

机器翻译论文选辑

科学技术文献出版社

083

8

机器翻译论文选辑

(限国内发行)

编辑者：中国科学技术情报研究所

出版者：科学技术文献出版社

印刷者：中国科学技术情报研究所印刷厂

新华书店北京发行所发行 各地新华书店经销

开本 $787 \times 1092 \cdot \frac{1}{16}$ 12印张 307千字

科技新书目：117—14

统一书号：17176·173 定价：1.10元

1979年7月出版 印数：12,600册

出版者说明

机器翻译是科技情报工作现代化中的一个重要环节。近年来，机器翻译研究无论在国内或是在国际上都有很大发展。关心和研究机器翻译的人日益增多。但是，有关的中文资料却很难找到。现将散在一些杂志上的部分论文汇集成册，以满足读者需要。为了保持原作的本来面目，一般未进行加工。

书末附有两篇有关最新发展情况的译文，这是过去未发表过的。

目 录

1) 俄汉机器翻译中的词序问题及其解决办法.....	(1)
2) 机器翻译浅说.....	(8)
3) 机器翻译研究——庆祝建国十周年.....	(14)
4) 俄汉机器翻译初步试验成功.....	(24)
5) 俄汉机器翻译试验.....	(25)
6) 机器词典.....	(42)
7) 机器语法.....	(61)
8) 从俄语到汉语.....	(78)
9) 俄汉机器翻译规则系统新旧方案比较.....	(91)
10) 机器翻译及其发展情况.....	(115)
11) 机器翻译和文字改革.....	(137)
12) 机器翻译中的词序问题.....	(145)
13) 俄汉机器翻译中关于标点符号的分析.....	(155)
附录:	
1. 七十年代的机器翻译.....	(168)
2. 机器翻译和机助翻译的现状.....	(176)

俄汉机器翻译中的词序问题及其解决办法*

刘 涌 泉

(中国科学院语言研究所)

在谈正题之前，先介绍一下俄汉机器翻译工作的一般情况。这项研究工作从1958年1月才正式开始进行。工作中使用的方法和苏联科学院精密仪器与计算技术研究所采用的方法一样。所根据的材料主要是取自数学文献。

俄汉机器翻译系统由下面两个独立的部分组成：俄文句子的分析和汉文句子的综合。第一部分的作用是分析俄文句子，得到必要的词汇语法信息，为综合准备条件；第二部分的作用是根据分析所得的信息组成汉文句子。

汉文句子综合部分有以下两个线路图表：“改变词序”和“翻译形态特征”。第一个线路图表是综合部分中的基本图表，篇幅相当大，而且很复杂。句子的全部句法加工都在这个图表中进行。第二个线路图表比第一个小得多，它主要是对担负“谓语”作用的动词加工：俄语动词的形态成分在这里由汉语的词汇成分代替。例如，“Он написал** статью”，（“他写成了论文”）。

以上就是俄汉机器翻译的一般情况。现在转回报告本题。

报告的第一部分将阐述机器翻译中词序的概念和意义。第二部分将具体地叙述俄汉机器翻译中的词序问题如何解决。

一 机器翻译中词序的概念和意义

词序是一种句法范畴，因此，应该把它了解为由一个词或一串词组成的各个和造句法有关的单位的次序，而不应该了解为作为词汇学和形态学单位的“单个词”的次序。

当我们谈改变词序时，我们所指的是改变造句法单位（定语、补语等等）的次序，而不是改变词汇学或形态学单位（名词、形容词等等）的次序。

“词序”这一概念应用在机器翻译上还具有更广泛的意义。词序不仅意味着简单句中各成分的位置，而且还意味着更大的句法单位（复合句中的分句）的位置。

从上面我们可以看出，“词序”这个术语不是很精密的，它带有很大的假定性。

词序在各种不同的语言中起着不同的作用。在一些语言里它只是次要的语法手段，但在另一些语言中它却起着非常重要的或甚至是主要的作用。

词序的重要性，在两种不同类型的语言之间进行翻译时，表现得特别明显。它是研究工作中的一个重心，它决定着机器翻译能否成功。

为了节约时间，关于词序在机器翻译中的意义不再多说了。本报告的第二部分自然会加

* 本文原是作者在1958年5月15—21日莫斯科举行的机器翻译会议上所作的报告。为了保持原来面目，在此发表时，只是在极个别的地方改动了一下。

** 黑体字是我们举例说明的地方。下同。

以具体说明。

二 俄汉机器翻译中词序问题的解决

俄语和汉语是两种不同类型的语言。俄语是综合型、屈折型的,而汉语是分析型的,形态成分很不发达的。因此,两种语言各有不少独特的结构特点。词序方面也有许多显著的差别。

从俄文译成汉文时,有时不仅需要改变定语、同位语和状语的次序(正如一些别的机器翻译系统所作的那样),而且还要改变主语、补语、插入语、独立短语和从句的次序。换言之,差不多所有的造句法单位都有可能改变次序。

为了说明俄汉语词序上有哪些差别及俄汉机器翻译中词序改变过程是多么复杂,我们现在在举几个例子看看*。

1) “Какую пользу может принести этот прием?”

