

Units of Meaning in Translation — How to Make Real Use of Corpus Evidence

Pernilla Danielsson

Centre for Corpus Research
Department of English
University of Birmingham
United Kingdom
pernilla@ccl.bham.ac.uk

Introduction

Over the last decades, the use of corpora has grown in popularity as more and more researchers and professionals in the language industry have discovered its potential. With new methodologies and data, corpus linguistics has made us aware of language patterns that were previously hidden in obscurity. The field of translation is only one of many language subjects to be fascinated by the insights which have come from authentic language data.

The impact of corpora in translation studies is still only in its infancy. As Baker (1999) states: “Work in this area began in an exploratory fashion in the early nineties and is only now beginning to yield some concrete findings, albeit on a relatively small scale”. What separates translation studies from other fields is the way it can make use of several different types of corpora. Aligned parallel corpora, monolingual translated corpora as well as monolingual original corpora may all be used as aids in the translation process.

This article will begin by looking at aligned parallel texts although using this type of data has various shortcomings. Despite its great value in finding equivalents, it not be the most useful resource for a translator. A parallel corpus will contain all the difficulties that a translator is faced with, such as structure mirroring the source language and odd word-choices. This sometimes gives the text a foreign feel. There are no obvious solutions to problems with quality of the target text, but this paper will propose the monolingual corpus as a superior source of information. Examples will be given in Swedish and

English, taken from parallel texts, comparable texts and, most importantly, large monolingual corpora, such as the Swedish Language Bank and the Bank of English.

2. Parallel texts

The unit of analysis in language is often said to be the single word. However, the relevant units, the units of meaning, are often beyond the word and instead take the form of multi-word units. This is a fact that most translators are already familiar with. When using the single word as the unit of analysis, we are faced with the problems of apparent ambiguity. I write *apparent*, as ambiguity is primarily a problem in the automatic treatment of language, or for students in the field of linguistics. Rarely will a translator discover a section of text which is also ambiguous to the eyes of a professional. The problems of ambiguity only come into the translator's work when we need to look up something in a bilingual dictionary or from lexical databases. Suddenly, the single word found in the text can correspond with a whole list of possible equivalents in the dictionary. This is not a new observation; the short-comings of bilingual dictionaries have been referred to in many previous publications. However, if in combination with the bilingual dictionary we also have parallel texts, we get access to a sea of disambiguating decisions already made by previous translators. The examples below show the Swedish word *mål* which appears ambiguous when translating into English¹. The word corresponds to *aims*, *goals* and *objectives*, in both singular and plural forms.

of the Union ' s regional policy [[aims]] . The report shows that
av hur unionens regionalpolitiska [[mål]] uppnåtts . Rapporten visar

contribution to this great common [[goal]] . Madam President , I will
bidrag till detta stora gemensamma [[mål]] . Fru talman ! Jag vill öka

You have fulfilled your [[objective]] for today , which is to answer ..
Ni har lyckats uppnå dagens [[mål]] : att besvara samtliga frågor .

Example 1. The Swedish word *mål* in parallel texts, corresponding to 'aims', 'goal' and 'objective'.

¹ Crayton Walker's work on native German speakers' difficulties in differentiating between *objectives*, *goals*, *target* and *aims* in English gave the inspiration to use this example here.

Parallel *texts* are very attractive as they compare the translated *text* with its original (Barlow 1996). By making use of computational algorithms, such as sentence alignment and parallel concordances, it enables the user to see every sentence with its corresponding translation (for more on alignment see Church and Gale 1993 or Danielsson and Ridings 1997).

The idea of using previous translation as a source of information for new translation is already well established. Systems such as Translation Memories (TMs) have been on the market for over a decade and are fairly successful. The problems that have been reported by users of TMs are mainly that quality control is difficult to maintain when used in a large network and that only certain type of document actually benefit from these types of sentence based replacements.

