FIRST STEPS IN MECHANICAL TRANSLATION

John Hutchins
(University of East Anglia, Norwich, UK)
WJHutchins@compuserve.com

MT Summit VI
San Diego, California, 1997

Rene Descartes
letter to Marin Mersenne
20 November 1629

Thus the sole utility I can see arising from this invention is for writing: if a large dictionary is printed in all the languages in which one wants to be understood, where common characters are put for each primitive word corresponding to the sense and not to the syllables, such as the same character for *aymer*, *amare* and *filein*, those having the dictionary and knowing the grammar would be able, through seeking all these characters one after another, to understand in their language what is being written.

JOH. J. BECHERI, Spirenſis

CHARACTER,

Pro

NOTITIA LINguarum Univerſali.

INVENTUM STEGANOGRAPHICUM HACTENUS INauditum quo quilibet ſuam Legendo vernaculam diverſas imò omnes Linguas, unius etiam diei informatione, explicare ac intelligere poteſt.

FRANCOFVRTI,
Sumpt. Johannis Wilh. Ammonii
& Wilhelmi Serlini,
Typis JOHANNIS GEORGII SPÖRLIN.
ANNO M. DC. LXI.

Coniunctivus

Præsens

51 Activum		51 Passivum	
93	Amem, Monecam, Legam, Audiam	123	Amer, Monear, Legar, Audiar
94	ames, moneas, legas, audias	124	ameris, moneare, legaris, audiaris
95	amet, moneat, legat, audiat	125	ametur, moneatur, legatur, audiatur
96	amemus moneamus legamus audiamus	126	amemur, moneamur, legamur, audiamur
97	ameris moneatis legatis audiatis	127	ameminí, moneaminí, legaminí, audiaminí
98	ament moneant legant audiant	128	amentur, moneantur, legantur, audiantur.

Imperfectum

99	Amarem, Monerem, legerem, Auditem	129	Amarer, Monerer, Legerer, Audirer
100	amares, moneres, legeres, audires	130	amareris, monereris, legereris, audireris
101	amaret, moneret, legeret, audiret	131	amaretur, moneretur, legeretur, audiretur
102	amaremus, moneremus, legeremus, audiremus	132	amaremur, moneremur, legeremur, audiremur
103	amaretis, moneretis legeretis, audiretis	133	amaremini, moneremini, legeremini, audiremini
104	amarent, monerent, legerent, audirent	134	amarentur, monerentur, legerentur, audirentur

Perfectum

105	Amaverim, Monuerim, Legerim, Audiverim	135	Amatus, Monitus, Lectus, Auditus sim vel fuerim
106	amaveris, monueris, legeris audiveris	136	amatus, monitus, lectus auditus sis
107	amaverit monuerit, legerit, audiverit	137	amatus monitus lectus auditus sit
108	amaverimus monuerimus legerimus audiverimus	138	amati, moniti lecti auditi simus
109	amaveritis, monueritis, legeritis, audiveritis	139	amati, moniti lecti auditi sitis
110	amaverint, monuerint, legerint audiverint	140	amati, moniti, lecti, auditi sint

Plusquamperfectum

111	Amavissem, Monuissem, Legissem Audivissem	141	Amatus, Monitus, Lectus, Auditus essem vel fuissem
112	amavisses monuisses legisses, audivisses	142	amatus, monitus, lectus auditus esses
113	amavisset, monuisset, legisset audivisset	143	amatus monitus lectus auditus esset
114	amavissemus monuissemus, legissem⁹ audivissem⁹	144	amati, moniti, lecti, auditi essemus
115	amavisetis, monuissetis, legissetis, audivissetis,	145	amati, moniti lecti, auditi essetis
116	amavissent, monuissent, legissent, audivissent	146	amati, moniti, lecti, auditi essent

Futurum

117	Amavero, Monuero, Legero, Audivero	147	Amatus, Monitus, Lectus, Auditus ero vel fuero
118	amaveris, monueris, legeris audiveris	148	amatus, monitus lectus, auditus eris
119	amaverit, monuerit, legerit, audiverit	149	amatus, monitus, lectus, auditus erit
120	amaverimus, monuerimus legerimus audiverim⁹	150	amati, moniti, lecti, auditi erimus
121	amaveritis, monueritis, legeritis, audiveritis	151	amati, moniti, lecti, auditi eritis
122	amaverint, monuerint, legerint audiverint	152	amati, moniti, lecti auditi erunt

Infinitivus

Præsens

Activum		Passivum	
153	Amare, Monere, Legere, Audire	163	Amari, Moneri, Legi, Audiri