(什么)(好处)(可以)(带来)(这个)(方法)

汉语译文:“这个方法可以带来什么好处?”

2) “Какую из этих двух производных она символизирует?”

(哪个)(之中的)(这)(两个)(导数)(它)(代表)

汉语译文:“它代表这两个导数之中的哪个?”

3) “……приращение функции, имеющей производную в точке x ,
(变量)(函数)(具有)(导数)(在)(点)(x)

может быть представлено в виде……”

(可以)(标作)

汉语译文:“……在点 x 具有导数〔的〕**函数〔的〕变量可以标作……”

4) “Как цифровой код сообщения на русском языке преобразуется в
(如何)(数码)(信息)(俄)(文)(变成)

цифровой код сообщения на китайском языке?”

(数码)(信息)(汉)(文)

汉语译文:“俄文信息〔的〕数码如何变成汉文信息〔的〕数码?”

5) “Производная этой функции равна, как мы знаем, $f'(x) \varphi'$
(导数)(这个)(函数)(等于)(正如)(我们)(所知)

(t)……”

汉语译文:“正如我们所知,这个函数〔的〕导数等于 $f'(x) \varphi'(t)$ ……”

6) “Конечно, во всех случаях требуется оценка той погрешности
(当然)(在)(一切)(情况)(需要)(估计)(那个)(误差)

$o/\Delta x$, которую мы допускаем, заменяя приращение функции Δy

($o/\Delta x$)(我们)(造成)(代替)(变量)(函数)

ее дифференциалом dy .”

(微分)

* 数学方面的例子引自А. Я. 辛钦著《数学分析简明教程》。

**方括号中的字是在翻译形态成分的过程中添加的,下同。

汉语译文：“当然，在一切情况〔都〕*需要估计我们〔在〕〔用〕微分 dy 代替函数 Δy 〔的〕变量〔时〕造成〔的〕那个误差 $o/\Delta x$ 。”

为了改变上述各种类型的词序，在俄汉机器翻译系统中设有一个“改变词序”的线路图表。这个图表根据词典**和分析部分所提供的信息进行工作。

“改变词序”这个线路图表又由下面五个相互连系并依次进行工作的次线路图表组成：“定语和同位语”、“状语”、“主语”、“补语”和“付句”。前四个次图表改变简单句与复合句各分句中的句子成分的次序，后一个次图表改变复合句中付句的次序。

在分析这个线路图表之前，先谈一谈编制这个图表的一些基本原则：

（一）改变词序时，首先应该确定句中改变词序的基本轴心。这个基本轴心是谓语。主语、补语和一部分状语的位置都依靠谓语而定。定语的轴心是它的被定语。

（二）各次图表的排列次序及每一次图表中各步骤的次序非常重要。不重视这种次序，就要带来很多翻译上的错误。例如：“Какую из этих двух производных она символизирует?”（“它代表这两个导数之中的哪个？”）。如果直接补语的加工先于间接补语，则这句话中的间接补语（из этих двух производных）就无法在左面找到它所补充说明的那个词（какую）。

为了工作顺序安排得合理，带有“特别副词状语”特征的词在“定语同位语”次图表中加工；带有“补语”特征并处于定语词组、形动词短语或副动词短语之中的词在各相应的次图表中进行加工。详细情况见下面。

（三）与上述两点相联系的是：必须使定语和被定语或其他相互联系的造句法单位连成为一个整体（组合）。达到这一目的办法是：所有加入这个组合的词都给与一个共同的号码。不这样做，不是得出歪曲原文的译文，就是得到词的无意义的堆积。如：“……приращение функции, имеющей производную в точке x ……”就会译成“函数〔的〕具有变量量）（函数）（具有）（导数）〔在〕〔点〕〔 x 〕点导数在〔 x 〕”。

（四）在改变词序的过程中要调整标点符号——在把俄文译成汉文时，有时需要增加或去掉标点符号。这项工作宜于在改变词序的过程中进行。例如：“Производная этой функции равна, как мы знаем, $f'(x)\varphi'(t)$ ……”（“正如我们所知，这个函数的导数等于 $f'(x)\varphi'(t)$ ……”）。为了去掉原文插入语左边的逗号及把它右边的逗号附在它身上，在改变词序时增加以下两个步骤：

（1）把该词（插入语）左边的“逗号”削掉。

（2）该词右边的“逗号”，给予和该词同一的号码。

（五）在改变词序的过程中有时需要添加汉字以翻译俄语的形态成分——在把俄文译成汉文时，由第二格名词形成的非一致定语应该移至被定语之前；同时还需要在定语和被定语之间添加“的”字，以表达俄语名词里由格的范畴所表示的属性。例如：“дифференциал（微分）функции（函数）”这个词按下列方式加工：

