The Georgetown-IBM experiment demonstrated in January 1954

W. John Hutchins
jhutchins@beeb.net; http://ourworld.compuserve.com/homepages/WJHutchins

Abstract. The public demonstration of a Russian-English machine translation system in New York in January 1954 – a collaboration of IBM and Georgetown University – caused a great deal of public interest and much controversy. Although a small-scale experiment of just 250 words and six 'grammar' rules it raised expectations of automatic systems capable of high quality translation in the near future. This paper describes the system, its background, its impact and its implications.

1 The impact

On the 8th January 1954, the front page of the *New York Times* carried a report of a demonstration the previous day at the headquarters of International Business Machines (IBM) in New York under the headline "Russian is turned into English by a fast electronic translator":

> A public demonstration of what is believed to be the first successful use of a machine to translate meaningful texts from one language to another took place here yesterday afternoon. This may be the cumulation of centuries of search by scholars for "a mechanical translator."

Similar reports appeared the same day in many other American newspapers (*New York Herald Tribune, Christian Science Monitor, Washington Herald Tribune, Los Angeles Times*) and in the following months in popular magazines (*Newsweek, Time, Science, Science News Letter, Discovery, Chemical Week, Chemical Engineering News, Electrical Engineering, Mechanical World, Computers and Automation*, etc.) It was probably the most widespread and influential publicity that MT has ever received. The experiment was a joint effort by two staff members of IBM, Cuthbert Hurd and Peter Sheridan, and two members of the Institute of Languages and Linguistics at Georgetown University, Leon Dostert and Paul Garvin.

2 The background

Léon Dostert had been invited to the first conference on machine translation two years before in June 1952. He had been invited for his experience with mechanical aids for translation. Dostert had been Eisenhower's personal interpreter during the war, had been liaison officer to De Gaulle, and had worked for the Office of Strategic Services (predecessor of the Central Intelligence Agency). After the war he designed and installed the system of simultaneous interpretation used during the Nuremberg war crimes tribunal, and afterwards at the United Nations. In 1949 he was invited to Georgetown University to establish the Institute of Languages and Linguistics at the University's School of Foreign Service for training linguists and translators primarily for government service [10].

Dostert went to the conference as a sceptic but returned as an enthusiast determined to explore the possibilities of machine translation. It was his conviction that MT needed to demonstrate its feasibility in a practical experiment. For obvious political reasons Dostert decided that the demonstration should translate from Russian into English; the lack of knowledge about activities in the Soviet Union was already a major concern in US government circles.

Dostert contacted a personal acquaintance, Thomas J. Watson, founder of IBM, and they agreed to collaborate. The project was headed by Cuthbert Hurd, director of the Applied Sciences Division at IBM, and Dostert himself. The linguistic side of the experiment was the work of Garvin, a Czech linguist (associate professor) at the Institute – see Montgomery [6] for a biography. The computer programming was done by Peter Sheridan, staff member of IBM.

The Georgetown pair decided to demonstrate translations on a small number of sentences from organic chemistry and some others on general topics, which would illustrate some grammatical and morphological problems and give some idea of what might be feasible in the future. The experiment was to be small, with a vocabulary of just 250 lexical items (stems and endings) and a limited set of just six rules.

3 The demonstration

Reports of the demonstration appeared under headlines such as "Electronic brain translates Russian", "The bilingual machine", "Robot brain translates Russian into King's English", and "Polyglot brainchild" – at the time computers were commonly referred to as 'electronic brains' and 'giant brains' (because of their huge bulk).

The newspapermen were much impressed:

> In the demonstration, a girl operator typed out on a keyboard the following Russian text in English characters: "Mi pyeryedayem mislyi posryedstvom ryechi". The machine printed a translation almost simultaneously: "We transmit thoughts by means of speech." The operator did not know Russian. Again she types out the meaningless (to her) Russian words: "Vyelyichyina ugla opryedyelyayatsya otnoshyenyiyem dlyini dugi k radyiusu." And the machine translated it as: "Magnitude of angle is determined by the relation of length of arc to radius." (New York Times)

It appears that the demonstration began with the organic chemistry sentences. Some of these were reported, e.g.

