[*Mechanical Translation*, vol.2, no.1, July 1955; pp. 15-24]

translation of russian technical literature by machine notes on preliminary experiments*

James W. Perry, School of Library Science, Western Reserve University, Cleveland, Ohio

The Russian alphabet, the Russian words encountered in scientific and technical material and the Russian grammar differ greatly from their English counterparts. In order to read scientific or technical Russian, it is necessary to have the meaning of a large number of Russian words stored in the memory. In translating Russian, the corresponding English words must be supplied by the memory accurately and quickly.

Automatic electronic equipment can be designed so as to have a memory capacity sufficient for translating Russian scientific and technical material. Machine memory, supplemented by appropriate selecting mechanisms, provide the basis for effecting word-by-word translation of Russian.

Preliminary experiments have been performed in which machine translation was simulated. One person copied the individual words from samples of Russian text on separate pieces of paper and the writer took the words at random and supplied separate translations for each word. The text was then recreated by restoring the words to the order in the Russian original. The crude translation so obtained was then evaluated by persons having scientific background but no knowledge of Russian.

The results obtained were unexpectedly good and justify the conclusion that even this most primitive form of machine translation enables persons knowing no Russian to understand, to a surprising extent, the subject matter of the Russian original. This understanding is far better than would be provided by numerous index entries to the text material. In fact, some sentences were understood with complete accuracy.

These experiments indicate that a practical, experimental approach to further development of machine translation should yield very useful results. The quality of translations produced by machine can be greatly improved by designing the machine system so that at least the simpler principles of Russian grammar are exploited. How to do this to best advantage is a problem which will require considerable experimentation.

introduction

English-speaking scientists who undertake to learn to read scientific and technical papers in the Russian language encounter a number of difficulties. The most obvious of these is the alphabet which consists for the most part of strange, exotic looking letters.

Mastery of the alphabet does little more than open the door to further difficulties. Although an Indo-European language, Russian is a member of the Slavic group. The words that constitute the backbone of the Russian language bear so little similarity to corresponding English words that a heavy burden is imposed on the memory when acquiring the vocabulary needed to read scientific and technical material. It is true that the purely technical and scientific terminology of modern Russian is, in large degree, derived from the same basic words—Latin, Greek, German or French—as are the corresponding English terms. However, in adopting words of foreign origin, the Russian language employs numerous suffixes, which, though used for the most part in a logical fashion, nevertheless require considerable effort to impress on the memory.

Finally, the grammar is a source of so many difficulties that it often becomes a barrier to learning to read the language.

Grammar difficulties are not due to a lack of logical structure in the Russian language. On the contrary, the basic rules of Russian grammar can, to a large degree, be stated in a simple, straightforward fashion. Inflectional endings play a dominating role in Russian grammar; they alone account for much of the discouragement one so often encounters.

In spite of some strange grammatical features, the basic structure of sentences in Russian and English is similar. Perhaps the most important similarity is the word order, which is so nearly the same that, once the corresponding English words have been written under the successive words in a Russian sentence, very often no rearrangement is needed to produce understandable English sentences and minor rearrangement suffices to provide good idiomatic English.

When the Russian endings are not taken into account, a word-by-word translation often proves deficient with respect to simple English connec-

*This is a slightly revised version of a paper originally written in September, 1952 and given limited circulation in mimeographed form. Mr. Perry was then with the Center for International Studies at M.I.T.

tives such as "of" and "to." In spite of these shortcomings word-by-word translations of Russian technical material have a surprisingly high degree of intelligibility, as will be evident from the experiments described below.

experimental method and results

In these experiments, paragraphs were selected at random from Russian texts on physics, chemistry and astronomy. The lines in the paragraphs were numbered as were also the words in each line. Each individual word in the Russian text was copied on a separate piece of paper along with the two numbers which identified the line and the position of the word in the line. The slips were then shuffled so as to place them in random order. Randomizing the Russian words had the purpose of preventing the writer from interpreting the meaning of the word in the light of the context. After this had been done by an assistant who knew no Russian, the writer supplied one, or if necessary more than one, English word as a translation for each Russian word on an individual basis without knowing how the Russian sentences had been worded. This operation of translating individual words one by one could be accomplished by an appropriately designed automatic electronic machine in whose memory units a Russian-English dictionary in properly encoded form had been recorded.

