Semiotics at Work: Technical Communication and Translation in a Multilingual Corporate Environment

Arne Larsson, Espoo
Magnus Merkel, Linköping

Abstract

In the paper an attempt is made to find a unifying approach to the study of the translator's praxis, assuming that translation is guided by certain, recognizable, semiotic processes. Computational, corpus-based methods intended to aid in the research of large text bases are introduced. Alignment of text segments from files in different languages contained in a corpus, where these text files are known to be mutual translations is described. Text encoding in order to allow comparison of the results of translation studies performed by different scholars is also demonstrated. One goal is to establish qualitative and quantitative variables, on the sentential as well as the textual level, which would permit generalizations about the concrete procedures performed by professional translators in authentic work situations, e.g. in multi-lingual corporate environments.

Empirical, descriptive methods

Today large amounts of texts sit on the hard disks of computers in companies and organizations, but exact, empirical, detailed, descriptive information telling us what translators actually do when they translate is not abundant. A natural solution to this dilemma is the collection of evidence from existing texts included in aligned bilingual corpora. The purpose of text alignment is to establish *version complexes*[1], i.e. sets of corresponding elements in the source and target texts.

The Alignment Tool (LinAlign)

At the Department of Computer Science at Linköping university an Alignment program was developed in 1993[2]. The program (called LinAlign) creates translation memories of a source and target text, that is, it links a sentence in the original with a corresponding sentence in the target document. There are different techniques to accomplish the alignment of segments. Most notable has been the statistical approach,

[1]This concept was introduced by Wollin (1981), followed by Platzack (1983).
[2]The major part of the programming has been done by Bernt Nilsson.

Proceedings of NODALIDA 1993, pages 155-163

which the LinAlign tool also adheres to. The best-known statistical algorithm is the one developed by Gale & Church (1991). LinAlign uses a much simpler method than Gale's & Church's program, but in a test described below its performance is equal to theirs, if not better.

The algorithm is based on three assumptions of the source and target texts.

1. The source and target texts are similarly ordered.
2. If two sentences in one text are combined to one sentence in the other text, it is always adjacent sentences that have been joined.
3. The alignment is based on paragraph and sentence lengths (number of characters).

Apart from 1–1 relations, LinAlign also handles 1–2 and 2–1 relations (one source sentence – two target sentences, two source sentences – one target sentence).

Below is a sample of the output from the LinAlign program:

```
f1:21.1 Specify the amount of time before you receive
  messages about printer problems.
f2:21.1 Ange efter hur lång tid ett meddelande rörande
  skrivarproblem ska visas.

f1:22.1 Select the default printer.
f2:22.1 Välj standardskrivaren.

f1:23.1 The following sections explain how to perform
  each of these tasks.
f2:23.1 Följande avsnitt förklarar hur du vidtar dessa
  åtgärder.
```

The code before each segment gives information about each document and its respective paragraph and sentence ordering. In the example above `f1:21.1` indicates that the segment is taken from the target language (f1), the 21st paragraph (`:21`) and the first sentence of that paragraph (`.1`)

To illustrate the way the algorithm works when there are an unequal number of sentences in the corresponding paragraphs, let us consider the example below. The first part of the example is a help text that is described if LinAlign is run in Debugging mode.

```
searching for sentences to join...
f1:444.1 & f2:444.1 + f2:444.2 = 6
f1:444.2 & f2:444.2 + f2:444.3 = 38
 -> f2:444.1 + f2:444.2
f1:444.1 To cancel a selection, you can use mouse or
  keyboard techniques, or the Select Files command.
f2:444.1 Du använder musen eller tangentbordet för att
  avbryta markeringar.
f2:444.2 Du kan också använda kommandot Markera filer.
```

```
f1:444.2 You can cancel one selected file or a group of
   selected files.
f2:444.3 Du kan avbryta markeringen av såväl en enskild
   fil som hela grupper.
```

