

STUDIES IN MECHANICAL TRANSLATION

No. 6

REPORT ON RESEARCH RESULTS IN MECHANICAL TRANSLATION IN CONNECTION WITH  
THE ROCKEFELLER GRANT FOR THE SUMMER QUARTER 1952

by

ERWIN REIFLER

Department of Far Eastern and Slavic Languages and Literature

University of Washington

INDEX

1. The Pre-Editorial Work Necessary In German
2. Human Pre-Editor Versus Mechanical Pre-Editor
3. Extemporized Substantive Composita
4. The Mechanical Determination of the Constituents of Extemporized Substantive Composita.
5. A Joint Far Eastern and Electrical Engineering Project.
6. The Development of MT In Small Scale Models.
7. Death of the German Pre-Editor.

August 30, 1952

To: Professor George E. Taylor

From: Erwin Reifler

Re: Report on Research Results in MT in connection with the Rockefeller  
Grant for the Summer Quarter 1952.

Dear George:

The Rockefeller grant for research in MT during the summer quarter 1952 of which I have been the recipient was, as you know, given for the study of THE MINIMUM GRAPHIC MODIFICATION OF SOURCE TEXTS NECESSARY FOR THE MECHANICAL RESHUFFLING OF WORD ORDER AND THE MECHANICAL DETERMINATION OF INCIDENT NON-GRAMMATICAL MEANING and for the study of ADJUSTED MODEL TARGET LANGUAGES IN MT. This research has now already produced some significant results.

1. The Pre-Editorial Work Necessary in German

I had previously worked out two tentative systems of a universal supplementary signalization of source texts, one using capitalization, the other diacritic marks (cf. my STUDIES IN MT No.3, pp.9-15). For the second of the two research subjects it was necessary to have a number of translations into some other languages of professor Stuart C.Dodd's paper on "Model English." Four such translations have been made to date, namely into Russian, Chinese, Japanese and German.

Being very familiar with German, and since this language is comparatively rich in the formal indication of grammatical meaning and thus graphio-grammatically more distinctive, I naturally began with German. The first question to be answered was what modification of my universal system of

supplementary signalization was necessary for this language. This depended on the degrees of graphio-semantic explicitness characteristic of German. It became soon clear that this language in comparison with, for example, Chinese required much less pre-editorial work.

## 2. Human Pre-Editor Versus Mechanical Pre-Editor

After the minimum pre-editorial work necessary for German had been established, the next question to be settled was which portions of this pre-editorial work could be mechanized. Drs. Bar-Hillel, Bull, and Oswald had at the MT conference in June stressed that pre-editing could be almost entirely eliminated and Dr. Oswald had actually said, "A recognition of syntactic connection can be built into the 'memory' of machines of the high speed computer type" (WORD-BY-WORD TRANSLATIONS, p.3). I was skeptical about this and, though unable to deny the possibility, I pointed out that if it were possible for some languages and some types of relevant grammatical information, it may not be possible for others; and, even where it was possible, the engineering required may turn out to be so complex, time-consuming, and expensive that a human pre-editor may be considered preferable. (STUDIES IN MT, No. 3, pp. 10-12)

## 3. Extemporized Substantive Composita

But at least for one aspect of German even Dr. Oswald did not see any other solution than the cooperation of a German pre-editor, namely in the case of German extemporized substantive composita (MICROSEMANTICS, p.6). I, therefore, concentrated on this problem in order to find out whether or not a mechanization was possible here. If it was not possible, if a pre-editor had to be retained for the indication of the "seam" of these unpredictable substantive compositions, then there was the question whether the

cooperation of a human agent in the determination of the "seam" of substantive composita could be used for the simultaneous supplementary signalization of other relevant grammatical meanings by the use of different "seam" signals (STUDIES IN MT, No. 3, p. 11d).

#### 4. The Mechanical Determination of the Constituents of Extemporized Substantive Composita

As you will see from the accompanying papers, I have succeeded in solving the problem of the mechanical identification of the constituents of all substantive composita which are not "memorized" by the "capital memory" (mechanical dictionary into which all forms with a capital initial are fed automatically) but whose constituents occur there, and in thus eliminating the pre-editor from this section of German MT. On August 22 I worked out the mechanical determination of all German substantive composita which do not contain an "X"-factor, and on August 27 I completed working out the mechanical matching steps for both left-to-right and right-to left matching of the constituents of substantive composita with an "X"-factor. The recognition that there are only four possible matching situations for all types of substantive composita with an "X"-factor together with the planned utilisation of the distinction given to source text forms by the occurrence or absence of space (the space interval which separates free forms) and of the mechanical action of the space bar, and in conjunction with a "capital memory" whose contents are especially planned, make the mechanization of this aspect of German MT possible. Moreover, in drawing up the principals for the planning of the "capital memory", the peculiarities of source-target semantics have been considered, so that a large chunk of the problem of incident non-grammatical meaning of substantives will be taken care of. Fortunately, my solution is

applicable to all languages which have the same problem. Examples from English, Russian, and French will be given below.

