

COMPUTER TRANSLATION : A DIFFERENT APPROACH

In recent years a number of nations have spent literally millions of dollars on automatic machine translation projects. Now that almost all of these projects have been abandoned it may seem gratuitous to describe them as naive and illconsidered but, at least in retrospect, it is difficult not to feel that they reflect a failure to appreciate the difficulty of the task, that they were undertaken partly at least simply because the computer was available, and that they represent yet another typical misuse of its very considerable capacities. But that dream is now almost dead and we are perhaps ready for a more realistic approach to this problem. With this in mind we here offer a very general description of a program developed in 1973⁽¹⁾ and known as PROTRAN for *PRO*visional *TRAN*slation. It attempts much less but has the considerable merit of actually attaining its goal which is simply that of assisting the translator in his very difficult task. Our work was done with quite primitive equipment but obviously can be carried out even more easily using more up to date hardware.

This program was first used to translate Ib Ostenfeld, *Søren Kierkegaards Psykologi* from Danish into English. The original, running to only 79 pages, was first published by Rhodos in Copenhagen in 1972 and the translation is expected to appear as the first volume of a new

Kierkegaard monograph series. The present report employs examples from this application throughout. Figure 1 shows five sample lines from the original text and Figure 2 the computer version of the same. Figure 3 gives a short excerpt from the dictionary while Figure 4 provides an example of the final computer printout. Figure 5 shows the finished translation of these same lines.

The first step in this very simple method was to keypunch the text as it appears in the original but, as shown in Figure 1, with line numbers and without carry over hyphenation. Some of the conventions employed, for example, for some punctuation and for special Danish vowels, are quite arbitrary and hence of little interest but some at least can be deduced by comparing Figures 1 and 2. It is however perhaps worth noting that capitalization is indicated at the end of the word with an asterik (or other such character) to simplify the problem of alphabetization.

The second step was to have the computer produce an alphabetized list of all word-types in the text together with their frequencies. This was output on hard copy for checking and on punch cards in preparation for the next step. The words were punched on the card left-justified beginning in column 1 and the frequencies right-justified beginning in column 27 as shown in the left part of Figure 3.

The next step was to create a Danish-to-English dictionary for the computer. This involved identifying and ordering the various probable English translations of these words and punching them on the appropriate card beginning in columns 30, 45, and 60, where necessary

with more than one word in a field. In making these decisions we kept in mind the character and style of the work and the fact that it was a psychiatric study. Thus, for example, we decided that "af" was more likely to be used in the sense of "of" than of "out of", that "saa" would usually be used in the sense of "so" rather than "saw", "there", or "that", etc., and ordered and punched these words accordingly. Of course, the order or, for that matter, the translations themselves could easily be changed for, say, a text in geography where other senses of these words are likely to be most common. For a small project such as the one we are describing this is perhaps most easily done by repunching the relevant cards but for work is part of a continuing project such changes could perhaps be done as part of the program.

The fourth step was to have the computer produce its final printout which spaces out the Danish words and prints their alternative translations directly under each as shown in Figure 4. Finally, of course, the last step is to go through the printout selecting or, more probably, creating the best possible translation. Sometimes this involves rejecting all the suggested words but there are relatively few cases in which the meaning of the original cannot be immediately grasped simply by scanning these substitutions.

It is obvious that the computer printout itself yields only a literal and completely wooden translation and that our method does not cope with such difficult problems as idioms, figures of speech, etc. But it is equally obvious that, accepted for what it is, it has a number of important advantages. The reader will no doubt see some of

these for himself but we note others which may not be so obvious.

Even the user who has only a rudimentary knowledge of the original language can use this method to gain a first-hand impression of the contents of a work and, if it proves interesting, to make a rough or provisional translation. Certainly it will enable such a person to gain a good understanding of the text without the necessity of constant and repeated recourse to the dictionary. Despite this, it has actually proven an excellent way of learning vocabulary and, as Figure 5 shows, can help such a person to produce an entirely acceptable finished translation.