* 这个“都”字是在进行最后一步工作（修辞）时加上的。

** 机器翻译使用的词典除提供译文以外，还指出一系列特征，（如一个词属于哪一词类，属于哪一组别，是否为动名词，是否要求一定的格等特征），为俄文的进一步分析和组成汉文句子提供信息。

- 1/2, 12/ 检查分句中是否有具有“第二格名词作为非一致定语”特征的词;
- 2/3, 5/ 检查该词有无“动名词”特征;
- 5/6, 8/ 检查该词右边是否有另一个具有“第二格名词作为非一致定语”特征的词;
- 8/9, 11/ 检查该词左边是否有另一个具有“名词”特征的词;
- 9/0, 0/ 把该词移至找到的那个词之前, 给与该词以找到的那个词同一的号码, 并在找到的那个词之前加“的”字。

于是, 就可得到: “函数〔的〕微分”。

从上面第四第五两点可以看出, 这个线路图表的功用不仅是改变词序, 而且还调整标点符号和翻译一部分俄语形态特征。

现在分别地谈一谈“改变词序”这一线路图表之中的各个次线路图表。

在“定语和同位语”次线路图表中先进行“同位语”的加工, 后进行“定语”的加工。

数学文献中的同位语绝大多数由公式组成。处于名词后的公式又常常是同位语, 只有很少的一些是定语*。同位语不移位, 但定语要移至被定语前。

区别同位语或定语, 没有什么困难, 只需检查一下前一个词或前几个词, 就可得到明确的结果。如果公式在 точка, переменная, функция, приращение, величина, производная, дифференциал, катет, угол, треугольник 这一类型的名词之后, 则是同位语。如公式不在这类名词之后, 则是定语。例如:

- 1) “В прямоугольном треугольнике МТР катет ТР равен другому МР……”
“在直角三角形МТР中, 边ТР等于另一边МР……”

- 2) “……в частности, производная порядка n

$$y^{(n)} = n! a_0$$

есть даже многочлен степени 0, т. е. постоянное число……”

“……特别是n级导数

$$y^{(n)} = n! a_0$$

甚至是零次的多项式, 换句话说就是常数……”

这个次图表的“定语”部分对带有下列特征的词进行加工: “特别副词状语”, “一致定语”, “由无定式动词表示的非一致定语”, “由第二格名词表示的非一致定语”, “由前置词短语表示的非一致定语”, “由后置形容词表示的独立定语”, “形动词短语”**。

以上各种定语中只有第一种, 即“一致定语”在俄语中是前置的 (俄语倒装词序的一些情况除外), 而其他的几种都是后置的 (“разного рода вопросы”一类型的组合另外考虑)。

然而, 大家知道, 汉语的定语一般都是前置的。因此, 必须把俄语中所有的后置定语移至被定语之前。如果被定语前后都有定语, 后置定语通常不仅移至被定语前, 而且还移至前置定语之前。例如:

- 1) “Цифровой код сообщения на русском языке”

(数码) (信息) (俄) (文)

“俄文〔的〕信息〔的〕数码”

* 为了机器翻译的方便, 有些补语我们看作是定语。

** 和一般的分法不同, 我们没把“由后置形容词表示的独立定语”和“形动词短语”算作“一致定语”。

- 2) “Приращение функции, имеющей производную в точке x ”
(变量) (函数) (具有) (导数) (在) (点) (x)

“在点 x 具有导数〔的〕函数〔的〕变量”

应该指出,在这个次图表中还要对带有“特别副词状语”特征的词,对带有“直接补语”特征并位于动名词定语之后的词,以及对形动词短语中带有“直接补语”、“间接补语”或“状语”特征的词进行加工。有这种特征的词如不在此处加工,就要得到错误的译文。例如:

- 1) “Выражение (4) для производной служит прежде всего очень удобным
(式) (4) (对于...(导数) (是) (首) (先) (很) (方便的)
来说)

обозначением ее.....”

(记号) (它的)

“式(4)对于导数来说首先是它的很方便的记号.....”

- 2) “Дело защиты мира священо.”

(事业) (保卫) (和平) (神圣的)

“保卫和平〔的〕事业〔是〕神圣的”。

- 3) “погрешность, вводимая такой заменой”

(误差) (引起) (这种) (代替)

“〔由〕这种代替引起〔的〕误差”。

同时还必须说明,带有上述特征的词在这里加工之后,就不再到其他次图表中加工。

状语次图表对带有下列特征的词进行加工:“副词状语”,“状语+第五格名词”,“带前置词的状态”,“插入语”和“副动词短语”。

值得特别提出的是,俄语数学文献中任何类型的状语如果位于谓语或动词性质的词(形动词、副动词等)之后,则在译成汉语时绝大多数都移到谓语或动词性质的词之前。例如:

- 1) “Функция y дифференцируема в точке x .”

