STUDIES IN MECHANICAL TRANSLATION

No. 9

ELIMINATION AND MECHANICAL DETERMINATION OF FORMS WITH DUAL NATIONALITY

(Preliminary Report On My Recent Research Results In MT, Through February 3, 1953).

by

ERWIN REIFLER

Department Of Far Eastern And Slavic Languages And Literature

University Of Washington

S T U D I E S I N M T

****

NO 9

INDEX

****

****

****

****

****

****

****

****

**

STUDIES IN MECHANICAL TRANSLATION

NO 9.

Feb. 3,1953.

Far Eastern

To: Professor George E. Taylor

From: Erwin Reifler

Re: Preliminary Report On My Recent Research Results In Mechanical Translation Through February 3.1953.

Dear George:

9.1. In my SIMT No 7 (THE MECHANICAL DETERMINATION OF THE CONSTITUENTS OF GERMAN SUBSTANTIVE COMPOSITA) I twice referred to the so-called "Ur-"Problem Solution (SIMT No 7/5Db and 10/II and III) which I promised to describe in a subsequent paper. This problem is only part of the wider problem of the mechanical determination of incident non-grammatical meaning in unseparated clue-sets (polypod clue-sets; cf. SIMT No.5, p. 12). Before this problem could be discussed and a solution formulated, a farther refinement of MT terminology was necessary.

9.2. In my forthcoming paper SIMT No 10 (THE MECHANICAL DETERMINATION OF INCIDENT NON-GRAMMATICAL MEANING IN POLYPOD CLUE-SETS) this new terminology, the "Ur-"Problem Solution and the whole problem of "memorization" versus "Synthetization" of polypods will be discussed in greater detail. It will also include a second possible solution to the problem of "divergent polypods" (composita whose target equivalents can not be inferred from the target equivalents of their constituents, for example "Ur/Laub", leave, "Ur/teil", judgement, "Mit/gift", dowry, "Hoch/zeit", wedding, etc. Cf.SIMT No 7/4(2)C and 5(2) B,C,D).

9.3. I am glad to be able to report that I have, at last, succeeded in solving the fascinating problem of "internationals" or rather "diglots" — that is, words which, apart from minor differences of an orthographic and/or grammatical nature in some cases, occur in the source language as well as in the target language in the same graphic form and meaning. If all diglot monopods (not composed free source forms) are "memorized", then my SIMT No 7 has solved the whole problem of the mechanical dissection of all substantive polypods (composed free source forms), and my SIMT No 8 the whole problem of the mechanical determination of all operational form classes. But if all "internationals" are transferred, then the problem of the mechanical dissection of

hybrid polypods and the filtering out and the transfer of the international constituents arises. We have then to distinguish between two kinds of source forms :

A. Monoglots — that is, those which graphically and semantically are not shared by, or are not intelligible in, the English target language. Examples are German "Wald" (forest), "Komponist" (composer), "Kolportage" (hawking by itinerant vendors), "schwer"(heavy), "majorenn" (of full age), "kolportieren" (to hawk), "Taste* (not "taste", but "key of a musical instrument or of a typewriter), "krank" (not "crank", but "sick"), etc.etc.

B. Diglots — that is, those which graphically and semantically are shared by, or are intelligible in, the English target language, these are again subdivided into :

a) Indigenous Diglots, for example German "Horn", "Gold", "Fisch","Goldfisch" "Glas" (glass), "Goldfischglas", "Wolf, "Hunger", "Wolfshunger",etc.

b) Immigrant Diglots (mostly of Romance and Greek origin), for example "Differential", "Quotient", "Differentialquotient", "Mission", "Journal", "Missionsjournal" "Orthographie", "Mikroskop", "Musik", "Literatur", "Musikliteratur", Theater, "Insulation" "Material", "Insulationsmaterial", "fundamental", "lapidar", "konstitutionell","sozial" "demokratisch", "sozialdemokratisch", "musikalisch", "theatralisch", "organisieren", "zivilisieren", etc.etc.