But it can also be useful even for those who are relatively familiar with both the original and target languages. By presenting all probable translations of each word at one and the same time it enables such a person to see all the possibilities and discover the intended sense much more quickly and accurately than he could without such assistance. Further, it can help him to maintain consistency in his finished work. Of course, the same word should not always be translated in the same way but the printout reminds the translator of his own "specified" or "standard" translations and thus helps him to see whether one of these or some other word is most appropriate. Indeed, given this method, it is a simple matter to index any number of words in the original and thus be able to check and compare all instances of their translation. Thus the method is a tool which can bring a new level of consistency and self-consciousness to the art of translation.

Finally, this method has one other advantage which can be indicated yet more simply. All the computer operations in the original project described above were completed on a total budget of less than fifty dollars.

Alastair McKinnon
McGill University

Per Bondesen
Århus Universitet

January, 1977

NOTES

- (1) This work was done at the University of Edinburgh where Bondesen was working in the Department of Linguistics and McKinnon was an Honorary Fellow in the Institute for Advanced Studies in the Humanities. The actual programming was done by Neil Hamilton-Smith of the Edinburgh Regional Computing Centre.

Disse faa Bemærkninger om Kierkegaards Stil skulde med tilstrækkelige Argumenter have overbevist om, at Forfatterskabet ikke paa noget Tidspunkt røber Tegn paa, at han producerede i nogen sygelig Sindsforfatning, end mindre at han producerede som han gjorde *paa Grund af* en saadan Forfatning. Vi maa redde Kierkegaards Forfat-

Fig. 1. Sample Lines (ll. 20-25, p. 26) of Original Danish Text

20 DISSE* FAA BEMÆRKNINGER* OM KIERKEGAARDS* STIL* SKULDE
 21MED TILSTRÆKKElige ARGUMENTER* HAVE OVERBEVIST OM, AT FORFATTERSKABET*
 22IKKE PAA NOGET TIDSPUNKT* RØBER TEGN* PAA, AT
 23HAN PRODUCEREDE I NOGEN SYGELIG SINDSFORFATNING*, END
 24MINDRE AT HAN PRODUCEREDE SOM HAN GJORDE PAA GRUND* AF
 25EN SAADAN FORFATNING*. VI* MAA REDDE KIERKEGAARDS* FORFATTERY*

FIG. 2 COMPUTER VERSION OF SAMPLE LINES

DIGTEREN*	1	*THE POET	
DIGTERISKE	3	POETICAL	
DIGTERS*	1	THE POET'S	
DIGTERSIND*	1	*THE MIND OF	THE POET
DIGTING*	2	POETRY	
DILEMMA*	1	DILEMMA	
DIREKTE	1	DIRECT	
DIRIGERE	1	CONDUCT	
DIRIGERER	1	CONDUCT	
DISCRIMEN	1	DISCRIMINATION	
DISKONTINUERLIG	1	DISCONTINUOUS	
DISKONTINUERLIG	1	DISCONTINUOUSLY	
DISKONTINUITET*	1	DISCONTINUITY	
DISKUSSION*	3	DISCUSSION	
DISSE	20	THESE	

