Two precursors of machine translation: Artsrouni and Trojanskij

John Hutchins

[Website: http://ourworld.compuserve.com/homepages/WJHutchins]

1. Introduction

When machine translation (MT) began to become popular in the late 1950s and early 1960s, people started looking for historical antecedents. Some traced its origins back to the 17th century, with the ideas of Descartes, Leibniz and other philosophers and scholars about universal, philosophical and 'logical' languages. However, while 'logical' languages can be seen as prerequisites for digital computer programs and while 'universal' languages can be regarded as forerunners of auxiliary languages such as Esperanto, as precursors of universal classification systems such as those of libraries and biological taxonomies, and even as early forms of interlinguas in machine translation itself, these proposals should not be considered in any way as constituting embryonic automatic translation systems.

One of the most worked out proposals for a 'universal' interlingua, the work of a German chemist Johann Joachim Becher in 1661, was publicised in a 1962 reprint as a program for machine translation (Becher 1962). However, it was in fact no more than a Latin vocabulary with a numerical notation. In principle, after similar glossaries for other languages had been provided (which Becher did not do), it could be used for the 'mechanical' conversion of lexical items in one language into 'equivalent' lexical items in another language. Of course, Becher did not and could not, given the state of technology at the time, even suggest how a mechanism could be constructed to do such conversion. The most detailed of all the 'philosophical' languages was the 'real character' of John Wilkins developed for the newly founded Royal Society of London and published in 1668. It was the most detailed attempt in the seventeenth century to construct a rational 'universal' notation for common concepts (Wilkins 1668). It was in effect a proposal for an interlingua, but it was not a proposal for a method of automatic translation.

In the eighteenth and nineteenth centuries there were many proposed 'universal languages' (the primary source is still the volume by Couturat and Leau, 1903), but the best known are probably Volapük (by Johann Martin Schleyer) and Esperanto (by Ludwig Zamenhof). However, none were put forward as interlinguas for the automatic translation between natural languages – they were proposed as auxiliary languages to be learnt and used instead of natural languages (Large, 1985). Occasionally, there were proposals for codes to enable users to use dictionaries for 'mechanical' translation, e.g. the *Zifferngrammatik* by Rieger cited by Couturat and Leau, but as in Becher's project in the seventeenth century it was to be used by a human acting in a machine-like manner.

Proposals for machines to perform dictionary consultation or translation did not come until technological developments in the early twentieth century. The first we know of is the model of a proposed typewriter-translator demonstrated by an Estonian in February 1924 (Mel'čuk and Ravič 1967: 26) – presumably a kind of mechanical

dictionary.[1] There may have been other such proposals, but the first definite legitimate forerunners of machine translation are contained in the two patents applied for simultaneously in 1933, in France and Russia. In both cases the patents were for electromechanical devices capable of being used as translation dictionaries. The patent in France submitted by Georges Artsrouni was for a general-purpose machine which could also function, with some additional equipment, as a mechanical multilingual dictionary. The patent in Russia by Petr Trojanskij was also for a mechanical dictionary for use in multilingual translation, but he went much further with his proposals for coding and interpreting grammatical functions using 'universal' symbols and with ideas on the basic configuration of a complete 'translating machine'.

2. Georges B. Artsrouni

The patent granted to Georges Artsrouni on 22 July 1933 was for what he called a 'mechanical brain' (*cerveau mécanique*), a general-purpose device with many potential applications. It seems that he had been working on his invention since 1929 and had completed construction by 1932. (The source for information about Artsrouni and his machine is the article by Michael Corbé (1960), an American working for UNESCO, who apparently met Artsrouni towards the end of his life.)
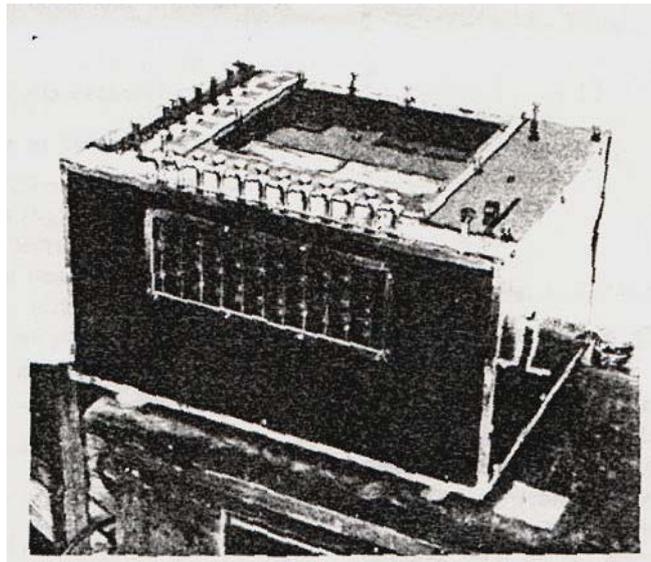

Fig. 1. Artsrouni's machine (*from* Corbé 1960)

Georges Artsrouni was a French engineer of Armenian extraction who had been a student at a school in St.Petersburg. His 'mechanical brain' was not primarily a calculator – unlike its computer successors, often called 'electronic brains' in the 1950s – but a general-purpose storage device with facilities for retrieving and printing stored information. Artsrouni suggested applications such as the automatic production of railway timetables, of telephone directories, of commercial telegraph codes, of banking statements, and even of anthropometric records. It was claimed to be

---

[1] The source is a report by the Estonian newspaper "Vaba Maa", which was reproduced (in Russian) in the proceedings of a conference on MT held in Tallin in 1962. The newspaper reported the 'demonstration' of a "model of a translating typewriter" by its inventor A.Vakher (Vaxer), who proposed to develop a prototype – of which, however, no more was heard, according to the editors of the conference proceedings.

particularly suitable for cryptography, for deciphering and encrypting messages, and finally it was claimed to be a device for translating languages[2].