The example describes paragraph 444 of a particular translation text, showing both the source and target texts. In the English text there are two sentences, but in the Swedish there are three target sentences. The help text above the translation pairs helps us to understand the way the algorithm works. The program has to determine whether it is the first and the second Swedish sentence that should be joined as the translation of the first English sentence, or if the second and third Swedish sentences should be taken as the translation of the second English sentence. Based on the number of characters in the sentences the different options are compared and the one with the closest match is selected. In the example above LinAlign values the cost of regarding the first and second Swedish as the translation of the first English sentence as the cheapest alternative (i.e. the shortest "sentence distance") and therefore these two sentences are joined in a 1–2 relation.

The sentence distance measure is computed by the following formula:

sentence distance = P(l1 + l2 – olfactor)

where P is the proportional measurement of the two texts,
l1 is the length of sentence 1 measured in characters,
l2 is the length of sentence 2 measured in characters,
and *olfactor* is the overlap factor that is used to capture the fact that two sentences joined together becomes longer than a corresponding single sentence (default value is 15)

Alignment test

A test of the LinAlign tool when run on a manually translated text, showed that out of 624 sentences, it failed on only four sentences. The test was done on an English-Swedish corpus consisting of a chapter from a manual for a computer program. Church & Gale (1991) reported that their tool when tested on a similarly sized English-French material failed on 22 sentences out of 621. It is of course impossible to draw any conclusions on the quality of the tools from such small and different test materials.

However, one interesting factor found when we analyzed the source text with a tool for measuring recurrence was that 23 sentence types were repeated between 2 to 19 times in the text. (The Recurrence Analyzer is developed at the same department as LinAlign and results from analyses of technical documentation can be found in Merkel (1992).) A recurrence test on the target text revealed that out of these 23 sentence types 20 had

been translated with consistent translations. The three sentence types (all with the frequency 2) that had different translations could have had consistent translations, without impeding readability. In the following example, the three source sentences are shown together with the alternative target sentences.

Recurrent source sentences with different translations:

1. The options available in the dialog box below may vary, depending on the network you are using.
1a. Vilka alternativ som finns i dialogrutan nedan beror på vilket nätverk du använder.
1b. Tillgängliga alternativ i dialogrutan beror på vilket nätverk du använder.
2. Select the port you want to assign the printer to.
2a. Markera den port du har anslutit skrivaren till.
2b. Välj vilken port du vill ansluta skrivaren till.
3. Select the port you want to use.
3a. Välj den port du vill använda.
3b. Markera den port du vill använda.

In other words, there was nothing special in the context that demanded variation. It was just what the translator had chosen at a certain point in the translation process, unaware of the fact that the exact sentence occurred at a different text segment.

It would be interesting to take this analysis methodology one step further by analyzing the variation in the target text on a much larger scale. For example, how widespread are these phenomena in different types of text? Furthermore, to what extent can segments with explicit cohesive markers (Halliday & Hasan 1976, Källgren 1979) be reused in different local contexts in, for example, technical documentation and legal treaties? And will consistent use of memory-based translation make certain translations "worse" in the aspects of text binding? These are questions that can only be answered if huge masses of translated texts are aligned and analyzed in detail.

Language independence

Two text fragments from a technical manual in Finnish and Swedish were also aligned using the LinAlign tool, thus demonstrating the language independence of this statistical method. The actual aligned segments are similar to the English-Swedish ones above. For reasons of space, they will not be reproduced here.

Other alignment methods

Morphology-based alignment is used in a computer-based workstation for the lexicographer (Picchi et al. 1992). Aligned parse trees from a

dependency grammar parser were proposed for machine translation purposes in Sadler (1989). A method for alignment of words as well as sentences was presented in Kay & Röscheisen (1993).