FIG. 3 EXCERPT FROM DANISH-TO-ENGLISH DICTIONARY

20	DISSE*	FAA	BEM_RKNINGER*	OM	KIERKEGAARDS*	STIL*	SKULDE			
20	THESE	GET	THE REMARKS	ABOUT OF		STYLE	SHOULD			
20		FEW		ON IN						
20				BY IF						
21	MED	TILSTR_KFELIGE	ARGUMENTER*	HAVE	OVERBEVIST	OM	, AT	FORFATTERSKABET*		
21	WITH BY	SUFFICIENT	ARGUMENTS	HAVE	CONVINCED	ABOUT OF	, TO	AUTHORSHIP		
21	IN TOO	ENOUGH				ON IN	, THAT			
21						BY IF	,			
21										
22	IKKE	PAA	NOGET	TIDSPUNKT*	RIBER	TEGN*	PAA	, AT		
22	NOT	ON UPON	SOME	TIME	REVEALS	SIGN	ON UPON	, TO		
22	NO	IN AT	A LITTLE	MOMENT			IN AT	, THAT		
22		IN	ANY				IN	,		
22										
23	HAN	PRODUCENEVE	I	NOGEN	SYGELIG	SINDSFORFATNING*	, END			
23	HE	PRODUCED	IN	SOME	SICKLY	STATE	, THAN			
23			ON	ANY	MORBID	OF MIND	,			
23			AT	SOMEONE	UNHEALTHY		,			
23										
24	SMINDRE	AT	HAN	PRODUCEREDE	SOM	HAN	GJORDE	IPAA	GRUND*	AF
24	SMALLER	TO	HE	PRODUCED	WHO	HE	DID	ION UPON	REASON	OF, BY, FROM, FOR
24		THAT			WHOM			IN AT		WITH
24					AS			IN		OUT OF, OFF
24										
25	EN	SAADAN	FORFATNING*	. VI*	HAA	REODE	KIERKEGAARDS*	FORFATIERY*		
25	A	SO	CONDITION	. ME	MAY	SAVE		REPUTATION AS A		
25	AN	SUCH	CONSTITUTION	.	MUST	RESCUE		WRITER		
25	ONE	IN THIS WAY		.	HAVE TO					
25										

Fig. 4. Final Printout of Sample Lines (photo-reproduced)

20	DISSE*	FAA	BEM, RKNINGER*	OM		KIERKEGAARDS*	STIL*	SKULDE		
20	THESE	GET	THE REMARKS	ABOUT OF			STYLE	SHOULD		
20		FEW		ON IN						
20				BY IF						
21	MED	TILSTR, KHELIGE	ARGUMENTER*	HAVE	OVERBEVIST	OM	, AT	FORFATTERSKABET*		
21	"ITH BY	SUFFICIENT	ARGUMENTS	HAVE	CONVINCED	ABOUT OF	, TO	AUTHORSHIP		
21	IN TOO	ENOUGH				ON IN	, THAT			
21						BY IF	,			
21										
22	IKKE	PAA	NOGET	TIDSPUNKT*	RIBER	TEGN*	PAA	, AT		
22	NOT	ON UPON	SOME	TIME	REVEALS	SIGN	ON UPON	, TO		
22	NO	IN AT	A LITTLE	MOMENT			IN AT	, THAT		
22		IN	ANY				IN	,		
22										
23	HAN	PRODUCEREDE	I	NOGEN	SYGELIG	SINDSFORFATNING*	, END			
23	HE	PRODUCED	IN	SOME	SICKLY	STATE	, THAN			
23			ON	ANY	MORBID	OF MIND	,			
23			AT	SOMEONE	UNHEALTHY		,			
23										
24	MINDRE	AT	HAN	PRODUCEREDE	SOM	HAN	GJORDE	IPAA	GRUND*	AF
24	SMALLER	TO	HE	PRODUCED	WHO	HE	DID	ION UPON	REASON	OF, BY, FROM, FOR
24		THAT			WHOM			IIN AT		WITH
24					AS			IIN		OUT OF, OFF
24										
25	EN	SAADAN	FORFATNING*	.	VI*	MAA	REDDE	KIERKEGAARDS*	FORFATTERY*	
25	A	SO	CONDITION	.	WE	MAY	SAVE		REPUTATION AS A	
25	AN	SUCH	CONSTITUTION	.		MUST	RESCUE		WRITER	
25	GNE	IN THIS WAY		.		HAVE TO				
25										

Fig. 4. Final Printout of Sample Lines

The arguments contained in these few remarks on Kierkegaard's style should have convinced the reader that his literary production shows no signs that he wrote from an unhealthy state of mind and, even less, that he wrote because of such a state. We must free Kierkegaard's reputation as a writer ...

Fig. 5. Finished Translation of Sample Lines.