At the Paris Universal Exhibition in 1937 the device attracted much attention, several thousand demonstrations were given and it received a prize (*diplôme de grand prix*) in the section for data processing (*mécanographie*). A number of state organizations were impressed by its versatility and entered into provisional contracts with the inventor for the development of prototypes designed for their particular requirements. The French post office, for example, wanted a machine for postal cheque accounting; the railway administration envisaged a machine for printing tickets to various destinations; the Ministry of Defence wanted a 'brain' for registration and processing of prisoners of war. None of these plans came to fruition after the Occupation of France in 1940, and it meant the end for the 'mechanical brain'. By 1960, according to Corbé, there remained just two prototypes in two different versions.

From the beginning, Artsrouni saw one of its main applications as a mechanical dictionary for producing crude word for word translations. In his 1933 description, he stated explicitly that his 'brain' could be adapted for the "translation of a foreign language into one of the three other languages recorded in it", and that even if "the existing model could operate only on these four languages… the number of languages and the number of words contained in the dictionary for each language could be without limit." (Corbé 1960)

As a mechanical dictionary, the 'brain' had four basic components: a 'memory' of words in the four languages (*bande de réponse*), an input device consisting of a keyboard activating a reading head (*mécanisme de repérage*), a search mechanism (*sélecteur*), and an output mechanism (*mécanisme de sortie*) activated in its turn also by the reading head. The four components were driven by a motor, and the whole apparatus was contained in a rectangular box measuring 25x40x21 cm.

The memory was the core of the device. It consisted of a paper band 40 cm wide, which could be up to 40 meters in length, moving over two rolling drums and held in position by perforations on the edges. The dictionary entries were recorded in normal orthographic form (i.e. not coded) line by line in five columns. The first column was for the source language word (or term), the other columns for equivalents in other languages and for other useful information. Using a Varityper, the band could contain up to 40,000 lines, which could be doubled if both sides of the band were used. For even greater capacity, Artsrouni proposed that entries could be printed in two different colours (red and blue) superimposed on each other on the same lines, and read by switching from one to the other by changing filters. Since the machine could use several bands, and since the width of the bands could also be increased, the amount of dictionary information could be infinite – as Artsrouni maintained, it was limited only by the effort required to record the data itself. As a further feature, the device was furnished with a recording mechanism permitting the user to modify the contents of the memory by suppressing some lines and adding others. Such modifications would be easy to make because the sequence of entries could be perfectly arbitrary.

Input, search and output took place on a board on the top of the machine (see Fig.1). At the bottom of the board, nearest the operator, came the input keyboard; immediately above it came a row of lettered cogs to display the search word and at the

---

[2] We may note that the assumption that methods for encrypting and deciphering could be applied to translating was also made by Warren Weaver in his celebrated memorandum of July 1949 (Weaver 1949).

top, furthest from the operator, came a row of slits for displaying the five columns of a line selected from the memory device. The word (or term) to be found (i.e. translated) was input at the keyboard and, through a linked mechanism, displayed on the row of cogs (the reading head) – apparently up to a maximum of ten letters. Corrections to input could be made by pressing a button to set the reading head to neutral and by inputting again.

The input word was linked to the dictionary memory by the 'selector' search mechanism. This also consisted of a band (paper or metal) rotating on two moving drums. The band contained all the words (terms) that could be selected and searched for, listed in the same order as recorded in the memory; however, in this case, words were coded in the form of perforations (i.e. similar to those on the paper tapes of the 1950s and 1960s).

The selector mechanism located, via the perforation band, the corresponding word (term) in the memory. The whole line (five columns) was then displayed in the row of five slits at the top of the operator board. These slits represented the output mechanism: the first slit showed the source word and the four others the translations and other information. (The number of slits could be increased to 15 if desired.) The slits were provided with windows of red and blue glass, allowing users to select either blue or red entries. As well as this visual display of results, the 'brain' could be provided with a printer to obtain typed output. Even more ambitiously, Artsrouni envisaged spoken output by a special mechanism (which was not described by Corbé (1960), but which presumably would have involved pre-recording on a tape.)

In brief, translation proceeded in five stages:
- the word to be translated was keyed in, which activated automatic movement of the cogs on the reading head;
- the motor set into motion simultaneously the bands of the selector mechanism and the bands of the memory device;
- both bands halted when the perforations of the selector matched exactly those indicated by the reading head;
- the slits opened automatically and the results of the search were read visually by the operator, or typed out, or produced as sounds;
- the sought term was erased and the same cycle began again for the next term.