One of the major drawbacks with parallel texts is their labour-intensive compilation (see for example Johansson & Hofland (1994). This, in conjunction with copyright issues and difficulties in obtaining the two language versions in electronic form, is to blame for the fact that only a few are currently freely available. To make matters worse, the few available parallel corpora are usually very small in size, at least compared to their monolingual counterparts. They will therefore, only cover a small part of the vocabulary for each language. For the same reason that Translation Memory may not be the optimal translation tool, parallel texts tend to not always give us ‘accurate’ equivalence information. What has once been erroneously introduced as a unit of translation may become justified by its appearance in such a language resource or tool.

Mr President , I am participating **in this discussion** in order to support the Herr talman ! Jag deltar **i denna debatt** för att stödja min kollegas , Ma ...

So we are asking **in this question** , Commissioner , for confirmation... Därför ber vi **i denna fråga** , herr kommissionär , om bekräfte ...

report and the aid report share common ground **in this White Paper** naturligtvis också en gemensam grundval **i denna vitbok** .

Example 2. The word *denna* in a Swedish English parallel text of debates from the European parliament.

The example above is taken from a small collection of English-Swedish parallel texts, these consist of debates from the European parliament. The output is produced using Barlow's ParaConc, a parallel concordancer that arranges the data according to a node word (Barlow 1995, 1996). Whereas each sentence in the example is grammatically correct, they do not appear the same as in debates from the Swedish *riksdag* (as will be discussed further below). The difference lies in the phraseology, as Swedish original documents tend to use the phrase '*i den här frågan*' rather than '*i denna fråga*'. This is a minor observation; it only becomes important when one difference adds to another and the text starts to get a foreign feel to it. By simply re-using information from other translated texts, we will not be able to get out of this vicious circle. But if bilingual dictionaries cannot suffice to give us all information and translation memories have problems with quality control, how then can we avoid these pitfalls? In this article we will propose the use of monolingual corpora as the most important source of data for translators. By comparing the target text with text created originally in the same language, we may spot many of the differences.

3. Why do translated texts not look like original texts?

Baker (1995, 1999) claims that many researchers are undermining the value of translated texts; many corpus linguists do not wish to include translated text into a corpus of modern-day English, for example. This is based on the observation that translated texts do not always conform to the patterns and structures of original texts. While this may be true, it still does not justify a complete exclusion of translated texts from monolingual corpora, as many languages today are dependent on translated texts or monolingual corpora. Take the Swedish language as an example; out of all detective stories published in Sweden 1999, 66% were translations², mainly from English, this was seen as an important step forward by Swedish authors. Whereas the detective genre may be an exception in its domination by foreign authors, most libraries in Sweden have at least 40% translated novels in their collections. This means that two out of five Swedes, picking up a bed-side book, are likely to read translated text; a figure that should not be

² The figures taken from an article in the *Aftonbladet*, July 31, 1999.

ignored when compiling language corpora. Gellerstam (1986) has shown how word meanings are influenced by translations. Words such as ‘drama’, which originally was only used in theatre contexts, started to appear in phrases such as ‘dramatic changes’ (*dramatiska förändringar*). This was first seen in translated texts, but is now also found in original Swedish texts. But despite the influence translated text has on the language, they are often given a lower status. So, what are the differences that make some people regard translations as inferior texts?

To illustrate these differences, we will here perform a comparative study of two very small text collections, namely 10 Swedish original documents taken from debates in the Swedish *riksdag* (*'parliament'*). These will be compared with 15 documents translated into Swedish, taken from debates in the European parliament. Although small, it will here suffice to highlight certain differences. The important factor is that both collection consists of debates, and could therefore be expected to use equivalent words and structures. Translation companies may in fact perform such tests on their own texts in comparison with original texts. The comparison is performed using the software *Keywords*, implemented by Mike Scott as part of the WordSmith tools (Scott 2001) which runs on most Windows machines. The tool *Keywords* compares the relative frequency of both text collections and highlights words which are significantly over or under-used in each respective collection.