Perfectum

154	Amavisse, Monuisse, Legisse, Audivisse	164	Amatum, Monitum, Lectum, Auditum esse vel fuisse

Futurum

155	Amaturum esse vel amaturum ire Monitum esse vel monitum iri Lecturum esse vel lectum iri Auditurum esse vel auditum iri	165	Amatum, Monitum, Lectum, Auditum iri

Gerundia

156	Amandi, Monendi, Legendi Audiendi		
157	Amando, Monendo, Legendo, Audiendo		
158	Amandum, Monendum, Legendum, Audiendum		

Supina

159	Amatum, Monitum, Lectum, Auditum		
160	Amatu, Monitu, Lectu, Auditu		

Europæi Sequentia adjiciant.
168. Incomparativo.
169. Insuperlativo.
170. Adverbialiter.
171. Derivativo ut Negutiorsum
172. Purus numerus.
173. Impersonaliter sumptum, à verbum, sin verisimeli, in genere Neutro & Nominativo Singulari intelligatur.

Participium

Præsens

Activum		Passivum	
161	Amans, Monens, Legens, Audiens		

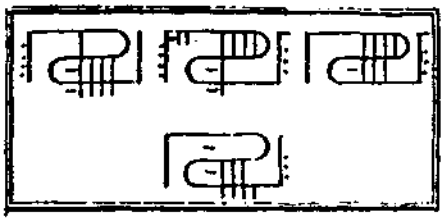
Perfectum

		166	Amatus, Monitus, Lectus, Auditus, ta, tum

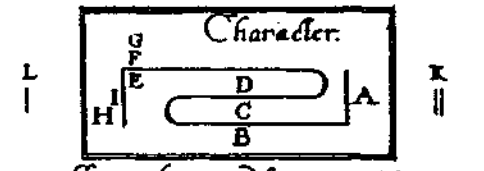
Futurum

162	Amaturus, Moniturus, Lecturus Auditurus, ra, rum.	167	Amandus, Monendus, Legendus Audiendus, da, dum.

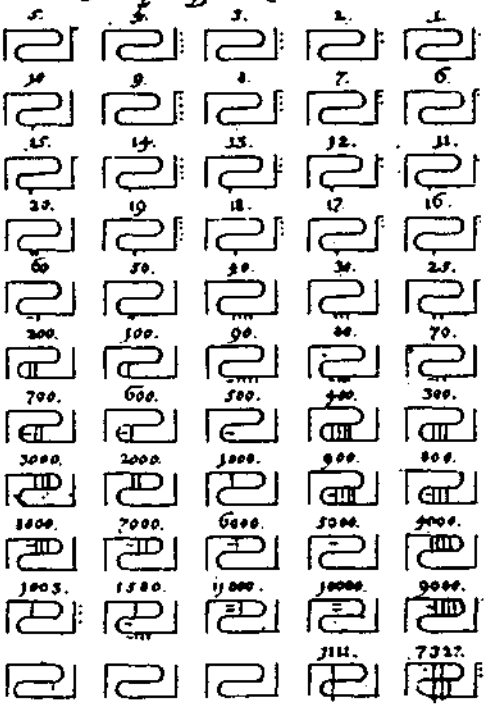
NB. Hæc Tabula sub finem Descriptionis operis, pernecessaria est. R. H.

