The Georgetown-IBM experiment demonstrated in January 1954

- John Hutchins

[Reference: http://www.hutchinsweb.me.uk/GU-IBM-2005.pdf]

Front page headlines

- Russian is turned into English by a fast electronic translator (New York Times)
  - 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.”
- Robot translates nimbly (Christian Science Monitor)
- It’s all done by machine (New York Herald Tribune)
  - A huge electronic “brain” with a 250-word vocabulary translated mouth-filling Russian sentences yesterday into simple English in less than ten seconds.
- Robot brain translates Russian into King’s English (Washington Times Herald)
- The bilingual machine (Newsweek)
- Polyglot brainchild (Chemical Weekly)

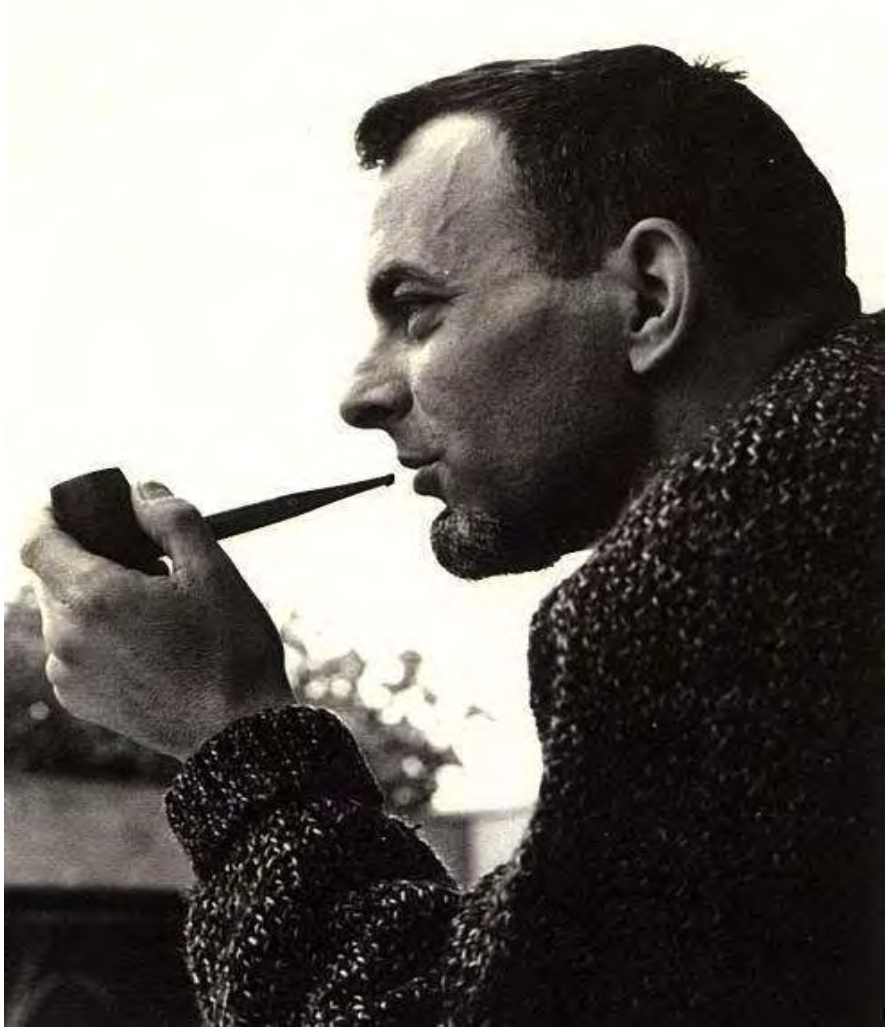
Origins of the experiment

- 1947: first discussions by Warren Weaver and Andrew Booth
- 1949 July: Weaver's memorandum
- 1952 June: Leon Dostert went to first MT conference, at MIT, invited for his expertise in setting up simultaneous interpretation systems at Nuremberg trial and at United Nations
- he went as a sceptic, but left convinced that MT needed a public demonstration of its feasibility
- contacted Thomas J.Watson of IBM, for project collaboration
- project led by Dostert (Georgetown) and Cuthbert Hurd (IBM)
- Peter Sheridan of IBM -- programming
- Paul Garvin of Georgetown -- language analysis
- small experiment: 250 words, 6 rules, limited sentences (just over 60)

Hurd, Dostert and Watson at the interface

IBM'S WATSON (right) AND FRIENDS:† For a mathematical wizard . . .