1	EUROPEISKA	('European')
2	SKALL	(auxiliary verb, 'will')
3	RÅDET	('the council')
4	KOMMISSIONEN	('the commission')
5	ORDFÖRANDE	('chairman')
6	ÄNDRINGSFÖRSLA+	('amendments')
7	PARLAMENTET	('parliament')
8	UNIONEN	('the union')
9	BETÄNKANDE	('report')
10	EUROPAPARLAMEN+	('European parliament')
11	MEDLEMSSTATERN+	('Member States')
12	UNIONENS	('the union's')
13	BETÄNKANDET	('the report')
14	DENNA	('this')
15	AV	('of')
16	EUROPA	('Europe')
17	DETTA	('This')

Example 3. Words more significant in translated EU parliament debates, when compared to Swedish *riksdags* debates

1	SKA	(aux verb, 'will')
2	ANF	(abbreviation)
3	M	(abbreviation, political party)
4	NBSP	(abbreviation)
5	FP	(abbreviation, political party)
6	LARS	(name)
7	S	(abbreviation, political party)
8	SVERIGE	('Sweden')
9	STATSRÅDET	('counsellor')
10	KD	(abbreviation, political party)
24	VÄLDIGT	('very')
46	LITE	('little')

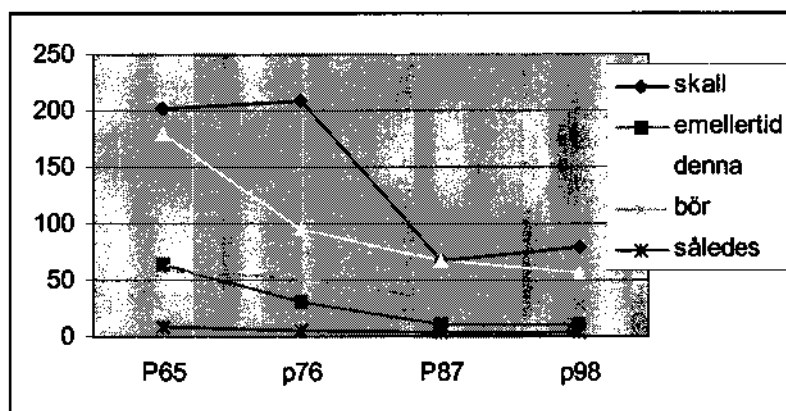
Example 4. Keywords significant to original Swedish debates when compared with translated debates.

Amongst the words that are more widely found in the documents from the EU parliamentary debates than in the Swedish original documents, we find *European*, *parliament*, *the council* and *the commission*. At the other end of the lists, example 4, we find words that are more significant in the original documents, such as *Sverige* ('Sweden'), *statsrådet* ('the counsellor'), as well as several names and abbreviations of political parties. These words tend to signal "aboutness", i.e. they are directly linked to the genre and text type of the document. As such, they highlight differences in the content of the documents rather than in the language.

For our study, it is substantially more interesting to find words such as *skall* ('shall, will'), *emellertid* ('however', a formal word), and *huruvida* ('whether', a formal word), which are non genre and type specific words. These words are singled out as significant for the translations, i.e. they tend to be used much more in these translated texts than in the original Swedish debates. They should be contrasted by words such as *ska* ('shall, will'), *lite* ('little'), *väldigt* ('very'), and *här* ('here'), which are significantly more commonly used in the original texts; Why is this?

This question may be partially answered by using a monitor corpus. The term *monitor corpus* was first introduced by Sinclair (1987) in reference to a dynamic, as opposed to static corpus, in which one may study the changing nature of language. The best monitor

corpus of Swedish is the Swedish language Bank, which has a sub-corpus from each of the year 1965, 1976, 1987 and 1998, i.e. every 11th year.



Example 5. A graph over the diminished use of specific words in Swedish.

The graph in example 5 represents the relative frequency of words in each corpus. What should be apparent in the data is that the words *skall*, *emellertid* and *således* have all become much less frequent over time. We may claim that the words have an archaic ring to them, or they sound rather too formal.

This observation points to a tendency for using words that are out of fashion. If we compare these findings to previous observations in translation studies, we may see that they conform to general trends. Baker (1999) shows how many words in their Translational English Corpus refer to an archaic meaning of the word, as in the case of 'gay' which tends to mean 'happy' in translated texts, but has come to mean 'homosexual' in modern-day original texts.