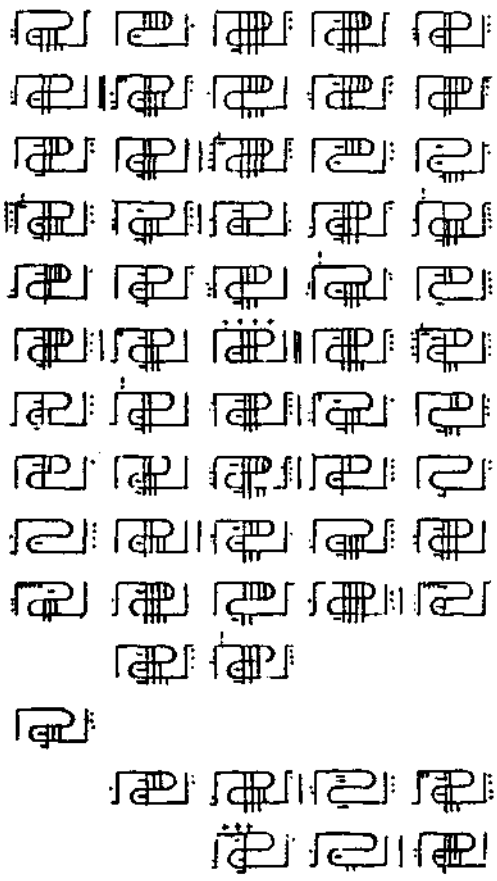


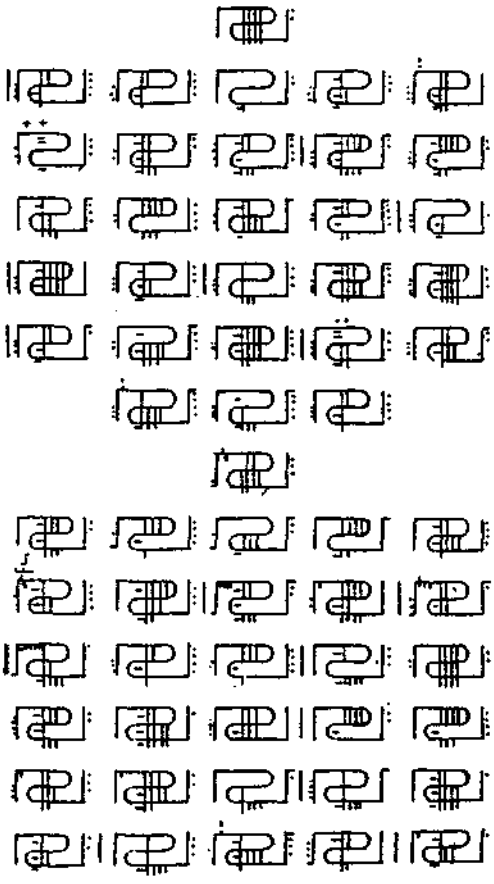
Huc spectant Exempla pro Numeri notitia et post hæc Tabula pro Variationibus Sensuum.



Exempla pro Numeri notitia.



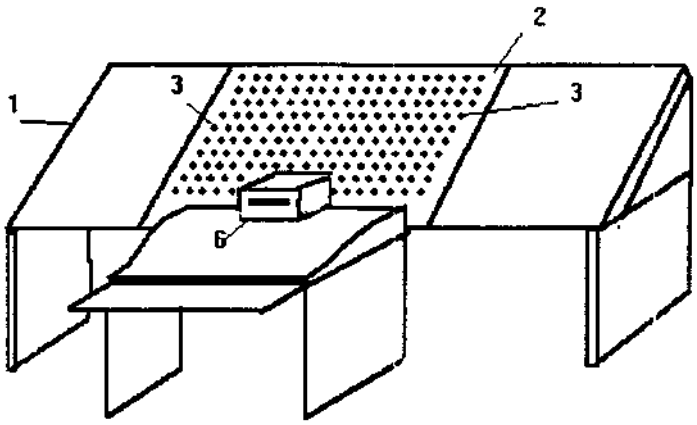




Georges Artsrouni and Petr Petrovich Troyanskii patents, 1933

Artsrouni's invention intended for: the production of timetables and telephone books, accounting, deciphering and encrypting messages, and translating.

Troyanskii's mechanical translator



Symbolic characters based on Esperanto, e.g.: -*aj* indicated a plural noun, -*n* a direct object case, -*ir* a verb form.

Three-stage process: (1) Human 'editor' prepared source text, reducing to stems and identifying 'logical' symbols; (2) mechanical conversion, giving equivalent sequence of words, retaining logical symbols; (3) Human 'editor' converted strings of stems and symbols into target text.

Warren Weaver
letter to Norbert Wiener
4 March 1947

Recognizing fully, even though necessarily vaguely, the semantic difficulties because of multiple meanings, etc., I have wondered if it were unthinkable to design a computer which would translate. Even if it would translate only scientific material (where the semantic difficulties are very notably less), and even if it did produce an inelegant (but intelligible) result, it would seem to me worth while.

Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography ... one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

Richard H. Richens and Andrew D. Booth

The idea of using punched cards for automatic translation arose as a spin-off, fuelled by my realisation as editor of an abstract journal *(Plant Breeding Abstracts)* that linguists conversant with the grammar of a foreign language and ignorant of the subject matter provided much worse translations than scientists conversant with the subject matter but hazy about the grammar

(German)

Wenn in ein*em gross*eren Gebiet zwei Form*en neben*einander leb*en, ohne sich zu vermisch*en, so gehör*en sie verschieden*en Form*en*kreis*en an.

(English)

if in a/one *d* large (more) area two form *m* beside one another live *z* without self to/too mix z, so belong/hear *p z z* different *m* form *m* circle *m* at.

(French)

Il n'est pas étonn*ant de constat*er que les hormone*s de croissance ag*issent sur certain*es espèce*s, alors qu'elles sont in*opér*antes sur d'autre*s, si l'on song*e à la grand*e spécificité des ces substance*s.