It was claimed that the selector and the memory could operate at a speed of 60 seconds for 40,000 lines. If the search began midway on the band this speed would be doubled. In fact, a special braking and acceleration device was suggested that could reduce the search of a full band to 10 or 15 seconds. According to Corbé, even these speeds applied to the oldest model. In a later model the friction between the reading head and the selector could be eliminated by the use of cathode lamps for display instead of the mechanical cogs, and the search speed could be further reduced to three seconds.

When he put forward his invention, Artsrouni was not thinking of fully automatic translation and certainly not of high quality translation. He was no linguist, and had no awareness of problems of polysemy, idioms, or syntactic ambiguity. But he did believe that his device could be used for producing quick rough translations. He thought that operators could use a telegraphic style for input and output, and a 'telegraphic language' could act as an intermediary language – allowing people not knowing each other's languages to convey simple messages. It would not replace translators, but it could aid communication. In addition, Artsrouni envisaged a dictionary of 'phrases' rather than words, and thus the possibility of more accurate

translations. The problem of course would be the size of the 'memory', the cost of compiling the dictionary, and the slow operating speed.

In the opinion of Corbé (1960), the overall conception bore considerable similarities to the photoscopic disk device under development during the 1950s by Gilbert King for the USAF Mark I and Mark II translating machines.[3] In both cases, the devices were non-electronic storage mechanisms which could be used for large dictionaries in translation systems and capable of producing simple word-for-word rough 'translations'. The principal difference, of course, was that King's device was linked to an electronic computer for the search, retrieval and manipulation of data. Artsrouni's device operated purely mechanically.

When MT research began in France in the late 1950s, it was believed by many researchers that digital computers, as then constructed, were not suitable for mechanical translation, and that what was needed was special-purpose translation machines. In an account of the work at IBM France, Corbé and Tabory (1962) stated that "As far as our equipment requirements are concerned, they are definitely oriented towards a very large memory provided with a relatively rudimentary logic limited to a series of essentially identical look-ups." They regarded the ideas of Artsrouni for such a device as still valid: "The possibility of building such a specialised device was indicated about thirty years ago by G.B. Artsrouni to whom we are very much indebted in this respect." In a footnote, it is stated that the inventor "died a few months after the publication of this article [Corbé 1960][4], leaving a considerable body of unfinished research" (Corbé and Tabory 1962: 640).

3. Petr Petrovič Trojanskij

Like Artsrouni, Trojanskij was unknown until the late 1950s. When the first MT experiments in the Soviet Union were published, his proposals were briefly reported by L.I.Žirkov (1956), the title page of the 1933 patent was reproduced in a report by Panov et al. (1956)[5], and then in 1959 the Academy of Sciences published a substantial collection of his writings (Bel'skaja et al. 1959).[6] This 'discovery' prompted Yehoshua Bar-Hillel to refer to Trojanskij as the "Charles Babbage of machine translation" (Bar-Hillel 1960). Just as Babbage had constructed an early form of calculating machine, using the technology of the nineteenth century, and had made suggestions about programming, Petr Petrovič Trojanskij (1894-1950) had described how a translating machine might be constructed using the electromechanical technology of the 1930s and 1940s. His proposals went much further than describing a mechanical dictionary – although this was the focus of the patent itself – since he

---

[3] For information about Gilbert King, his photoscopic store and its use in the IBM translators see Hutchins (2000)

[4] Artsrouni must then have died in late 1960. No date or place of birth are indicated by Corbé and Tabory, although reference (in Corbé 1960) to schooling in pre-Revolutionary St. Petersburg, suggests that he may have been born about 1900, and possibly in St. Petersburg itself.

[5] The first mentions of Trojanskij in non-Russian publications were an abstract of Žirkov's article which appeared in the journal *Mechanical Translation* 3(3), December 1956: 91 [unfortunately misdating the patent to 1938]; and a translation of the report by Panov and others, as "La traduction automatique" in *Recherches International à la Lumière du Marxisme* 7 (1958): 162-193.

[6] The collection was edited by three members of the MT group at the Institute of Precision Mechanics and Computer Technology (*Institut točnoj mexaniki i vyčislitel'noj texniki*) of the USSR Academy of Sciences: I.K.Bel'skaja, L.N.Korolev and D.Ju.Panov. The texts were not translated into English; consequently, knowledge of Trojanskij's pioneering efforts has remained relatively neglected outside Russia. (In an effort to remedy this neglect, translated extracts and commentaries from the 1959 collection have been published by Hutchins and Lovtskij, 2000.)

described in other documents of the time how the overall 'translation processes' could operate on the basis of 'universal' symbols for the coding and interpretation of grammatical functions.

Trojanskij was born in January 1894 into the large family of a railway repair-shop worker in Orenburg, South Urals. After many hardships he managed to start studies at the University of St.Petersburg, but they were interrupted by the First World War. After the Russian revolution in 1917 he studied in the 'Institute of Red Professors' (*Institut krasnoj professory*), an institute established by Lenin in 1921 with the task of preparing suitably qualified (and no doubt politically acceptable) people for teaching in higher education in the sensitive fields of economics, philosophy and history. Trojanskij's membership of this institute indicates that at this date he must have been a committed Communist. Later he taught at other higher educational establishments and he participated in the compilation of the Great Soviet Encyclopedia.