The clear writing campaign, which is active in Sweden as in several other European countries, works towards simplifying bureaucratic language, making it more available and understandable for people outside the politic arena (klarspråksgruppen, <http://justitie.regeringen.se/klarsprak/>). On their website they include a blacklist of words that are not to be used, all of them bearing an archaic and much too formal feel. One item on their agenda has been the modernisation of official language to a more present-day

Swedish; a message that has not yet left Stockholm for Luxemburg and Strasbourg. In regards to the word *skall* and *ska* (*shall*), it states that the longer form sounds formal and should be used only in legal texts and very formal documents, but otherwise the shorter form is preferred. Compared to the debates from the Swedish riksdag, we find that whereas *skall* is used in translated EU texts, *ska* is used in the original Swedish debates.

N	cluster	Freq.	Translation
1	i den här frågan	9	(in this question)
2	att det här är	8	(that this is)
3	det här är ett	7	(this is a)
4	den här typen av	6	(this type of)
5	tycker att det här	6	(thinks this)
6	det här är en	5	(this is a)
7	i det här fallet	5	(in this case)
10	i den här kammaren ⁴		(in this chamber)
16	i den här budgeten	3	(in this budget)
17	i den här debatten	3	(in this debate)
18	i det här landet	3	(in this country)

Example 6. Repetitive phrases around the word *här*

Whereas many of the words overused in the translated text can be explained by their too formal status or old-fashioned nature, this argument cannot be turned around to answer the question of why some words are more often used in the original texts. It is not a list of ‘hip’ new words entering the language. A word such as *lite* ('little') has hardly moved in the frequency list over a period of 40 years, neither has *här* ('here'). Instead, the reasons become obvious as we start to study the patterning around these words. These significant words are used repetitively in various larger units, such as "*det händer för lite*" ('not enough is happening') and "*det gäller inte minst*" ('not least in/with...'). In the case of *här*, this was part of the corresponding unit referred to in example 1 above, instead of '*i denna fråga*', original texts tend to use '*i den här frågan*', both corresponding to the English phrase 'in this question'.

4. Multi-word Units of Meaning

Language is a unique object of study as it has the ability to redefine itself from within; the way we used a word today may not be fashionable tomorrow. Language also offers

more than one way of expressing something, which means that it is often left unnoticed to translators that they have a problem. But, as we humans are creatures of habit, we have made it our style to repeat the same patterns and structures, making it the signature of native-like speech (Kjellmer 1987). The problem with translation is of course that when you come up with a grammatically correct utterance, there appears to be no need to search for a conventional pattern. The result sometimes leaves much to be asked for and creates the labelling of translated texts as inferior texts.

Although everything in a language can be said in a variety of ways, there is always a preferred way of saying it. In a study by Erman and Warren (2000), they showed that as much as 55% of the words in a text belongs to some sort of prefabricated chunks. Why then, does a translator translating into his or her native language not always come up with these patterns? Most likely, this is due to the way we store information in our mind. When asked about a word we tend not to first come up with the patterning, but instead start from the words as a single item. When asked about the word 'scruff', we tend to start by defining it as 'a scruff is someone who is very untidy'. This is despite the overwhelming evidence from the corpus that the word is mainly used in the phrase 'to take someone/something by the scruff of the neck'. Even when prompted for multi-word units, we tend to only come up with those that begin with the word in question (Wray 2002). If we hear *kith* our mind will fill in '*and kin*'. But if we hear *eye*, we do not tend to think of '*to keep an eye on something*' or '*to turn a blind eye to something*'.

In translation, the source text tends to gear us towards word choices and structures that may not have been our preferred option, if we had created the text from scratch. This is where the monolingual corpus comes into use. Where our minds may hide relevant patterns, a corpus will flatten the language and display all patterns at once. As such, it may be the most important translation aid of all.

The following excerpt comes from a press release published by Birmingham University, shortly before the last football World Cup in the year 2000. It states:

With more than 30,000 staff and students from all over the world, the traditionally reserved University of Birmingham is gearing up for an unprecedented outbreak of World Cup fever.

