for such "serendipitous" learning: corpora allow unpredictable, incidental learning in that the user may notice and explore unknown or unfamiliar uses in a concordance and go off at a tangent to follow them up.

In spring 2004, I began compiling a corpus of English-language tourism brochures with the aim of using it to teach students how the competent use of electronic text corpora in conjunction with corpus analysis tools can help both the trainee translator and the professional translator to become better language service providers by enhancing both the quality of their work and their productivity, particularly when translating special field texts into a foreign language. (Many translators of non-literary texts in Finland frequently translate into their L2).

By September 2004, with the help of a student assistant, I had compiled a corpus amounting to 670,000 words. The Tourism Corpus contains mainly texts from brochures from the British Isles and from North America, especially Canada. When compiling the corpus, a major reason for including Canadian brochures was that they contain descriptions of activities that are often featured in Finnish source texts - e.g. snowshoe treks, skiing, snowmobile trips, wilderness adventures - which are rarely mentioned in British brochures. The file names were labelled with one of the following codes: BI, CA, US, so that the user can immediately identify whether a concordance line is from the British Isles, Canada, or the United States.

Corpus analysis tools enable users to investigate and manipulate the information contained within a corpus in a variety of ways. For example, most corpus analysis packages comprise a "concordancer", which will find all the occurrences of a search word, or search pattern, and display them in the centre of the screen, together with a span of co-text – a so-called Key Word In Context (KWIC) display. Figure 1 shows a KWIC display of 8 of the 70 concordance lines containing the words *discovery* or *discoveries* generated by *WordSmith Tools* from the Tourism Corpus.

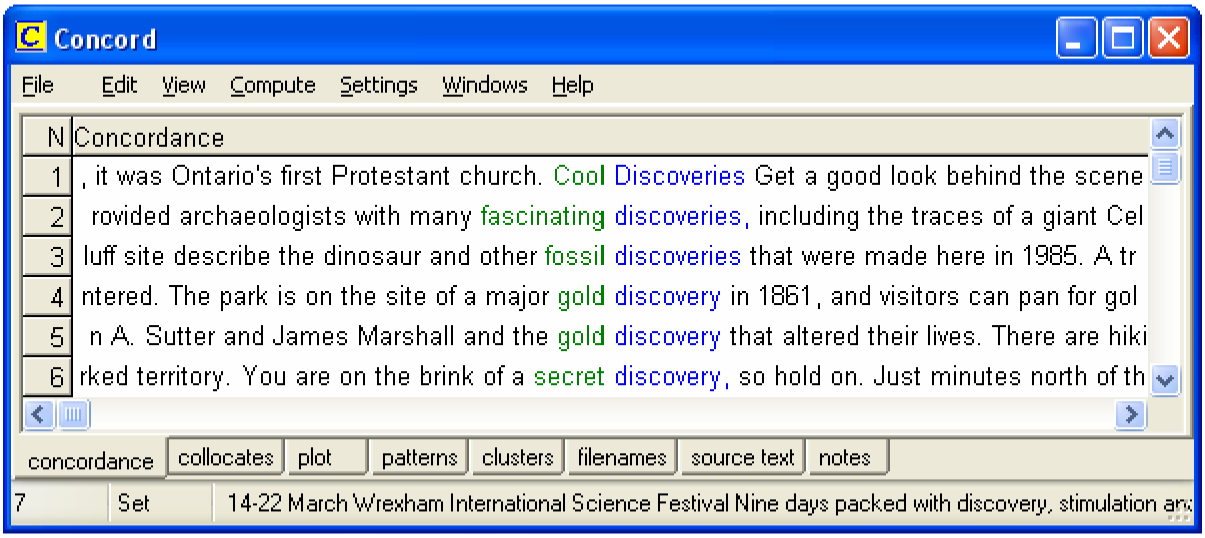


```
 N Concordance
36  nal Park. Day 7 Jasper to Banff A full day of discovery awaits you as you travel down the I
37  week adventure. This spectacular journey of discovery and adventure is for the purist. For
38 Sainte-Marie among the Hurons. Journeys of Discovery are themed tours that offer behind-
39 ing.' Experience this electrifying musicfest of discovery March 3 to 6. Hot popcorn with you
40 ent, Tynwald. A visit here is also a voyage of discovery, an opportunity to learn a little abo
41  , but within these shores there is a wealth of discoveries to be made. This is a very differe
42 within these shores there is such a wealth of discoveries to be made. This is a very differe
43 eyond the Sixtieth Parallel, there's a world of discovery and adventure to explore in forest I

70   Set   ntains, you will love this three week adventure. This spectacular journey of discovery and adventure is
```

Figure 1: Some of the concordance lines generated by WordSmith Tools for the search pattern *discovery/discoveries*

You can manipulate the order of the concordance lines: for example if your search word is a noun, you can ask the concordancer to sort the words immediately preceding the search word in alphabetical order, which may help you to find suitable adjectives that collocate with the search word, as shown in figure 2.

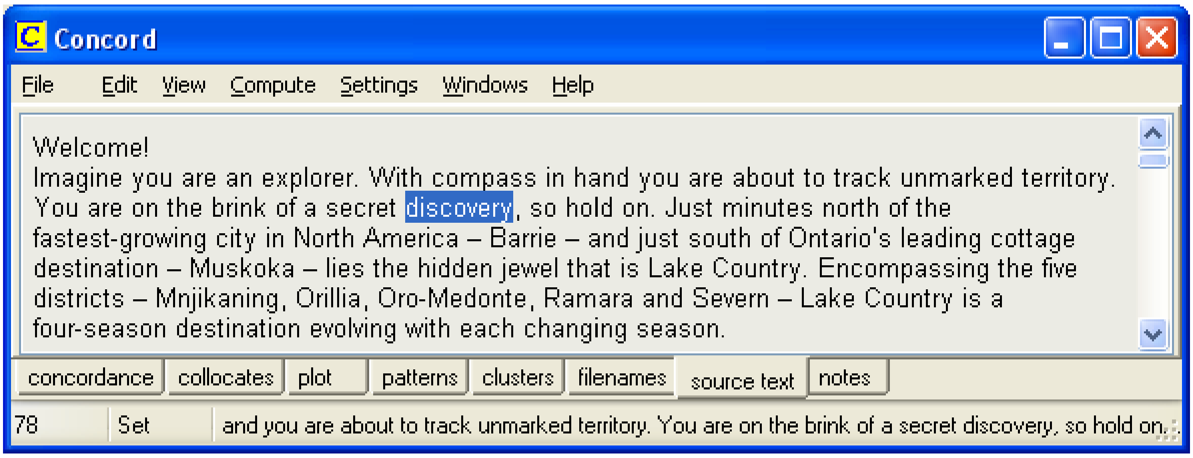

Concord
File Edit View Compute Settings Windows Help
concordance collocates plot patterns clusters filenames source text notes

N	Concordance
1	, it was Ontario's first Protestant church. Cool Discoveries Get a good look behind the scene
2	rovided archaeologists with many fascinating discoveries, including the traces of a giant Cel
3	luff site describe the dinosaur and other fossil discoveries that were made here in 1985. A tr
4	ntered. The park is on the site of a major gold discovery in 1861, and visitors can pan for gol
5	n A. Sutter and James Marshall and the gold discovery that altered their lives. There are hiki
6	rked territory. You are on the brink of a secret discovery, so hold on. Just minutes north of th

7 Set 14-22 March Wrexham International Science Festival Nine days packed with discovery, stimulation an

Figure 2: Edited KWIC display generated by WordSmith Tools for the search pattern *discovery/discoveries*

By double-clicking on a line, you can view it in its full context, as in figure 3, which displays line 6 of figure 2 in a fuller context.