However, Trojanskij's life-long ambition became the development of his 'translating machine', or (as the title of his 1933 patent describes it) "a machine for selecting and typing words when translating from one language into another or several others simultaneously." The somewhat curious title is attributable at least in part to the fact that the invention was classified by the patent office as a novel method of typesetting.

Fig.2. Petr Petrovič Trojanskij (*from* Bel'skaja et al. 1959)

4. Trojanskij's patented device, 1933

The patent submitted on 5 September 1933 describes a machine consisting of "a smooth sloping desk, over which moving easily and freely in different directions is a

belt provided with perforations which position the belt in front of an aperture" (see Fig. 3). This broad belt was a large dictionary, with entries in six languages in parallel columns. The operator located a word of the source language and moved the belt to display in the aperture the corresponding word of the target language. The operator would then type in a code indicating the grammatical category or role of the word in question – codes that Trojanskij referred to as 'signs for logical parsing' – and the combination of target word and code were then photographed onto a tape. Then the next source word would be located and 'translated' in the same way. From a tape of the target language words in sequence, a typist would then produce a 'coherent text' for a reviser to substitute the correct morphological forms for each word based on the assigned codes. As a final stage a 'literary editor' would produce the final target text. Unlike the operator and the reviser, who needed to know only their own languages (those of the source and the target respectively), the editor would need to know both languages in order, in Trojanskij's words, "to extract the meaning of the translation, to choose synonyms, to polish the unevenness, i.e. to do general literary finishing."

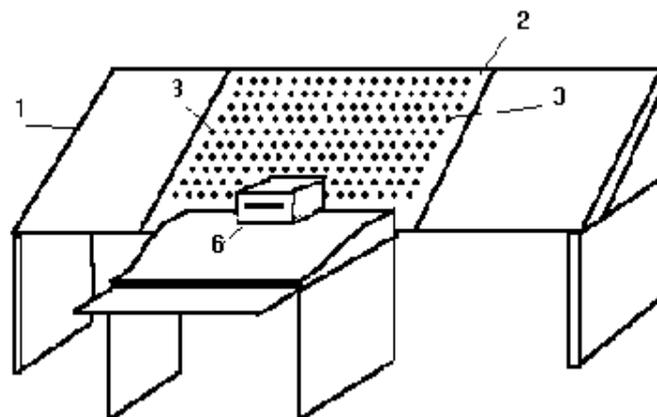

Fig. 3. Trojanskij's translating machine (*from* Bel'skaja et al. 1959)

5. Logical parsing and ambiguities

What sets Trojanskij's proposal apart from his contemporary Artsrouni was that he went beyond the mechanization of the dictionary by his clear enunciation of some basic processes of translation and by his proposals for 'logical parsing symbols'. These symbols were intended to represent 'universal' grammatical relationships, therefore applicable to any language and when translating between any languages.

The symbols were initially taken from Esperanto, as were also some 200 'ancillary words'. Examples of the 'logical parsing' symbols are given by Trojanskij in a description of his invention that was written in September 1933 to accompany the patent application: *j* ("plural"); *n* (indicating "verbal government, direct objects, accusative case"); *de* ("dependence of a declinable part of speech form on another declinable form", "agent in passive voice"); *per* ("instrument in passive voice…instrumental case"); *e* ("adverb" or "verbal adverb"); *oni* ("impersonal form of verb"); *a* ("adjective in a predicate, expressing some kind of auxiliary verb" or "participle"); *i* ("indefinite declination"); *as, is, os* ("the present, the past and future tenses of verbs in indicative mood"); etc.

Trojanskij says very little about the 'ancillary words', but it would appear that he intended to use the Esperanto words for conjunctions, prepositions, pronouns, etc. (as examples he gives *kies, kie, kien, alkiu, kial, kiam, kiom da, kioma* – i.e. *whose, where, where to, to whom, why, when, how much, how many*). These and the 'logical

parsing' symbols were therefore intended as basic elements of what would now be called an 'intermediary representation'. He believed these symbols were precise and unambiguous, "based on scientific principles" and could be used by anyone without difficulty. He did, however, recognise that there would have to be some means of dealing with synonyms and homonyms in the dictionary (which he called the 'glossary field') on the moving belt. His answer was to simply display alternative possible translations for synonyms:

Speak : govorit', razgovorit': parler, causer : sprechen
Swift : bystryj, skoryj : rapide, vite, soudain : schnell

and to distinguish homonyms by adding 'explications of meaning in parentheses':

Kosa (pesčanaja) spit (of sand)[7]
Kosa (devič'ja) (maiden's) plait
Kosa (dlja kos'by) (farmer's) scythe
Perevod (po službe) transfer (of duty)
Perevod (sočinenija) translation (writing)
Perevod (denežnyj) remittance (monetary)
etc.

More complex problems of lexical transfer did not occur to him – and nor did they to many other pioneers of MT in the early 1950s.