> The quality of coal is determined by calory content
>
> Starch is produced by mechanical method from potatoes.

but the journalists were clearly much more impressed by those on other topics:

> And then just to give the electronics a real workout, brief statements about politics, law, mathematics, chemistry, metallurgy, communications, and military affairs were submitted in the Soviet language... (Christian Science Monitor)

All the reports recognised the small scale of the experiment but they also reported future predictions from Dostert:

> "Those in charge of this experiment," the professor continued, "now consider it to be definitely established that meaning conversion through electronic language translation is feasible." [and] the professor forecast that "five, perhaps three, years hence, interlingual meaning conversion by electronic process in important functional areas of several languages may well be an accomplished fact." (Christian Science Monitor)

He made other projections and predictions:

> 100 rules would be needed to govern 20,000 words for free translation… Eventually, the machine will be able to translate from Russian: "She taxied her plane on the apron and then went home to do housework." In such a sentence with double-meaning words, the machine will be able to tell what meaning of apron and taxi would be needed in that particular context. (New York Herald Tribune)

Whether these were Dostert's words is not known, but obviously expectations were high. Some wider implications, both for linguistics as well as for translation, were also expressed. Neil Macdonald [4] gave a sound assessment:

> Linguists will be able to study a language in the way that a physicist studies material in physics, with very few human prejudices and preconceptions... The technical literature of Germany,

> Russia, France, and the English-speaking countries will be made available to scientists of other countries as it emerges from the presses... But of course, it must be emphasized that a vast amount of work is still needed, to render mechanically translatable more languages and wider areas of a language. For 250 words and 6 syntactical structures are simply a "Kitty Hawk" flight. [4]

A number of reports picked up the observation made by the developers that the expensive 701 computer was "'overdesigned' for language translation; it has too many functions not essential to this task that were built in to solve problems in astronomy and physics" [7]. It was expected that MT would require special-purpose machines.

4 The processes

Most of the reports are illustrated with a photograph of a punched card with a Russian sentence; and many have photographs of the machines and of the Georgetown and IBM personnel. But they gave few hints of how the system worked.

The most common references were to rules for inversion, all using the example of Russian *gyeneral mayor*, which has to come out in English as *major general*. One gave some idea of the computer program:

> The switch is assured in advance by attaching the rule sign 21 to the Russian *gyeneral* in the bilingual glossary which is stored in the machine, and by attaching the rule-sign 110 to the Russian *mayor*. The stored instructions, along with the glossary, say "whenever you read a rule sign 110 in the glossary, go back and look for a rule-sign 21. If you find 21, print the two words that follow it in reverse order (Journal of Franklin Institute, March 1954)

A few explained how rules selected between alternative translations:

> The word root "ugl" in Russian means either "angle" or "coal" depending upon its suffix. This root is stored in the form of electrical impulses on a magnetic drum together with its English meanings and the Garvin rules of syntax and context which determine its meaning. The code is so set up so that when the machine gets electrical impulses via the punched cards that read "ugla" it translates it as "angle", when "uglya" the translation is "coal". (New York Herald Tribune)

It is doubtful whether newspaper readers would have gained much understanding from these brief explanations. However, some of the weeklies went into much more detail. Macdonald's report [4] included a list of the six rules, a flowchart of the program for dictionary lookup and a table illustrating the operation of the rules on a sample sentence.

5 The computer

An illuminating account of the computational aspects of the experiment is given in the contemporary article by Peter Sheridan [9]. As the first substantial attempt at non-numerical programming, every aspect of the process had involved entering quite unknown territory. Decisions had to be made on how alphabetic characters were to be coded, how the Russian letters were to be transliterated, how the Russian vocabulary was to be stored on the magnetic drum, how the 'syntactic' codes were to operate and how they were to be stored, how much information was to go on each punched card, etc. Detailed flow charts were drawn up for what today would be simple and straightforward operations, such as the identification of words and their matching against dictionary entries.

The IBM 701-type machine had been developed for military applications and was first installed in April 1953. Like other computers of the day its main tasks were the solution of problems in nuclear physics, rocket trajectories, weather forecasting, etc. It was hired out initially at $15,000 per month, and later sold at $500,000 – and was at that time only one of about 100 general-purpose computers in existence. Its huge size was impressive; it was likened to "an assortment of 11 complicated electronic units, not unlike modern kitchen ranges, connected by cables to function as a unit" and "which occupy roughly the same area as a tennis court." [7]. A similar-sized machine, the 702, was also developed for business applications. Its successor in late 1955 was the 704 model, a substantial improvement on the 701 and which sold in large numbers.