The numbers on the slips were next used to sort the individual words back into the original order (work slips arranged in order are reproduced below in an appendix). The English words were then copied off to produce the equivalent of a machine translation.

In the all important step of supplying an English translation for individual Russian words, no consideration was given to inflectional endings, with exception of certain irregular verb forms whose frequent occurrence would justify their being included in the dictionary as separate entries. The participles of verbs were also treated as though they were separate dictionary entries.

No consideration was given to case endings of nouns, pronouns and adjectives, nor to the tense endings of verbs. This means, first of all, that no distinction was made between the singular and plural of nouns. Furthermore the translation provided no hint that a Russian noun in the genitive case stands in a dependent relationship to another noun. Thus the phrase струйки фонтана was interpreted after machine translation as "little jet fountain" rather than as "a fountain's little jets," a more appropriate translation, which would have required account to be taken of the fact that фонтана was in the genitive singular case. The writer's assistants also pointed out that the interpretation of the machine translation would have been simpler if the plural of nounshad been indicated and if it had not been necessary to rely on the context to select those nouns which indicate the means or agency used to accomplish various actions. Interpreted in terms of Russian grammar, this latter observation means that it would be advisable for machine operations to take the instrumental case into consideration.*

In spite of these limitations—and other less obvious ones—the rough translations exhibited a high degree of intelligibility. To establish this point, two of the writer's assistants who had had training in physics (Miss Patricia Fergus) and chemistry (Mrs. Anna M. Reid) were requested to edit the rough translation produced by simulated machine operations so as to indicate how they would interpret its meaning. The results of their editorial interpretations are presented in the pages which follow, along with a rather literal translation of the Russian text prepared by the author as a check.

discussion of results

The practical usefulness of machine translation is, of course, the most important point we have to consider. As is evident from the results, such translation, even in a primitively simple form, provides an astonishing degree of insight into Russian technical and scientific material. Such insight is more than sufficient to allow decisions to be made as to the pertinency of a document to a given study. At the very least, therefore, machine translation provides a basis for selecting out documents to be investigated in further detail.

*K. E. Harper documents this conclusion in his paper "The Mechanical Translation of Russian—A Preliminary Report," *Modern Language Forum,* Vol. 38, No. 3-4, pages 12-29 (Sept.-Dec. 1953). See also his chapter "A Preliminary Study of Russian," in *Machine Translation of Languages,* Ed. by Locke, W. N. and Booth, A. D., Technology Press and John Wiley and Sons, 1955 (New York), pages 66-85.

SAMPLE I — PHYSICS

Пьезоэлектрические и пироэлектрические явления. Поляризация кристаллического диэлектрика может происходить не только под действием электрического поля, но в случае некоторых кристаллов (из числа не имеющих центра симметрии) поляризация может быть вызвана механическим, а также и термическим воздействием. Электрическую поляризацию кристалла, вызванную его растяжением или сжатием, называют пьезоэлектрическим эффектом, а поляризацию, происходящую при изменении температуры, называют пироэлектрическим эффектом.

— K. A. Putilov Kurs fiziki (Moscow) 1943, p. 150

imulated machine translation

Piezoelectrical and {thermoelectrical / pyroelectrical} phenomenon.

Polarization crystalline dielectric (noun) is able occur not only under action electrical field but in case certain crystal {from / of} number not {having / possessing} center symmetry) polarization is able to be {called forth / caused} mechanical {and / but} also and thermal action (on). Electrical polarization crystal {called forth / caused} {it / his / its} {pulling apart / tension} of compression {call (verb) / name (verb)} piezoelectrical effect {and / but} polarization {occurring / taking place} {at / during / near} change temperature {call (verb) / name (verb)} {thermoelectrical / puroelectrical} effect.

machine translation

Edited by Miss Patricia Fergus

PIEZOELECTRICAL AND THERMOELECTRICAL PHENOMENON. Polarization of a crystalline dielectric can occur not only under the action of an electrical field but in the case of certain crystals (a number of which do not possess center of symmetry) polarization can be caused by mechanical and also by thermal action. Electrical polarization of a crystal, caused by its tension or compression is called piezoelectrical effect and polarization taking place during a change in temperature is called thermoelectrical effect.

direct translation of russian original

J. W. Perry

PIEZOELECTRICAL AND THERMOELECTRICAL PHENOMENA. The polarization of a crystalline dielectric may occur not only under the influence of the electric field, but in the case of certain crystals (from the group not possessing a center of symmetry) the polarization may be caused by mechanical, and also even by thermal action. The electrical polarization of a crystal, when caused by its being under tension or compression, is called the piezoelectric effect, and polarization, occurring on change of temperature, is called thermoelectrical effect.

