Towards a multilingual corpus for contrastive analysis and translation studies

Stig Johansson
Department of British and American Studies, University of Oslo
Towards a multilingual corpus for contrastive analysis and translation studies
(draft, October 1999; please do not quote)

Stig Johansson
University of Oslo

1. Introduction

In the course of the last few years, a research team at the universities of Oslo and Bergen has
developed a bilingual corpus for contrastive analysis and translation studies: the English-
Norwegian Parallel Corpus (ENPC). Because of the way the corpus is structured, it can be
used both as a translation corpus, with original texts and translations, and as a comparable
corpus, with matched original texts in each language. The design makes the corpus into a
flexible tool for asking a range of research questions in contrastive analysis and translation
studies (see Johansson 1998).

In addition to compiling the corpus, the ENPC project has developed tools for the
building and use of bilingual corpora, especially the Translation Corpus Aligner (Hofland
1996, Hofland and Johansson 1998) and the Translation Corpus Explorer (Ebeling 1998). The
corpus has already been used for a number of studies on lexis, syntax, and discourse.¹

2. From the ENPC to the OMC

If we want to gain insight into language and translation generally, and at the same time
highlight the characteristics of each language, it is desirable to extend the comparison beyond
language pairs. The ENPC project has therefore been extended to include translations of the
English original texts into three other languages: German, Dutch, and Portuguese. The
extended corpus will be called the Oslo Multilingual Corpus (OMC). Both the alignment
program and the browser have been developed to handle more languages.

Together with the Norwegian translations and the translations into Swedish and
Finnish assembled by related projects in Sweden (Lund/Göteborg) and Finland
(Jyväskylä/Savonlinna), we can then compare across six languages using the English original
texts as a starting-point. See Figure 1.

![Multilingual comparison of English originals and their translations](image)

Figure 1 Multilingual comparison of English originals and their translations

Plans are also under way to include translations into French, in cooperation with research
teams in Belgium and France.

The expansion will focus especially on the triple English-Norwegian-German, for all
of which we are collecting comparable original texts and translations into the other two
languages. Figure 2 shows the design of the trilingual project which is now under way. This will allow all the types of research questions which were built into the design of the ENPC, i.e.:

- comparison of original texts and translations across languages;
- comparison of original texts across languages;
- comparison of translations across languages;
- comparison of original texts and translations within the same language.

Figure 2 Multidimensional comparison of English, German, and Norwegian

The rest of this paper will report on an exploratory study making use of sixteen original English fiction texts and their translations into German and Norwegian. All the texts are extracts of 10,000-15,000 words taken from novels first published in the last couple of decades. With one or two exceptions, each author or translator is represented by no more than a single text.4

3. The English verb spend and its correspondences in German and Norwegian

The starting-point for the study was the observation by Gellerstam (1996: 59) of the over-use of the Swedish verb *tillbringa* in texts translated from English, presumably caused by the translators' attempts to render the English verb *spend*. However, as Gellerstam's material was limited to Swedish texts, he could not include a comparison with the English original. This is what we set out to do here. Both German and Norwegian have close counterparts of Swedish *tillbringa*, and it will be interesting to see to what extent our results are compatible with the findings of Gellerstam.
3.1 Overview of results

The material yielded 68 examples after instances to do with the spending of money had been excluded. Also left out were a couple of instances of *spend* in the passive. The rest are all active and have to do with the spending of time. Table 1 gives a survey of the different forms of complementation in the original texts, and we see that, with two exceptions, the examples contain a temporal NP followed by further specification in the form of an adverbial phrase or an *ing*-clause denoting: spending time in a place (ADV\textsubscript{place}), with somebody (ADV\textsubscript{accomp}), in a particular manner (ADV\textsubscript{manner}), doing something (V-*ing*), or a combination of these.

Table 1 The distribution of complementation patterns of *spend* in 16 English fiction texts

<table>
<thead>
<tr>
<th>Complementation Pattern</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>spend</em> + NP\textsubscript{temp}</td>
<td>2</td>
</tr>
<tr>
<td><em>spend</em> + NP\textsubscript{temp} + ADV\textsubscript{place}</td>
<td>21</td>
</tr>
<tr>
<td>ADV\textsubscript{accomp}</td>
<td>6</td>
</tr>
<tr>
<td>ADV\textsubscript{manner}</td>
<td>4</td>
</tr>
<tr>
<td>ADV\textsubscript{place} + ADV\textsubscript{accomp}</td>
<td>2</td>
</tr>
<tr>
<td>ADV\textsubscript{place} + ADV\textsubscript{manner}</td>
<td>1</td>
</tr>
<tr>
<td>ADV\textsubscript{place} + V-<em>ing</em></td>
<td>2</td>
</tr>
<tr>
<td><em>spend</em> + NP\textsubscript{temp} + V-<em>ing</em></td>
<td>28</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>68</strong></td>
</tr>
</tbody>
</table>

Table 2 summarizes the distribution of the German and Norwegian translations. Although there is a variety of translations, the German and Norwegian translators have chosen similar strategies, very often opting for a transitive verb which yields the same overall structure as in the English original. In the discussion below we will first illustrate such congruent translations and then move on to less direct translations.