(English)

v not is not/step astonish *v* of establish *v* that/which? *v* hormone *m* of growth act *m* on certain *m* species *m,* then that/which? *v* not operate *m* on of other *m* if *v* one dream/consider *z* to *v* great *v* specificity of those substance *m.*

SWAC (Standards Western Automatic Computer) demonstration, Los Angeles, May 1949

(report, New York Times, 31 May 1949):

A new type of "electric brain" calculating machine capable not only of performing complex mathematical problems but even of translating foreign languages, is under construction here... While the exact scope of the machine will have in the translating field has not been decided, the scientists working on it say it would be quite possible to make it encompass the 60,000 words of the Webster Collegiate Dictionary with equivalents for each word in as many as three foreign languages.

When a foreign word for translation is fed into the machine, in the form of an electro-mathematical symbol on a tape or card, the machine will run through its "memory" and if it finds that symbol as record, will automatically emit a predetermined equivalent - the English word.

This admittedly will amount to a crude word-for-word translation, lacking syntax, but will nevertheless be extremely valuable, the designers say, for such purposes as scientists' translations of foreign technical papers in which vocabulary is far more of a problem than syntax.

(editorial, New York Times, 1 June 1949):

How is the machine to decide if the French word "pont" is to be translated as "bridge" or "deck" or to know that "operation" in German means a surgical operation? All the machine can do is to simplify the task of looking up words in a dictionary and setting down their English equivalents on a tape, so that the translator still has to frame the proper sentences and give the words their contextual meaning.

We are still far from the machine into which we will pour cards and pull out great poetry or great novels. In fact we shall never reach that stage.

Max Zeldner
letter on SWAC demonstration, June 1949

(New York Herald Tribune, 26 June 1949):

... One need not be a language specialist to realize the futility of translating intelligently from one language to another by machine...

Let us take to the machine... the fifty-five hebrew words that make up the famous 23rd Psalm. Let us also assume that the intricate and efficient mechanism could translate correctly all the Hebrew verb forms, and every noun with its pronominal suffixes, the possessive and construct state; and that it could, moreover, give the correct meaning of the jussive, cohortative and consecutive. We then press the proper buttons on the translating machine and out comes -

"Lord my shepherd no I will lack. In dwellings of grass he will cause me to lie down; on waters of resting places he will guide me. My life he will turn back; he will lead me in circles of justness so that his name. Also because I will go in valley great darkness no I will fear bad because Thou with me your tribe and your support they will console me. Thou shalt arrange before me a table opposite my enemies; Thou madest fat with oil my head my goblet saturation. But good and kindness he will chase me all days of my life; and I shall rest in the house of Lord to length days."

Surely, only those in a "snake pit" shall follow such machine-made translations...

Warren Weaver's memorandum
July 1949

contextual analysis:

... the concept of a translation process in which, in determining the meaning for a word, account is taken of the immediate (2N word) context."

theorem of McCulloch and Pitts (1943):

"states that a robot (or a computer) constructed with regenerative loops of a certain formal character is capable of deducing any legitimate conclusion from a finite set of premises...

insofar as written language is an expression of logical character, this theorem assures one that the problem is at least formally solvable."

cryptography:

"If we have useful methods for solving almost any cryptographic problem, may it not be that with proper interpretation we already have useful methods for translation?"

universal language:

Think, by analogy, of individuals living in a series of tall closed towers, all erected over a common foundation. When they try to communicate with one another, they shout back and forth, each from his own closed tower. It is difficult to make the sound penetrate even the nearest towers, and communication proceeds very poorly indeed. But, when an individual goes down his tower, he finds himself in a great open basement, common to all the towers. Here he establishes easy and useful communication with the persons who have also descended from their towers.

Thus it may be true that the way to translate from Chinese to Arabic, or from Russian to Portuguese, is not to attempt the direct route, shouting from tower to tower. Perhaps the way is to descend, from each language, down to the common base of human communication - the real but as yet undiscovered universal language - and then re-emerge by whatever particular route is convenient.

A translator's reaction to MT in 1949

... translation is an art; something which at every step involves personal choice between uncodifiable alternatives; not merely direct substitutions of equated sets of symbols but choices of values dependent for their soundness on the whole antecedent education and personality of the translator.

the resulting literary style would be atrocious and fuller of "howlers" and false values than the worst that any human translator produces.

[computers] might be made to turn out a rough draft which a competent editor versed in the subject matter, though unacquainted with the *foreign language*, could then pull into shape.

[But] the machine would need to be given an enormous memory not only for variants of terms but for rules of accidence and syntax.

(J.E.Holmström: **Report on interlingual scientific and technical dictionaries.** Paris: Unesco, 1951.)