The use of Esperanto elements would have been widely accepted at the time. The inventor of Esperanto, Zamenhof, was a Russian-speaking Jew from the borders of Poland and Russia, who had internationalist pacifist ideals of a socialist tendency that appealed to many in Eastern Europe. In the first years of the Soviet Union, Esperanto had a large following and received official support; there were even moves to introduce the obligatory teaching of Esperanto in schools. With state patronage, the Soviet Esperanto Union (SEU) was established in 1921 to foster contacts with Esperantists and internationalists abroad, and within a few years claimed thousands of members. The president of SEU was Ernst Karlovič Drezen, a prominent and influential Bolshevik during the revolution and the early years of the Soviet Union, who advocated Esperanto for the translation of documents, for instruction in technical schools, and for use by international correspondents and union officials. Stalin (who during his Siberian exile had learnt Esperanto) appointed Drezen to head a commission for the creation of an international scientific-technical terminological code as "the basic means for the organization of labor, knowledge and thought", and "[i]n order to promote the 'greatest internationalism in form and content', Drezen based most of the code's root words, affixes, prepositions, conjunctions, and grammatical rules on Esperanto." (Smith 1998: 154-156)

In this context, Trojanskij's advocacy of Esperanto as the interlingual basis for a device to aid communication among the multilingual population of the Soviet Union, to spread technology and managerial efficiency, and to propagate the proclamations of the Soviet government to this wider community, would clearly have had the highest level of potential support in the early 1930s. Unfortunately, by the late 1930s the situation had changed. The internationalist utopianism of the first Soviet years was replaced by the Stalinist policy of 'socialism in one country', and Esperantists were now suspected of collaboration with enemies of the Soviet Union and many of them were executed or imprisoned during the Stalinist terror.

6. Reception of Trojanskij's patent, 1939 and 1944

Whether for this reason or not, Trojanskij's patent proposals were ignored. Obviously disappointed that the authorities had not taken up his idea, he approached

---

[7] In Trojanskij's document the target language equivalents are German; here they have been translated.

the USSR Academy of Sciences in 1939 in order to obtain some assistance with the linguistic aspects of his invention, in order to construct an experimental model.[8] But he did not receive the support he had hoped for. As Žirkov reported some years later, "the invention… was received by linguists with profound scepticism; it was considered impractical and quite unnecessary" (Žirkov 1956). No decision was reached in the next few years, until a meeting at the Institute of Automation and Telemechanics (*Institut avtomatiki i telemexaniki*) of the Academy of Sciences on 31 July 1944 where it was discussed by a group of eminent linguists and specialists in mechanics and electrical engineering. The engineers agreed with the linguists in rejecting the very possibility of mechanical translation, "talking about synonyms and subtle nuances of meaning". The proposed model of his translating machine was not therefore built. However, Trojanskij was able to give a demonstration with a translation into French of the Russian sentence

> Rešajuščie opyty mexaničeskogo perevoda, kotoryx my ožidali v tečenie dvux mesjacev, osuščestvilis' v Moskve segodnja v 4 č.30 m.

The sentence was provided with 'logical symbols', which were no longer based on Esperanto (for the obvious political reasons mentioned above) but indicated by numerical codes:

> Rešajuščie 51 opyty 1-5 mexaničeskogo 551-6 perevoda 51-6…', etc.

The Russian words (stems) were next replaced by French equivalents, giving:

> expériment 1-5 décisif 51 traduction 51-6 mécanique 551-6 que 091 nous 01 avons 02-1 attendus 02-1* pendant 0902-1 deux 068 mois 05068 ont 02-1 eu 2-1* lieu 2-1** à 67 m-o-s-c-o-u- 67* aujourd'hui 68 à 67 quatre 68 heure 568 trente 68 minute 568 stop.

From this output from the machine, the editor dictated directly to a typist:

> Les expériments décisifs de la traduction mécanique que nous avons attendus pendant deux mois ont eu lieu à Moscou aujourd'hui à quatre heures trente minutes.

7. Trojanskij's later elaborations, 1947

After his negative reception by the academicians, Trojanskij devoted the following years to further developments of the technical aspects of the system and to answering his critics (particularly the linguists). He wrote a long paper, completed apparently in 1947, devoted to an expansion of the linguistic aspects of his ideas and demonstrating how easy and time-saving his method could be in comparison with human translation.

This essay emphasised, in particular, the labour- and cost-saving advantages of "dividing the translation process into three separate operations": (1) a 'monolingual' operation, from the original to the 'logical parsing form' – performed by someone needing to know only the one language, (2) a 'bilingual' replacement operation, from the 'logical form' in one language to the 'logical form' in another – performed by his machine, and (3) a 'monolingual' operation, from the 'logical form' into the "natural, i.e. fully national-grammatical form" – performed by someone knowing only the one language. Actually, Trojanskij admits there would need to be another 'monolingual' process by an 'editor' to produce a stylistically acceptable translation, but it should be noted that Trojanskij had now dropped the requirement that this 'literary editor' should know both source and target languages. Evidently he now thought the 'editor' would be able, just from knowledge only of the target language, to select synonyms and to find the right idioms. To have required bilingual knowledge would obviously

---

[8] By this date, he was calling himself Smirnov-Trojanskij, apparently adopting his wife's surname – for unknown reasons. It is under this name that he is often referred to it in the literature, particularly outside Russia.

have weakened his case, particularly with respect to translation into the minority languages of the Soviet Union.