Paul Garvin



The ‘giant brains’

- Developed as mechanical (non-electronic) calculators during World War II
- ENIAC (Moore School, Univ. Pennsylvania, 1943-1946)
- EDVAC (Univ. Pennsylvania, first von Neumann-type stored program computer, 1944-1945, operational 1951)
- EDSAC (Univ. Cambridge, 1949)
- UNIVAC (first commercial computer, 1951)
- IBM 701 (launched 1952)

IBM 701

- Announced in 1952, originally called Defense Calculator
- most (19) went to US Defense Department or aerospace customers
- initially not sold outright; rental fees $15,000 a month; later sold at $500,000
- primary memory (Williams cathode ray tubes) stored up to 4096 36-bit words; intermediate storage (magnetic drum)
- 2,000 multiplications per second
- machine installed at IBM headquarters (New York), 31 December 1952
- IBM 702 for business users
- IBM 650, a smaller version: general-purpose, drum-based computer; delivered in 1954; eventually over one thousand installations (rental $3,500 a month)
- IBM 704 successor of 701 (late 1955)
- IBM 709 successor (upgrade) of 704 (January 1957); replaced by ‘transistorized’ 7090 (late 1959)

IBM 701 at New York headquarters

– "filling a room as big as a tennis court" (New York Herald Tribune)

Programming

- no higher level languages
  - the first (FORTRAN) came in 1956
  - first languages for ‘non-numerical’ computing in 1957 (COMIT) and in 1960 (LISP)
- all programming in either machine code, i.e. binary digits
- or in ‘assembly language’
- IBM 701 had 33 operations such as: addition, subtraction, multiplication, division, shifting, transfers
- 7-bit code used for each character

Punched cards

- 72-column cards (forerunner of ‘standard’ IBM 80-column cards), maximum capacity of 72 upper-case letters; converted to binary code at 150 per minute

- 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)

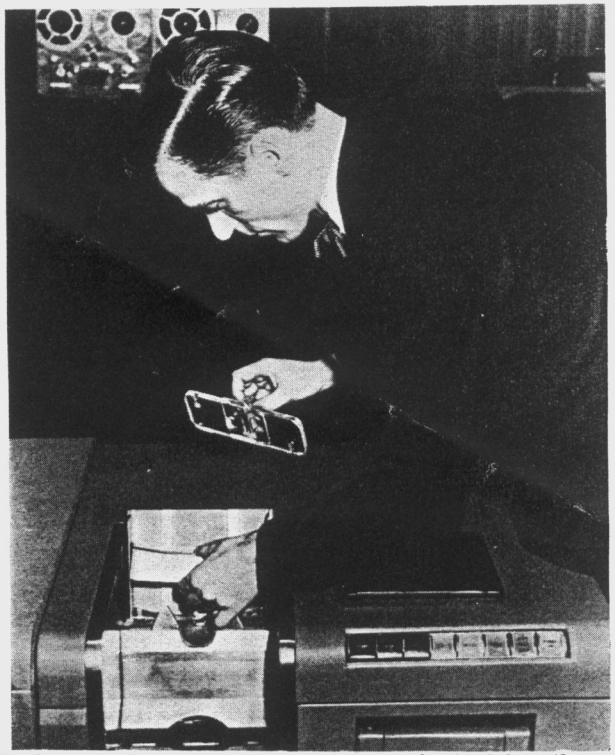
Punched card input

Sentences in Russian are punched into standard cards for feeding into the electronic data processing machine for translation into English