##### 5. A Joint Far Eastern and Electrical Engineering Project

In several conferences with Professor W. Ryland Hill of the Electrical Engineering Department of our University the feasibility from the engineering point of view was established. As a matter of fact, the Electrical Engineering Department intends to submit to the Rockefeller Foundation a program of a project for the construction of a model of this mechanization of a partial problem of MT. An important companion to this project is a research concerning the German general vocabulary in order to determine the contents of the "capital memory", what substantive constituents cannot occur as last constituents of composed proper names, all possible "X"-factors, all "X"-factor forms, all ascertainable substantive composita, whose target meaning can not be inferred from the meaning of the target equivalents of their constituents (cf. German "Hochzeit"), etc. The details are described in the accompanying paper. This German language research, is not only important for the engineering project mentioned above, but for German MT as a whole. It will, moreover, supply the experience for similar ventures in other Languages.

As a matter of fact, I should like to submit a simultaneous general vocabulary project on the lines described above for German, English, and Russian. For one of the advantages of mechanizing the determination of the constituents of substantive composita is the fact that as a result the number of substantive composite which have to be entered into the "capital memory" can be very much reduced. Composita like English "seashore" (substantive plus substantive), "highland" (adjective plus substantive),

"afterthought" (adverb plus substantive) and "cutthroat" (verb plus substantive) need not be "memorized" at all because not only will their constituents all occur in the English MT substantive memory, but their target meanings can be inferred from the meaning of the target equivalents of their constituents. The same holds true for Russian composita like "nebosvod" (substantive plus substantive), literally "sky vault", meaning "firmament"; "novoselenets" (adjective plus substantive), literally and freely "new settler"; "posleslovie" (adverb plus substantive), literally "afterword", meaning "epilogue," and "bezoblačnost'" (preposition plus substantive), literally "without cloudiness", meaning "cloudlessness". A Russian example demonstrating the most complicated situation of a substantive compositum with an "X"-factor problem permitting two mechanical dissections, but, nevertheless, yielding a unique and simultaneously correct answer after the third matching step, will be given in the accompanying paper.

No doubt, this solution of the mechanical determination of the constituents of substantive composita is also applicable to the French language. But in this language mostly only those substantive composita permit a formal and source-target semantic determination through a mechanical synthesis of their constituents which--sometimes with minor idio-orthographical and idio-morphological changes--occur also in German, Russian, and English (for example "tele/gramme, tele/graphs, tele/graphic, tele/scope, micro/cosma, micro/cosme, micro/graphie, micro/metre, micro/scope," etc.). Thus the engineering feasibility of my solution can be demonstrated also for French by the model suggested for the other three languages.

#### 6. The Development of MT in Small Scale Models

A number of different glotto-mechanical problems have to be solved before MT can materialize. At the conference the view was expressed that experiment-

tation with existing machines may yield answers to certain MT problems. I fully agree with Dr.A. Don Booth that "the machine does what you can do with a pencil and paper" (p. 250 of the transcript). The question was also raised whether we should begin with the existing machines or, from the start, think in terms of new machines. I believe that MT research should not from the very outset be sidetracked or narrowed down by what machines built for different aims can do. I suggest that the ultimately most satisfactory and most economical way of developing MT will be one in which for every solution of a distinctive problem--a solution, however, in which all other distinctive problems have been fully considered--a model is built on a limited scale in order to prove the engineering feasibility and in order to discover engineering shortcuts. Every such model has then to be reviewed in the light of all other models designed for the solution of the remaining distinctive problems. Once all problems have been solved and all solutions tested in such models, the model of a complete MT mechanism should be built on a small scale demonstrating the feasibility of MT for all possible types of graphio-semantic-mechanical problems that either permit a mechanical solution or whose mechanical solution is considered worth-while.

Since my solution for the mechanical determination of the constituents of German substantive composita is equally applicable to English and Russian substantive composita, it will only be necessary to build one, single substantive Constituent Determination model with a "memory" containing the building material of, at least, one sample compositum for each of the possible types of substantive composite of these languages.

7. Death of the German Pre-Editor (August 30, 2 a.m.)

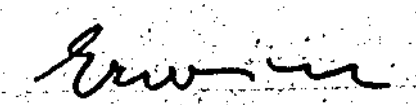
Once the problem of the mechanical determination of the constituents of German (long established and extemporized) substantive composite was solved,

there remained only one question, namely: which other of the few graphio-  
semantically not distinctive grammatical meanings a pre-editor would have to  
indicate in the German source text could **be** mechanically determined by  
translating into mechanical storing and synthesizing steps the thought processes  
by which a human being (either native German or a student of German) abstracts  
synthetic grammatical meanings—that is, meanings which have to be inferred  
by the consideration of more than one source text form?

By 2 o'clock, this morning (August 30th) I was convinced that all those  
types of such meanings which are relevant for German-English MT could be  
mechanically determined and, moreover, that the fundamental engineering  
problem was only one of a complex wiring, switching, and storing system.  
This I shall attempt to show in consecutive papers.

Thus Drs. Bar-Hillel, Bull, and Oswald were right after all, at least  
as far as the German language is concerned. With this confirmation, the  
German pre-editor faded completely into oblivion.

Respectfully submitted

A handwritten signature in black ink, appearing to be 'G. S. ...', written on a dotted grid background.