THE GEORGETOWN-IBM EXPERIMENT OF 1954: AN EVALUATION IN RETROSPECT*

PAUL L. GARVIN

Enough time has elapsed and sufficient other work has been attempted in machine translation since 1954 to allow an appraisal of this much-talked-about demonstration in the light of the experience since gained.

Whatever its implications may have been in terms of publicizing and stirring up interest in the problem, from a research standpoint the purpose of the verbal program underlying the Georgetown-IBM experiment of 7 January 1954 was to test the feasibility of machine translation by devising a maximally simple but realistic set of translation rules that were also programmable. The actual execution of the program on the 701 computer turned out to be an interesting exercise in nonmathematical programming, but showed nothing about translation beyond what was already contained in the verbal rules.

The verbal program was simple because the translation algorithm consisted of a few severely limited rules, each containing a simple recognition routine with one or two simple commands. It was realistic because the rules dealt with genuine decision problems, based on the identification of the two fundamental types of translation decisions: selection decisions and arrangement decisions.

The limitations of the translation algorithm were dual: the search span of the recognition routine was restricted to the immediately adjacent item[1] to the left or right; the command routine was restricted, for selection decisions, to a choice from among two equivalents, for arrangement decisions, to a rearrangement of the translations of two immediately adjacent items.

The translation program was applied to one Russian sentence at a time: the lookup would bring the glossary entries corresponding to the items of the sentence into the working storage, where the algorithm would go into effect.[2]

The requirements of simplicity and realism were reconciled on the basis of an analy-

* Work on this paper was done under the sponsorship of the AF Office of Scientific Research of the Office of Aerospace Research under Contract No. AF 49 (638) - 1128.

[1] The term "item was introduced to designate Russian words or word partials, as opposed to the term "word" which was reserved for computer words. The term "decision point" was introduced to designate an item for which the program has to make a translation decision, the term "decision cue" (or "cue") to designate an item which is considered the relevant condition for making a certain decision.

[2] A statement of the verbal program, the transliteration table, an excerpt from the machine glossary, as well as a selection from the original test sentences, are contained in the Appendix.

sis of the logical structure of a few translation problems. The different variables entering into each problem were isolated, and the rules were then designed to deal each with one particular variable, leaving the remaining aspects of the problem unsolved, or giving an arbitrary solution. In a number of cases, for instance, where the correct choice would have required the operation of rules which were not included in this simple program, a translation appropriate to the input sentences was arbitrarily placed into the glossary. The underlying assumption was that additional rules covering this residue could be written later, without invalidating the rules included in the experiment.

Thus, the translation of Russian case suffixes was analyzed into two decision steps: a first-order decision to determine whether or not to translate the suffix by a preposition, and a second-order decision to choose the particular preposition where one is required. In the experiment, only the first-order decision was implemented, and for only a few suffixes; the second-order decision was ignored by arbitrarily assigning a simple English prepositional translation to each suffix (namely, that which impressionistically seemed the most frequent). This was done by applying rule 3: case suffixes with other than accusatival function were translated by zero whenever a Russian preposition or adjectival suffix preceded the item in question, they were translated by a preposition when this condition did not apply, and in the latter instance, the order of the translations of stem and suffix (the English noun or adjective, and preposition, respectively) was then inverted.

The same rule was used to effect the translation decision for first-person plural forms of verbs, which is analogous to the first-order decision for case suffixes: the verb form was translated without using a pronoun in English whenever a pronoun was present in the Russian text (sentence 32).

Another method of simplifying the translation decision was to limit the cue distance (i.e., the distance between decision cue and decision point) and cue location arbitrarily to conform to the one-word search span, while realistically defining the decision cue in terms of grammatical conditions. An instance of this was the application of rule 3 to the translation of the case suffixes -a, -я. For the appropriate nouns these were interpreted as animate accusatives and translated by zero, whenever they were preceded by a transitive verb form (sentence 40).

A further simplification of certain selection decisions affecting the translation of prepositions, verbs, and nouns, was brought about by not only restricting the cue distance but also limiting the scope of the decision itself to a choice between two equivalents.