The 701 could perform 33 distinct operations: addition, subtraction, multiplication, division, shifting, transfers, etc. – all coded in 'assembly language'. Multiplication was performed at 2,000 per second. It consisted of two types of storage. Electrostatic (high-speed) storage was in the form of a bank of cathode ray tubes; each unit could accommodate up to 2048 "full words", where a "full word" comprised 35 bits (binary digits) and one sign bit – 36 bits in all. Each full word could be split (stored) as two "half words", each of 17 bits and one sign bit. Although the 701 had two electrostatic units, only one was used in the MT experiment. Average access time was 12 microseconds. The second type of storage (with lower access speed, 40 milliseconds) was a magnetic drum unit comprising four 'addressable' drums, each accommodating up to 2048 'full words'. The magnetic drum was used to store dictionary information; the reading and writing rate was 800 words per second.

Input to the 701 was by card reader. Information from 80 column cards – only 72 columns were used, so each card had a maximum capacity of 72 upper case (capital letter) alphabetic or numeric characters – could be read and converted to internal binary code at a rate 150 per minute. Output was by a line printer (also capital letters only) at a rate of 150 lines per minute.

The program used a seven-bit code for characters: six bits for distinguishing 40 alphanumeric and other characters, plus one sign bit used for various tests (see below). This means that each "full word" location could contain up to five characters.

The Russian-English dictionary was input by punched cards and stored on the (low-speed) magnetic drum. The Russian word and the English equivalents (two maximum) were stored on consecutive locations, separated by 'full words' containing zeros. They were followed by the so-called diacritics on consecutive drum locations. Each 'word' included a 'sign bit', either + or -, which indicated whether the entry was for a stem or for an ending, respectively.

Sentences were read into the electrostatic storage, separated by strings of zero-filled 'words'. The input words were then each looked up in the drum storage, first by consultation of a "thumb index" which gave the address (location) of the first word in the dictionary with the same initial letter. The lookup routine searched for the longest matching string of characters (whether complete word or stem plus hyphen), extracted the (two) English equivalents onto a separate area of the store, and copied the diacritics onto another area of the store. A special area was also set aside for the temporary (erasable) location of word-endings. Each of these areas and addresses would have to be specified either directly (specifically by store address) or indirectly (using variables) in the program (called Lexical Syntax Subprogram). Sheridan describes the operations of comparison in terms of successive and repeated processes of logical multiplication, addition and subtraction using 'masks' (sequences of binary digits). When a 'diacritic' indicated that either the first English equivalent or the second English equivalent was to be selected, then the program went back to the addresses to the separate store area, and transferred the one selected to a (temporary) print-out area of the electrostatic store.

6 The six rules

Before the system was given to Sheridan of IBM for programming, it was tested by hand on a set of cards by people who did not know Russian ([2], [4], [5]. The sentences were written in Russian characters on the cards. The test involved finding the corresponding cards for each word and following the instructions. The instructions for the Operational Syntax Subprogram were formulated in language such as:

> Rule 1. Rearrangement. If first code is '110', is third code associated with preceding complete word equal to '21'? If so, reverse order of appearance of words in output (i.e., word carrying '21' should follow that carrying '110') – otherwise, retain order. In both cases English equivalent I associated with '110' is adopted.
> Rule 2. Choice-Following text. If first code is '121', is second code of the following complete, subdivided or partial (root or ending) word equal to '221' or '222'? If it is '221', adopt English equivalent I of word carrying '121'; if it is '222', adopt English equivalent II. In both cases, retain order of appearance of output words.
> Rule 3. Choice-Rearrangement. If first code is '131', is third code of preceding complete word or either portion (root or ending) of preceding subdivided word equal to '23'? If so, adopt English equivalent II of word carrying '131', and retain order of appearance of words in output – if not, adopt English equivalent I and reverse order of appearance of words in output.
> Rule 4. Choice-Previous text. If first code is '141', is second code of preceding complete word or either portion (root or ending) of preceding subdivided word equal to '241' or '242'? If it is '241', adopt English equivalent I of word carrying '141'; if it is '242' adopt English equivalent II. In both cases, retain order of appearance of words in output.
> Rule 5. Choice-Omission. If first code is '151', is third code of following complete word or either portion (root or ending) of following subdivided word equal to '25'? If so, adopt English equivalent II of word carrying '151'; if not, adopt English equivalent I. In both cases, retain order of appearance of words in output.
> Rule 6. Subdivision. If first code associated with a Russian dictionary word is '***', then adopt English equivalent I of alternative English language equivalents, retaining order of appearance of output with respect to previous word.