SAMPLE II — CHEMISTRY

Осахаривание клетчатки начинает применяться в технике. Для этого отбросы деревообделочных заводов нагревают под давлением с 0,1%-ным раствором серной кислоты; полученный таким путем сироп перерабатывают на винный спирт. По другому способу осахаривание производится на холоду действием очень крепкой (уд. вес 1,21) соляной кислоты. После удаления кислоты остается твердый продукт, применяемый как кормовое средство.

— B. A. Parlov Kurs organicheskoy Khimii (Moscow) 1943, p.253

simulated machine translation

Saccharification cellulose begin {use (verb) / employ} {in / into / at} {technology / technique}. For what waste product {wood processing / wood working} {plant (industrial) / factory} heat (verb) {under / below (preposition)} pressure. {with / from} 0.1% solution {sulfuric / sulfate} acid; obtained such {means / way} syrup {process (verb) / convert (verb)} {on / at} {wine (adj.) / tartaric} alcohol. {According to / Along / In accord with} other process saccharification {accomplish / carry out} {on / at} cold action very strong (sp. weight 1.21) {salt (adj.) / hydrochloric} acid. After removal acid, remain (verb) solid product being used {as / how} {food (adj.) / forage (adj.)} {medium (noun). / means (noun).}

machine translation

Edited by Mrs. Anna M. Reid

Saccharification of cellulose begins to employ technique. For that, the waste products of wood processing plants are heated under pressure with a 0.1% sulfuric acid solution. The syrup thus obtained may be converted on to wine alcohol. According to other processes, saccharification may be accomplished by cold action of very strong hydrochloric acid (sp. gr. 1.21). After removal of the acid, the solid product remaining is used as a food material.

direct translation of russian original

J. W. Perry

The saccharification of cellulose is beginning to be employed in technology. For this purpose, waste products of wood-working plants are heated under pressure with 0.1% solution of H_2SO_4; the syrup obtained in this way is processed into alcohol. According to another process the saccharification is carried out in the cold by the action of very strong (sp. gr. 1.21) hydrochloric acid. After removal of the acid there remains a solid product, which is used as a feed stuff.

SAMPLE III — MATHEMATICS

На рис. 12 вычерчены параболы, по которым движутся тела, брошенные со скоростью 10 м/сек под углами к вертикальной линии в 15°, 30°, 45°, 60°. Так располагаются струйки фонтана, выбрасываемые по всем направлениям из точки А. Огибающая всех этих струек, нанесенная на чертеже пунктиром, тоже парабола. Это и очертание головы кометы.

– S. V. Orlov Priroda komet (Moscow) 1943, p. 50

simulated machine translation

{On / Onto / At} Fig. 12 {traced / mapped out / drawn} parabola {according to / along / in accord with}

which move {thrown / deserted} {with / from} velocity 10 m./sec. {under / below}

angle {to / toward} vertical line {into / in / at} 15°, 30°, 45°, 60°.

Thus {locate / place (verb) / distribute} {groove / little jet} fountain {being ejected / being thrown out}

{according to / along / in accord with} all direction {of / from} point A. {Deflecting / Bending}

all these {groove / little jet} {applied / brought / plotted} {on / onto / at} {sketch / drawing / plan} dotted

line also parabola. This and {is (in fact) / are (in fact)} {outline / contour / form}

head comet.

machine translation

Edited by Miss Patricia Fergus

On Fig. 12 a parabola is drawn according to which a body moves, thrown with the velocity of 10 m/sec and making angles of 15°, 30°, 45°, 60° with the vertical line. Thus a little jet fountain is being thrown out in all directions from point A. Deflecting all these little jets, plotted on the graph, the dotted line also forms a parabola. This is, in fact, the outline of the head comet.

direct translation of russian original

J. W. Perry

In Fig. 12 are plotted the parabolas, along which bodies move when ejected with a velocity of 10 m/sec at angles of 15°, 30°, 45° and 60° to the vertical. Thus are distributed a fountain's little jets, when they are ejected in all directions from point A. The envelope of deflection of all these little jets has been plotted on the sketch as a dotted line, and it is also a parabola. And this is in fact the contour of the head of a comet.