First research, 1949-1950

Erwin Reifler at University of Washington (Seattle)

first proposals for pre- and post-editing

Abraham Kaplan at RAND Corporation: statistical investigations

"most practical context is ... one word on each side, increased to two if one of the context words is a particle"

Victor Oswald (UCLA) and Stuart Fletcher (NBS) analysis of German syntax

(original German):

Bevor wir diese Definition im Einzelnen zergliedern, wollen wir einige Beispiele von Mengen betrachten, die uns anschauliches Material zum Verständnis der Definition liefern sollen.

(rearranged):

Bevor wir zergliedern diese Definition im Einzelnen, wir wollen betrachten einige Beispiele von Mengen, die sollen liefern uns anschauliches Material zum Verständnis der Definition.

(English word for word translation):

Before we analyze this definition in detail we want-to regard some examples of sets, which shall furnish us perceptible material for-the understanding of-the definition.

Bar-Hillel

Appointed to Massachusetts Institute of Technology, May 1951

Visits all US sites involved in MT, October 1951

Writes survey report, 1951

Convenes first MT conference, June 1952

Yehoshua Bar-Hillel
state of mechanical translation 1951

It seems obvious that fully automatic MT, i.e. one without human intervention between putting the foreign text into the reading organ of the mechanical translator and reading off its output, is achievable only at the price of accuracy.

[W]ith a lowering of the target, there appear less ambitious aims the achievement of which is still theoretically and practically viable

For those targets in which high accuracy is a *conditio sine qua non,* pure MT has to be given up in favor of mixed MT, i.e. a translation process in which a human brain intervenes.

the post-editor's task is ... rather easy, if one assumes that a machine has eliminated all the grammatical ambiguities and part or all of the so-called "idioms," and has rearranged the text in the TL order

Whereas specific MT [i.e. only two languages] will, in all probability, continue to be mainly an application of trial-and-error investigations, general MT will require establishment of a *universal,* or at least *general grammar,* perhaps even the construction of a whole artificial exchange-language.

The usual combination of metaphysical preconceptions, Aristotelian logic, and complete ignorance of any knowledge with respect to the so-called exotic languages is not a very promising mixture. Empirical open-mindedness, mathematical logic, and modern structural linguistics may perhaps prove to be a better one.

There are situations, where perhaps a restricted vocabulary or a restricted number of sentence-patterns, or perhaps both, are used or might be used.

The theoretical difficulties of such a type of MT are clearly less formidable.

More important, perhaps, might be the possibility of restricting, by voluntary convention, the richness of expression in writing abstracts of technical papers... to such a degree that sentence-pattern translation might easily and quickly applied...

First MT Conference
Massachusetts Institute of Technology
17 to 20 June 1952

Convened by Yehoshua Bar-Hillel
Financial support from Rockefeller Foundation

Participants:

Y. Bar-Hillel (MIT)
Jerome B.Wiesner (Research Laboratory of Electronics, MIT)
Jay W.Forrester (Digital Computing Laboratory. MIT)
James W.Perry (Center of International Studies, MIT)
William N.Locke (Dept. of Modern Languages, MIT)
Vernon Tate (Director of Libraries, MIT)
Dudley A.Buck (Dept. of Electrical Engineering, MIT)
Olaf Helmer (Rand Corporation)
Harry D.Huskey (NBS Institute for Numerical Analysis)
William E.Bull (Dept.Spanish, UCLA)
Victor A.Oswald (Dept. German, UCLA)
Erwin Reifler (University of Washington, Seattle)
Stuart Dodd (University of Washington, Seattle)
Duncan Harkin (Department of Defense, Washington DC)
A.Craig Reynolds (IBM Endicott Laboratories)
Leon Dostert (Georgetown University)
Victor H. Yngve (University of Chicago)
Andrew D.Booth (Birkbeck College, London).

First MT conference – topics (1)

Pre-editing

goal of pre-editing (Reifler):

"a graphic supplementation of the conventional form of the foreign message which raises its graphic-semantic explicitness to the level necessary for a mechanical translation."

dealing not only with "morphological and syntactical ambiguities" but also with

> "the rearrangement of the FL [i.e. source] text in accordance with a standard order in the TL following a set of instructions available to him in his own language."

Reifler's proposed device to support pre-editing:

When the pre-editor dials the conventional graphic form of the foreign message into the translation mechanism, it would first pass through the mechanical dictionary. Whenever in terms of the target language no multiple meanings are involved, the dictionary mechanism would not intervene and the dialled material would move on to the next stage in the translation process. Otherwise a device would call the attention of the pre-editor to the fact that multiple meanings are involved and the dictionary entry concerned would appear on a screen. The pre-editor would then select the meaning required by the context and dial the distinctive graphic symbol representative of this meaning and supplied by the dictionary.