“函数 y 在点 x 是可微的”。

- 2) “Функция, имеющая производную в точке x ”

“在点 x 具有导数的函数.....”

应该说明:这个次图表在对副动词短语进行加工时,同时也对副动词短语之中的直接和间接补语进行加工。在这个次图表加过工的词不再到以下的次图表中加工。

现在来谈“主语”次图表。这个次图表对带有“主语”特征的词进行加工,并把一些置在谓语后边的主语移至谓语之前。谓语之后的主语,如其前面无谓语副词或 *иметься*、*существовать*、*требоваться* 一类型的动词,或其后面无 *который* 类型的形容词-连词,则应移至谓语之前。

这个次图表比较简单,其所以如此,首先是由于俄汉语中的主语,尤其是在数学文献中,一般是在谓语之前;其次是由于在翻译过程中句子成分可以酌量改变,这就是说,俄语中的主语在译成汉语时不一定非译成主语不可。例如:1) “Во всех случаях требуется оценка погрешности”,“在一切情况都需要估计误差”;2) “Трудно решить эту задачу”,“很难解决这个问题”。第一句话中的主语 *оценка* 在译成汉语时变成了谓语的一部分,而第二句话中的主语 *решить* 则变成了一个纯粹的谓语。

“补语”次图表先对“间接补语”加工，后对“直接补语”加工。这个次图表中的“间接补语”又按照下列顺序进行加工：间接补语+第二格→间接补语+第三格→间接补语+第四格（带前置词的）→间接补语+第五格→间接补语+第六格。

翻译补语时，句子成分的改变，更是常见的。这种现象同时又支配着词序的改变。

例如，

1) “Я пишу карандашом”，“我〔用〕铅笔写字”；

（我）（写字）（铅笔）

2) “Я занимаюсь машинным переводом”。“我研究机器翻译”。

（我）（研究）（机器）（翻译）

在前一种情况下，间接补语译作“行为方法状语”，因而移至谓语之前。在后一种情况下，间接补语译作“补语”（亦即“宾语”），因而原位不动。（译成“行为方法状语”，还是译成“补语”，这靠动词来区别：заниматься 一类型的动词在词典里给与特殊组别的代码。）

前面几个次图表中已加过工的补语就不必再在这个次图表进行加工。

“副句”次图表在现阶段只对定语副句加工。对于翻译数学文献来说，这一点基本上足够了。

在句中各分句范围内的工作都完成之后，这个次图表才开始工作。与其他几个次图表不同，这个次图表不是在分句范围之内进行工作，而是在整个句子范围之内进行工作。下面把这个次图表简单地说一下。

开始时，先检查句中有无带“形容词-连词”特征的词，（即который类型的词）；如果有，则这个次图表就开始按下列方式工作：从该词起，往右直到分句末，所有的词都给与和该词同一的号码，并置“的”字于分句末；向该词左面找带有“名词”特征的词（即它的被定语），把它（即找到的词）移至该词（即分句）的后面。（这里，把被定语移至定语后，从技术操作上讲较经济。）

现在举几个例子说明简单句和复合句中的词序如何改变。

1) Какую пользу может принести этот прием?

词的代码： 1 2 3 4 5 6 7

原文的成分特征：（定）（补）（谓）（谓）（定）（主）

意义：（什么）（好处）（可以）（带来）（这个）（方法）

在第一个次图表加工后，这句话就变成：“ $\boxed{1 \rightarrow 2} * + 2 \ 3 \ 4 \ \boxed{5 \rightarrow 6} + 6 \ 7$ ”。第二

个次图表不工作，因为没有可加工的。第三个次图表加工后的结果是：“ $\boxed{1 \rightarrow 2} + 2$

$\boxed{5 \rightarrow 6} + 6 \ 3 \ 4 \ 7$ ”。第四个次图表加工之后得到：“ $\boxed{5 \rightarrow 6} + 6 \ 3 \ 4 \ \boxed{1 \rightarrow 2} + 2$

7”。译文是“这个方法可以带来什么好处？”

* 方块中的号码是加工后改变了的号码， $\boxed{1 \rightarrow 2}$ 表示：1 改变成 2。下同。

2) Машина сохраняет ту скорость, какую
词的代码: 1 2 3 4 5 6
原文的成分特征: (主) (谓) (定) (补) (形容词-连词)
意义: (机器) (保持) (那种) (速度)

она имеет в тот момент.
词的代码: 7 8 9 10 11 12
原文的成分特征: (主) (谓) (前置词) (定) (状)
意义: (它) (具有) (在) (那个) (时刻)