A major 'accusation' – presumably encountered from critics at the Academy of Sciences – had been that his machine was no more than an automatic dictionary. He answered by arguing that "of the three operations comprising its technological processes, namely the operations A-A^1, A^1-B^1 and B^1-B, only the second one (A^1-B^1) represents the translation function proper, as it embraces different languages and only it formulates translation itself from language to language." By contrast the two monolingual processes were "pre-translation and post-translation procedures", and the bilingual translation process was "performed directly by the machine". Furthermore, he claimed that the output from the machine (i.e. words plus logical symbols) could be understood, with practice, by recipients[9]. "Therefore my machine is a true translation machine."

While Trojanskij stressed the central role of the bilingual dictionary he insisted also that the monolingual operations using his 'logical parsing symbols' were essential parts of his method. He saw no difference with other types of mechanisation:

> Any mechanisation of work processes introduces… its own regularities… When using a machine of any kind… the material to be processed… must first be adapted and reduced always to a form suitable for processing on this particular machine… Such preliminary treatment of the material to be processed on the translating machine is exactly what the operation of logical parsing is. Logical parsing is namely an integral part of machine translation technology.

His main arguments for the 'monolingual' method were, however, economic. The costs of translation would be reduced because there was no need for bilingual translators; the bulk of the work was to be done by people knowing only one language. He claimed – but without offering any proof – that costs could be reduced to "one percent of the former level", since "it is obvious that the higher the number of languages involved… the cheaper will cost translation in each separate language." He stressed particularly the need for large-volume translation in the multilingual context of the Soviet Union. There will never be enough translators: "there remains mass translation work which has to be handed over to the machine… because we simply cannot keep up with this mass, with its ever-growing volume."[10] And in the spirit of the age, he drew an analogy with the mechanization of manual tasks: "Seeds can be sown by hand, and sown not badly. But there exist tractor-drawn sowing machines." There will always be a place for non-mechanized crafts:

> There exist craftsmen who, without machines, using only hand tools, are capable of making precision watch mechanisms. But they would be ridiculed if they denied the necessity of using machines in watch making.

Evidently, Trojanskij envisaged a future where translating machines would be part of everyday life.

In his 1947 essay Trojanskij also gave further elaboration and justification for the use of his 'logical symbols'. He asserted that his

> 25 universal international symbols of logical parsing for all languages, used in various combinations which number about one hundred, …are capable of rendering without exception all relations and the slightest shades of human thought expressed in words and notions, and ensure absolutely exact translation into other languages without distortion of meaning.

In addition, he had now a further argument in favour of the universal nature of his 'logical symbols': the fact that they were not tied to surface syntactic sequences.

---

[9] Similar arguments were used by some MT pioneers in the 1950s, believing that MT-ese (or 'pidgin' translation) could be comprehensible when recipients are familiar with the subject matter.

[10] The benefit was particularly pertinent at the time, with Stalin's drive for technological development, integration of the non-Russian minorities into the Union, and the centralization of bureaucratic power.

Therefore, "they release the phrase from a rigid word order, thanks to the potentialities inherent in them".

By now (1947), he appears to have dropped the use of the 'auxiliary' Esperanto words entirely, and indeed, the Esperanto origin of his symbols is not now explicitly mentioned,[11] although they would have been readily identified by those familiar with the language from his example translations, e.g.

> Le parti périt s'il commence à cacher ses erreurs
> Le parti-o périr-as si il commencer-as cacher-i son-ajn l'erreur-ojn
> Partija-o pogibat'-as esli on načinat'-as skryvat'-i svoi-ajn ošibka-ojn
> Partija pogibaet esli ona načinaet skryvat' svoi ošibki

As an alternative to the use of the symbols as suffixes, Trojanskij puts forward the use of what he calls 'footnotes' as abbreviated forms of representation:

> Le^1 $parti^1$ $périt^2$ s'^3 il $commence^4$ $à^5$ $cacher^5$ ses^6 $erreurs^7$; Le^1 $tableau^1$ du^8 $monde^8$ $montre^4$ $comment$ la^1 $matière^1$ se^9 $meut^9$, comment la^1 $matière^1$ $pense^4$.
> where: ^{1}-o, ^{2}-ir-as, ^{3}si, ^{4}-er-as, ^{5}-er-I, ^{6}son-ajn, ^{7}-r-ojn, ^{8}de-o, ^{9}se mouvoir-as

(The result is reminiscent of the indexed version of 1944 (or earlier) reported by Žirkov in 1956.)

In a number of respects his 'logical parsing' resembles the kind of interlingual syntactic representation found in later MT work, e.g. where 'semantic cases' are used. Indeed, the commentary by Izabella K. Bel'skaja which accompanied this paper in the 1959 collection points to a number of analogies between Trojanskij's methods and those of the MT pioneers in the 1950s.[12]

He answered also the charge that his logical symbols were difficult to use. On the contrary, he thought analysis into his symbols would be easier than the parsing Russian children had to do:

> If it is acceptable for school children to carry out the far more complicated school analysis … "by parts of speech" and "by parts of sentences", then the notation of my logical parsing is surely considerably simpler, especially for a grown-up person literate in his own language.