Heads of school and department managers will be turning a **blind eye** as employees take time out from their duties to watch their home nation compete in the tournament.

Example 7. Excerpt from a press release before the Football World Cup 2000.

The highlighted unit is *turning a blind eye* to something³. This is a phrase that dates back to Admiral Lord Nelson and the battle of Copenhagen, where Nelson refused to obey orders and give up the battle. He knew he could win and decided to put his monocular to his blind eye, thereby not seeing the signals to turn back. This is a phraseology that is linked to a specific historical event, mainly relevant to British speakers, and therefore cannot be expected to have an equal phrase in other languages. This makes it difficult but certainly not impossible to translate.

eye, Nelson is unable to "turn a blind eye" to the signal ordering him to
who he claimed were turning a blind eye to the role of right-wing
or, at the very least, turn a blind eye to it. I mean, the breakdown of
that my country will not turn a blind eye toward the plight of human
accuses the department of turning a blind eye to simple and cost-free
and said federal OSHA turned a blind eye to the situation. Meantime,
reports the KGB sometimes turns a blind eye for a cut. <p> To keep Soviet
How many times can we all turn a blind eye to the poverty and hunger and
asked. 'As long as Americans turn a blind eye to the poverty that is creating
can muster. I promise never to turn a blind eye to the suffering that some are
paper accused the West of turning a blind eye to what it called Israeli
make it illegal for owners to turn a blind eye to any activity on their
States military advisers turned a blind eye to these death squads. US and
and obscenity - she simply turns a blind eye to it altogether. <p>

Example 8. Concordance of *turn a blind eye*.

Today, this phrase is used without any apparent reference to the old battle, simply meaning that that the subject, for various reasons, choose to ignore something.

In Swedish we have several ways of expressing the act of choosing to ignore something. The most frequent is "*blunda för*" ('close our eyes to something') which occurs 66 times in the 20-million Swedish Parole corpus. A less frequent phrase is the biblical phrase "*se genom fingrarna*", alternatively "*se mellan fingrarna*", which together is represented 16

³ A more extensive analysis of the word *eye* may be found in Sinclair (1991).

times in the Parole corpus. Whereas the original biblical phrase was '*se genom fingarna med*' the corpus shows that the preposition *med* is losing its importance.

all respekt . Men vi skall **inte** blunda för att den just på grund av den snabba vårt samhälle . Men vi **kan inte** blunda för att ungefär i samma takt som detta aktum inom några år . **Ingen kan** blunda för den anpassning som EMU innebär på va till regeringen **kan** han **inte** blunda för att var femte s-väljare trots s en del problem , man **kan inte** blunda för det . För att bli accepterad måste idag . Samtidigt **kan** man **inte** blunda för att datatekniken utvecklas snabbt . Detta innebär att vi **inte** får blunda för den framtid som många unga åt fel håll men man ska **inte** blunda för hur verkligheten ser ut , tänkte ganiserat samarbete . Genom att blunda för den ekonomiska saneringens s melodi ? Ska vi fortsätta att blunda för sanningen ? Är det inte bättre att ejnikov . -Det går det **inte** att blunda för , sa nöjde tränaren Olof ycket litet parti . Vi **kan inte** blunda för att omvärlden förändrats och att r . Men det går tyvärr **inte** att blunda för verkligheten . Även det i sig yckat projekt " . " Vi **kan inte** blunda för problemen i Östra nordstan " , gång i centrum : - Vi **kan inte** blunda för problemen i Östra Nordstan . Att

Example 9. A few of the 66 occurrences of '*blunda för*' in the Swedish Parole corpus.