Thus, the translation of the preposition к was effected by rule 2 as determined by certain governed nouns, and other aspects of the translation decision were ignored (sentences 4, 19, 40). Conversely, rule 3 was used to translate a noun as determined by the immediately preceding governing verb (sentence 31), or by a modifying adjective (sentences 15-17). The definite article was selected by rule 5 in a few cases in which the Russian noun in question preceded a noun in the genitive, corresponding

to the English construction *N of N,* in which an article is frequently required for the first of the two nouns (sentences 19, 20, 27-29).[3]

One arrangement decision in addition to that required for case suffix translation was made: rule 1 was used to invert the order of the translations of a verb and its immediately following subject (sentences 2, 7, 11, 13, 33-34, 45).

Finally, one idiom translation was attempted: rules 3 and 5 were used to translate a three-word Russian idiom by its two-word English equivalent (sentence 26). This was done by choosing the second English word as the equivalent of the second Russian word by rule 5, with the third Russian word considered the cue, and by choosing zero as the equivalent of the third Russian word by rule 3, with the second Russian word considered the cue (for the term "cue", see fn. 2).

The program utilized a dictionary lookup for calling the translation algorithm in the following manner:

The suffixes for which translation decisions were made, and the stems from which they had to be detached, were each entered in the glossary separately. A stem-suffix splitting subroutine, called the "hyphen rule", was included in the lookup. It was applied only to the so-called subdivided items, i.e., the items involved in the above suffix-translation decisions; all other glossary items were entered undivided.

All entries, whether they represented undivided items or the portions of subdivided items, were listed in a single alphabetic sequence.

The five rules of the translation algorithm were operated by a set of two-digit and three-digit numerical code symbols, called diacritics, attached to the glossary entries. The first of the digits was used to indicate whether the diacritic was assigned to a decision-point entry or a decision-cue entry. The second digit indicated the number of the rule to be applied, and the third digit, used only for some decision-cue diacritics, marked which of two choices was to be made (for terms, see fn. 2).

One limitation was imposed by the convenience of the computer program, namely that a particular glossary entry was allowed to contain no more than two three-digit diacritics and one two-digit diacritic.

The general characteristics of the 1954 experiment can be summarized as follows:

(1) The scope of the translation program was clearly specified. Any sentence meeting its narrow specifications could be translated, provided the required entries were present in the glossary. The glossary could be expanded without difficulty and the program made to operate on it, provided the new entries were limited to items to which the previously established code diacritics could be assigned.

(2) The lookup routine was designed for maximum efficiency of the translation algorithm, in that the splitting subroutine was applied only to those cases where it would serve to simplify the operation of the rules, and not to all grammatically possible cases.

(3) The translation algorithm was based on the collocation of decision points and

[3] This solution was suggested by A. A. Hill.

decision cues, rather than directly on the linguistic factors involved, although the decision points and cues themselves were established by linguistic analysis. The same rule was thus used to solve problems of different linguistic structure, but with similar decision structure; rule 3, for instance, was used to translate case suffixes, to choose the translation of nouns on the basis of the verbs governing them, to translate verbs with or without pronouns, and was also utilized in the one idiom translation.

(4) The word length of a sentence turned out to be operationally trivial, since the rules allowed the translation of consecutive strings of similar constructions, provided they were within the specifications of the algorithm.

(5) Selection and arrangement were confirmed as the basic algorithmic operations. "Omission" and "insertion" emerged as simple variants of the selection problem: omission amounted to the choice of a zero equivalent; insertion to the choice of a two-or-more word equivalent for a single input word.

The importance of the 1954 experiment lies in the fact that it formed a significant first step in a continuing research process which is first now nearing completion. This first step consisted in providing an essentially correct formulation of the problem of machine translation which can be succinctly stated as follows:

(1) The machine translation problem is basically a decision problem.

(2) The two fundamental types of decisions are selection decisions and arrangement decisions.