经第一个次图表加工后,这句话变成: “1 2 3→4 + 4 5 6 7 8 9 10→11 + 11 12”。第二个次图表加工的结果是: “1 2 3→4 + 4 5 6 7 9→11→8 + 10→11→8 + 11→8 + 8 12”。第三第四两个次图表加工后无任何变更。第五个次图表加工之后得到: “1 2 6 + 7→6 + 9→11→8→6 + 10→11→8→6 + 11→8→6 + 8→6 + 的 3→4 + 4 12”。译文是“机器保持它在那个时刻具有的那种速度”^{*}。

3) Конечно, во всех случаях требуется оценка
词的代码: 1 2 3 4 5 6 7
原文的成分特征: (状) (前置词) (定) (状) (谓) (主)
意义: (当然) (在) (一切) (情况) (需要) (估计)
той погрешности Δ_x , которую мы допускаем,
词的代码: 8 9 10 11 12 13 14 15
原文的成分特征: (定) (补) (同) (形容词-连词) (主) (谓)
意义: (那个) (误差) (我们) (造成)
заменяя приращение функции Δ_y ее дифференциалом dy .
词的代码: 16 17 18 19 20 21 22 23
原文的成分特征: (状) (直补) (定) (同) (定) (间补) (同)
意义: (代替) (变量) (函数) (微分)

在第一个次图表加工后,这句话变成: “1 2 3 4→5 + 5 6 7 8→9 + 9 + 10→9 | 11 12 13 14 15 16 18→17 + 19→18→17 + 的 + 17 20→21 + 21 + 22→21 23”。第二个次图表加工的结果是: “1 2 3→5 + 4→5 + 5 6 7 8→9 + 9 + 10→9 | 11 12 13 + 在 + 用 + 20→21→16→14 + 21→16→14 + 22→21→16→14 + 16→14 + 18→17→16→14 + 19→18→17→16→14 + 的 + 17→16→14 + 时 + 14 23”。第三第四两个次图表加工后无任何变动。第五个次图表加工以后得到: “1 2 3→5 + 4→5 + 5 6 7 12 + 13→12 + 在

* 说明: 代码5 (代表逗号) 在第五次图表加工过程中已削掉; какую一词在多义词典里没有给予译文, 因而, 尽管加工后还存在, 但对译文无妨。

+ 用 + $\boxed{20 \rightarrow 21 \rightarrow 16 \rightarrow 14 \rightarrow 12}$ + $\boxed{21 \rightarrow 16 \rightarrow 14 \rightarrow 12}$ + $\boxed{22 \rightarrow 21 \rightarrow 16 \rightarrow 14 \rightarrow 12}$ + $\boxed{16 \rightarrow 14 \rightarrow 12}$
 + $\boxed{18 \rightarrow 17 \rightarrow 16 \rightarrow 14 \rightarrow 12}$ + $\boxed{19 \rightarrow 18 \rightarrow 17 \rightarrow 16 \rightarrow 14 \rightarrow 12}$ + 的 + $\boxed{17 \rightarrow 16 \rightarrow 14 \rightarrow 12}$ + 时 +
 $\boxed{14 \rightarrow 12}$ + 的 $\boxed{8 \rightarrow 9}$ + 9 + $\boxed{10 \rightarrow 9}$ 23”。译文是：“当然，在一切情况〔都〕需要估计我们〔在〕〔用〕微分 d_v 代替函数 Δ_v 〔的〕变量〔时〕造成〔的〕那个误差。 $(\Delta_x)^*$ ”。

最后应该指出，还有一系列问题在现阶段没有解决。例如：

1) Дифференциалом функции называется главная линейная часть ее прираще-
 ния.

2) Надо изучать историю и грамматику русского языка.

第一个例子中的困难在于：在改变词序之后，得到不能令人满意的译文（“它的变量的主要线性部分叫作函数的微分”）；“它的”一词在改变词序后出现在它所代替的那个词（函数）之前了。

第二个例子中的困难是：定语“русского языка”即可以有一个被定语“грамматику”，又可以有两个被定语“историю”和“грамматику”。

此外，定语副句译法中的一些问题也还没有得到很好的解决。

尽管如此，现在仍可以这样说：俄汉机器翻译中词序改变的问题原则上已经解决了。当然，现有的这个线路图表今后还得进一步修改和补充。

〔原载《语言研究》第四期，1959年〕

机 器 翻 译 浅 说

刘 涌 泉

一 什么是机器翻译

机器翻译，也叫自动化翻译。从字面上就可看出它的意义：这是一种由机器（具体地说就是电子计算机）根据一定程序自动进行的翻译。

机器翻译，作为一门学科，它是语言学、数学和计算技术三门科学的产物。语言学家提供适合于机器进行工作的词典和语法规则；数学家把语言学家提供的材料代码化和程序化，即变成机器语言；计算技术专家造出便于进行翻译的计算机。由此可见，缺少上述任何一方，机器翻译就不能实现；机器翻译实现得好与不好，也完全取决于上述三方面的共同努力。