According to Sheridan, the rules formulated in this manner were easily converted into program code for the 701 computer. Sheridan's account [9] includes a flowchart of the processes of dictionary lookup and 'operational syntax'.

7 The sentences

The most detailed account of the linguistic operations is given by Garvin [3] in a retrospective evaluation of the experiment and its significance. He includes 137 dictionary entries for words, stems and endings; and 49 of the original Russian sentences (in a transliteration scheme devised for the experiment)

Most of the "more than 60" sentences in the demonstration concerned topics of organic chemistry. These were intended to illustrate the variety of sentence patterns which the rules could deal with, and nouns and verbs occurring in different roles. Some examples are:

(1) (a) They prepare TNT; (b) They prepare TNT out of coal; (c) TNT is prepared out of coal; (d) TNT is prepared out of stony coal; (e) They prepare ammonite; (f) They prepare ammonite out of saltpeter; (g) Ammonite is prepared out of saltpeter.

(2) (a) They obtain gasoline out of crude oil; (b) Gasoline is obtained out of crude oil; (c) They obtain dynamite from nitroglycerine; (d) Ammonite is obtained from saltpeter; (e) Iron is obtained out of ore; (f) They obtain iron out of ore; (g) Copper is obtained out of ore.

(3) (a) They produce alcohol out of potatoes; (b). Alcohol is produced out of potatoes; (c). They produce starch out of potatoes; (d). Starch is produced out of potatoes; (e) Starch is produced by mechanical method from potatoes.

(4) (a) The quality of coal is determined by calory content; (b). The price of potatoes is determined by the market; (c). Calory content determines the quality of coal; (d) Calory content determines the quality of crude oil; (e) The quality of crude oil is determined by calory content; (f) The quality of saltpeter is determined by chemical methods.

There were also, as the newspapers reported, a number (about 20) of sentences of relative greater complexity on other topics. It is notable that the journalists picked on these sentences in their reports rather than the less interesting chemistry examples.

(5) Magnitude of angle is determined by the relation of length of arc to radius.
(6) Angle of site is determined by optical measurement.
(7) We transmit thoughts by means of speech.
(8) Military court sentenced the sergeant to deprival of civil rights.
(9) A commander gets information over a telegraph.
(10) Penal law constitutes an important section of legislation.
(11) Vladimir appears for work late in the morning.
(12) International understanding constitutes an important factor in decision of political questions.

8 The dictionary and the linguistic operations

Dictionary entries (for both stems and endings) included three codes. The first code, Program Initiating Diacritic (PID) was one of '110', '121', '131', '141'or '151'. The second and third codes were Choice Determining Diacritics (CDD). The second code (CDD_1) was one of '221', '222', '241', '242'. The third code (CDD_2) was one of '21', '23', '25'.

The operations can be illustrated with the following table for one of the journalists' favourite sentences: *Magnitude of angle is determined by the relation of length of arc to radius* (5: translation of величина угла определяется отношением длины дуги к радиусу). The table is adapted from [9]; it was reproduced in [4], and [7].