(3) For the automatic implementation of a translation decision, the algorithm has to have the capability for recognizing the decision points and the appropriate decision cues.

The research derived from this formulation has therefore been focused on the detection of the recognition criteria needed for the identification of the decision points and decision cues. This approach to the decision problem is based on an understanding of syntactic and semantic structure which increases as our empirical treatment of it develops.

The Bunker-Ramo Corporation

APPENDIX: DOCUMENTATION OF THE 1954 EXPERIMENT

1. *Verbal Program*

LOOKUP

Match each item of the input sentence consecutively against items stored at the head of glossary entries. Apply hyphen rule whenever necessary.

Hyphen rule. If the lookup does not find a match for all the letters of an input item with a complete item in the glossary, try first for a match of the initial letters with a left partial (stem, as indicated in the glossary by a following hyphen), then try for a match of the remaining letters with a right partial (suffix, as indicated in the glossary by a preceding hyphen).

Bring matched glossary entries into working storage in the order of the input.

ALGORITHM

Calling the rules. Scan the diacritic field of the dictionary entries in working storage consecutively from left to right until you find the first decision-point diacritic, as indicated by a numeral 1 in the first digit position, and operate the rule indicated by the second digit of the diacritic. Then return to scanning for diacritics, beginning with the entry immediately to the right of where you left off.

Rule 1. Look for cue diacritic 21 in the diacritic part of a complete-item entry immediately to the left of the decision point.

Yes — invert the order of the translations of the items concerned.

No — retain order.

Rule 2. If the decision point is a complete item, look for cue diacritics 221 or 222 in the diacritic field of a complete-item entry, or of either partial entry for a subdivided item, immediately to the right of the decision point. If the decision point is a left partial, look for cue diacritics in the corresponding right-partial entry. Select as follows:

221 — choose the first equivalent of the decision-point entry.

222 — choose the second equivalent of the decision-point entry.

Rule 3. If the decision point is a left partial, look for cue diacritic 23 in the diacritic field of a complete-item entry, or of either partial entry for a subdivided item, immediately to the left of the decision point. If the decision point is a right partial, look for cue diacritic 23 in the diacritic field of a corresponding left-partial entry.

Yes — choose the second equivalent of the decision-point entry.

No — choose the first equivalent of the decision-point entry, then invert order as follows: if the decision point is a complete item or a left partial, place its translation before that of the item immediately to the left of it; if the decision point is a right partial, invert the order of the translations of the right and left partials.

Rule 4. Look for cue diacritics 241 or 242 in the diacritic field of a complete-item entry or of either partial entry for a subdivided item, immediately to the left of the decision point. Select as follows:

241 — choose the first equivalent of the decision-point entry.

242 — choose the second equivalent of the decision-point entry.

Rule 5. Look for cue diacritic 25 in the diacritic field of a complete-item entry, or of either partial entry for a subdivided item, immediately to the right of the decision point.

Yes — choose the second equivalent of the decision-point entry.

No — choose the first equivalent of the decision-point entry.

2. *Transliteration table*

А A	Ж ZH	М M	С S	Ц TS	Ь J
Б B	З Z	Н N	Т T	Ч CH	Э E
В V	И YI	О O	У U	Ш SH	Ю YU
Г G	Й Y	П P	Ф F	Щ SHCH	Я YA
Д D	К K	Р R	Х X	Ы I	Ъ W
Е YE	Л L				

3. *Excerpt From Glossary*

ENTRY	EQUIVALENTS	CODES		
-A	OF	131	222	25
	--	132	222	25
-AMYI	BY	131	222	
	--	132	222	
BO-	BATTLE	222		
BOLJSH-	A LARGE			
	LARGE			
BOYETS	FIGHTER	242		
BYENZYIN	GASOLINE	241	21	
BYETON-	CONCRETE			
DLYIN-	LENGTH			
DOBIVAYUT	THEY OBTAIN	110		
DOMA	AT HOME	151	241	
	HOUSES	152	241	
DOROGI	ROADS	241		
DUG-	ARC			
DYINAMYIT	DYNAMITE	241	21	
FAKTOR-	FACTOR			
FYEDYERATSYIYA	A FEDERATION			
	THE FEDERATION			
GRAZHDANSK-	CIVIL			
-I	OF	131	25	
	--	132	25	
-IM	BY	131	23	
	--	132	23	
-IMYI	BY	131	222	23
	--	132	222	23
-IX	OF	131	222	23