In fact, he obviously thought it was so straightforward that at one point in his paper he asserts that eventually

> "*logical parsing itself can be automated* if we build a special machine which with completely accurate printed texts will do work that… is carried out in the name of 'logic'. It is even possible in the machine to contend with such subtleties as having one and the same word capable of being both a verb and a noun… Then its particular syntactic position, role and relations with other parts of the sentence will indicate to the *mechanism* in what *sense* the given word is used, namely: in the noun sense or in the verb sense." [Trojanskij's emphases.]

Undoubtedly, he underestimated the difficulties of analysis and interpretation, but in this respect he may have been no more naïve than some of his successors in the 1950s and early 1960s. The problems of multiple meanings of words had been recognised by him in the document attached to his patent application in 1933, when he had proposed tables of homonyms to aid in lexical selection. In his 1947 paper he provides further examples. Basically, his solution for homonyms was the one adopted by many early MT pioneers: to print out all the possibilities "so that the editor has only to strike out the superfluous synonyms". As for idioms, he suggests that they are "either replaced by non-idiomatic expressions, or … left untranslated …, or (finally) a suitable idiom is selected…"

---

[11] It may be noted also that the Russian editors in 1959 did not draw attention to the Esperanto origins of his symbols – no doubt for political reasons.
[12] For details see Hutchins and Lovtskij (2000)

In this connection he made the important point that understanding (interpretation) also requires knowledge of the subject.

> My monolingual translation methodology does not necessarily get rid of special training for understanding specialist texts in one's own native language. Strange as it may seem, this circumstance was blamed on my monolingual translation methodology and my translating machine. But this, you know, is a general feature of any speciality: it has to be learnt.

In this respect, Trojanskij anticipated many of the later objections to the usability of MT output. Experts with knowledge of the subject matter can (and do) make good use of crude 'automatic translations' in order to extract the essential messages of texts.

8. Further technical developments of the 'translating machine'

From the time of his patent in 1933, Trojanskij worked continuously on the technical improvements for his invention. The original wholly mechanical parts were gradually replaced by electro-mechanical devices, as far as possible made from widely available off-the-shelf components. The editors of Trojanskij's papers in 1959 believed[13] these technical improvements to be far ahead of his time, e.g. photo-electric coding and reading of items on the 'glossary field', and a set-up similar to a telecommunications network where several operators could work simultaneously on different texts and different languages without interfering with each other.

Most intriguing of all, he proposed a "portable translation machine for personal use." This was to consist of a folding screen on which parts of the glossary field and lines of logical parsing symbols could be displayed using a "photo-projection device". The operator was to select words in the source language and decide on the appropriate 'logical symbols'; the machine would then find equivalents in another language, and the results (words plus logical symbols) would be projected onto the screen. These would then be 'read' by electrical pulses and displayed on receivers. As a further refinement, the results could be transmitted "by cable or radio" to other receivers at a distance.

9. Artsrouni and Trojanskij as precursors

Problems of computerized dictionaries (content, structure, storage, creation, maintenance, etc.) preoccupied all the pioneer researchers of MT in the 1950s (cf. Hutchins 1986) – for obvious reasons, since a good bilingual dictionary is at the core of any translation system, and the current computers presented major difficulties of memory, input, and external storage. In this sense, both Artsrouni's and Trojanskij's proposed devices were definite precursors – not for their technical solutions but (as Corbé and Tabory acknowledged) for their general methods. Artsrouni did not, however, go beyond the dictionary; he did not consider problems of synonymy and homonymy or of rearranging output to conform to the syntax of the target language. Trojanskij, by contrast, tackled the analysis of words into stems and sets of general ('universal') grammatical functions – in fact, his ideas anticipated the stem-splitting approach of Richens and Booth in 1948 (cf. Hutchins 1997).

More significant, however, must be Trojanskij's conception of a mechanizable translation process in three stages (analysis, transfer, and synthesis), and one based to some extent on 'universal' linguistic elements – in so far as Esperanto can be regarded as 'universal'. He recognized some of the basic problems of translation (homonyms, synonyms and idioms, the problems of analysis into abstract 'symbols', the need for post-editing), and he stated clearly some of the major advantages of mechanization (in

---

[13] Comments on technical aspects were provided by D.Ju.Panov and L.N.Korolev.

particular with multilingual output). All this was many years before the similar ideas of the MT pioneers in the 1950s (Hutchins 1997).