Russian input	English equivalents Eng_1	Eng_2	1st code (PID)	2nd code (CDD_1)	3rd code (CDD_2)	rule
vyelyichyina	magnitude	---	***	***	**	6
ugl-	coal	angle	121	***	25	2
-a	of	---	131	222	25	3
opryedyelyayetsya	is determined	---	***	***	**	6
otnoshyenyi-	relation	the relation	151	***	**	5
-yem	by	---	131	***	**	3
dlyin-	length	---	***	***	**	6
-i	of	---	131	***	25	3
dug-	arc	---	***	***	**	6
-i	of	---	131	***	25	3
k	to	for	121	***	23	2
radyius-	radius	---	***	221	**	6
-u	to	---	131	***	**	3

The first word величина ('vyelyichyina') has just one English equivalent (*magnitude*) and its PID (***) refers to rule 6 – i.e. the result is simply copied out and there is no change of word order. The second entry is the stem 'ugl-', which initiates rule 2 by PID '121' searching for code '221' or'222' in the CDD_1 of the following entry; it finds '222', therefore the second equivalent (Eng2) of 'ugl-' is chosen (*angle*). The next entry, the suffix '-a' (of угла) with PID '131', triggers rule 3 searching for '23' in the CDD_2 of the preceding entry, which since it is not there prompts selection of the first equivalent (Eng_1 *of*) and reversal of word order (i.e. *of angle*). The next entry is the verb form 'opryedyelyayetsya' with PID '***', hence rule 6, selection of first equivalent (*is determined*) and no change of word order. The next word (отношением) has been subdivided: the stem ('otnoshyenyi-') initiates rule 5 (PID

'151') searching for code '25' in the CDD_2 of the following entry (i.e. its ending) or of the next following word (stem or ending). The '25' is found in the ending '-i' of the word (длины), so the second equivalent (Eng_2) of 'otnoshyenyi-' is selected (*the relation*) and the word order is retained. The process now continues with the next entry after 'otnoshyenyi-', i.e. its instrumental ending (-yem'), where the PID '131' initiates rule 3, a search for '23' in preceding entries. None is found so the first equivalent (Eng_1: *by*) is chosen and word order is reversed (i.e. producing *by the relation*). Next comes the entry 'dug-' (stem of дуги) with PID '***', i.e. selection of Eng_1 (*arc*), and no change of order; then its ending ('-i') with PID '131' (rule 3) searching for '23' in preceding entries and failing, so Eng_1 (*of*) is chosen and word order is reversed (i.e. *of arc*). The process now comes to the preposition 'k' which has two equivalents – out of the many possible translations of the Russian word – viz. *to* and *for*. Rule 2 (PID '121') searches for '221' or '222' in the CDD_1 of the following stem or ending, and finds '221' in the relevant CDD of 'radyius-'; thus, the first equivalent (Eng_1: *to*) is selected. The entry for 'radyius-' (PID '***') initiates no change. Finally, the PID '131' of its ending '-u' searches for '23' in one of the two preceding entries, finds it in the entry for the preposition 'k', selects the second equivalent (Eng_2), i.e. blank, and retains word order.

This detailed explication of the process gives rise to contradictory impressions. On the one hand there is the complexity of coding for maintaining or inverting word order. On the other there is the care to establish rules and codes of some generality, i.e. not specific to particular sentences and constructions. More specific comments follow.

Firstly, the inclusion of the full verb ('opryedyelyayetsya') was followed by Garvin for all the sentences listed above (1)-(4) which have similar Russian reflexive forms translated in English by passives ('is prepared', 'is obtained', 'is produced', 'is determined') – each occurring only in the singular. The same option is made for their corresponding non-reflexive forms (translated as 'they prepare', 'they obtain', 'they produce' – all plural – and 'determines'). Thus problems of verb morphology and choice of translation are avoided.

Russian has no articles, so the insertion of *the*, *a* and *an* in English are problems for any MT system. Garvin has reduced them by including very few in the sample sentences. They are 'the price' and 'the quality' in several sentences, which, as in this example, are derived by rule '151' and a search for '25' in a following entry. The other example is 'the science', which however initiates rule '141' and a search for code '23'. The different treatment does not reveal a clear methodology.

The example above illustrates also the method of distinguishing homonyms which newspapers reported, namely the selection of *angle* or *coal* for 'ugl-'. Rule 2 (initiated by PID '121') searches for '221' or '222' in the following entry. When it is '-a', '221' is found and the result is the choice of *angle*; when it is '-ya', '222' is found and the choice is *coal*. In fact, the original Russian is not strictly a homonym, there are two separate words: угол (*corner* or *angle*) and уголь (*coal*). Garvin's procedure is based on the fact that the genitive for угол is угола and the genitive for уголь is уголя.