Concord
File Edit View Compute Settings Windows Help
concordance collocates plot patterns clusters filenames source text notes

Welcome!
Imagine you are an explorer. With compass in hand you are about to track unmarked territory.
You are on the brink of a secret discovery, so hold on. Just minutes north of the
fastest-growing city in North America – Barrie – and just south of Ontario's leading cottage
destination – Muskoka – lies the hidden jewel that is Lake Country. Encompassing the five
districts – Mnjikaning, Orillia, Oro-Medonte, Ramara and Severn – Lake Country is a
four-season destination evolving with each changing season.

78 Set and you are about to track unmarked territory. You are on the brink of a secret discovery, so hold on.

Figure 3: Display showing a concordance line in fuller context

Chance Discoveries

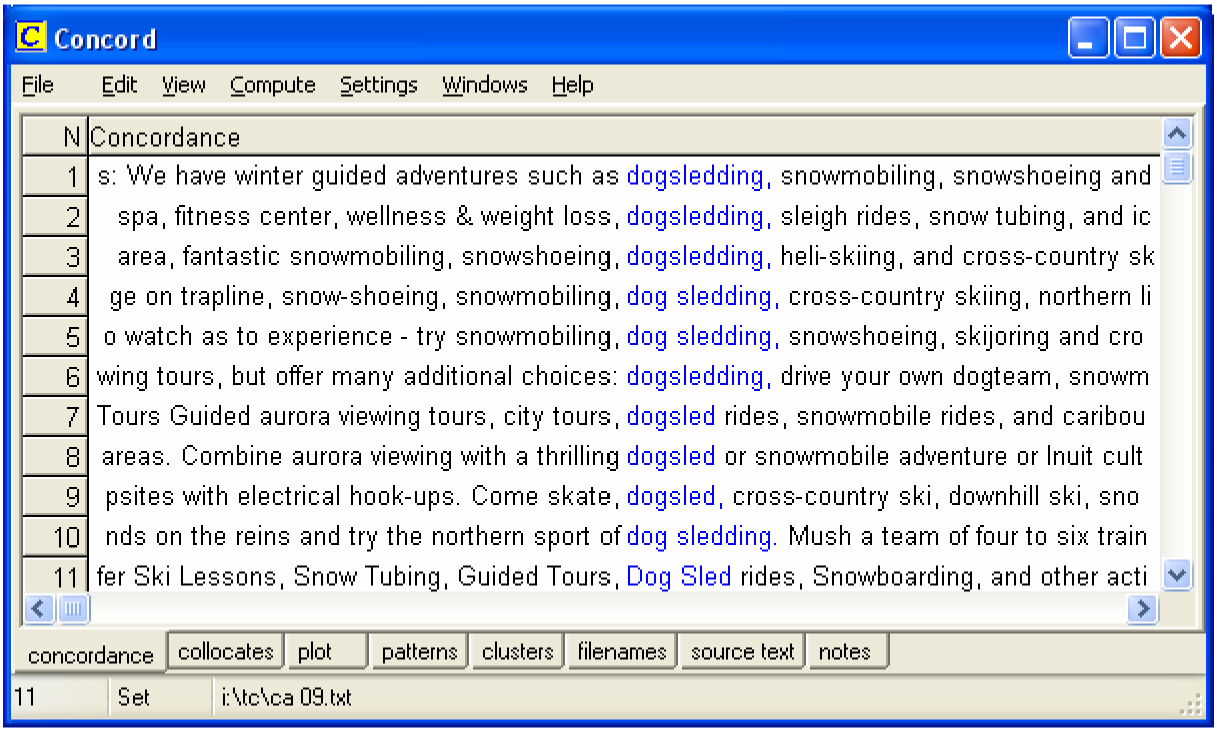
Bernardini (2000, 2004) is one of the leading advocates of using corpora for "discovery learning" and has encouraged advanced students of English to browse the 100-million-word British National Corpus (BNC) in open-ended, exploratory ways. In Bernardini (2001), the author describes a "journey of discovery" that she herself undertook with the BNC.

Bernardini’s students have also exploited a variety of other corpora in addition to the BNC – larger and smaller, general and specific, monolingual and bilingual – and have been guided to progress from more convergent activities to autonomous browsing (Bernardini, 2002).

Zanettin (2001), describing how a relatively small corpus (250,000 words) of British newspaper articles was used as a translation aid by Italian students translating mainly from their mother tongue into English, shows that some information relevant to the translation assignment resulted from chance discoveries.

In Wilkinson (2005a), I illustrated how a specialised monolingual target-language corpus can be of great help to the translator in confirming intuitive decisions, in verifying or rejecting decisions based on other tools such as dictionaries, in obtaining information about collocates, and in reinforcing knowledge of normal target language patterns. I also touched briefly on the potential of corpora to throw up previously unknown information that may be relevant to the translation assignment at hand or may come in handy for future assignments.

The KWIC display in figure 4 illustrates some of the concordance lines generated for the search pattern *dogsle*/dog sle*/dog-sle*.* The translator is looking for a translation equivalent for the Finnish term *koiravaljakkoajelu*. After hunting through traditional translation aids, the translator has come up with the terms *dog sled*, *dog sledge* & *dog sleigh*, each of which is also often written with hyphens or as one word. The corpus helps in deciding on which of these alternatives to use. The original KWIC display contained 22 hits for *dog sled*, 27 hits for *dogsled*, and 6 hits for *dog-sled*, with no hits at all for *dog sledge* or *dog sleigh* or variations thereof. Moreover there were 68 hits for *dogsledding*, often written also as two words.



```
 N Concordance
 1 s: We have winter guided adventures such as dogsledding, snowmobiling, snowshoeing and
 2   spa, fitness center, wellness & weight loss, dogsledding, sleigh rides, snow tubing, and ic
 3   area, fantastic snowmobiling, snowshoeing, dogsledding, heli-skiing, and cross-country sk
 4  ge on trapline, snow-shoeing, snowmobiling, dog sledding, cross-country skiing, northern li
 5  o watch as to experience - try snowmobiling, dog sledding, snowshoeing, skijoring and cro
 6 wing tours, but offer many additional choices: dogsledding, drive your own dogteam, snowm
 7 Tours Guided aurora viewing tours, city tours, dogsled rides, snowmobile rides, and caribou
 8  areas. Combine aurora viewing with a thrilling dogsled or snowmobile adventure or Inuit cult
 9  psites with electrical hook-ups. Come skate, dogsled, cross-country ski, downhill ski, sno
10  nds on the reins and try the northern sport of dog sledding. Mush a team of four to six train
11 fer Ski Lessons, Snow Tubing, Guided Tours, Dog Sled rides, Snowboarding, and other acti

11        Set        i:\tc\ca 09.txt
```