9.4.1. As a result of the solution of this problem these "internationals", which do not require translation, need not to be recorded in the mechanical memory, but can be mechanically transferred to the target side in their original alphabetization (cf. SIMT No 7/8(8) and SIMT No 8/12.10) . This means a further substantial reduction in the storage requirements.

9.4.2. Five problems had first to be solved to make this possible, namely :

A. The development of a third matching procedure,

B. The abstraction of the grammatical information from the internationals before they are transferred,

C. The neutralization of analogic "-s-",

D. The problem of hybrid polypods.

E. The "X-"factor problem (cf. SIMT No 7/2) in these hybrids.

A. A converging (hereafter called CV) matching procedure in addition to the previously developed left-to-right (hereafter LR; cf. SIMT No 7/9A) and right-to-left (hereafter RL; cf. SIMT No 7/9B) matching procedures had to be worked out in order to extract monoglot inner constituents from certain hybrid polypods (cf. Db/ff and c/gg,hh,jj below).

B. The extraction of the grammatical information of both "monopodic"(non-composed) and "polypodic" (composed) "internationals" in the form of endings and/or the "umlaut", such as "genitive", "plural", "infinitive", "present participle", "active present indicative third person singular" (which in the case of the target English "internationals" requires a special ending, cf. "organizes", "past tense", etc. (cf. German "Hungers" , "Protektorats", "Glases", "Wölfe", "Gläser" "Nationen", "garantieren", "kondensierend", "konserviert" , "konfrontierte"), before these "internationals" are transferred. This grammatical information is not only important for the pinpointing of the incident grammatical meaning of the "international" itself and of other co-occurrent free source forms. It is also important for the proper "naturalization" of the "international" in the target society: "Hungers" has to appear in English as "Hunger's" or "of the Hunger", "Protektorats" at least as "Protektorat's" or "of the Protektorat", "Glases" at least as "Glas's" or "of the Glas", "Wölfe" and "Gläser" — in which the MT mechanism may (but need not) also transfer the umlaut — as "Wolfes" and "(Glases" (i.e. "glasses"), "Nationen" as "Nationes", "garantieren" as "garant", "kondensierend" as "kondensing", "konserviert" — depending on the pinpointing results obtained on the source side before the transfer — as either "konserves" or "konserved", "konfrontierte" as "konfronted". The pre-transfer extraction of this grammatical meaning is possible because of the absolutely "lawful behaviour" of German "internationals" (no exception to any grammatical rule).

C. The neutralization of the analogic "-s-" at the end of certain feminine "international" right-bound forms (i.e. "Missions/museum") which in the English target language risks being interpreted as a plural ending.

D. The problem of hybrid polypods. Here we have to distinguish :

a) Hybrid Dipods (polypods with only two constituents). In these there are only two possibilities of sequential arrangement. Using M for "monoglot" and D for "diglot", these are :

aa) MD

bb) DM

M B

aa) MD, for example "Stahl/production" (steel production). Here LR matching identifies the monoglot first constituent ("Stahl"), thus determines the left boundary of the unidentifiable following rest and recognizes the latter as an international signal sequence to be transferred.

D M

bb) DM, for example "Integral/rechnung" (integral calculus). Here RL matching identifies the monoglot last constituent ("rechnung"), thus determines the right boundary of the unidentifiable preceding rest and recognizes the latter as an international signal sequence to be transferred.

NOTE : When I worked out the RL matching process, I did not yet quite realize the full range of its importance for MT (cf. the NOTE at the head of SIMT No 7/9B). It is this matching process which, in conjunction with LR matching, makes possible the solution of the problems presented by diglot-monoglot dipods and by hybrid tripods, tetrapods. pentapods, etc. (see b and c below).

b) Hybrid Tripods (polypods with three constituents). In these there are six possibilities of sequential arrangement, namely:

aa) MDD

bb) DDM

cc) MMD

dd) DMM

ee) MDM

ff) DMD.