Table 2 Correspondence patterns for *spend* in German and Norwegian translations of 16 English fiction texts

<table>
<thead>
<tr>
<th>German</th>
<th>Norwegian</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>spend</em> + NP\textsubscript{temp}</td>
<td>bleiben</td>
</tr>
<tr>
<td></td>
<td>sein</td>
</tr>
<tr>
<td><em>spend</em> + NP\textsubscript{temp} + ADV\textsubscript{place}</td>
<td>verbringen (13)</td>
</tr>
<tr>
<td>intr/refl verb (7)</td>
<td>intr verb (7)</td>
</tr>
<tr>
<td>other (1)</td>
<td>other (3)</td>
</tr>
<tr>
<td>ADV\textsubscript{accomp}</td>
<td>verbringen (4)</td>
</tr>
<tr>
<td>intr verb (1)</td>
<td>intr verb (2)</td>
</tr>
<tr>
<td>other (1)</td>
<td>other (2)</td>
</tr>
<tr>
<td>ADV\textsubscript{manner}</td>
<td>verbringen (2)</td>
</tr>
<tr>
<td>zubringen (1)</td>
<td></td>
</tr>
<tr>
<td>nutzen (1)</td>
<td></td>
</tr>
<tr>
<td>verwenden (1)</td>
<td></td>
</tr>
<tr>
<td>intr verb (1)</td>
<td></td>
</tr>
<tr>
<td>intr verb (1)</td>
<td></td>
</tr>
</tbody>
</table>
3.2 Congruent translations

A congruent translation has the same overall structure as the English original, i.e. a transitive verb plus a temporal NP in object position. The verbs chosen most often are *verbringen/zubringen* and *tilbringe*, which mean 'pass (time)'.

(1) He liked Sir Bernard Hemmings, but it was an open secret inside "Five" that the old man was ill and *spending less and less time in the office*. (FF1)

Er mochte Sir Bernard Hemmings, aber es war in "Fünf" ein offenes Geheimnis, daß der alte Mann krank war und *immer weniger Zeit im Büro verbrachte*.

Han likte Sir Bernhard Hemmings, men det var en åpen hemmelighet i "Fem" at den gamle mann var syk og *tilbrakte mindre og mindre tid på kontoret*.

Where there is an *ing*-clause, this is rendered both in the German and the Norwegian translation by an infinitive construction, usually denoting manner.

Less frequently, a congruent translation has a verb meaning 'use (time)': *nutzen* or *verwenden* in German, *bruke* in Norwegian

(2) I *spent most of the time sobbing in the protecting darkness of the great cathedral*, only half conscious of the endless stream of tourists shuffling past. (ABR1)


*Jeg tilbrakte det meste av tiden med å hulke i det beskyttende mørket i den store katedralen*, bare halvt oppmerksom på den endeløse strømmen av turister som subbet forbi.

(3) "Look Brian, I've *spent two years on that investigation*. (FF1)

"Hören Sie, Brian, ich habe *zwei Jahre auf diese Nachforschungen verwendet*. "Hør nå, Brian. Jeg har *brukt to år på denne etterforskningen*.

(4) I actually spend time thinking about this. (MA1)

*Ich verwende tatsächlich Zeit darauf, über diese Frage nachzudenken.*

*Jeg bruker faktisk tid på å tenke ut dette.*
This type of translation is found where there is a further specification in the form of an adverbial phrase or an \textit{ing}-clause.

The translations bring out the fact that the English verb may include elements both of passing and using time. In opting for a congruent translation, the German and Norwegian translators must choose between one meaning or the other. But in close to half the cases the translators have preferred some kind of restructuring.