As a native speaker of Swedish, I was immediately drawn to the phrase '*blunda för*' as a translation to '*turning a blind eye*'. This may be due to the link between *blind* and *blunda*, or perhaps because of the reference of eye. The phrase '*se genom/mellan fingrarna*' is not triggered by the source phrase. However, the interesting difference between the two patterns is highlighted in the example above. In the concordance lines, '*blunda för*' is frequently preceded by the negation *inte* (*not*), often also in combination with the verb *kan* ('can'). The phrase '*blunda för*' may be said to have a negative connotation, as it most frequently occurs in negated statements, we talk about not being able to close our eyes to something. By using '*blunda för*' we are subliminally sending out the message that we cannot do this, i.e. this is wrong. In the press release above, the manager of the university did not wish to place blame, but instead encourage the staff to participate in the festivities of a football World Cup. As such, the phrase '*blunda för*' would not be appropriate.

inte snarare än man **kunde** se mellan fingrarna med det . Markus den skojar , en skulle hon säkert **inte** se genom fingrarna med hans förolämpning . Eller , omvärlden har **tenderat att** se genom fingrarna på det kroatiska avgörandet . Han tillhålls att **inte** se mellan fingrarna på mord på flyende hutuer , e om att hyresnämnden **ska** se mellan fingrarna . Hans motivering är att ingen , idigt tycks myndigheterna se genom fingrarna vad gäller de arbetsgivare vilka ppas på att Italien **skall** se genom fingrarna med initiativet . Italien är . **kan** tillståndsenheten se mellan fingrarna . Då utdelas bara en varning . lla skäl , **fortsätter att** se mellan fingrarna . Kvinnor kan klä sig hur som

att Gatubolaget **verkar** se mellan fingrarna så fort det gäller arrangemang
ck ju från att dom **skulle** se mellan fingrarna med en så känd politiker som jag
Samhället kan inte längre se genom fingrarna med grovt olaglig verksamhet !
som helst och inte tänker se mellan fingrarna med vad Unionen uppfattar som
erna **tycks i** många stater se mellan fingrarna på denna uppenbart mycket

Example 10. All occurrences of "*se genom/mellan fingrarna*" in the Swedish 20-million words PAROLE corpus

The opposite can be said of "*se genom/mellan fingrarna*". Although it is only a tendency, we may argue that this phrase often has positive connotations as it used to state that someone should '*see through their fingers*', or '*turn a blind eye*', to something.

The use of monolingual corpora is vital for spotting these patterns. Simply relying on introspection without any authentic language data is most likely to be misleading as it does not give us any evidence about usage (Sinclair 1991). And a text without conventional pattern tends to sound odd.

Another example of how monolingual corpora can provide important information about longer units is where the units are deceptively similar. When talking about single words that look the same in different languages, we talk about "false friends". This term is not frequently used for multi-word units, probably because there are fewer examples of them.

Both English and Swedish languages have a unit which is '*so called/så kallad*' which looks deceptively similar. As such, we may assume that they form perfect translation equivalents. However, units tend to be coloured by their most prominent use in their respective language, this may vary from language to language. In Swedish, the phrase "*så kallad*" is often used to introduce new or foreign terminology, as shown in example 11, and is therefore it frequently occurring in academic texts.

smugit in några viner så kallade bag-in-box som rymmer fyra normal
och fiskar. Det är en så kallad "booze cruise" - med fri tillgång till
av inteventionsstudie med så kallad "cross-over" teknik, där patienten
Från 2004 kan ett förbud mot så kallad indirect reklam vara infört.

Example 11. Usage of the phrase '*så kallad*' in Swedish.

Turn to the English corpus and we find another tendency. Here, the phrase "so called" tends to be used to signal a disbelief in what comes after, as illustrated in the examples below.

highlight the weakness of the so-called recovery" <p> The trend in the young people are written off as a so-called lost generation" <p> This is an I often questioned the so-called special Queensland spirit when I of journalists and academics and so-called prominent feminists are". <p>ve been written off by their gay so-called 'friends'. I've witnessed the

Example 12. Use of the English Phrase "so-called"

By translating a Swedish use of *så kallad* into *so called*, you may signal to the reader that you do not believe in the term you are introducing, a rather unfortunate situation as it infers the opposite result as to which you intended.

Conclusion

In this paper, we focused on how to make a text sound native-like even when it is the result of a translation. Various methods and corpus types were illustrated. The most well-known type to the translation community, the parallel corpora, were avoided, mainly due to the fact that they seem to nourish rather than eliminate problems with non-idiomatic language. To find some of the most frequent differences between original Swedish text and Translated texts we made use of the tool *Keywords*. The comparison gave us a few points where translation is likely to differ from original texts, such as in the use of archaic or too formal words, the use of "glossary" translations, and in the phraseology.