Figure 4: Edited display of some of the 118 concordance lines generated by WordSmith Tools for the search pattern *dogsle*/dog-sle*/dog sle**

But what is particularly interesting in the above KWIC display is the large amount of previously 'unknown' information the translator might acquire when browsing through it. Lines 1,3 & 4 contain references to *snowshoeing*; lines 2 & 11 mention *snow tubing*; lines 7 & 8 *aurora viewing*; line 3 mentions *heli-skiing*, line 5 *skijoring*, and line 9 *electrical hook-ups*. All of these may lead to further exploration by viewing in fuller context or by entering new search patterns.

So we can see that the search pattern *dogsle*/dog-sle*/dog sle** provides a rich source of paths to explore for those wishing to embark on a journey of discovery à la Bernardini. And indeed, in my translation courses in Savonlinna, I encourage students to explore interesting leads generated by corpus searches, record potentially-useful discoveries, and share their findings in class. Recent student-discoveries include *interpretive centre*, *float plane*, *seaplane*, *perimeter trail, sport fishing, fly-in fishing, fly-in resort, illuminated skiing loop, bridleway, skijoring* & *wildlife viewing*.

Of course, such serendipitous learning also occurs when consulting texts in printed form – when you encounter interesting 'leads' in the text, you can follow them up in other sources. However, digitalised texts allow such leads to be explored much more rapidly and systematically.

Web searches also allow for serendipitous learning. However, although the Web can be an invaluable mine of information, especially for discerning translators who have honed their search skills, it can sometimes be slow, due to the time that is often required for separating the wheat from the chaff resulting from the numerous ‘unreliable hits’ that are generated. A well-designed specialised target-language corpus can probably in many cases be a more efficient and reliable tool for serendipitous learning as well as for searches that are more ‘targeted’.

Advanced Searching

Unfortunately the professional translator striving to meet a deadline for a brief, or indeed the translation student trying to meet a deadline for a teacher-set assignment, often does not have the luxury of making leisurely journeys of discovery due to time pressures. Therefore it is necessary to develop other strategies for discovering "unknowns", or at least "lesser knowns" and "partially knowns". As Varantola (2002, p.180) points out, search strategies must sometimes be elaborate, and if no adequate search string or term springs to mind, translators need to think of indirect ways of finding what they are looking for.

Examples of creative searching techniques are given in Wilkinson (2005b). Such so-called "fuzzy" searches can increase the likelihood of "accidentally" finding relevant information. The *Advanced Search* feature of *WordSmith Tools* is especially useful for implementing creative searches. In my experience, newcomers to corpus analysis tools tend to under-use this feature, and therefore I shall provide a couple of examples of how it can help in discovering potential translation candidates.

The *Advanced Search* feature facilitates concordancing with contextually-relevant search words. It 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.

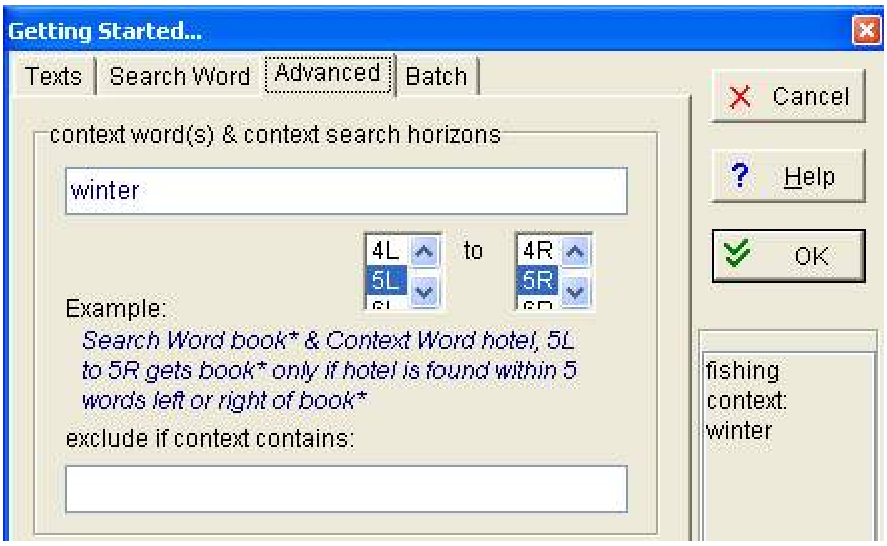
Example I: Fantastic Fishing

Suppose a Finnish translator needs to find a translation equivalent in English for the verb *pilkkiä* and/or the noun *pilkkiminen*. These occur frequently in Finnish tourist brochures. The translator knows that they mean "fishing through a hole cut in the lake ice".

A bilingual dictionary may suggest words like *jig* / *jigger* / *jigging*. If one checks, for example, *jigging* in a monolingual dictionary or on the net, one will find that this refers to the technique of jerking a jig (a small artificial lure) or other bait up and down in the water. This does not convey the fact that the activity of *pilkkiminen* takes place in winter and through the ice.

The translator might decide to go for a translation like "fishing with a jig through a hole cut in the ice." This would probably be a feasible translation, but might be a bit long-winded if the term appears repeatedly in your text.

When translating tourism-related texts, student translators at the Savonlinna campus of Joensuu University can use *WordSmith Tools* together with the Tourism Corpus to search for translation equivalents. In this case they could enter *fishing* as their main search word, and then click on the *Advanced* tab and enter *winter* as their context word, setting the context search horizons as they see fit. In figure 5, the horizons have been set so that concordance lines will be generated whenever *winter* appears within 5 words to the left or right of *fishing*.



Getting Started...