根据翻译材料的不同，机器翻译可分作书面翻译和语音翻译。前者根据字母或其他书写符号进行翻译，后者根据说话的语音进行翻译。

* 说明：1) 代码11和15（都代表逗号）在加工过程中已削掉；2) “ee”一词在这里没有译文；3) “都”这个字是在进行最后一步工作（修辞）时加上的。

机器翻译按照参加翻译的语言多少，又可分作一对一的翻译和多种语言之间的翻译。一对一的翻译是从一种语言译成另一种语言，例如英俄机器翻译、俄汉机器翻译，等等。多种语言之间的翻译是借助媒介语同时可以从多种语言译成多种语言。

二 机器进行翻译是否可能

用机器进行翻译，看来好象是不可能的，但是进一步分析分析，就会得到肯定的结论。

1) 从语言方面来看：语言是一个一定的符号系统，它不是不可捉摸的。这个系统中的各个成分（词）以及这些成分的组合（句）都有其物质基础——语音以及反映在书面上的书写符号。由于有物质（语音和书写符号）作基础，因而就完全有可能利用现代技术来代替人的劳动进行翻译。再进一步说，语言之中的词虽然不少，但常用的并不多，句子的表现法虽然种类繁多，但是从它们当中也能分析出一定的类型来。如果估计到科技文献的语言特点：

a) 用词少，b) 绝大多数的词是单义词，c) 句子结构比较简单，d) 词序比较固定，e) 修辞色彩不浓，那么，就更不成问题了。

2) 从技术条件上来看：作为二十世纪科学技术中最伟大的成就之一的电子计算机，能够以非常高的速度进行数字运算和逻辑运算，运算的差错率只不过几万分之一。新式的电子计算机每秒钟可进行十万到二十万次运算。同时，人们现在还在研究制造每秒钟运算百万次或几百万次的计算机。由于有了这样的机器，机器翻译也就有了技术的保证。

3) 从翻译工作本身来看：人进行翻译时所需要的东西——词典和语法规则，机器在进行翻译之前已经得到。人进行翻译时所经过的过程，机器也同样遵守履行：先查词典得到词的意义和一些基本的语法特征（如词类等）；如果查到的词不止有一个意义，那么就要根据上下文取所需要的意义；在词汇意义和基本语法特征弄清楚之后，就要进一步明确各个词的语法意义（看一个动词是第几人称，什么时间；看一个名词是单数还是复数，是第几格；看句中每个词起什么句法作用）。在这以后，根据译入语的要求组成译文（这里包括改变词序，翻译原文词的一些形态特征以修辞）。

机器翻译可以实现，已经得到了不止一次的证明（例如，英译俄、法译俄。详见下面）。因此说，机器翻译已经不是一些热心者的幻想，它已经成了现实。现在的问题已经不再是可能与否的问题，而是走哪条路最好的问题了。

三 机器翻译的历史发展简况

机器翻译这个科目的历史已经有十多个年头了。1946年，英美开始注意这个问题。从1955年起，苏联开展了这个科目的研究工作。我国对这个问题正式进行研究开始于1958年初。截至目前为止，各国的注意力主要是在书面翻译方面，而且绝大多数只限制在科学技术文献的范围之内，并曾进行过几次试验；（1954年美国曾进行过一次简单的俄英机器翻译试验；机器内只存了250个词和一些简单的语法规则。1955年苏联进行了一次较复杂的英俄机器翻译试验；词典中有952个词，语法也比较复杂；这次所得的结果比美国进了一大步。另外，1957年苏联又进行了一次法俄机器翻译试验。）不过都还是同一语系的语言之间的对译。

近几年来，机器翻译工作以苏联开展得最快最好。上述的试验可以证明这一点。另外还可以举出一系列事实作为证明：1) 从研究单位的数目来看——1955年只有苏联科学院精密仪器与计算技术研究所和数学研究所两个单位研究机器翻译问题，然而现在，除上述两个单位外，又有语言研究所和列宁格勒大学等十来个单位对这个问题进行研究。2) 从研究的语言数目来看——过去只研究英俄和法俄翻译问题，而现在则已扩展到研究从汉、英、法、德、日等二十几种语言译成俄语的问题。3) 从翻译范围上看——由研究数学文献的翻译已扩展

到研究物理、化学、生物学、原子能、电子学 and 政论文章的翻译。4) 另外, 还在媒介语问题(即多种语言之间的翻译问题)的研究上进行了很多工作。

四 机器翻译的远景

随着科学技术的进展和语言研究的提高, 机器翻译在不久的将来就可得到更大的成绩, 从而可以在实际工作中得到应用, 以代替繁重的翻译工作。

机器翻译未来的情景可以作如下的简单描述:

1) 书面翻译方面——某一种文字的书籍放在机器的一端, 开动机器后, 机器的自读装置(利用光学方法或其他方法)自动地将文字变成信号送进机器, 机器便开始对送进来的信号加工, 分析综合, 将译出语的信号转换成译入语的信号, 然后通过输出装置将译入语的信号改变成文字, 送到新式印刷机上, 自动印刷和包装。这样, 一本本另一种文字的书籍便产生了。

当然, 我们还可以做到这样: 如果我们想把俄文译成汉文, 我们只需把机器上的俄文输出纽和汉文输入纽打开; 如果需要把汉文译成法文, 那么就把汉文输出纽和法文输入纽打开。以此类推。

2) 语音翻译方面——语音翻译机器将包括录音、语音分析、语音代码的变换(即翻译过程)、语音组合和放音等几个主要部分。由于是用半导体元件组成, 它的体积不会很大。将这样一架小机器摆在不同语言的人中间, 开动机器后, 人们就可各用自己的语言交谈。不同于一般交谈的地方只是中间有一个不大的间隙。

五 机器翻译如何具体实现

事先的准备: 1) 语言学工作——编出适用于机器的词典和语法规则(见附表一和二); 2) 数学程序工作——将词典和语法规则代码化和程序化; 3) 将所要翻译的材料代码化, 即通过穿孔纸带用数码代替每个字母。

整个翻译的过程简单描述如下: a) 输入——将打好的穿孔纸带上所表示的数码送进机器中去; b) 查词典——原文中词的代码和词典中词的代码进行比较, 相等时就算找到。单义词查到后就可从词典上得到意义(即译文代码)和一些语法特征; 可组成成语的词和多义词在词典中得到“可组成成语”和“多义词”特征, 要进入成语词典和多义词典进一步加工; c) 语法加工——先进行译出语的分析, 后进行译入语的组合。这里要通过十几道工序进行形态和句法加工(见附表三); d) 输出——把代码变成译入语的文字, 通过打字机印出。

六 机器翻译和语言学

最后再谈谈机器翻译和语言学的关系。它们的关系可从以下几个方面看出:

1) 语言学是机器翻译研究的基础——机器翻译是语言学(应用语言学)的一支, 它虽有自己一些特点, 但毕竟是以语言学为基础的。机器翻译工作中使用的一套理论和方法基本上是取自语言学(例如句段理论等), 同时, 机器翻译今后还将经常不断地吸取一般语言学中的新成就。不过这里需要说明一句: 机器翻译在利用一般语言学成就时通常不是机械地搬用, 而是适应着自己的需要对它们加以某些改造。

2) 机器翻译能促使语言学精密化——在科学技术大发展的今天, 语言科学由于和其他精密科学(如电声学)的结合, 逐步走向精密化。机器翻译的出现, 更促进了语言学的精密化。

机器翻译要求进行精确的语言对比研究, 找出两种语言之间词汇和语法的一致现象和不

一致现象以及消除这些不一致现象的单一方法。模棱两可的解决办法对机器是无效的。

机器翻译要求发展语言统计学：为了在机器翻译系统中合理安排词和语法范畴的次序，必须要统计词的使用频率及语法范畴和格式的出现频率。为了解决语音翻译问题，还要统计语音出现频率。

语音机器翻译要求进行大量的语言试验（而且不仅是一种语言的），求出语音质和量方面的全部数据，以备分析翻译过程中输入进来的语音和组合输出去的另一种语音之用。

3) 机器翻译可以检验、修订和丰富语言学——机器翻译实践证明：语言学中的一些成规和概念应该修订和补充。现在举两个例子来说明：a) 传统语法中的某些规范制定得不够合理。机器翻译对俄语传统的词类划分作了一些调整，如代词根据它的功用和形态特征归并到名词和形容词两类词中。b) 我们对俄语前置词的研究不够全面。直到目前为止，人们在论述前置词时，通常只是注意到它们要求第几格，有什么词汇意义，没有指出过它们可以有什么句法功能。机器翻译不仅照顾到它们的意义和要求什么格，还指出了它们的句法功能。例如前置词“O”的句法功能：不引起状语；和某些动词连用时引起补语，否则，引起定语。（要知道，不明确前置词的句法功能、定语、状语、补语在机器中是难以分辨出来的）。

机器翻译研究工作中使用了一些数学方法和概念，如集论、统计学和信息论的一些方法等。事实证明：有些数学方法不仅对机器翻译适用，而且对语言学的其他方面也照样适用。例如，苏联库拉吉娜同志曾利用集论研究语法结构；赛尔格波利斯基同志曾利用概率论研究确定语言亲属的问题。另外，还有人曾把信息论方法应用在构词法的研究上。