Phraseology may be the most important, but also the most difficult, part to control when translating. Not only because it is difficult to find the correct units, but also because the awareness of what creates idiomatic language is often low even amongst native speakers. In this paper, it has been suggested that a monolingual corpus can provide vital information about how to make a text look idiomatic. When a foreign text is luring you to take less conventional paths in the creation of your target language, a check with the monolingual corpora could convince you to choose a more idiomatic pattern. Although not yet vastly explored amongst translators, these language resources are available in

large quantities. As soon as you have started to use one, you will realise, you have been hooked!

References

- Baker, M. 1999. The role of corpora in investigating the linguistic behaviour of professional translators. *International Journal of Corpus Linguistic* 4(2): 281-298.
- Baker, M. 1995. Corpora in Translation Studies. An overview and some suggestions for future research. *Target* 7(2): 223-243.
- Barlow, M. 1996. Analysing Parallel Texts with ParaConc. In: Proceedings from ALLC - ACH 1996, University of Bergen, Norway.
- Barlow, M. 1995. ParaConc: A concordancer for parallel texts. In: *Computers and Texts*. No 10, pp 14-16.
- Gale, W and Church, K. 1993. A program for aligning sentences in bilingual corpora. In: *Computational Linguistics* 19(1), pp 75-101.
- Danielsson, P & Ridings, D. 1997. Practical Presentation of a "Vanilla" aligner. In Reyle, U. and Rohrer, C. (eds) Presented at the TELRI Workshop on Alignment and Exploitation of Texts. Institute Jozef Stefan. Ljubljana.
- Erman, B. and Warren, B. 2000. The idiom-principle and the open-choice principle. In: *Text* 20(1), pp29-62.
- Gellerstam, M. 1986. Translationese in Swedish Novels translated from English. In: Wollin, L. & Landquist, H. (eds.) *Translation Studies in Scandinavia*. Lund: Gleerup.
- Johansson, S. and Hofland, K. 1994. Towards an English-Norwegian Parallel corpus. In: Fries, U. Tottie, G. and Schneider, P. (eds.) *Creating and Using English Language Resources*. Amsterdam: Rodopi, pp 25-37.
- Kjellmer, G. 1987. Aspects of English Collocations. In Meijs, W. (ed.) *Corpus Linguistics and Beyond*. Amsterdam: Rodopi
- Scott, M. 2001, "Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith Tools suite of computer programs", in M. Ghadessy, A. Henry & R.L. Roseberry (eds.) *Small corpus studies and ELT: theory and practice*. Amsterdam: Benjamins, pp. 47-67
- Sinclair, J. 1987. *Looking Up : An account of the COBUILD Project in lexical computing*, Collins ELT, 1987
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stefansson, K 2003. 'The job as seen by a Swedish translator' [http://www.europarl.eu.int/trans/2001-trad sv.pdf](http://www.europarl.eu.int/trans/2001-trad_sv.pdf). [as displayed October 10, 2003]
- Teubert, W.
- Wray, A. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

The **corpora** used or referred to in this paper were:

The Bank of English; www.cobuild.collins.co.uk

The Swedish Language Bank; <http://spraakbanken.gu.se/lb/konk/>

The Swedish Parole Corpus; <http://spraakbanken.gu.se/lb/parole/>

Translational English Corpus:

<http://www.monabaker.com/tsresources/TranslationalEnglishCorpus.htm>

For English monolingual corpora, see also

The British National Corpus; <http://www.natcorp.ox.ac.uk/>

Collection of debates from the European Parliament, www.euoparl.eu.int

Collection of debates from the Swedish riksdag, www.riskdagen.se

More information about the **programs** mentioned in this article can be found on the following suites:

WORDSMITH; <http://www.lexically.net/wordsmith/index.html>

ParaConc; <http://www.athel.com>

LookUp (software for the Bank of English); www.cobuild.collins.co.uk