Proposes that authors should write in MT-ese, i.e. regular language

Reifler's new orthography:

> "all nouns would have... a capital first letter..., all principal verbs ... a capital second letter and all attributive adjectives ... a capital third letter..."
>
> e.g. German *er hegt die fromme Hoffnung*
> as "er hEgt die frOmme Hoffnung"

First MT conference – topics (2)

Post-editing

How Bar-Hillel saw the problem:

> For a particular sentence, say of 10 words length, this can easily result in possible combinations of words in the target language extending to several thousands of more or less meaningful combinations
> (as cited by [Reynolds 1954]).

> If one takes into account the fact that the post-editor will receive instructions, in his own language, for handling certain strange-looking combinations, that certain words with many possible translations might reoccur in the passage quite frequently in this same meaning so that time-consuming decisions will not have to be repeated, and so on, it should be clear that the burden on the post-editor will not be too heavy.

The challenge:

> the "task of finding a good combination of the mechanical methods (and perhaps others), either for human or machine translation, should prove to be interesting not only for [MT] but also for the theoretical linguist."

First MT conference – topics (3)

Idiomatic expressions

Bar-Hillel's methods for dealing the German 'idiomatic' phrase *es gibt...*
translated as either *there is/are*
or as *he gives, she gives* or *it gives*

1) the dictionary includes *there* as an additional 'correlate' of *es,*
and *is* and *are* as additional correlates of *gibt.*
But: "it works too well", generating not just:
there is/are and *he/she/it gives*
but also *he/she/it is.*
(thus *she is a doll* as translation of *es gibt eine Puppe.)*

2) provide a phrase dictionary,
with *there is (are)* as translation of *es gibt*
as an alternative to literal version
(need to apply grammatical rules before consulting phrase dictionary, e.g. for questions: *gibt es einen Unterschied?)*

task of the post-editor to recognise when *he/she/it gives...*
should be replaced by *there is* or *there are.*

3) give post-editor responsibility for recognising idiomatic usage,
i.e. all 'idioms' would be translated literally word for word,
post-editor or reader would have to 'know' that
raw translation *it (he, she) gives*
might be replaced in some contexts by *there is (are).*
(problem: monolingual reader with no knowledge of the source)

Problem of missing idioms:
The main danger in not having sufficient idioms in a phrase-dictionary is not in the fact that some literal translations would be jibberish, it lies in the fact that some of these translations will make sense but the wrong sense and the post-editor will be unable to find this out.

First MT conference – topics (4)

Controlled language

Stuart Dodd (University of Washington, Seattle) proposed
Model English:

> "standardization of English syntax as a means of simplifying the use of English either as a source language or as a target language"

a) regularisation of verb forms,
e.g. *She did be loved* instead *of She was loved;*

b) nominative forms of pronouns,
e.g. / *will send he to she;*

c) standardisation of word order,
e.g. adverbs always before verbs,
direct objects before indirect;

d) use of words in one meaning only,
e.g. *tank* to mean only water tank,

e) obligatory use of qualifiers,
e.g. always *army tank.*

Reifler proposed use not only in source texts to ease problems of analysis but also in MT output to reduce task of post-editor:.

> "either restrict post-editorial interpretation to a minimum, or it may even make it completely superfluous."

First MT conference – topics (5)

Universals

Bar-Hillel had argued (in 1951):

> 'general MT' (into more than one target language) would necessarily "require the establishment of a universal, or at least general, grammar, perhaps even the construction of a whole artificial language."

Reifler saw answer in comparative-historical linguistics to identify 'real universals' but also 'pseudo-universals':

> by arbitrarily attributing grammatical meanings to linguistic forms which they, in fact, do not have, namely by changing the structure of a language, we may, for instance, within the limitations of intelligibility, so modify the grammar of a language as to bring it more in line with the grammar of other languages.

Mandarin Chinese:

t'a^{1} *tsou*^{3}*-ti k'uai*^{4} (English *he walks quickly)*
where *k'uai*^{4} = "to be quickness" or "to be quick"
tsou^{3}*-ti* = "walk's" or "of walk";
a literal translation : "he is quickness of walk".

In other contexts *tsou*^{3}*-ti* = "walking", therefore (with arbitrary equation of *-ti* and English *-ing):*

> We may therefore render the Mandarin sentence by "he walk-ing quick". This is bad English, but perfectly intelligible and, because it permits a word-to-word translation, has the great advantage of simplifying the mechanical correlation problem.

First MT conference – topics (6)

Sublanguages

Bar-Hillel (1951) mentioned restricted languages:

e.g. Basic English, artificial languages (Esperanto), air traffic control, meteorology

Victor Oswald proposed micro-glossaries

> an alternative arrangement is possible: to replace the battery of specialists by a series of permanent micro-glossaries, each of which would provide no more than two-to-one, and a preponderance of one-to-one, TL equivalents.