Texts | Search Word | Advanced | Batch

context word(s) & context search horizons

winter

4L 5L to 4R 5R

Example:

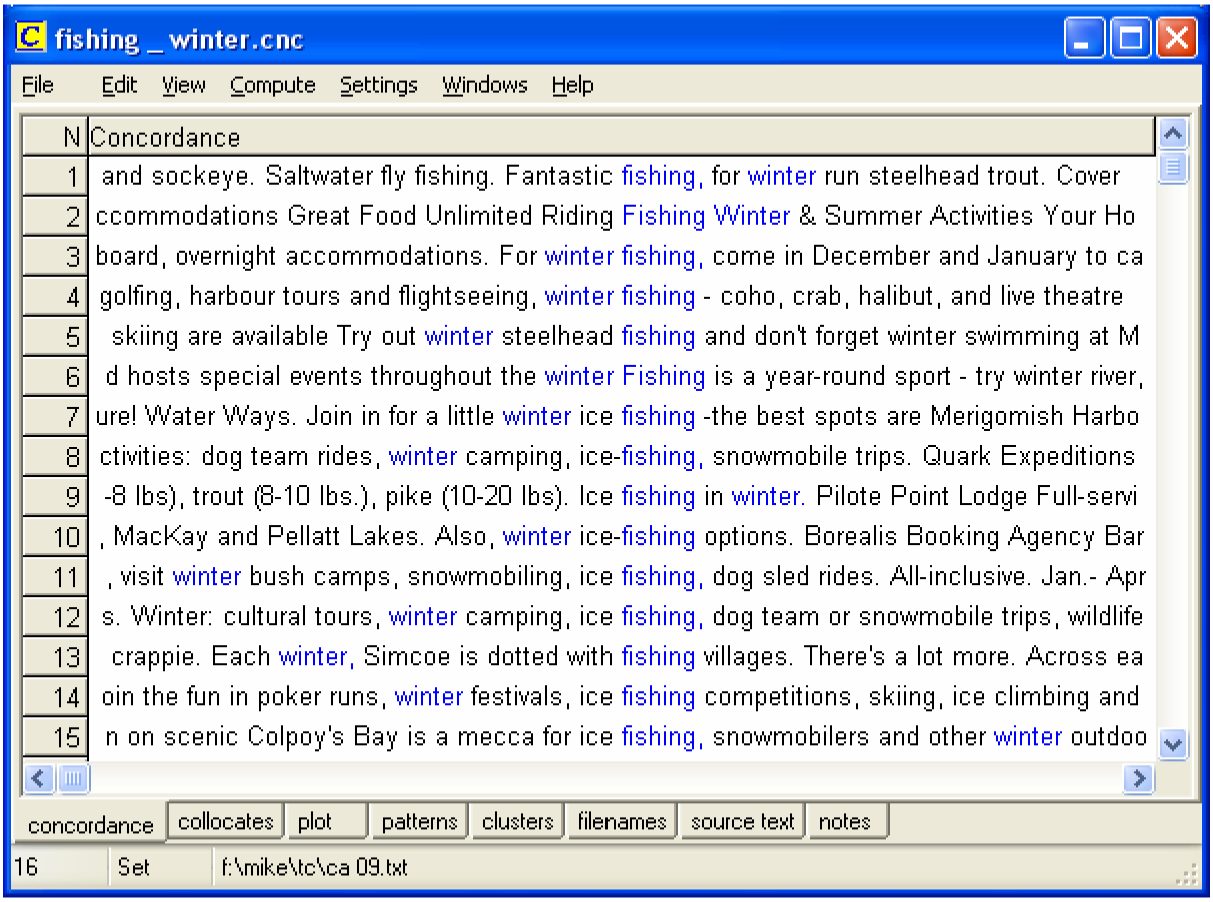
Search Word book & Context Word hotel, 5L to 5R gets book* only if hotel is found within 5 words left or right of book**

exclude if context contains:

fishing
context:
winter

Figure 5:Search query using WordSmith Tools for *fishing* with *winter* as the context word

The resulting KWIC display is shown in figure 6. The translator will quickly notice the occurrences of the term *ice fishing* in lines 7-12 & 14-15. A follow-up search for *ice fishing/ice-fishing* without a specified context word will produce many more concordance lines, which can be explored and viewed in their full text context.



```
 N Concordance
 1  and sockeye. Saltwater fly fishing. Fantastic fishing, for winter run steelhead trout. Cover
 2 ccommodations Great Food Unlimited Riding Fishing Winter & Summer Activities Your Ho
 3 board, overnight accommodations. For winter fishing, come in December and January to ca
 4  golfing, harbour tours and flightseeing, winter fishing - coho, crab, halibut, and live theatre
 5   skiing are available Try out winter steelhead fishing and don't forget winter swimming at M
 6  d hosts special events throughout the winter Fishing is a year-round sport - try winter river,
 7 ure! Water Ways. Join in for a little winter ice fishing -the best spots are Merigomish Harbo
 8  ctivities: dog team rides, winter camping, ice-fishing, snowmobile trips. Quark Expeditions
 9  -8 lbs), trout (8-10 lbs.), pike (10-20 lbs). Ice fishing in winter. Pilote Point Lodge Full-servi
10 , MacKay and Pellatt Lakes. Also, winter ice-fishing options. Borealis Booking Agency Bar
11  , visit winter bush camps, snowmobiling, ice fishing, dog sled rides. All-inclusive. Jan.- Apr
12  s. Winter: cultural tours, winter camping, ice fishing, dog team or snowmobile trips, wildlife
13   crappie. Each winter, Simcoe is dotted with fishing villages. There's a lot more. Across ea
14  oin the fun in poker runs, winter festivals, ice fishing competitions, skiing, ice climbing and
15  n on scenic Colpoy's Bay is a mecca for ice fishing, snowmobilers and other winter outdoo
16      Set      f:\mike\tc\ca 09.txt
```

Figure 6: KWIC display for the search word *fishing* with *winter* as the context word

A follow up using other resources will quickly confirm that this is a good equivalent for the Finnish term. For example Wikipedia gives the following definition: "Ice fishing is the sport of catching fish with lines and hooks or spears through an opening in the ice on a frozen body of water. Fisherman may sit on a stool on the open expanse of a frozen lake or sit in a heated cabin on the ice with bunks and amenities". Nevertheless the translator may decide that since the term *ice fishing* appears only in Canadian brochures, target audience readers who are from other countries may not be familiar with this concept, and therefore a longer explanation may be needed the first time this term appears, after which the more concise translation of *ice fishing* can be utilised.

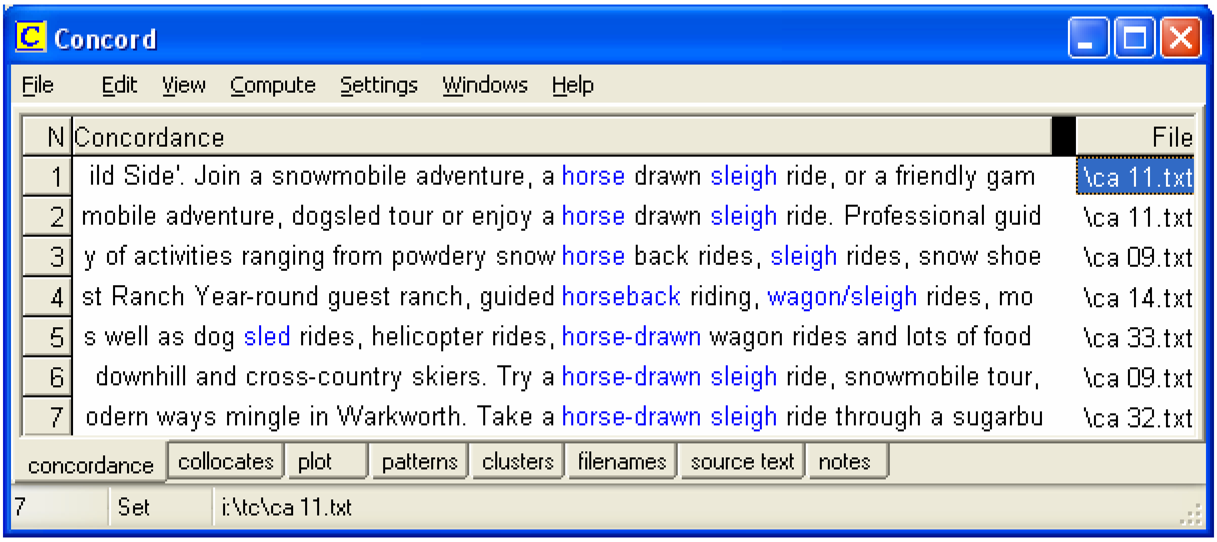
Example 2: Jingle Bells

Suppose our Finnish translators need to find an equivalent in English for *hevosrekiajelu* as in the following phrase from an authentic commission:

Koiravaljakko- ja ***hevosrekiajelut*** *tilauksesta*. (Dogsled rides and **horsesled rides** by prior booking).

I described earlier how a search of the Tourism Corpus provides evidence that *dogsled*, *dog-sled* and *dog sled* are all feasible candidates for *koiravaljakko*. But what about when the vehicle is pulled by a horse? Can the translator use *horse sled rides*. The translator may recall the well-known Xmas song *Jingle Bells*: "Oh what fun it is to run in a one-horse open sleigh". So would it be better to go for *horse-sleigh rides*? A search for *horse** (which finds all words beginning with the letters *horse*) generates almost 450 concordance lines – rather many to browse through. However, by sorting to the centre, the translator quickly sees that there are no occurrences of *horsesled*, *horsesledge* or *horsesleigh*, even when written with hyphens; and, by sorting alphabetically to the right, the translator also quickly sees there are no occurrences of *horse sled*, *horse sledge* or *horse sleigh*, even though a search of the Internet would produce hits for all of these, and especially for *horse sleigh rides*.

Having ruled out what is not used in the corpus, the translator now needs to find out what, if anything, is used. Once again, use can be made of the *Advanced Search* feature. By entering *horse** as the search pattern and *sle** as the context word, the KWIC display shown in Figure 7 is generated.



N	Concordance	File
1	ild Side'. Join a snowmobile adventure, a horse drawn sleigh ride, or a friendly gam	\ca 11.txt
2	mobile adventure, dogsled tour or enjoy a horse drawn sleigh ride. Professional guid	\ca 11.txt
3	y of activities ranging from powdery snow horse back rides, sleigh rides, snow shoe	\ca 09.txt
4	st Ranch Year-round guest ranch, guided horseback riding, wagon/sleigh rides, mo	\ca 14.txt
5	s well as dog sled rides, helicopter rides, horse-drawn wagon rides and lots of food	\ca 33.txt
6	downhill and cross-country skiers. Try a horse-drawn sleigh ride, snowmobile tour,	\ca 09.txt
7	odern ways mingle in Warkworth. Take a horse-drawn sleigh ride through a sugarbu	\ca 32.txt

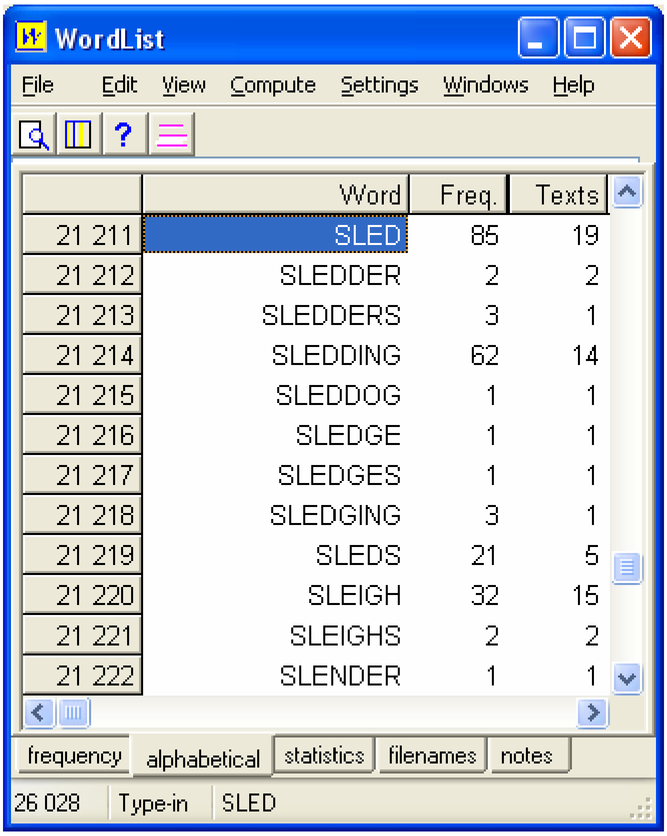
7 Set i:\tc\ca 11.txt

Figure 7: KWIC display for the search pattern
*horse** with *sle** as the context pattern

Now it is easy to notice several occurrences of *horse-drawn sleigh ride*. A cross-check on the Internet for this term would produce hundreds of hits, even when the search is restricted to Canadian sites or UK sites. The translator may now, if s/he has time, go off at a tangent and explore the corpus for other occurrences of *horse-drawn* vehicles, and discover *buggies*, *carriages* and *wagons*. Or alternatively s/he may investigate the usage of apparent synonyms such as *sled*, *sledge* and *sleigh*.

Sleds, Sledges and Sleighs

When encountering synonyms or partial synonyms in a corpus, the translator can employ a variety of exploration techniques to try to differentiate between them. For example, most corpus analysis programs include a word-list tool, which can show all the words in the corpus displayed in alphabetical order or in frequency order. Figure 8 shows part of the word-list display generated from the Tourism Corpus and sorted alphabetically.