Data input

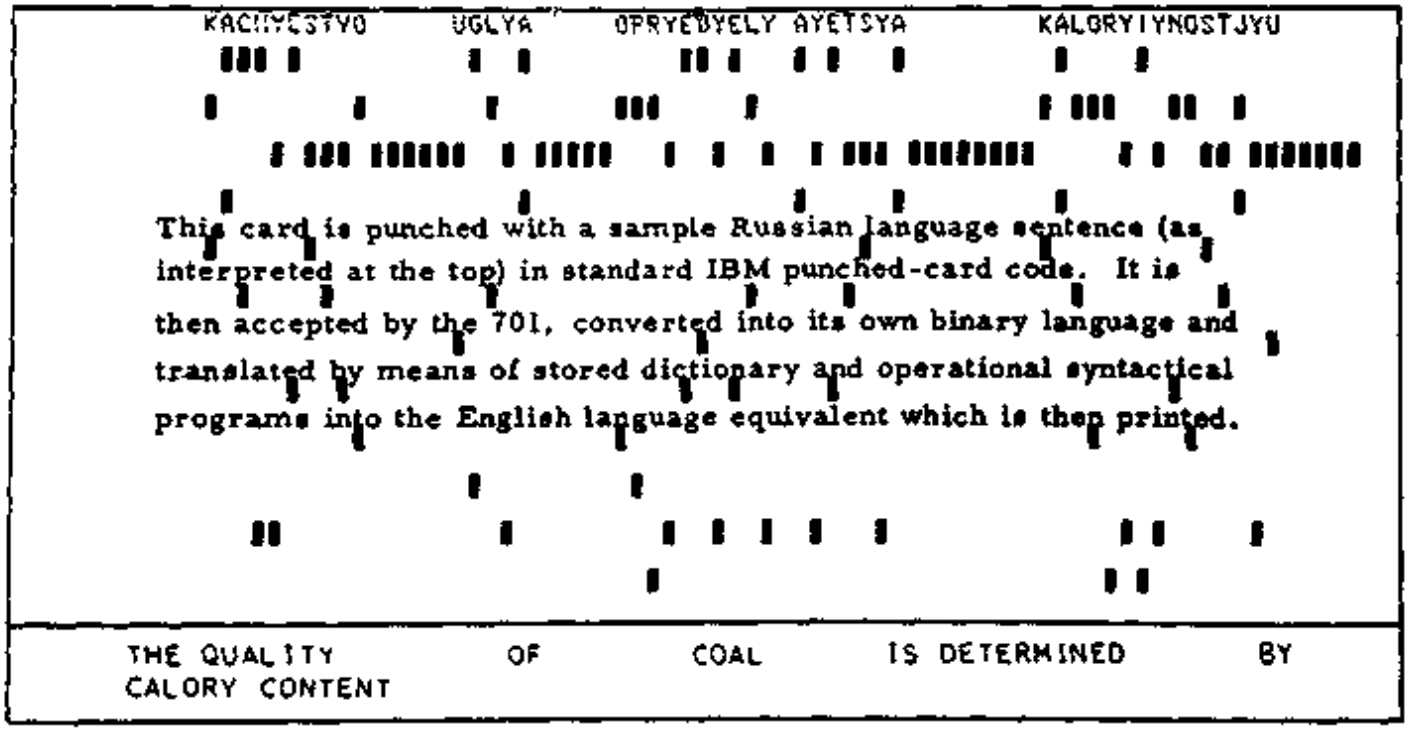
- Russian-English dictionary input on punched cards and stored in intermediate, low speed (magnetic drum) memory
- sentences on punched cards read into electrostatic (high speed) store
- input words of sentences each looked up in drum storage (using longest match principle)
- extracted English equivalents (with codes) copied into separate area of electrostatic store
- then sentence processed using codes

Data input



Cards containing sentences in Russian are inserted into the card reading unit. The reading device of the magnetic drum unit then "thumbs through" the dictionary record on it and comes up with the translation and pertinent syntax data

Sentence input on punch card



Specimen punched card and below a strip with translation, printed within a few seconds

Output [as reported]

- For nine silent seconds the machine mulled over the message. Then its automatic typewriter pounded out the English translation. (Newsweek)
- The "brain" didn't even strain its superlative versatility and flicked out its interpretation with a nonchalant attitude of assumed intellectual achievement. (Christian Science Monitor)
- To do all this the machine performs about 60,000 operations a sentence. During the demonstration yesterday it had two “nervous breakdowns”. Random errors crept in which automatically shut the 701 down. “She” didn't cry. (New York Herald Tribune)
- As lights flashed and motors whirred inside the “brain” the instrument’s automatic type-writer swiftly translated statements on politics, law, science and military affairs. Once the Russian words were fed to the machine no human mind intervened. (New York Herald Tribune)

- output on line printer at 150 lines per minute

Output on line printer

A continuous sheet of English-worded sentences comes up on the printing unit seconds after sentences, in Russian, have been fed into the computer