The example also illustrates Garvin's approach to the ambiguity (or rather multiple possible translations) of prepositions. Garvin reduces the problem by providing just two equivalents, which are selected by occurrences of '221' or '222' in the ending of the following noun – which can certainly be justified linguistically.

A number of sentences contain instrumental phrases (*by chemical process*, *by mechanical method*). Each are generated by the production of the preposition *by* from the case ending of the adjective and then translating the following noun as if it had no case ending. This would be regarded then and now as abnormal, since adjectival forms are held to be dependent on the nouns they are modifying, so we would expect *by* to be generated from the noun ending. For example, the phrase *by optical measurement*:

Russian word	Eng_1	Eng_2	PID	CDD_1	CDD_2
optichesk-	optical		***	***	**
-yim	by	---	131	***	23
yizmyeryenyi-	measurement		***	***	**
-yem	by	---	131	***	**

The entry 'optichesk-' with PID '***' is printed out (*optical*); the suffix '-yim' initiates rule '131', fails to find CDD_2 '23', outputs Eng_1 (*by*) and inverts word order (i.e. *by optical*). The next entry 'yizmyeryenyi-' has one equivalent (*measurement*); and its suffix '-yim' invokes rule '131', finds a CDD_2 '23' in the preceding subdivided word (the entry for '-yim'), selects Eng_2 ('---') and retains word order.

From these descriptions it is apparent that much of the variety of sentences is derived from a fairly restricted set of interlocking rules and codes operating on fixed patterns into which nouns and phrase constructions are slotted.

Many of the operations are quite clearly specific to the particular words and sentences in the sample, and the rules are applied as appropriate in specific instances – this is particularly obvious in the non-chemistry sentences. There was no attempt to align rules with specific linguistic features. In particular, there was no analysis in terms of grammatical categories (noun, verb, adjective) and no representation of either agreement relations, or dependency relations, or phrase/clause structures.

9 The consequences

The Russians had seen reports of the January event and the time was propitious (after the 'thaw' following Stalin's death) for the Russians to engage in the development of computers and their applications. The Institute of Precise Mechanics and Computer Technology had just completed development of BESM, the first working Russian computer, and machine translation was to be among its first applications, under the direction of Yurij Panov. By early 1956 it was ready to demonstrate a prototype system, which in many respects followed the design of the IBM-Georgetown system, with a basic set of rules for substitution, movement and morphological splitting [8].

At Georgetown itself, however, despite the widespread publicity there was no official support for further research until a grant in June 1956 from the National Science Foundation, stimulated, as it seemed at the time, by the Soviet interest [5]. The funds were in fact from the Central Intelligence Agency – Dostert had worked for its predecessor (Office for Strategic Services) and was a good friend of its director Allen Dulles. A full-scale project for Russian-English translation was organized with more than twenty researchers [5]. Initially two groups were set up: one for developing a dictionary, the other for linguistic analysis. After examining the coding of the 1954 experiment for a few months, the group decided to abandon continuation on these lines – a fact often forgotten by later critics of the Georgetown activity. There emerged a considerable divergence of opinions, and Dostert decided to give each of the proposed methods a chance to show its capability in 'free competition'. By January 1957 there were four groups, known as 'code-matching', 'syntactic analysis', 'general analysis', and 'sentence-by-sentence'. The first group, headed by Ariadne Lukjanow, assigned codes to dictionary entries which indicated grammatical and association functions and which were compared and matched during analysis. The second group under Garvin developed a method of dependency syntactic analysis later known as 'fulcrum method'. The third group under Michael Zarechnak formulated a method of sentence analysis at various levels (morphological, syntagmatic, syntactic), i.e. a variant of 'phrase structure' analysis. The fourth 'group' was a one-man project of French-English translation by A.F.R.Brown where procedures developed first for one sentence were tested on another, more procedures were added, tested on another sentence, further procedures were added, tested, and so forth. In due course, Lukjanow and Garvin left the Georgetown project to continue elsewhere, and the 'general analysis' method was adopted together with Brown's computational techniques [5], [10], [11].