M D D

aa) MDD, for example "Reichs/uranium/monopol" (imperial uranium monopoly.

Here LR matching identifies the monoglot first constituent ("Reichs"), thus determines the left boundary of the following unidentifiable rest recognizes the latter as an international signal sequence to be transferred.

D D M

bb) DDM, for example "Gold/fisch/pflege" (goldfish care). Here RL matching identifies the monoglot last constituent ("pflege") , thus determines the right boundary of the preceding unidentifiable rest and recognizes the latter as an international signal sequence to be transferred.

M M D

cc) MMD, for example "Reichs/salz/monopol" (imperial salt monopoly) . Here LR matching dissects the monoglot left complex into its two constituents, thus determines the left boundary of the following unidentifiable rest ("monopol") and recognizes the latter as an international signal sequence to be transferred.

D M M

dd) DMM, for example "Hand/arbeits/stunde" (hand work hour). Here RL matching dissects the monoglot right complex into its two constituents, thus determines the right boundary of the preceding unidentifiable rest ("Hand") and recognizes the latter as an international signal sequence to be transferred.

M D M

ee) MDM, for example "Schwer/industrie/ausstellung"(heavy industry exhibition). Here both LR and RL matching comes into play. LR matching identifies the monoglot first constituent ("Schwer") and thus determines the left boundary of the unidentifiable inner signal sequence, whereas RL matching identifies the monoglot last constituent ("ausstellung") and thus determines the right boundary of the unidentifiable inner signal sequence which thus is ultimately recognized as an international signal sequence to be transferred.

D M D

ff) DMD, for example "Horn/bläser/konzert" (horn blower concert) It is this type of hybrid polypods which constituted the greatest obstacle to the solution of the problem of "internationals". Since neither the first nor the last constituent has a memory equivalent, neither LR nor RL matching, nor both together, will suffice to recognize them as "internationals" and to recognize and identify the monoglot inner constituent. Here a third matching procedure had to be worked out, namely a "converging" process which first identifies the inner constituent ("bläser") and thus determines the inner boundaries of the two external "internationals".

c) Hybrid polypods with more than three constituents (tetrapods,pentapods,etc The correct mechanical dissection of hybrid polypods with more than three constituents does — after the solution of the problem of hybrid tripods — not present any further difficulty. We shall exemplify this in the following with hybrid tetrapods. In these there are ten possibilities of sequential arrangement, namely:

aa) MDDD

bb) DDDM

cc) MMMD

dd) DMMM

ee) MMDD

ff) DDMM

gg) MDMD

hh) DMDM

ii) MDDM

jj) DMMD.

M D D D

aa) MDDD for example "Reichs/uranium/monopol/skandal" (scandal of the imperial uranium monopoly) . Here LR matching identifies the monoglot first constituent ("Reichs") and thus recognizes the following unidentifiable rest as an international signal sequence to be transferred.

D D D M

bb) DDDM, for example "Gold/fisch/glas/ständer" (goldfish glass stand). Here RL matching identifies the monoglot last constituent ("ständer") and thus recognizes the preceding identifiable rest as an international signal sequence to be transferred

M M M D

cc) MMMD for example "Welt/handel/geld/hunger" (world commerce money hunger). Here LR matching dissects the left monoglot complex into its three consti - tuents and thus recognizes the following unidentifiable rest ("hunger") as an international signal sequence to be transferred.

D M M M

dd) DMMM, for example "Hand/arbeits/stunden/lohn" (hand work hour wage) . Here RL matching dissects the monoglot right complex into its three constituents and thus recognizes the preceding unidentifiable rest ("Hand") as an international signal sequence to be transferred.

M M D D

ee) MMDD, for example "Reichs/salz/monopol/statistik" (imperial salt monopoly statistics). Hero LR matching dissects the monoglot left complexinto its two constituents and thus recognizes the following unidentifiable rest as an international signal sequence to be transferred.