Text encoding

When the alignment of the Finnish and Swedish texts was completed, the text fragments were marked up according to the function of the *primary sentential constituents*. The following abbreviations were used (for additional details and examples, see Platzack 1983: 249 ff, Larsson 1993):

FV	Inflected verb (finit verb)	**OO**	Direct object
IA	Content adverbial (innehållsadv)	**SP**	Predicate NP (subjektiv predikatsfyllnad)
IO	Indirect object	**SS**	Subject
IV	Infinitives (infinit verb)	**SA**	Added clause

Operations performed by the translator

Eight different types of operations, which the translator may apply were identified by Wollin (1981), viz. addition, convergence, deletion, divergence, functional modification, mixing, structural identity, and transposition (Platzack 1983: 256 ff.). Four of these operations were used in the text fragments (1) and (2) below:

(1) *Structural identity*

Ennen asennustyön aloittamista	**IA**	**IA**	Innan installationsarbetet påbörjas
tehdään	**FV**	**FV**	utarbetas
asennussuunnitelma	**OO**	**OO**	installationsplan
noudattaen tässä kirjassa ja asennettavien laitteiden käyttökirjoissa annettuja ohjeita.	**IA**	**IA**	med ledning av föreliggande handbok och anvisningarna i handböckerna för den teleutrustning som skall installeras.

(2) *Addition (SA), functional modification and transposition (OO=>SS)*

Apuna suunnitelman teossa	**IA**	**SS**	Bifogade planeringsblanketter
voidaan	**FV**	**FV**	kan
käyttää	**IV**	**IV**	användas
liitteenä olevia suunnitelmalomakkeita.	**OO**	**IA**	som hjälp vid utarbetande av planen
		SA	(jfr bilagorna).

An important distinction is the one between obligatory versus optional operations. Here, operations that are absolutely necessary for the formation of *grammatical structures* in the target language are called

obligatory (e.g. insertion of articles and prepositions in the Swedish, which are non existing in the Finnish; 'correct' order between primary constituents and word order within constituents), otherwise they are called *optional.*

Professional translators' performance

Focusing on the syntactic level alone will render a somewhat shallow picture of the complicated processes of translator performance.[1] Therefore, certain *textual variables* were used to supplement the *sentential variables* (i.e., the operations on version complexes outlined above) in order to achieve enhanced explanatory power.

The following textual variables were chosen, because contrastive studies of Finnish and Japanese, vs. Anglo-American writers indicate specific problematic differences (Kohl et al. 1993, Mauranen 1992, Ventola 1992) with respect to 1) reference items, reference chains and text coherence; 2) theme and rheme, thematic progression, choice of connectors; 3) reflexive expressions, 'text about text'; 4) signals of propositional relationships: making the point, stating opinions; 5) types of strategic moves; 6) culture as discourse. Similar differences can be expected between other language pairs as well.

The choice was also guided by the existence of methods for the successful study of the variables involved (Källgren et al. 1977, Källgren 1979, Sigurd 1987).

Linguistic preferences govern translation

As can be seen from the semantic network (table 1), no less than six of the referents are implicit in the Finnish source text (marked *if* in P5, P6, P7, P10, P11 and P13), versus two implicit items in the Swedish target (marked *is* in P10 and P14, one of which is the predicate *är* 'is, are').

The Finnish tendency toward implicitness together with late introduction of referents (Mauranen 1992: 109) explain fairly well, why the translator has made use of the structure changing operations *addition, functional modification* and *transposition* in (2). The reason for these manipulations is that the target language community requires more explicit referents and prefers an earlier introduction of these.

---

[1]Platzack (1983:266) stressed the need for obtaining additional information concerning the mutual influences between various properties of the languages involved in the translation, and the frequency of application of different operations.

Today, we have evidence to the effect that the appropriate use of textual connectors will make a text easier to read, more logical, more convincing, and add to the writer's credibility (Mauranen 1992: 187). Computer tractable, well structured thesauri will facilitate decisions about *what* to actually add in order to achieve enhanced *connectivity*.