10 The assessments

A year after the demonstration, Dostert [2] gave an assessment of the significance of the experiment, and suggested future ideas for MT development. In his opinion, the experiment (a) "has given practical results by doing spontaneous, authentic, and clear translation", (b) showed that "the necessity of pre- and post-editing has not been verified", (c) demonstrated that "the primary problem in mechanical translation… is a problem of linguistic analysis…", (d) formulated "the basis for broader systematic lexical coding", defining "four specific areas of meaning determination… from which fruitful results may be expected", (e) developed a "functional coding system, permitting the preparation of functional, subfunctional and technical lexicons… reducing the magnitude of the coding requirements and thereby… the extent of storage needs", and (f) provided a "theory for the development of a general code for the mechanical formulation of multilingual syntax operations". These were major claims, and clearly not justified on the basis of this first small-scale experiment. Rather these were expectations Dostert had for later research.

The retrospective assessment by Garvin [3] was much more modest than Dostert's. In this somewhat defensive account of his early essay in MT research, Garvin freely admitted the shortcomings of the experiment. The limitations were the consequence of restricting the algorithm to "a few severely limited rules, each containing a simple recognition routine with one or two simple commands." Nevertheless, in Garvin's view, the experiment 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."

Garvin summarised the limitations as: the restriction of the search span to immediately adjacent items, the restriction of target words to just two possibilities, and the restriction of rearrangements to two immediately adjacent items. The choice of target language equivalents was restricted to those which were idiomatic for the selected sentences only. The limitation of the procedure for Russian case endings was severe: either a case suffix was not translated at all or it was translated by one "arbitrarily assigned" English preposition. Further limitations were highlighted by Michael Zarechnak [11], a member of the Georgetown group. None of the Russian sentences had negative particles; all were declaratives; there were no interrogatives or compound sentences (coordinate or subordinate clauses); and nearly all the verbs were in the third person.

Does this mean that the experiment was fixed, a deception? Naturally members of the Georgetown group deny it – pointing out that the program "was thoughtfully specified and implemented; the program ran, the translation was generated according to the program, which was developed based on… linguistic principles." [6]. This was basically true, however, only for the chemistry sentences and the rules and dictionary entries which were applied for their translation. Further chemistry sentences could clearly have been treated, with an expansion of the dictionary – but only as long as the sentences conformed to the patterns of those in the sample. There are many chemistry sentences that would obviously not be covered by the rules. Although organic chemistry might constitute a sublanguage and its vocabulary might be captured in a 'micro-glossary' (as others advocated at the time) with few ambiguities, this program in 1954 did not cover all of the field, nor indeed a substantial proportion of it. As for the non-chemistry sentences, these were clearly produced by dictionary entries and codes specifically designed for this particular demonstration; and there could have been no question of expanding general coverage on the lines of this program – as indeed was found in the later research at Georgetown.

The limitations of the experiment made it possible for the output to be impressively idiomatic, and would have suggested to many observers (not only reporters) that continued experiments on the same lines would lead to systems with larger vocabularies and even better quality. On the other hand, the experiment drew attention to the importance of the linguistic problems, and in particular that translation was not a trivial task even for the largest computer, as a contemporary commentator, Ornstein [7], remarked:

> the formulation of the logic required to convert word meanings properly, even in a small segment of two languages, necessitated as many instructions to the computer as are required to simulate the flight of a guided missile.

In truth, neither Dostert nor Garvin claimed much more than that it was a first effort (a "Kitty Hawk" experiment) – not even a prototype system. In later years, they might well have agreed that the demonstration had been premature; certainly it was made public at a stage much earlier than other contemporary MT researchers would have contemplated. However, there was another probably more important aim for Dostert; it was to attract funds for further research at Georgetown, and it succeeded.

11 The implications

Undoubtedly, the wrong impression had been given that automatic translation of good quality was much closer than was in fact the case. Sponsorship and funding of MT research in the following years were more liberal (and unquestioning) than they ought to have been. The results in the next 10 years were inevitably disappointing, and as a consequence, the funders set up an investigation committee, ALPAC [1]. One of the principal arguments used by ALPAC was that MT output had to be extensively post-edited. They pointed out that the Georgetown-IBM output was of a quality that had no need of editing while that of later Georgetown systems did. The mistake of ALPAC was to ignore the preliminary nature of the 1954 experiment, that it had been specifically designed for a small sample of sentences, that it had not been a 'prototype' system but a 'showcase' intended to attract attention and funds, and that comparisons with full-scale MT systems were invalid.