从上面可以看出，机器翻译直接或间接地对语言学已经发生了一定的影响。但是，这还只是开始。随着机器翻译研究工作的进展，这影响定会逐渐扩大。

表一

词典举例（以俄汉机器翻译系统中的俄文词典为例）：

- | | |
|---------------|-----------------------|
| 1. в | “第4组前置词”，“可组成成语”，“多义” |
| 2. но | 但是，“不同等连词” |
| 3. бы | “零意义”，“第10组副词” |
| 4. закон | 规律，“名词，A组” |
| 5. зна- | 知道，“动词” |
| 6. сложн- | “形容词”，“多义” |
| 7. производн- | 导数，“名词，同形容词变化” |
| 8. работ- | 工作，“名词，F组” |
| 9. перв- | “形容词”，“可组成成语”，“多义” |

说明：

1) 词典里的实词以词干形式存入，如зна-实为знать，сложн-实为сложный。这样做，便于查词典，既省时间，又省存储单元。顺便提一下：我们所说的词干和一般的了解有些不同。我们的词干指的是词的不变化部分，如поня-ти-е，Функци-я。而一般所谓的词干是понят-ие，Функци-я。

2) 词查到之后可从词典上得到意义（实为译文词代码）和一些特征。表中带“ ”者为特征，不带“ ”者为“意义”。如果一个词在译文中不需有意义，则得“零意义”特征。

3) 为了辅助查词典，设有一个“削尾”图表进行削尾（即把词改变成词干形式以查词典）。词如在词典中查不到，就进入削尾图表，削去最后一个字母，然后去查词典，如再查不到，就再削一个…这样连削六次（估计到最大限度的词尾красив-ейшими），任何词都可查到。削完尾的词在词典中找到并

得到意义和一些特征之后，还要到削尾图表来对词尾进行加工，再得到一些语法特征。

4) 得到“可组成成语”或“多义”的词转入“成语词典”或“多义词典”进行加工，得到意义和语法特征。

成语词典举例

1. в первую очередь 首先，“状语”

2. во всех случаях 在一切情况下，“状语”

⋮

多义词典举例

24/a,25/查该词为“сложн-ый”

a/I,II/查下一词为“Функция”

译文 I: 复合，“定语”

II: 复杂的

表二

语法图表举例（以俄汉机器翻译的语法规则为例）：

1) “名词分析”图表：

14/15,0/查该名词格不确定。

15/16,19/查前一词为名词

16/17,18/查该词有无“复数二或四格”或“复数二、四或六格”特征

17/0,0/作出“复数二格”特征

18/0,0/作出“单数二格”特征

19/20,24/查前一词为数词

20/17,21/查找到的数词大于5

2) “改变定语次序”图表

24/25,43/查分句中是否有带“非一致定语+第二格名词”特征的词

25/26,29/查该词为“第一组名词”

29/30,33/查前面的词为“第一组名词”

33/34,36/查该词为“动名词”

36/37,39/查该词右方是否有另一带“非一致定语+第二格名词”特征的词

39/40,41/查该词左方是否有“名词”

40/0,0/把该词放在找到词的前面，给该词以找到词的号码，同时置“的”字于找到词的前面。

表3 机器翻译过程举例 (以俄汉机器翻译为例)

1		2	3	4	5	6	7	8	9	10
输入	俄文词典	Научная	работа	играет	большую	роль	в	деле	социалистического	строительства
	词典分析	18**形容词	8名词 E组	动词, 多义	56形容词	101名词 K组	第4组前置词 “可组成成语”	37名词 “多义” B组	205形容词	376名词 B组
俄文句子分析	删尾词典	-аа	-а	-ет	-ую	-а	-е	-ого	-а	
	成语词典	单数一格	“主语”	谓语	单数一四格	单数六格	单数二格	复数一四格	单数二格	
语句子分析	多义词典	***								
	形容词词典			75			1			
语法分析	公式和符号									
	标点符号									句子界限
汉文句子分析	动词词典			现在时						
	名词词典					单数四格				单数二格
汉文句子组	句法分析	一致定语		一致定语	直接补语	状语	一致定语	非一致定语		
	同位语和定语	1 + 2		4 + 5		8 + 9 移至 7 前, 并加 “的” 字				
汉文句子组	状态语					6 + (8 + 9) + 7 移至 3 前				
	主补副	Δ (此号表示不动)								
汉文句子组	副句					Δ				
	翻译形态成分									
汉文词典	修辞									
	输出	18科学	8工作	75有	56巨大	101作用	1在...中	37事业	206社会主义	376建设
输		科学工作在社会主义建设的事业中有巨大作用。								

* 这是词进入机器后的顺序号码。

** 这是译文代码, 是该词在汉文词典中的位置号码。

*** 因为无成语, 不需加工, 所以是空白。凡是不加工的, 都是空白, 下同。

(原载《中国语文》1958年12月号)