Technologically, of course, neither Artsrouni nor Trojanskij were precursors. We can say that Artsrouni's device (his 'mechanical brain') was a general-purpose machine which can certainly be regarded as a precursor of the electro-mechanical calculators and the first electronic computers of the 1950s. Trojanskij's machine was specifically designed for translation. Like Artsrouni, Trojanskij was working in the age before electronics; he may well have known about the electro-mechanical calculators in the United States in the 1930s, but obviously not those developed during the war, and he would have certainly known nothing of the electronic computers. As we have seen, Corbé saw Artsrouni's device as a precursor of Gilbert King's photoscopic store – and a similar claim was made by Panov and Korolev in their comments in the 1959 collection. In the 1950s and 1960s many thought that current general-purpose computers were inappropriate for MT and that special-purpose machines should be built. Today, however, the unbounded flexibility and massive storage capabilities of digital computers make such notions obsolete.

It is not known whether Artsrouni continued work on his 'translating machine' after the 1930s. As we have seen, however, Trojanskij worked on his ideas until his death. As an engineer, Trojanskij was undoubtedly more interested and more familiar with the mechanical aspects of his proposals than with the linguistic aspects; and like later MT pioneers with engineering backgrounds in the 1950s he tended to be over-optimistic about the prospects of rapid improvements in language processing:

> the first models of machines never turn out perfect, on the contrary, their design perfection, expansion of output, acceleration and improvement of their technological process is always reached gradually. It will suffice to recall the first makes of telephones, typewriters, radio receivers, aircraft, locomotives and many other machines and instruments. The same will happen to the translation machine. To its aid will come the natural development of linguistics and technology.

When Trojanskij died in 1950, the electronic computer was still virtually unknown in the Soviet Union, and indeed the first Soviet machines were not developed until a few years later, by S.A.Lebedev at the Institute of Precision Mechanics and Computation Technology, the MESM in 1951 and the BESM in 1955. If he had lived, it is quite probable that he would have been one of the first to exploit its potential as a translation machine. His ideas were already as mature as any of those who began experiments in the United States in the early 1950s and in the Soviet Union from 1954 onwards.

References

Bar-Hillel, Y. (1960): "The present status of automatic translation of languages", *Advances in Computers* 1.91-163.

Becher, J.J. (1962): *Zur mechanischen Sprachübersetzung: allgemeine Verschlüsselung der Sprache. Ein Programmierungsversuch aus dem Jahre 1661*. Stuttgart: Kohlhammer.

Bel'skaja, I.K., Korolev, L.N. and Panov, D.Ju., eds. (1959): *Perevodnaja mašina P. P. Trojanskogo: sbornik materialov o perevodnoj mašine dlja perevoda s odnogo jazyka na drugie, predložennoj P.P.Trojanskim v 1933 g.* Moskva: Izd.Akad.Nauk SSSR.

Corbé, M. (1960): "La machine à traduire française aura bientôt trente ans", *Automatisme* 5(3): 87-91.

Corbé, M. and Tabory, R. (1962): "Introduction to an automatic English syntax (by fragmentation)", *1961 International Conference on Machine Translation and Applied Language Analysis: Proceedings of the conference held at the National Physical Laboratory, Teddington, Middlesex, on 5th, 6th, 7th and 8th September*, 617-640. London: H.M.S.O.

Couturat, L. and Leau, L. (1903): *Histoire de la langue universelle*. Paris: Hachette.

Hutchins, J. (1986): *Machine translation: past, present, future*. Chichester: Ellis Horwood. New York: Halsted Press.

Hutchins, J. (1997): "From first conception to first demonstration: the nascent years of machine translation, 1947-1954. A chronology", *Machine Translation* 12 (3), 195-252.

Hutchins, J. (2000): "Gilbert W. King and the USAF Translator", *Early years in machine translation: memoirs and biographies of pioneers*, ed. W.J.Hutchins, 171-176. Amsterdam: John Benjamins.

Hutchins, J. and Lovtskij, E. (2000): "Petr Petrovich Troyanskii (1894-1950): a forgotten pioneer of mechanical translation", *Machine Translation* 15 (3).187-221.

Large, A. (1985): *The artificial language movement*. Oxford: Blackwell.

Mel'čuk, I.A. and Ravič, R.D. (1967): *Avtomatičeskij perevod 1949-1963: kritiko-bibliografičeskij spravočnik*. Moskva: Vsesojuznij Institut Naučnoj i Texničeskoj Informacii, Institut Jazykoznanija.

Panov, D.Ju., Ljapunov, A.A. and Muxin, I.S. (1956): *Avtomatizacija perevoda s odnogo jazyka na drugoj*. Moskva: AN SSSR.

Richens, R.H. and Booth, A.D. (1955): "Some methods of mechanized translation", *Machine translation of languages: fourteen essays*, ed. W.N.Locke and A.D.Booth, 24-46. Cambridge, Mass.: M.I.T.Press.

Smith, M.G. (1998): *Language and power in the creation of the USSR, 1917-1953*. Berlin: Mouton de Gruyter. (Contributions to the Sociology of Language 80.)

Weaver, W. (1949): Translation. Repr. in: *Machine translation of languages: fourteen essays*, ed. W.N.Locke and A.D.Booth, 15-23. Cambridge, Mass.: M.I.T.Press.

Wilkins, J. (1668): *An essay towards a real character, and a philosophical language*. London: Samuel Gellibrand and John Martyn.

Žirkov, L.I. (1956): "Granitsy primenimosti mašinnogo perevoda", *Voprosy Jazykoznanija* 5 (5). 121-124.