In 1954, when other MT groups saw reports of Dostert's demonstration they were disparaging or dismissive. They disliked three things. One was the conduct of research through newspapers; another was the exaggerated publicity given by journalists to an obviously incomplete system; and a third was the passing-off as true 'translations' sentences which could only have been extracted as wholes from computer memories. Other MT groups were far from even thinking of demonstrating their results – and were unprepared to do so for many years to come.

It was only the first of demonstrations by the Georgetown group. Later ones were undoubtedly more genuine – systems had not been 'doctored' – but the suspicion of other MT groups was that they were not all they appeared. Such suspicions continued to haunt the Georgetown group throughout its existence and have coloured the judgements of later commentators.

It does have to be admitted, however, that the Georgetown-IBM demonstration has not been the only example of a MT system being 'doctored' for a particular occasion. In subsequent years it was not uncommon for demonstrated systems to introduce grammar and vocabulary rules specifically to deal with the sentences of a particular text sample, with the aim of showing their system in the best possible light.

In recent years MT researchers have been much more circumspect when demonstrating experimental systems and have been less willing to indulge in speculations for journalists. The painful lessons of the Georgetown-IBM demonstration seem to have been learned by MT researchers. On the other hand, some vendors of systems have a more 'liberal' attitude: many MT systems are being publicised and sold (particularly on the internet) with equally exaggerated claims and perhaps with equally damaging impact for the future of machine translation.

The historical significance of the Georgetown-IBM demonstration remains that it was an actual implementation of machine translation on an operational computer. Before 1954, all previous work on MT had been theoretical. Considering the state of the art of electronic computation at the time, it is remarkable that anything resembling automatic translation was achieved at all. Despite all its limitations, the demonstration marked the beginning of MT as a research field seen to be worthy of financial support.

References

1. ALPAC: *Languages and machines: computers in translation and linguistics*. A report by the Automatic Language Processing Advisory Committee, Division of Behavioral Sciences, National Academy of Sciences, National Research Council. Washington, D.C.: National Academy of Sciences – National Research Council. (1966)
2. Dostert, Leon E. : 'The Georgetown-I.B.M. experiment', in Locke, W.N. and Booth, A.D. (eds.) *Machine translation of languages.* Cambridge, Mass.: M.I.T.Press (1955), 124-135.
3. Garvin, Paul: 'The Georgetown-IBM experiment of 1954: an evaluation in retrospect', *Papers in linguistics in honor of Dostert*. The Hague: Mouton (1967), 46-56; reprinted in his *On machine translation* . The Hague: Mouton (1972), 51-64.
4. Macdonald, Neil: 'Language translation by machine - a report of the first successful trial', *Computers and Automation* 3 (2), 1954, 6-10.
5. Macdonald, R.R.: 'The history of the project' in: *General report 1952-1963*, ed. R.R.Macdonald. Washington, DC: Georgetown University Machine Translation Project (1963), 3-17. (Georgetown University Occasional Papers on Machine Translation, 30)
6. Montgomery, Christine A.: 'Is FAHQ(M)T impossible? Memoirs of Paul Garvin and other MT colleagues', in: W. John Hutchins (ed.) *Early years in machine translation: memoirs and biographies of pioneers*. Amsterdam/Philadelphia: John Benjamins (2000), 97-110.
7. Ornstein, Jacob: 'Mechanical translation: new challenge to communication', *Science* 122, 21 (October 1955), 745-748.
8. Panov, Yu.N.: *Automatic translation*. Translated… by R.Kisch. London: Pergamon. (1960)
9. Sheridan, Peter: 'Research in language translation on the IBM type 701', *IBM Technical Newsletter* 9, 1955, 5-24.
10. Vasconcellos, Muriel: 'The Georgetown project and Leon Dostert: recollections of a young assistant', in: W. John Hutchins (ed.) *Early years in machine translation: memoirs and biographies of pioneers*. Amsterdam/Philadelphia: John Benjamins (2000), 87-96.
11. Zarechnak, Michael: 'The history of machine translation', in: Bozena Henisz-Dostert, R. Ross Macdonald and Michael Zarechnak, *Machine translation*. The Hague: Mouton (1979), 20-28.