Obviously, such further investigation may require the services of a skilled translator to assure that obscure—though important—points are not misunderstood.

The first example (see page 17) provides an instance in which misunderstanding regarding an important point crept into the machine translation. In editing Sample I (Physics), Miss Fergus made the first sentence read "Polarization of a crystalline dielectric can occur not only under the action of an electrical field but in the case of certain crystals (a *number of which do not possess center symmetry)* polarization can also occur by mechanical and also by thermal action." The italicized parenthetical statement is somewhat erroneous and would be better translated by "from the group not possessing a center of symmetry." The error was the result of the rather uncommon use of the Russian word число to mean "group" instead of "number." To eliminate this type of error, some of the rarer meanings of words would have to be included in the machine output.

Close inspection of the other examples of machine translation reveals similar misunderstandings, which do not, however, invalidate our previous conclusion that machine translation can provide an astonishing degree of insight into Russian scientific and technical material.

As already noted, machine translation could serve the very useful purpose of facilitating selection of documents pertinent to a given subject or problem. It is possible to imagine a system which would index Russian material without translating it and in this way provide a basis for machine searching by recently developed automatic equipment. To set up such a system, a list of key Russian words and phrases would have to be drawn up and these encoded so as to constitute an indexing system. The translating machine, when it encountered a key word or phrase would perform two operations simultaneously. One would be the translation of the word or phrase into English, the other the encoding of the key word or phrase so as to convert it into an index entry appropriate for machine searching operations. Once such a system was set up, it would permit a large volume of Russian material to be analyzed and correlated without the help of persons having the scientific and linguistic training necessary to read and understand Russian scientific and technical literature.

Another point to be remembered when estimating the value of a machine translation is its usefulness to a human translator as a rough draft from which he can prepare a completely accurate translation of documents whose importance warrants such attention. A rough draft prepared by machine translation can save much time and effort on the part of human translators.

The crude examples of machine translation presented above were produced with only a minimum of use of Russian grammar, namely the addition of a parenthetical notation—e.g. "noun," "verb," "adj."—to an English word to indicate the part of speech of its Russian counterpart. Such grammatical identification can be readily accomplished in machine translation, as the Russian language is so constructed that it is easy to distinguish between nouns, verbs, adjectives and other parts of speech. The young ladies who edited the crude translations remarked that it would have been helpful if more grammatical notations could have been included.

Many possibilities of exploiting the Russian grammar to improve the quality of machine translation await exploration. In particular, the elaborate Russian system of inflectional endings provides a wide range of leads to the structure and meaning of Russian sentences. When investigating these possibilities, the most practical approach would be to establish by experimentation which features of grammar can be most advantageously incorporated into a machine translation system.*

It is perhaps obvious that advantage is gained when the time and effort involved in using the output of a translative machine are decreased, but the expense of increased complexity of design and increased maintenance cost must be borne in mind. It would be easy to go beyond the point of diminishing returns in developing elaborate machines and elaborate machine translating methods, which might produce translations of better literary quality, but might fail to provide a profitable return on the increased investment.

*Much work has been done in this direction since the present paper was originally written. See especially Oettinger, A. G., *A Study for the Design of an Automatic Dictionary,* Harvard thesis 1954, also Harper, op. cit.

A good starting point for investigating the possibilities of exploiting Russian grammar to improve machine translation might be furnished by the more than 700 example sentences which which the writer used to illustrate the different points of grammar in his book *Scientific Russian,* Interscience Publishers, New York, 1950.