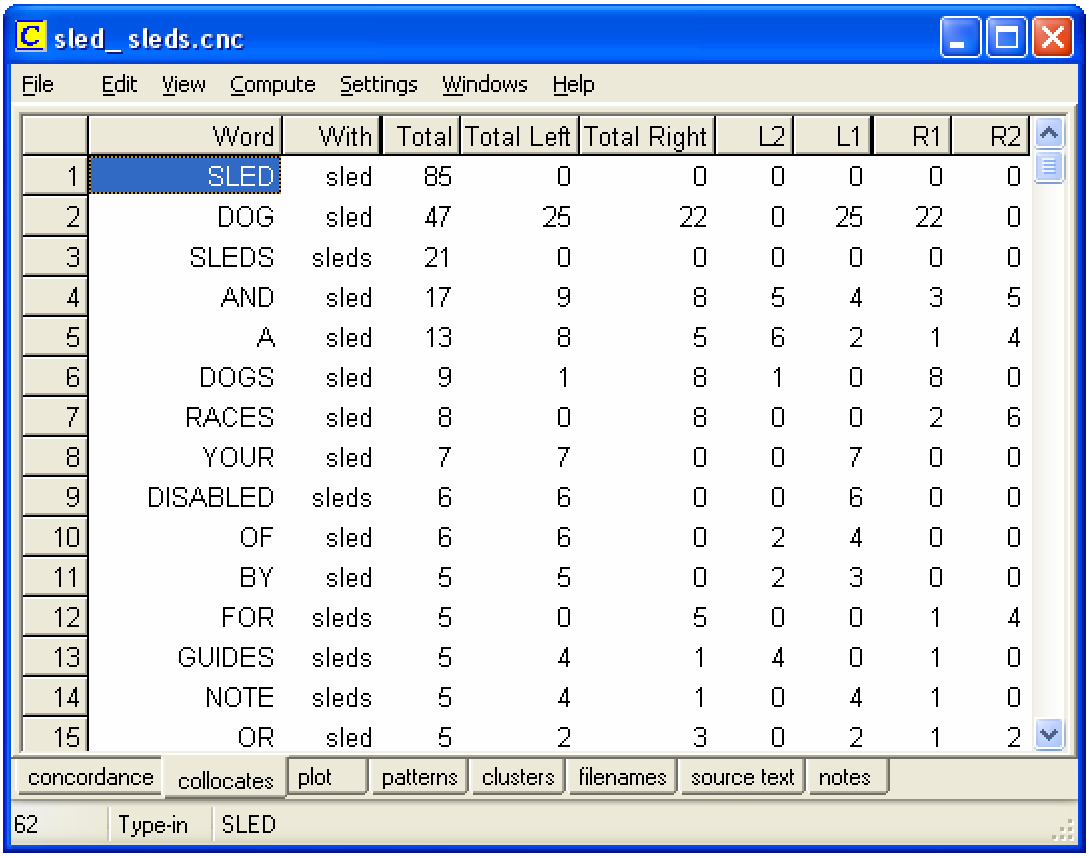
	Word	Freq.	Texts
21 211	SLED	85	19
21 212	SLEDDER	2	2
21 213	SLEDDERS	3	1
21 214	SLEDDING	62	14
21 215	SLEDDOG	1	1
21 216	SLEDGE	1	1
21 217	SLEDGES	1	1
21 218	SLEDGING	3	1
21 219	SLEDS	21	5
21 220	SLEIGH	32	15
21 221	SLEIGHS	2	2
21 222	SLENDER	1	1

26 028 Type-in SLED

Figure 8: WordList display for some of the words beginning with *sle**
in the Tourism Corpus

The bottom left hand corner shows that there are 26,028 different words (or tokens) in the corpus. The translator can see that *sled* occurs 85 times (and in a wide range of texts), *sleds* 21 times, *sleigh* & *sleighs* 34 times, and *sledge* & *sledges* only 2 times. Of course, such crude statistical measures should not be used as the sole determining factor for choosing between two or more apparent synonyms. And besides, although this kind of information about the frequency of such synonyms in a corpus throws some light on their usage, it does not help the translator to make subtle distinctions in regard to their meanings. For this, the collocation display is more useful. By examining collocations the translator can see common lexical and grammatical patterns of co-occurrence, which may be difficult to discern in the concordance lines, especially if there are lots of them.

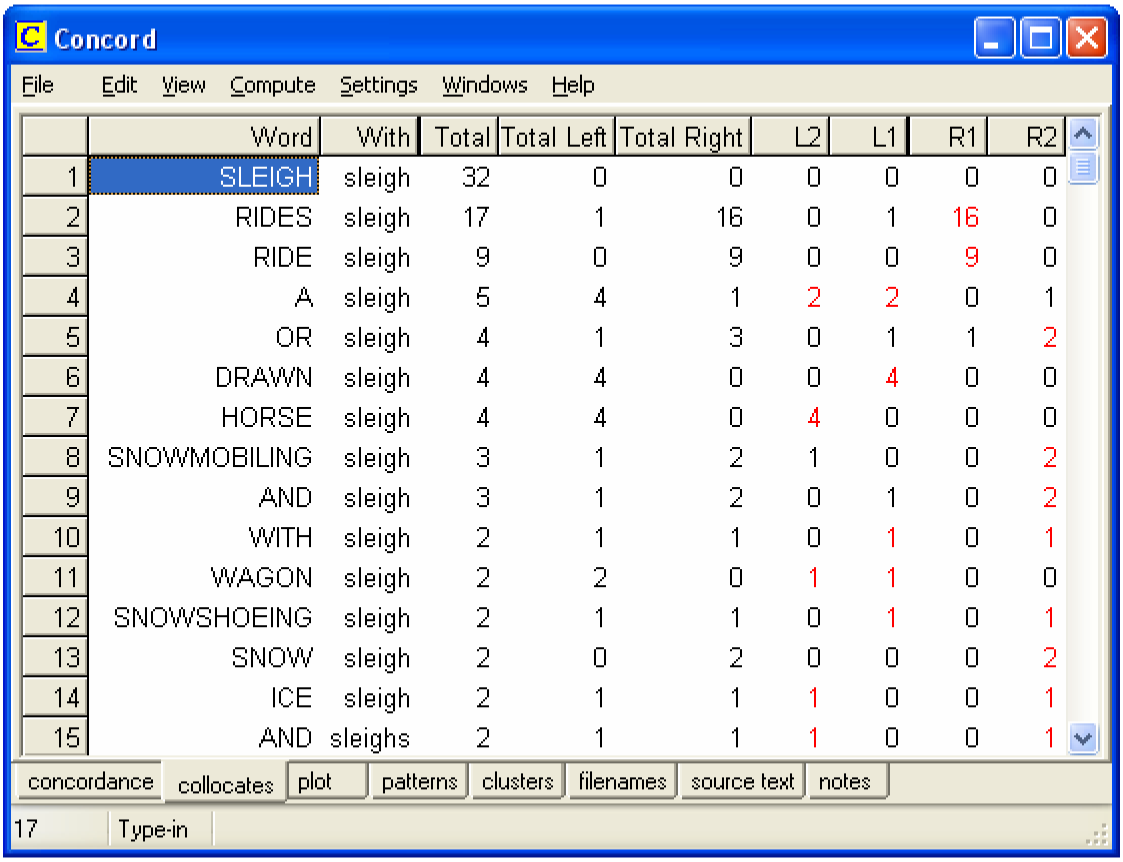
For example, figure 9 shows the words that collocate with *sled* & *sleds* in the Tourism Corpus, arranged in order of frequency. Of the 62 words that occur at least twice within the immediate neighbourhood of the search words (i.e. within a span of 2 to the left or 2 to the right) we see that *dog* is the most frequent collocate, occurring 47 times. We also see from the L1 & R1 columns that *dog* appears either immediately to the left or immediately to the right of *sled* & *sleds*. However there are no occurrences of *horse(s)* in the immediate neighbourhood of *sled* & *sleds*.