### 3.3 Restructuring

In the most common type of restructuring the temporal object is translated by an adverbial of duration combining with an intransitive verb:

(5) "But I spent the night at Rose's." (JSM1)  
"Aber ich hab \textit{heut nacht bei Rose geschlafen}."  
"Men jeg har \textit{jo ligget over hos Rose}." (lit. 'lie over')

(6) Since the age of eighteen, he'd spent an accumulated nine years in jail. (SG1)  
Seit seinem achtzehnten Lebensjahr hatte er \textit{alles in allem neun Jahre im Gefängnis verbracht}.  
Siden attenårsalderen hadde han \textit{sittet inne i tilsammen ni år.} (lit. 'sit inside')

In (5) we note that the Norwegian translator has chosen a phrasal verb that lexicalizes the notion 'stay the night', while the German translator has the intransitive verb \textit{schlafen} ('sleep') plus an adverbial. In (6) the German translator has opted for a congruent structure with \textit{verbringen}, and the Norwegian translator has again picked a phrasal verb, here lexicalizing the notion 'be in jail'. Other intransitive (or reflexive) verbs found in the material are: \textit{sich aufhalten} (lit. 'keep oneself'), \textit{bleiben} ('stay'), \textit{verweilen} ('stay'); \textit{bo} ('live, stay'), \textit{oppholde seg} (lit. 'keep oneself'), \textit{sitte} ('sit'), \textit{være} ('be').

Where there is a temporal object plus an \textit{ing}-clause, the temporal object is often translated by an adverbial of duration and the verb of the \textit{ing}-clause is 'raised' to become the predicator of the main clause. Here there is thus no verb corresponding to \textit{spend}.

(7) After leaving school at sixteen, Rawlings had \textit{spent ten years working with and under his Uncle Albert in the latter's hardware shop}.  
Nach seinem Schulabgang im Alter von sechzehn hatte Rawlings \textit{zehn Jahre in der Eisenwarenhandlung seines Onkels Albert gearbeitet}.  
Rawlings hadde sluttet på skolen da han var seksten år og siden \textit{arbeidet i ti år sammen med og under sin onkel Albert som drev jernwarehandel}.

(8) \textit{We spent a lot of the time driving}, in our low-slung, boat-sized … (MA1)  
\textit{Die meiste Zeit fuhren wir in unserem niedrigen, bootsförmigen Studebaker herum …}  
\textit{Mye av tiden kjørte vi bil, en lav Studebaker, …}  

(9) \textit{Nights on end she spends flying}, beyond the reach of all that threatens her by day.  
(ABR1)  
\textit{Ganze Nächte hindurch fliegt sie} dahin, unerreichbar für alles, das sie tagsüber bedroht.  
\textit{Natt etter nattflyr hun}, utenfor rekkevidde av alt det som truer henne om dagen.

(10) \textit{He spent pleasurable hours dithering} over questions of punctuation. (AT1)
Er grübelte vergnügliche Stunden lang über Interpunktionsprobleme nach.
Han tilbrakte koselige timer med å gruble over tegnsettingen.

In the last example the Norwegian translator has chosen a congruent translation rather than restructuring the clauses.
Less common ways of restructuring, the last three exemplifying the category 'other' in Table 2, are illustrated in:

(11) Jim Rawlings spent the hour between nine and ten that night sitting in another, smaller rented car outside Fontenoy House. (FF1)
Jim Rawlings verbrachte die Stunde zwischen neun und zehn Uhr an diesem Abend in einem anderen, kleineren Mietwagen vor dem Fontenoy House.
Jim Rawlings tilbrakte timen mellom ni og ti den kvelden bak rattet i en annen, mindre leiebil utenfor Fontenoy House.

(12) Hartmann had tried to emulate this activity when they kindly included him in their invitations to spend the evening in a pub … (AB1)
Hartmann hatte versucht, ihnen nachzueifern, wenn sie ihn freundlich aufforderten, mit ihnen den Abend in einem Pub zu verbringen …
Hartmann hadde forsøkt å etterligne dem når de var så hyggelige å invitere ham til en pubkveld … (lit. 'invite him to a pub night')

(13) After matric, Aunt Cathérine invites him to spend a long holiday in France. (ABR1)
Nach der Matura lädt Tante Cathérine ihn ein, ausgiebig Ferien in Frankreich zu machen. (lit. 'lavishly make holidays in France')
Etter studenteksamen inviterer Tante Cathérine ham på en lang ferie i Frankrike. (lit. 'invite him to a long holiday in France')

(14) We've seen microscopes before, but not at such length; we can spend a lot of time with them before getting tired of them. (MA1)
Wir haben schon Mikroskope gesehen, aber noch nie so lange; wir können uns sehr lange mit ihnen beschäftigen, ohne uns zu langweilen. (lit. 'we can occupy ourselves a lot with them')
Vi har sett mikroskoper før, men ikke så lenge om gangen, og vi blir ikke lei av å holde på med dem. (lit. 'we are not tired of keeping on with them')

In the translations of (11) the overall structure is preserved, but there is no verb corresponding to the predicat of the ing-clause, presumably because it is felt to be redundant. The Norwegian translator does without a verb corresponding to spend in (12) and (13), while the German translator has a congruent translation in one case (12) and a non-literal paraphrase in the other (13), where the meaning is aptly captured by the adverb ausgiebig (connected with ausgeben 'give out'; cf. the core meaning of spend). Other non-literal paraphrases are found in the translations of (14).