Vocabulary identified by statistical analysis:

> the data of all frequency counts fall into the same pattern, which means that a frequency count of any micro-segment of any language – say the nouns in German contexts pertaining to brain surgery – should give a parabolic curve where high-frequency elements ought to dispose of eighty-percent of all running nouns.

Familiarity with 80% of technical words in article sufficient for comprehension.

Furthermore, similar frequency distribution for non-technical words:

> In other words, brain surgeons writing on brain surgery are not only compelled to choose their technical nouns from a limited vocabulary, but their patterns of communication are so limited by practice and convention that even the range of non-technical nouns is predictable.

But Oswald cautious: "[t]heir ultimate efficiency remains untested, however, and it is possible that it might be prohibitively expensive to produce them."

First MT conference – topics (7)

Statistics

Enthusiasm for statistical analysis and data:
micro-semantics (Oswald)
disambiguation in context (Kaplan)
failure of traditional linguistics to provide data on vocabulary and syntax

statistical data from real texts could fill the gap

"A discussion of the means required ... showed clearly that the analysis could be facilitated by the use of punched cards" (Reynolds)

However:

William E. Bull: scepticism about frequency analyses -

> There exists no scientific method of establishing a limited vocabulary which will translate any predictable percentage of the content (not the volume) of heterogeneous material... A micro-vocabulary appears feasible only if one is dealing with a micro-subject, a field in which the number of objective entities and the number of possible actions are extremely limited. The number of such fields is, probably, insignificant...

The dictionary problem for MT (William E. Bull):

> The limitations of machine translation which we must face are, vocabularywise, the inadequacy of a closed and rigid system operating as the medium of translation with an ever-expanding, open continuum.

First MT conference – topics (7)

Grammatical analysis

Bar-Hillel: operational grammars (to go beyond word for word translation):

to identify and disambiguate grammatical categories

to analyse syntactic structures

foundations:

1) methods used for language teaching

2) Oswald and Fletcher (1951) analysis of German syntax (constituency analysis), designed for SWAC computer

3) Bar-Hillel: categorial grammar (based on Ajdukiewicz)

grammatical categories are combinations of basic categories *n* and *s,* defined in terms of potentiality to combine with other categories,

e.g. an intransitive verb is *s/(n),*

(because it can combine with a noun *(n)* to its left to form a sentence *(s))*

a transitive verb is *s/(n)[n]* or *s/(n)[s]*

(combines with a noun to the left and either a noun or a phrase (sentence) to the right)

cancellation rules:

β α/(β) replaced by α

α /[β] β replaced by α

β α/(β)[ω] ω replaced by α

(problem): "word-category-list... for English... some million and a half entries [and] The preparation of such a list is certainly not a simple task, since all possible occurrences of these words in all kinds of syntactic construction have to be envisaged."

(however): "a linguist with a staff of a few assistants and clerks should be able to provide [an operational system] for any language that has already been more or less exhaustively described – like English, German, or Russian – within a year or two."

First MT conference – topics (8)

Pivot language

Leon Dostert (Georgetown University, Washington, D.C.):

"general MT (mechanical translation from one into *many* languages)... should be so developed that one translates first from the input language into one 'pivot' language (which in our case will, most likely, be English) and from that pivot language into any one of the output languages desired"

assumed a natural language could be the pivot/interlingua, other possibilities were:

Esperanto, Basic English and Dodd's Model English.

Note no direct link made between a 'pivot' language and a 'universal' language, i.e. not assumed that interlingua must be language-independent (universal)

Universals in 1952:

(elements to assist analysis or transfer between languages of dissimilar structures)

derived from:

logical syntax (Bar-Hillel)

comparative-historical linguistics (Reifler)

First MT conference – topics (9)

Modest aims

The limitations (Bar-Hillel):

> completely automatic and autonomous mechanical translation with unique correlates to the original text is, in general, practically excluded, even with respect to scientific texts... This being so, machine translation means no more than *mechanical aids to translation.* Only some kind of *brain-machine partnership* is envisaged.

But importance (Bar-Hillel):

> Even if it should turn out that none of the possible machine-brain partnerships would be more effective than a human translator, in the sense that they will be neither quicker nor cheaper nor more exact than the human translator, under conditions existing today or in the near future, I would strongly advocate a continuation of this research. Electronic machines will doubtless become cheaper, human brains probably more expensive. A partnership that could not stand free competition today may well outbid its human competitors in some not too remote future.