Translation as choice and change

Text-linguistic methods will not only provide explanations of the translators use of certain operations, but also facilitate a systematic approach to active text planning and organization during the creative phase of writing or translation.

The underlying *elementary propositions* for the above technical manual fragments can be presented as a *semantic network*[1] where paths representing various, *logically possible* texts involving the factoms can be drawn.

TABLE 1: Semantic network for the technical manual text fragments (1) and (2).

		1	2	3	4	5	6	7	8	9	10	11	12	13
P1	Man börjar installations-arbetet (IA)		fs ⇳ ⇳	①										
P2	Man gör en installations-plan (IP)	●	⇳ ⇳ ⇳	fs ⇳ ⇳										
P3	P2 föregår P1	fs	(fs)	(fs)①										
P5	Handboken (HB)*if* ger ledning		●		fs		(f) ⇳ ⇳							
P6	HB har anvis-ningar (A)*if*					fs	⇳ ⇳	①						
P7	HB/A gäller (tele)*if*utrust-ning (TU)						fs		①					
P8	Man installerar TU							fs		①				
P9	P2 föregår P8								fs	①				
P10	(HB)*is*/*if* har planerings-blanketter (PB)				●					s				(s) ⇳⇳
P11	(PB)*if* ger hjälp					①				f	s		(fs)	⇳
P12	Hjälpen gäller P2						①				f	s	⇳⇳	⇳⇳
P13	(HB)*if* har bilagor										①	f	⇳ ⇳s	⇳⇳
P14	Bilagorna (är)*is* PB										①		f	(s)

f	Finnish text fragment	*if*	implicit in the Finnish text	
s	Swedish text fragment	*is*	implicit in the Swedish text	
⇳	Reader (translator) needs to backtrack			
①	Suggested paths through the network	●	The point we want to make	

[1]See Källgren (1979), Larsson (1993), Sigurd (1977, 1987), Wintraecken (1990) for additional details.

Building new texts using paths through a network

In planning and creation, paths may be chosen, which are considered optimal for the communicative task at hand. Unnecessary propositions may be left out. Below, the symbol ① represents suggested paths, and the black symbol ❶ the points we want to make. Two new versions in Swedish and two in English of the technical manual fragments are presented with the propositions reordered to avoid backtracking, and explicit referents inserted:

First draft (Swedish):
(P2)Gör en installationsplan (P5)med ledning av anvisningarna, (P3, P1)innan du börjar arbeta med installationen. (P10)Planeringsblanketterna (P11)hjälper dig (P12)att göra upp planen. (P6)Anvisningarna finns i handboken (P7)för teleutrustningen som (P9)skall (P8)installeras. (P13, P14)Se bilagorna.

A more official version may be needed (changes relative to the first draft are marked using underlining.):
(P2)Gör alltid upp en installationsplan (P5)med ledning av gällande anvisningar, (P3, P1)innan arbetet med installationen börjar. (P10)Nokia har tagit fram planeringsblanketter (P11)som hjälp_ (P12)vid upprättandet av installationsplanen. (P6)Anvisningarna finns i handboken (P7)för teleutrustningen som (P9)skall (P8)installeras. (P13, P14)Se bilaga 14-21.

We might even dare an attempt at an English version:
(P2)Make a plan of the installation (P5)according to the instructions, (P3, P1)before you start working on the installation. (P10)Forms (P11)help you (P12)make the plan. (P6)Instructions are in the manual (P7)for the telecommunications equipment (P9)under (P8)installation. (P13, P14)Please, refer to the Appendix.

Which we might want to edit later:
(P2)We strongly recommend, that you create a plan of the facilities (P5)according to the instructions given by the manufacturer, (P3, P1)before you start working on the setup of the equipment. (P10)Forms (P11), which help you (P12)with the installation planning, (P13, P14)are provided in Appendix 14-21. (P6)Instructions are in the manual (P7)for the telecommunications equipment (P9)to be (P8)installed.