D D M M

ff) DDMM, for example "Korn/monopol/amts/sitz" (corn monopoly office seat). Here RL matching dissects the monoglot right complex into its two constituents and thus recognizes the preceding unidentifiable rest as an international signal sequence to be transferred.

M D M D

gg) MDMD, for example "Welt/gold/handels/monopol" (world gold commerce monopoly). Here LR matching identifies the monoglot first constituent ("Welt") and determines the left boundary of the diglot second constituent. CV matching identifies the monoglot third constituent ("handels"), determines the right boundary of the diglot second ("gold") and the left boundary of the diglot last ("monopol") constituent and thus recognizes them as international signal sequences to be transferred.

D M D M

hh) DMDM, for example "Gold/handels/monopol/tagung" (gold commerce monopoly session). Here CV matching identifies the monoglot second constituent ("handels") and determines the right boundary of the diglot first ("Gold"),and the left boundary of the diglot third ("monopol") constituent. RL *mat*ching identifies the monoglot last constituent ("tagung") and determines the right boundary of the diglot third constituent ("monopol") . Both "gold" and "monopol" are thus recognized as international signal sequences to be transferred.

M D D M

ii) MDDM, for example "Welt/gold/produktions/bericht" (world gold production report). Here LR matching identifies the monoglot first ("Welt"), RL matching the monoglot last ("bericht") constituent and thus recognizes the inner unidentifiable rest as an international signal sequence to be transferred.

D M M D

jj) DMMD, for example "Horn/bläser/frühlings/konzert" (horn blower spring concert). Here CV matching identifies the monoglot second constituent ("bläser") and determines the right boundary of the diglot first constituent ("Horn") and the left boundary of the monoglot third constituent ("frühlings"). LR matching identifies the monoglot third constituent ("frühlings") and the left boundary of the diglot last constituent ("konzert"). The diglot first and last constituents are thus recognized as international signal sequences and transferred.

E. The "X"-factor problem in hybrid polypods:

a) The "X"-factor problem will not play any role in polypods with entirely indigenous constituents because "X-"factor forms which in such polypods may create an "X-"factor confusion in the target language will not be treated as transferable diglot constituents (i.e. bound forms). Indigenous diglot monopods are long established (i.e. are all ascertainable) and far less numerous than the (ever-growing number of) immigrant diglot monopods, They can, therefore, all be easily checked. For example the indigenous diglot "Winter" is an "X-"factor form ("-er"), but "Wint" does not mean anything in the English target language. "Winter-" is, consequently, a transferable right-bound form.

b) Nor is the "X-"factor problem likely to play any role in polypods all of whose constituents are immigrant diglots. For example the immigrant diglot constituent "Insulin-" in "Insulintoleranz" is an "X-"factor form ("-in-") and both "toleranz" (tolerance) and "intoleranz" (intolerance) are intelligible in the English target language. But "Insul-" does not mean anything in English. "Insulin-" is, therefore, a transferable right-bound form.

c) The "X-"factor problem may, however, be of some consequence in hybrid polypods. For example in "Dichter/inspiration" (poet's inspiration) the matching mechanism would as largest monoglot left constituent identify "Dichterin" (poetess) and as unidentifiable rest transfer "spiration", resulting in the unintelligible "poetess spiration". The "X-"factor problem requires, therefore, special consideration in the treatment of internationals" in MT.

9.5. Th« detailed presentation of the problem of "internationals" in MT, the description of the third matching procedure and of the fixed sequence of all matching procedures will be the subject of my forthcoming paper SIMT No 11 (THE TREATMENT OF INTERNATIONALS IN MT).

9.6. This solution of the problem of "internationals", especially as far as the mechanical dissection of hybrid polypods is concerned, supplements my SIMT No 7. On the other hand it makes necessary certain readjustments with respect to my paper SIMT No 8 (THE MT FORM CLASS FILTER SYSTEM) . These will also be described in SIMT No 11.

F I N I S