	Word	With	Total	Total Left	Total Right	L2	L1	R1	R2
1	SLED	sled	85	0	0	0	0	0	0
2	DOG	sled	47	25	22	0	25	22	0
3	SLEDS	sleds	21	0	0	0	0	0	0
4	AND	sled	17	9	8	5	4	3	5
5	A	sled	13	8	5	6	2	1	4
6	DOGS	sled	9	1	8	1	0	8	0
7	RACES	sled	8	0	8	0	0	2	6
8	YOUR	sled	7	7	0	0	7	0	0
9	DISABLED	sleds	6	6	0	0	6	0	0
10	OF	sled	6	6	0	2	4	0	0
11	BY	sled	5	5	0	2	3	0	0
12	FOR	sleds	5	0	5	0	0	1	4
13	GUIDES	sleds	5	4	1	4	0	1	0
14	NOTE	sleds	5	4	1	0	4	1	0
15	OR	sled	5	2	3	0	2	1	2

Figure 9: Collocation display for search pattern *dog/dogs*

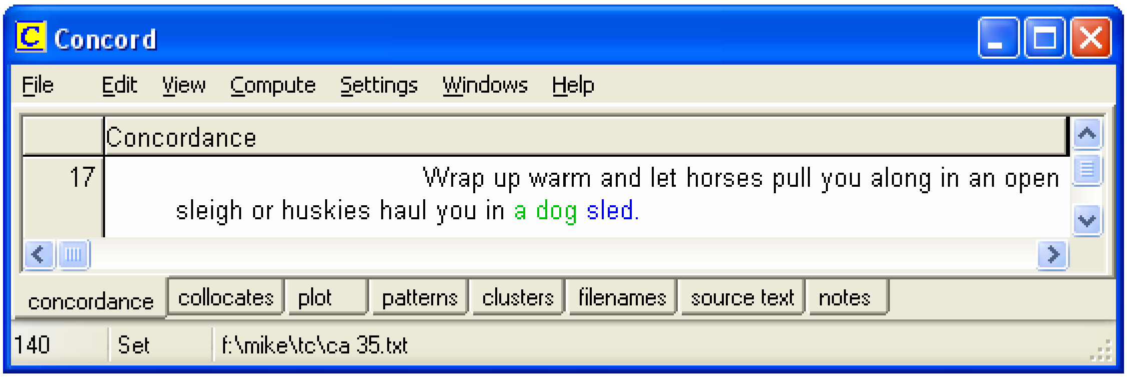
Figure 10 shows the words that collocate with *sleigh* & *sleighs*. Now we can see that *horse* appears seventh in the frequency list, mainly two to the left of the search words (L2), whereas *dog* no longer features in the top fifteen collocates.


Concord
File Edit View Compute Settings Windows Help
concordance collocates plot patterns clusters filenames source text notes
17 Type-in

	Word	With	Total	Total Left	Total Right	L2	L1	R1	R2
1	SLEIGH	sleigh	32	0	0	0	0	0	0
2	RIDES	sleigh	17	1	16	0	1	16	0
3	RIDE	sleigh	9	0	9	0	0	9	0
4	A	sleigh	5	4	1	2	2	0	1
5	OR	sleigh	4	1	3	0	1	1	2
6	DRAWN	sleigh	4	4	0	0	4	0	0
7	HORSE	sleigh	4	4	0	4	0	0	0
8	SNOWMOBILING	sleigh	3	1	2	1	0	0	2
9	AND	sleigh	3	1	2	0	1	0	2
10	WITH	sleigh	2	1	1	0	1	0	1
11	WAGON	sleigh	2	2	0	1	1	0	0
12	SNOWSHOEING	sleigh	2	1	1	0	1	0	1
13	SNOW	sleigh	2	0	2	0	0	0	2
14	ICE	sleigh	2	1	1	1	0	0	1
15	AND	sleighs	2	1	1	1	0	0	1

Figure 10: Collocation display for *sleigh/sleighs*.

A further investigation of the 140 concordance lines generated for the search pattern *sleigh/sleighs/sled/sleds* will further confirm that dogs are usually associated with pulling sleds while horses are associated with drawing sleighs (see figure 11); consequently the translator may well conclude that sleds are usually somewhat smaller, more light-weight vehicles than sleighs.


Concord
File Edit View Compute Settings Windows Help
Concordance
concordance collocates plot patterns clusters filenames source text notes
140 Set f:\mike\tc\ca 35.txt

17 Wrap up warm and let horses pull you along in an open sleigh or huskies haul you in a dog sled.

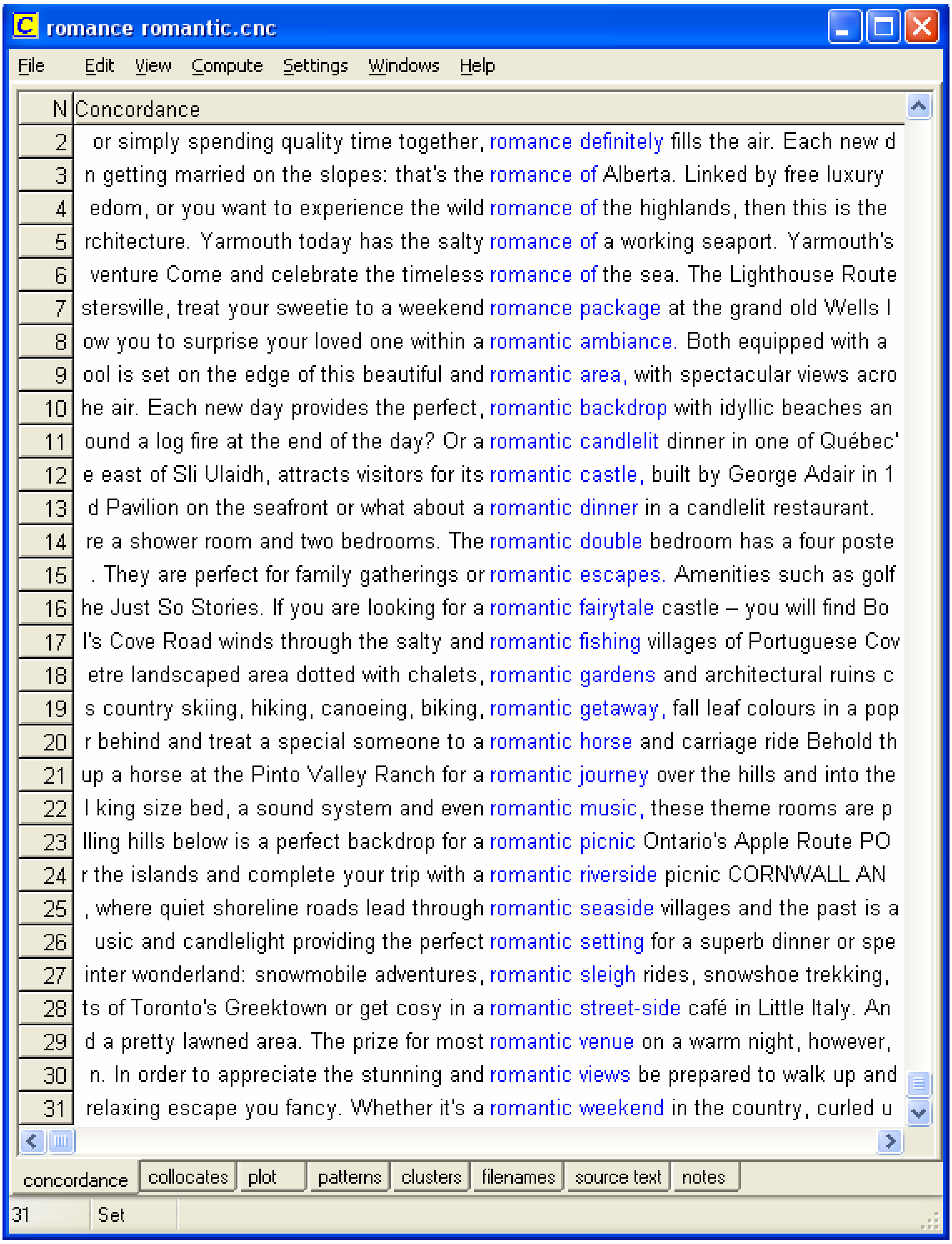
Figure 11: KWIC display for the search pattern *sleigh/sleighs/sled/sleds*