Expectations:

No fully automatic translation of near-human quality:
MT output would remain poor
post-editing essential
pre-editing and regularisation of input texts could ease analysis
statistical data on language could be valuable
disambiguation helped by identification of sublanguages
use of some regularities of syntax

Next steps

Leon Dostert:

"early creation of a pilot machine ... proving to the world not only the possibility but also the practicality of MT."

Georgetown University and IBM collaborate:
the first MT demonstration (January 1954)

(genuine computer implementation)

limitations:

250 words
6 rules of syntax
49 carefully selected sentences
search span restricted to immediately adjacent items
only two possible target language forms
rearrangement restricted to two adjacent items
no negatives
no imperatives
no coordinated or subordinated clauses
all verbs in third person
English article insertion for specific corpus

but the impressive idiomatic output
encouraged sponsorship
and: aim of MT as 'perfect' general-purpose systems with near-human quality output.

James Perry (MIT)
manual simulation of Russian-English translation
(September 1952)

On/Onto/At Fig.12 traced/mapped-out/drawn parabola according-to/along/in-accord-with which move thrown/deserted with/from velocity lOm/sec. under/below angle to/toward vertical line into/in/at 15°, 30°, 45°, 60°.

Industrial and Engineering Chemistry
(December 1952)

One wonders, though, of the outcome, should one of our more ardent and less responsible hunters of subversives get wind of these machines that may blithely swallow the words of an alien and repugnant tongue, and then spew forth what is possibly a strange and, therefore, suspect gibberish. Could any culpable and innocuous computer withstand the taunts and accusations leveled, perhaps, in its direction without manifesting an incoherent hum, suffering a twitch, or developing at least a mild paranoia? Our machines may now also need a built-in psychoanalytical circuit and couch for purposes of self-confession when mental doubts becloud their electronic tubes and snarl their recording devices.

Victor Yngve
MIT Research Laboratory of Electronics
(October 1953)
German into English simulation

Die CONVINCINGe CRITIQUE des CLASSICALen IDEA-OF-PROBABILITY IS eine der REMARKABLEen WORKS des AUTHORS. Er HAS BOTHen LAWe der GREATen NUMBERen ein DOUBLEes TO SHOWen: (1) wie sie IN seinem SYSTEM TO INTERPRETen ARE, (2) THAT sie THROUGH THISe INTERPRETATION NOT den CHARACTER von NOT-TRIVIALen DEMONSTRABLE PROPOSITIONen LOSEen. CORRESPONDS der EMPLOYEDen TROUBLE? I AM NOT SAFE, THAT es dem AUTHOR SUCCEEDED IS, den FIRSTen POINT so IN CLEARNESS TO SETen, THAT ALSO der UNEDUCATED READER WITH dem DESIRABLEen DEGREE-OF-EXACTNESS INFORMS wird.

IBM-Georgetown MT demonstration
7th January 1954

Neil Macdonald, *Computers and Automation,* February 1954:

> Many exciting possible developments are indicated by the success of the trial... 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.

New York Times, 8th January 1954:

... This may be the cumulation of centuries of search by scholars for a "mechanical translator". So far the system has a vocabulary of only 250 words. But there are no foreseeable limits to the number of words that the device can store or the number of languages it can be directed to translate.

... The "mechanical" part of the translation system, which is mostly electronic, is a standard commercial model of the largest International Business Machines "stock" computer.. the IBM Type 701.

... The "literary" part of the system is a mechanical model of language devised at Georgetown by Prof. Leon Dostert and Dr. Paul Garvin...

... Several short messages, within the 250-word range of the device, were tried. Included were brief statements in Russian about politics, law, mathematics, chemistry, metallurgy, communications and military affairs. The sentences were turned into good English without human intervention.

... 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.

... As soon as cards for Russian are completed, sets will be made for German and French. The other Slavic, Germanic and Romance languages can be set up at will.

IBM-Georgetown MT demonstration
7th January 1954

Christian Science Monitor, 11th January 1954:

It is expected... that within a few years there will be a number of "brains" translating all languages with equal aplomb and dispatch.

"The potential value of this experiment for the national interest in defense or in peace is readily seen," Prof. Leon Dostert... said...

"Those in charge of this experiment... now consider it to be definitely established that meaning conversion through electronic language translation is feasible." Although he emphasised it is not yet possible "to insert a Russian book at one end and come out with an English book at the other", 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, editorial, 13th January 1954:

Such an accomplishment, of course, is far from encompassing the several hundred thousand words which constitute a language. And with all the preparations for coping with syntax, one wonders if the results will not sometimes suggest the stiffness of the starch mentioned in one of the sentences as being produced by mechanical methods. Nevertheless, anything which gives promise of melting some of the difficulty which writers and speakers of different languages encounter in understanding each other – particularly as between English and Russian today – is certainly welcome.