Predictions - hardware

- It is expected by IBM and Georgetown University, which collaborated on this project, that within a few years there will be a number of "brains" translating all languages with equal aplomb and dispatch. (Christian Science Monitor)
- Dr.Dostert also foresaw the day when simpler and cheaper machines than the $500,000 I.B.M. super-calculator – the 701 – could be used. He said the 701 is “overdesigned” for the language translation problem and has many functions not necessary in this project but which were built in to solve problems in astronomy and physics. (New York Herald Tribune)
- Dr.Hurd said that the corporation would now design a machine particularly fit for translating rather than for general computing utility. Such a device should be ready within three to five years, when the Georgetown scholars believe they can complete the “literary” end of the system. (New York Times)
- Dr.Dostert said that it will not be too long -- possibly three to five years -- when automatic text-reading machines will feed in Russian sentences automatically into the machines without punched-card intervention. (New York Herald Tribune)

Predictions - language translation

- “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)
- 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)
- From the viewpoint of the electronic ‘brain’, the language translation also has tremendous significance. It has been learned, for instance, that the formulation of logic required to convert word meanings properly even in a small segment of two languages necessitates two and a half times as many instructions to the computer as are required to simulate the flight of a guided missile. (Journal of the Franklin Institute)

Dictionary entries and coding

- segmentation
  - undivided: определяется (opryedyelyayetsya) [is determined], утром (utrom) [in the morning], понимание (ponyimanyiye) [understanding]
  - divided: длины: dlyin- [length] and -i [of]
- multiple meaning
  - угл- (ugl-) [coal] or [angle]; решени- (ryeshyenyi-) [solution] or [decision]
  - на (na) [on] or [for]
- articles
  - цена (tsyena) [price] or [the price]; наука (nauka) [a science] or [the science]
- codes (diacritics):
  - PID (Program Initiating Diacritic), 3-digit code pointing to a rule
  - CDD_1 (Choice Determining Diacritic), 3-digit determining lexical selection
  - CDD_2 (Choice Determining Diacritic), 2-digit determining treatment of case endings

The six rules (informal)

- 1. Rearrangement: the order of words may be inverted.
- 2. The choice between target equivalents is determined by an indication ('diacritic') in the following word
- 3. The choice of target words is determined by an indication (‘diacritic’) in the preceding word
- 4. The choice is determined by the previous word or its ending, allowing for insertions (e.g. of articles)
- 5. Choice determined by following word, allowing for omissions.
- 6. No problems of selection: there is one-to-one equivalence of source and target words, and the word order of the source is followed.

The rules in code: in manual simulation and as given to Sheridan

- 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.

Sheridan's flowchart for syntax analysis (part of one rule)