However, although corpus analysis programs can help the translator to identify meaning differences between synonyms, partial synonyms or pseudo-synonyms, such tools are probably most effective in showing actual usage and in providing collocational information. In order to make more subtle semantic distinctions, the translator will usually need to turn to other sources, such as specialised dictionaries and encyclopaedias. For example, *Wikipedia* makes the following distinction between sleds, sledges and sleighs:

> *Sleds* are typically smaller and simpler than *sleighs*, though this is not always the case. Both are lightweight vehicles whereas a *sledge* is more usually a low and rough farm vehicle designed for heavy haulage of loads such as cordwood, stone or ice blocks.

Serendipity: Destiny... with a sense of humour!

In *Serendipity* – the 2001 movie – New Yorker Jonathan (John Cusack) “accidentally” meets Brit beauty Sara (Kate Beckinsale) and the couple spend a few blissful hours together. Due to a sequence of unlikely events, they are separated and go their separate ways, leaving future encounters to fate. After many twists and turns and chance discoveries, they finally re-discover one another. When searching through special field corpora, you may well make lots of exciting serendipitous discoveries, though if you are looking for romance and true love, you won't necessary find these... but then again, perhaps you might... see figure 12.



```
 N Concordance
 2  or simply spending quality time together, romance definitely fills the air. Each new d
 3 n getting married on the slopes: that's the romance of Alberta. Linked by free luxury
 4  edom, or you want to experience the wild romance of the highlands, then this is the
 5 rchitecture. Yarmouth today has the salty romance of a working seaport. Yarmouth's
 6  venture Come and celebrate the timeless romance of the sea. The Lighthouse Route
 7 stersville, treat your sweetie to a weekend romance package at the grand old Wells I
 8 ow you to surprise your loved one within a romantic ambiance. Both equipped with a
 9 ool is set on the edge of this beautiful and romantic area, with spectacular views acro
10 he air. Each new day provides the perfect, romantic backdrop with idyllic beaches an
11 ound a log fire at the end of the day? Or a romantic candlelit dinner in one of Québec'
12 e east of Sli Ulaidh, attracts visitors for its romantic castle, built by George Adair in 1
13  d Pavilion on the seafront or what about a romantic dinner in a candlelit restaurant.
14 re a shower room and two bedrooms. The romantic double bedroom has a four poste
15  . They are perfect for family gatherings or romantic escapes. Amenities such as golf
16 he Just So Stories. If you are looking for a romantic fairytale castle – you will find Bo
17 l's Cove Road winds through the salty and romantic fishing villages of Portuguese Cov
18  etre landscaped area dotted with chalets, romantic gardens and architectural ruins c
19 s country skiing, hiking, canoeing, biking, romantic getaway, fall leaf colours in a pop
20 r behind and treat a special someone to a romantic horse and carriage ride Behold th
21 up a horse at the Pinto Valley Ranch for a romantic journey over the hills and into the
22 l king size bed, a sound system and even romantic music, these theme rooms are p
23 lling hills below is a perfect backdrop for a romantic picnic Ontario's Apple Route PO
24 r the islands and complete your trip with a romantic riverside picnic CORNWALL AN
25 , where quiet shoreline roads lead through romantic seaside villages and the past is a
26  usic and candlelight providing the perfect romantic setting for a superb dinner or spe
27 inter wonderland: snowmobile adventures, romantic sleigh rides, snowshoe trekking,
28 ts of Toronto's Greektown or get cosy in a romantic street-side café in Little Italy. An
29 d a pretty lawned area. The prize for most romantic venue on a warm night, however,
30  n. In order to appreciate the stunning and romantic views be prepared to walk up and
31 relaxing escape you fancy. Whether it's a romantic weekend in the country, curled u
```

Figure 12: Some of the concordance lines generated by a search for *romance/romantic*

References:

Bernardini, Silvia (2000). "Systematising serendipity: Proposals for concordancing large corpora with language learners". In Lou Burnard and Tony McEnery (eds) *Rethinking language pedagogy from a corpus perspective* (pp.225-234). Frankfurt am Main: Peter Lang.

Bernardini, Silvia (2001). "'Spoilt for choice': A learner explores general language corpora". In *Learning with corpora*, edited by Guy Aston. Houston (TX): Athelstan / Bologna: CLUEB. 220-249.

Bernardini, Silvia. (2002). "Exploring new directions for discovery learning". In B. Kettemann & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July, 2000* (pp 165-182) Amsterdam: Rodopi.

Bernardini, Silvia. (2004). "Corpora in the classroom: an overview and some reflections on future developments". In *How to use corpora in language teaching.* Edited by John Sinclair. Amsterdam: Benjamins. 15-36.

Scott, Mike (2004). *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

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

Wilkinson (2005b). "Discovering Translation Equivalents in a Tourism Corpus by Means of Fuzzy Searching", in *Translation Journal*, Volume 9, No 4. Viewable online at: http://accurapid.com/journal/34corpus.htm

Wikipedia, the free encyclopedia. Online at http://en.wikipedia.org/wiki/Main_Page

Zanettin, Federico. 2001. "Swimming in Words: Corpora, Translation, and Language Learning". In Aston, Guy (ed). *Learning with corpora*. Houston, TX: Athelstan, 177-197.

BIOGRAPHY

Michael Wilkinson was born and brought up in Newcastle upon Tyne in the north-east 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 the Savonlinna campus of the University of Joensuu. Nowadays he mainly teaches courses in translation from Finnish to English, oral expression and liaison interpreting. His wife, Arja, is a professional translator, working mainly from Finnish into English.

Contact: michael.wilkinson@joensuu.fi