Conclusions

Aligned bilingual corpora can tell exactly *what* the translator does in terms of concrete syntactic operations. Text-linguistic methods explain *why* these operations were used. Moreover, operational and text-linguistic approaches facilitate systematic planning and organization of texts in a multi-lingual corporate environment. As a result, these methods form a useful complement to the goal oriented principles of "translatorisches Handeln" and "skopos" (Holz-Mänttäri 1982, Vermeer 1989). Future work will be focused on 1) automatic text alignment, 2) automatic tagging/parsing of aligned texts, 3) application of international standards, e.g. SGML, 4) tools for translators and writers.

References

Chesterman, A. 1989. *Readings in Translation Theory.* Loimaa.

Gale W. A. and K. W. Church. 1991. *A program for aligning sentences in bilingual corpora.* In: PROCEEDINGS FROM ACL–91, pp 177–184, Berkeley.

Halliday M. A. K. and Ruqaiya Hasan. 1976. *Cohesion in English.* Longman, London.

Holz-Mänttäri, J. 1982. *Vom translatorischen Handeln.* Turku.

Kay, M. and M. Röscheisen. 1993. *Text-Translation Alignment.* COMPUTATIONAL LINGUISTICS, Volume 19, Number 1, pp. 121–142, MIT Press.

Kohl J., R. Barclay, Th. Pinelli, M. Keene, J. Kennedy. 1993. *The Impact of Language and Culture on Technical Communication in Japan.* TECHNICAL COMMUNICATION, First Quarter 1993.

Källgren, G. 1979. *Innehåll i text.* ORD OCH STIL 11. Studentlitteratur, Lund.

Källgren, G., B. Sigurd and M. Westman. 1977. *Experiment med text.* Akademilitteratur, Stockholm.

Larsson, A. 1993. *Semiotic Aspects of Technical Translation and Communication.* In FACKSPRÅK OCH ÖVERSÄTTNINGSTEORI. VAKKI-SEMINARIUM XIII, Vasa universitet, Institutionen för språk, Vasa (forthcoming).

Mauranen, A. 1992. *Cultural differences in academic rhetoric. A textlinguistic study.* Dissertation, Birmingham.

Merkel, Magnus. 1992. *Recurrent Patterns in Technical Documentation.* RESEARCH REPORT, DEPT. OF COMPUTER AND INFORMATION SCIENCE, Linköping University.

Platzack, Ch. 1983. *Sex översättningar till svenska av Lewis Carolls "Alice in Wonderland".* In Engwall, G. and af Geijerstam, R. (eds), *Från språk till språk,* Studentlitteratur, Lund.

Picchi, E., C. Peters and E. Marinal. 1992. *The Pisa Lexicographic Workstation: The Bilingual Components.* In Tommola, H., K. Varantola, T. Salmi-Tolonen and J. Schopp (eds), *EURALEX '92 PROCEEDINGS I–II.* Papers submitted to the 5th EURALEX International congress on Lexicography. Tampere.

Sadler, V. 1989. *Working with Analogical Semantics: Disambiguation Techniques in DLT.* Dordrecht.

Sigurd, B. 1987. *High resolution referent grammar (HRG) in analysis and generation of text.* In WORKING PAPERS, Lund University, Dept. of Linguistics, Lund.

Ventola, E. 1992. *Reference and Theme and the Interplay of these two textual systems.* In Nyysönen, H. and L. Kuure (eds), *Acquisition of language – acquisition of culture.* AFINLA YEARBOOK. Jyväskylä.

Vermeer, H. J. 1989. *Skopos and commission in translational action.* In Chesterman, A., (ed.).

Wintraecken, J. 1990. *The NIAM information analysis method: theory and practice.* Dordrecht.

Wollin, L. 1981. *Svensk latinöversättning. I. Processen,* Dissertation, Lund.