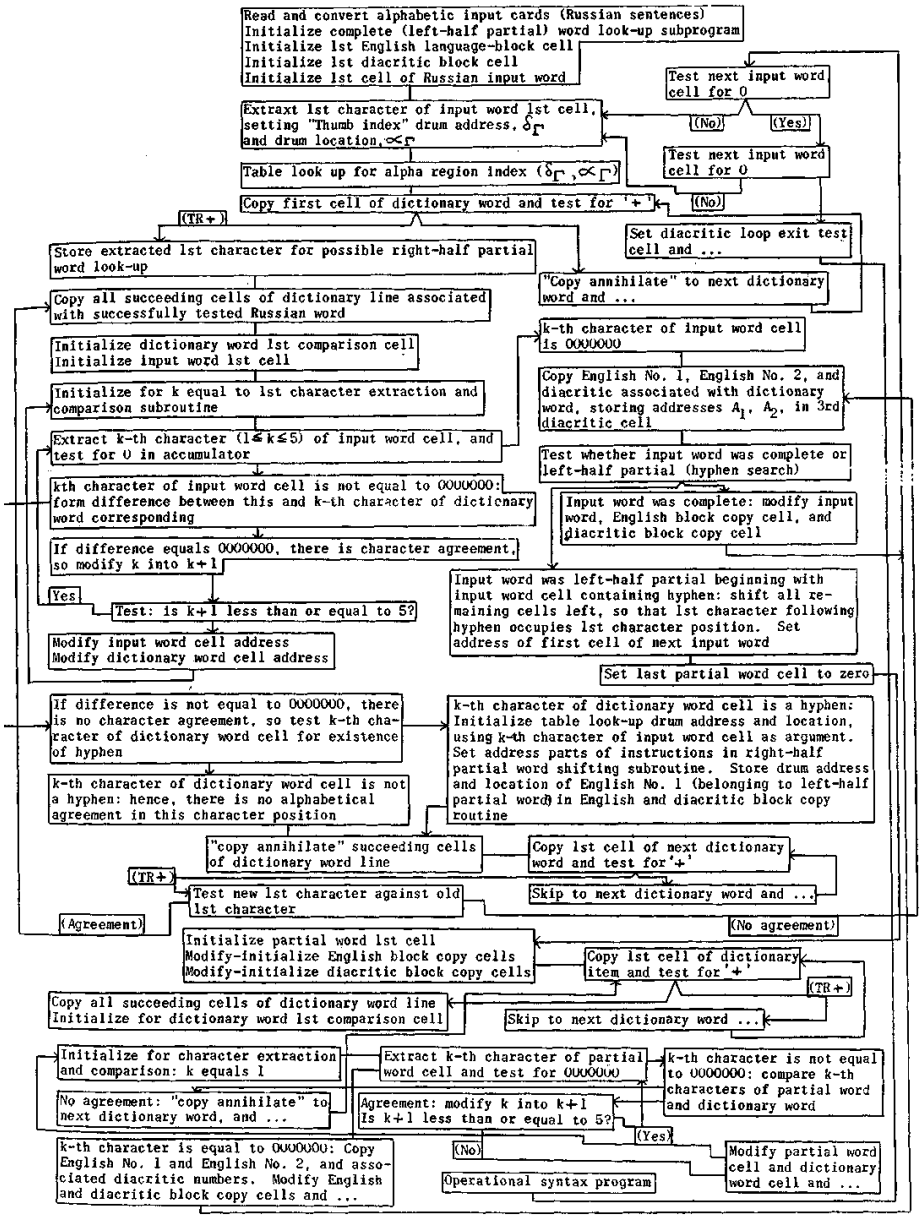


Figure 1 -- Dictionary Syntax Flow Chart

Some of the organic chemistry sentences

- (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.

More chemistry sentences (2)

- (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.

The non-chemistry sentences

- (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.

- Process will be illustrated with example sentence (5). Russian input:
- vyelyichyina ugla opryedyelyayetsya otnoshyenyiyem dlyini dugi k radiusu
- величина угла определяется отношением длины дуги к радиусу

Dictionary output for example sentence

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

The processing

- Line 1: *vyelyichyina* has just one equivalent (*magnitude*) and its PID (***) refers to rule 6, i.e. no change of order.
- Line 2: *ugl-* has two equivalents (*coal* and *angle*); its PID (121) refers to rule 2 searching for '221' or '222' in the CDD_1 of the following entry (viz. *-a*); it finds '222' and so the second equivalent (*angle*) is chosen.
- Line 3: *-a* has two equivalents (*of* and --, i.e. blank); its PID (131) refers to rule 3 searching for '23' in the CDD_2 of the preceding entry (viz. *ugl-*); since it is not found the first equivalent (*of*) is selected and the word order is inverted, producing *of angle.*
- Line 4: *opryedyelyayetsya* has one equivalent (*is determined* and --); its PID (***) refers to rule 6, i.e. no change of order
- Line 5: *otnoshyenyi-* has two equivalents (*relation* and *the relation*); its PID (151) refers to rule 5, searching for '25' in the CDD_2 of the following entry (i.e. its own ending) <u>or</u> in the CDD_2 of the ending of the next word (i.e. the ending *-i* of *dlyini*); this is where it is found, and so the second equivalent (*the relation*) is selected.
- Line 6: *-yem* (*by* or --) initiates rule 3 (PID '131'), searching for '23' in previous entries; none is found so the first equivalent (*by*) is chosen and the word order inverted (*by the relation*)
- etc.

Linguistic features of the experiment

- Avoidance of morphology problems
  - Russian reflexive forms (e.g. opryedyelyayetsya) entered as full forms and translated always by passives (e.g. *is determined*, *is prepared*, *is obtained*, *is produced*), and only in singular
  - Likewise, their corresponding non-reflexive forms are entered as full forms and translated as actives, usually in plural (*they prepare*, *they obtain*, *they produce*)
- Avoidance of problems of article insertion.
  - By including very few, and treatment by different codes
- Ambiguity of prepositions
  - only two equivalents, selected by noun endings
- Treatment of instrumental phrases (*by chemical process*, *by mechanical method*)
  - prepositions derived from ending of the adjective (instead of the noun)
- Chemistry sentences derived by interlocking rules operating on limited patterns
- Non-chemistry sentences contain mainly unique dictionary items and therefore unique outputs
- Finally: **no negatives, no interrogatives, no coordination (nouns or clauses), no subordinate phrases, no past tenses, no treatment of Russian copula**, etc.