ENTRY	EQUIVALENTS	CODES
	--	132 222 23
-IY	--	
-IYE	--	222
-JYU	BY	131
	--	132
K	TO	121 23
	FOR	122 23
KACHYESTVO	QUALITY	151 222
	THE QUALITY	152 222
KALORYIYNOST-	CALORY CONTENT	
KALORYIYNOSTJ	CALORY CONTENT	
KAMN-	STONE	
KAMYENN-	STONY	151
	--	152
KARTOFYEL-	POTATOES	
KLYINOM	BY A WEDGE	131
	IN WEDGE FORMATION	132
KRAXMAL	STARCH	21
KYIRPYICH-	BRICK	
KYISLORODN-	OXYGEN	
LYISHYENYI-	DEPRIVAL	221
MATYERYIAL-	MATERIAL	
MI	WE	23
MISLYI	THOUGHTS	
MNOG-	MANY	
MYEDJ	COPPER	21
MYEST-	PLACE	151 23
	SITE	152 23
MYEXANYICHESK-	MECHANICAL	242
MYEZHDUNARODN-	INTERNATIONAL	
NYIVYELYIROVANYI-	REVELING	
NA	ON	121 23
	FOR	122 23
NAUKA	A SCIENCE	242
	THE SCIENCE	242
NYEFT-	CRUDE OIL	
NYITROGLYITSYERYIN-	NITROGLYCERINE	
NYIVYELYIROVAHYI-	LEVELING	
O	ABOUT	141 23
	OF	142 23
OBRABOTKA	PROCESSING	
-OGO	OF	131 23
	--	132 23
-OM	BY	131
	--	132
OPRYEDYELYAYET	DETERMINES	
OPRYEDYELYAYETSYA	IS DETERMINED	
OPTYICHYESK-	OPTICAL	
OTDYEL-	SECTION	

ENTRY	EQUIVALENTS	CODES	
OTDYELYENYIYE	DIVISION	121	242
	SQUAD	122	242
OTNOSHYENYI-	RELATION	151	
	THE RELATION	152	
-OV	OF	131	222
	--	132	222
-OYE	--		
POGOD-	WEATHER		
POLUCHAYET	GETS		
POLYITYICHYESK-	POLITICAL		
PONYIMANYIYE	UNDERSTANDING	242	
POSLYEDN-	LAST		
	LATEST		
POSRYEDSTVOM	BY MEANS OF	23	
POVISHAYET	INCREASES	121	
	IMPROVES	122	
POZDNO	LATE		
PRAV	OF RIGHTS	131	
	RIGHTS	132	
PRAVO	RIGHT	141	242
	LAW	142	242
PROTSYESS-	PROCESS		
PRYI	AT	121	23
	IN	122	23
PRYIGOTOVLYAYETSYA	IS PREPARED	141	
	PREPARES SELF	142	
PRYIGOTOVLYAYUT	THEY PREPARE	110	
PRYIGOVORYIL	SENTENCED	23	
PRYIMYES-	ADMIXTURE		
PSHYENYITS-	WHEAT		
PUT-	PATH	141	
	METHOD	142	
PYERYEDAYEM	WE TRANSMIT	131	
	TRANSMIT	132	
PYERYEDAYET	TRANSMITS		
PYERYEGOVORI	NEGOTIATIONS	110	241
PYERYEMYIRYI-	AN ARMISTICE		
	THE ARMISTICE		
RABOT-	WORK	222	
RADYIOSTANTSYIYA	A RADIO STATION		
	THE RADIO STATION		
RADYIUS-	RADIUS	221	
RINK-	THE MARKET		
RUD-	ORE		
RYECH-	SPEECH		
RYESHYENYI-	SOLUTION	121	221
	DECISION	122	221
S	WITH	23	