4. Distribution in original and translated texts

As we have seen, both German and Norwegian translators are quite inventive in finding translations for structures with spend, but do they also overuse congruent translations? To examine this, let us compare the distribution of tilbringe in original texts in Norwegian and in texts translated from English; see Figure 3 and Table 3.
Figure 3 The distribution of English *spend* and Norwegian *tilbringe* ('spend') in original and translated fiction texts of the ENPC (30 texts of each type)

Table 3 The distribution of complementation patterns of Norwegian *tilbringe* ('spend') in original and translated fiction texts of the ENPC (30 texts of each type)

<table>
<thead>
<tr>
<th></th>
<th>Original</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>tilbringe</em> + NP_temp + ADV_place</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>ADV_accomp</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>ADV_manner</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>ADV_place + med + V-inf</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><em>tilbringe</em> + NP_temp + med/til + V-inf</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>22</strong></td>
<td><strong>53</strong></td>
</tr>
</tbody>
</table>

Figure 3 shows that *tilbringe* is more than twice as common in texts translated from English as in original Norwegian texts, a finding which agrees well with Gellerstam's observation on Swedish *tillbringa*. There is also a striking difference in the frequency of *spend* in original English texts and in texts translated from Norwegian. In both cases there is a clear translation effect. Norwegian *tilbringe* is stretched to accommodate English *spend*, leading to a more frequent use. On the other hand, the lack of a clear counterpart in the source language leads to a lower frequency of *spend* in English translations.

In Table 3, we see that the higher frequency for *tilbringe* in translations is associated with particular complementation patterns, notably those containing a specification in the form of an infinitive construction or an adverbal phrase denoting accompaniment or manner. In other words, by being able to compare not only original and translated texts in the same
language but also original and translated texts across languages, we can increase the delicacy of the study.

5. Possibilities and limitations

To conclude, the use of spend and its German and Norwegian translations illustrates how time may be construed differently in different languages: as consuming or passing time in English vs. as the duration of an event in German and Norwegian. At the same time, translators have a tendency to construe time in agreement with the source text. The differences are not absolute but reflect 'preferred ways of putting things', to borrow an expression from a paper by Graham Kennedy (1992).

In the study of 'preferred ways of putting things', there is no doubt that we need a corpus, whether we are concerned with lexis or syntax. In the case of a lexical study like the one reported here, the corpus material gives a far richer picture of cross-linguistic correspondences than a bilingual dictionary. More important, the design of the English-Norwegian Parallel Corpus and the new Oslo Multilingual Corpus makes it possible to ask new sorts of questions, and the comparison highlights both cross-linguistic differences and similarities and language-specific characteristics.

It is not necessary to choose between a translation corpus and a comparable corpus of matched original texts. Both can be accommodated within the same overall framework, and each can then be used to supplement and control the findings arrived at on the basis of the other. This is why the same corpus can be used both for contrastive analysis and translation studies. Building such a corpus is not unproblematic, however.

In the first place, the selection of texts is restricted by the types and quantity of texts that are translated. What is translated varies with the particular language pair, and with the direction of translation. For example, far more is translated from English into Norwegian than the other way around. The problem is compounded as the corpus is expanded to include more languages.

The compilation of the Oslo Multilingual Corpus will no doubt take some time, and it is uncertain whether it can be fully built up according to our design even for the triple English-Norwegian-German. But we are encouraged by what can be done with the material that has been compiled so far. Depending upon the topic of study, it will in any case often be necessary to go beyond the corpus. This is as true of contrastive studies as of language research in general.

Notes

1 For details on the ENPC, including publications, see the web page for the project: [http://www.hf.uio.no/iba/prosjekt/](http://www.hf.uio.no/iba/prosjekt/)

2 This is a joint project with the Department of Germanic Studies and the Section for Applied Linguistics at the University of Oslo. See the web page for this project: [http://www.hf.uio.no/german/sprik/index.html](http://www.hf.uio.no/german/sprik/index.html)

3 This figure was first drawn by Cathrine Fabricius-Hansen, University of Oslo. The figure has been simplified somewhat. All the lines should have bidirectional arrows, indicating that a comparison could be made in either direction.
4 In the examples given below, the original text is identified by the abbreviation given within parentheses. For more information on the texts, see the web page for the ENPC project (cf. note 1).

5 A related example in the Norwegian material contains the verb gjennomgå (lit. 'go through'): han hadde gjennomgått fireogtyve dårlige timer / he had spent a lousy twenty-four hours (SK1)

References