Was it fixed?

- The experiment/demonstration has often been dismissed as ‘fixed’, the output being predetermined
- The high quality of the English output was suspect
- Limited sentence patterns for chemistry examples, but nevertheless new dictionary entries and more sentences could be produced. In this sense, it was a first prototype of a rule-based MT system in a very limited domain.
- However, non-chemistry sentences were in effect produced by dictionary entries specific for these particular sentences -- further development was not possible on these lines
- Garvin readily admitted the limitations, and neither he nor Dostert ever claimed it was more than a first effort

Dostert's claims, 1955

- A year after the demonstration Dostert claimed that the experiment
- "has given practical results by doing spontaneous, authentic, and clear translation",
- showed that "the necessity of pre- and post-editing has not been verified",
- demonstrated that "the primary problem in mechanical translation… is a problem of linguistic analysis…",
- formulated "the basis for broader systematic lexical coding", defining "four specific areas of meaning determination… from which fruitful results may be expected",
- 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",
- provided a "theory for the development of a general code for the mechanical formulation of multilingual syntax operations".

Word by word MT

- Garvin claimed later (1967) that the experiment illustrated two fundamental types of translation decisions: selection and arrangement.
- Was it typical of the period? In essence, it was a word *by* word system with no structural analyses. [Not word *for* word!]
- Despite common impressions, word-by-word systems were rare in the 'first generation'. Most systems included multiple levels of syntactic analysis, and there were investigations of interlingua-based and transfer-based approaches.
- The closest approach to the GU-IBM experiment was the research on 'lexicographic' methods by Erwin Reifler which were later adopted by Gilbert King in the system for the US Air Force.

Impact

- In the United States
  - other researchers surprised by newspaper reports, they disliked the conduct of research through newspapers
  - they disliked the exaggerated publicity given to an obviously incomplete system
  - they disliked what they believed was the presentation as true ‘translations’ of output which appeared to have been extracted as wholes from computer memories.
  - they saw the demonstration as ‘premature’; other MT groups were far from publicising their results – and were unprepared to do so for many years to come.
  - damaged the credibility of later MT demonstrations (particularly those by GU)
- In the Soviet Union
  - reports seen at propitious time (‘thaw’ after Stalin’s death; cybernetics and computers encouraged); Institute for Precision Mechanics and Computer Technology developed first Russian computer (BESM); Panov began English-Russian MT on GU design, demonstrated early 1956.

Consequences at Georgetown

- Funding by CIA, but not until 1956 (Russian research was a factor)
- Abandoned 1954 approach. Divergences within team; four groups -
- Paul Garvin: fulcrum method (dependency structure analysis)
- Ariadne Lukjanow: code-matching (codes attached to dictionary entries compared and matched)
- Tony Brown: cyclical, cumulative (one sentence at a time); and SLC program
- Michael Zarechnak: multi-level analysis (morphological, syntagmatic, syntactic), GAT
- Toma: worked on programming (SERNA)
- operational model: GAT-SLC

From 1954 to ALPAC

- The GU-IBM experiment/demonstration gave the wrong impression 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.
- Inevitably, the results in the next 10 years were disappointing, and as a consequence, funders set up ALPAC
- ALPAC concluded that no MT of scientific text "in immediate prospect", and ended funding in US in favour of basic NLP
- One of the principal arguments used by ALPAC was that MT output had to be extensively post-edited; and that the GU-IBM output was not edited.
- ALPAC ignored fact that GU-IBM was a first limited experiment, not even a first prototype

Implications today

- the GU-IBM demonstration not the only example of a MT system being ‘doctored’ for a particular occasion.
- not uncommon for demonstrated systems to include grammar and vocabulary rules specifically to deal with the sentences of a particular text sample -- to show systems in the best possible light.
- Today researchers are more circumspect when demonstrating experimental systems, and less willing to indulge in speculations for journalists.
- However, some vendors of systems have a more ‘liberal’ attitude: many MT systems publicised and sold (particularly on the internet) with exaggerated claims -- and with probably equally damaging impact in the future.

Main sources

- 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.
- Sheridan, Peter: 'Research in language translation on the IBM type 701', *IBM Technical Newsletter* 9, 1955, 5-24.

- An expanded version (with photographs and copies of newspaper reports) will be made available on my website:
- http://ourworld.compuserve.com/homepages/WJHutchins
- source materials also on: http://www.mt-archive.info

_____