ENTRY	EQUIVALENTS	CODES		
SHTAT-	STATE	121		
	STATES	122		
SOOBSHCHYENYIYA	COMMUNICATIONS	241		
SOSTOYIT	CONSISTS			
SOYEDYINYENYI-	COMPOUND	121	242	
	COMPOUNDS	122	242	
SPROS-	THE DEMAND			
SPYIRT	ALCOHOL	21		
STROYATSYA	ARE CONSTRUCTED	141	242	25
	LINE UP	142	242	25
STROYITSYA	IS CONSTRUCTED	141	222	23
	LINES UP	142	222	23
SYELITR-	SALTPETER			
SYERZHANT-	A SERGEANT			
	THE SERGEANT			
TOL	T.N.T.	241	21	
TSYEL-	TARGET	131	25	
	--	132	25	
TSYENA	PRICE	151		
	THE PRICE	152		
-U	TO	131		
	--	132		
UGL-	COAL	121	25	
	ANGLE	122	25	
UGOL	ANGLE			
UGOLOVN-	PENAL	242		
UTROM	IN THE MORNING			
V	IN	122	23	
	TO	122	23	
VAZHN-	AN IMPORTANT			
	IMPORTANT			
VIRABATIVAYETSYA	IS PRODUCED			
VIRABATIVAYUT	THEY PRODUCE	110		
VLADYIMYIR	VLADIMIR	241		
VOPROS-	QUESTION	121		
	QUESTIONS	122		
VOYSKA	TROOPS	242		
VOZVISHYENYIYE	ELEVATION			
VYEDUTSYA	ARE CONDUCTED	21		
VYELYICHYINA	MAGNITUDE			
XYIMYI-	CHEMISTRY			
XYIMYICHYESK-	CHEMICAL	242		
-Y	OF	131	222	
	--	132	222	
-YA	OF	131	221	25
	--	132	221	25
YAVLYAYETSYA	APPEARS	141	23	
	CONSTITUTES	142	23	
-YAX	----	222		

ENTRYE	QUIVALENTS	CODES
-YE	TO	131 221
	--	132 221
-YEM	BY	131
	--	132
-YI	OF	131 25
	--	132 25
-YIM	BY	131 23
	--	132 23
-YIX	OF	131 222 23
	--	132 222 23
-YIYE	---	222
YIZ	OUT OF	23
YIZMYERYENYI-	MEASUREMENT	
YIZVYESTYIYA	BULLETINS	
-YU	TO	131
	--	132
ZAKONODATYELJSTV-	LEGISLATION	
ZHALOVANYIYE	SALARY	
ZHYELYEZO	IRON	21

4. *Selected Test Sentences*

1. PRYIGOTOVLYAYUT TOL
2. TOL PRYIGOTOVLYAYUT YIZ* UGLYA
3. TOL PRYIGOTOVLYAYETSYA YIZ UGLYA
4. BOYETS PRYIGOTOVLYAYETSYA K BOYU
5. KACHYESTVO UGLYA OPRYEDYELYAYETSYA KALORYIYNOSTJYU
6. TOL PRYIGOTOVLYAYETSYA YIZ KAMYENNOGO UGLYA
7. BYENZYIN DOBIVAYUT YIZ NYEFTYI
8. BYENZYIN DOBIVAYETSYA YIZ NYEFTYI
9. AMMONYIT PRYIGOTOVLYAYUT YIZ SYELYITRI
10. AMMONYIT PRYIGOTOVLYAYETSYA YIZ SYELYITRI
11. SPYIRT VIRABATIVAYUT YIZ KARTOFYELYA
12. SPYIRT VIRABATIVAYETSYA YIZ KARTOFYELYA
13. KRAXMAL VIRABATIVAYUT YIZ KARTOFYELYA
14. KRAXMAL VIRABATIVAYETSYA YIZ KARTOFYELYA
15. TOL PRYIGOTOVLYAYETSYA XYIMYICHYESKYIM PUTYEM YIZ KAMYENNOGO UGLYA
16. AMMONYIT PRYIGOTOVLYAYETSYA XYIMYICHYESKYIM PUTYEM YIZ SYELYITRI
17. KRAXMAL VIRABATIVAYETSYA MYEXANYICHYESKYIM PUTYEM YIZ KARTOFYELYA
18. TSYENA KARTOFYELYA OPRYEDYELYAYETSYA RINKOM
19. VYELYICHYINA UGLA OPRYEDYELYAYETSYA OTNOSHYENYIYEM DLYINI DUGI K RADYIUSU
20. KALORYIYNOSTJ OPRYEDYELYAYET KACHYESTVO UGLYA
21. OBRABOTKA POVISHAYET KACHYESTVO NYEFTYI

22. ZHYELYEZO DOBIVAYETSYA YIZ RUDI
23. MYEDJ DOBIVAYETSYA YIZ RUDI
24. DYINAMYIT PRYIGOTOVLYAYETSYA YIZ NYTROGLYITSYERINA S PRYIMYESJYU YINYERTNOGO MATERYIALA
25. VOZVISHYENYIYE OPRYEDYELYAYETSYA NYIVYELYIROVANYIYEM
26. UGOL MYESTA TSYELYI OPRYEDYELYAYETSYA OPTYICHYESKYIM YIZMYERYENYIYEM
27. TSYENA PSHYENYITSI OPRYEDYELYAYETSYA RINKOM
28. TSYENA PSHYENYITSI OPRYEDYELYAYETSYA SPROSOM
29. TSYENA KARTOFYELYA OPRYEDYELYAYETSYA SPROSOM
30. DOROGI STROYATSYA YIZ KAMNYA
31. VOYSKA STROYATSYA KLYINOM
32. MI PYERYEDAYEM MISLYIPOSRYEDSTVOM RYECHYI
33. ZHYELYEZO DOBIVAYUT YIZ RUDI
34. MYEDJ DOBIVAYUT YIZ RUDI
35. ZHYELYEZO DOBIVAYETSYA YIZ RUDI XYIMYICHESKYIM PROTSYESSOM
36. MYEDJ DOBIVAYETSYA YIZ RUDI XYIMYICHYESKYIM PROTSYESSOM
37. DYINAMYIT PRYIGOTOVLYAYETSYA XYIMYICHYESKYIM PUTYEM YIZ NYITROGLYITSYERYINA S PRYIMYESJYU YINYERTNOGO MATYERYIALA
38. DOMA STROYATSYA YIZ KYIRPYICHA
39. DOMA STROYATSYA YIZ BYETONA
40. VOYENNIY SUD PRYIGOVORYIL SYERZHANTA K LYISHYENYIYU GRAZHDANSKYIX PRAV
41. UGOLOVNOYE PRAVO YAVLYAYETSYA VAZHNIM OTDYELOM ZAKONODATYELJSTVA
42. NAUKA O KYISLORODNIX SOYEDYINYENYIYAX YAVLYAYETSYA VAZHNIM OTDYELOM XYIMYIYI
43. VLADYIMYIR YAVLYAYETSYA NA RABOTU POZDNO UTROM
44. MYEZHDUNARODNOYE PONYIMANYIYE YAVLYAYETSYA VAZHNIM FAKTOROM V RYESHYENYIYI POLYITYICHYESKYIX VOPROSOV
45. VYEDUTSYA PYERYEGOVORI O PYERYEMYIRYIYI
46. FYEDYERATSYIYA SOSTOYIT YIZ MNOGYIX SHTATOV
47. RADYIOSTANTSYIYA PYERYEDAYET POSLYEDNYIYE SOOBSHCHYENYIYA O POGODYE
48. RADYIOSTANTSYIYA PYERYEDAYET POSLYEDNYIYE POLYITYICHYESKYIYE YIZVYESTYIYA
49. VLADYIMYIR POLUCHAYET BOLJSHOYE ZHALOVANYIYE