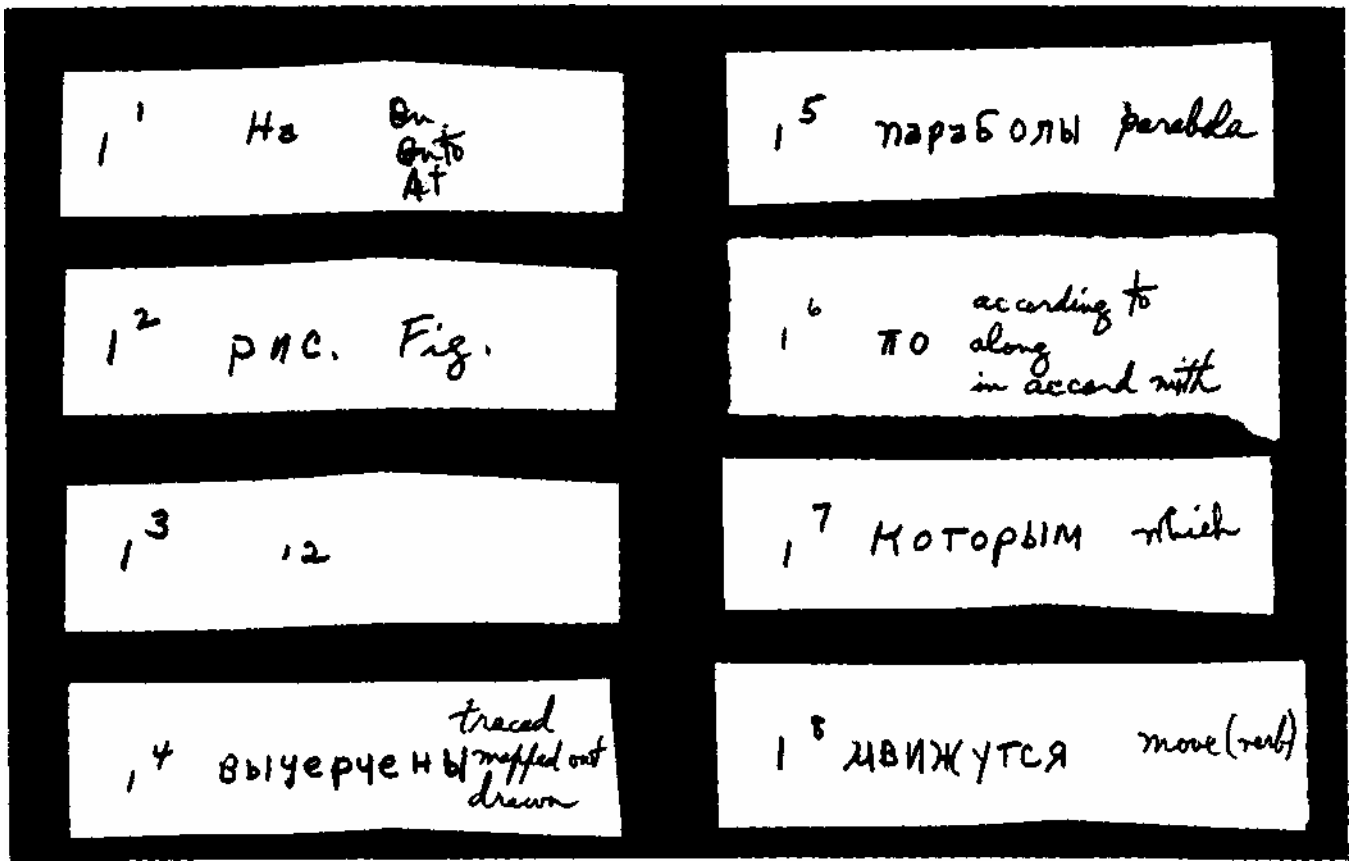
Certain news reports may have given the misleading impression that digital electronic equipment already in existence would be well suited for translating scientific and technical Russian. Discussions with experts in digital electronic machines indicate on the contrary that present machines would be grossly inefficient if used for translating but that techniques and sub-assemblies used in constructing digital computers can doubtless be used to construct a practical translating machine. Further investigation of the methodology of machine translation appears advisable before undertaking to design a translating machine. However, such an investigation, in order to remain within the realm of the practical, should take into account the limitations imposed by the present state of development of automatic electronic equipment.

conclusion

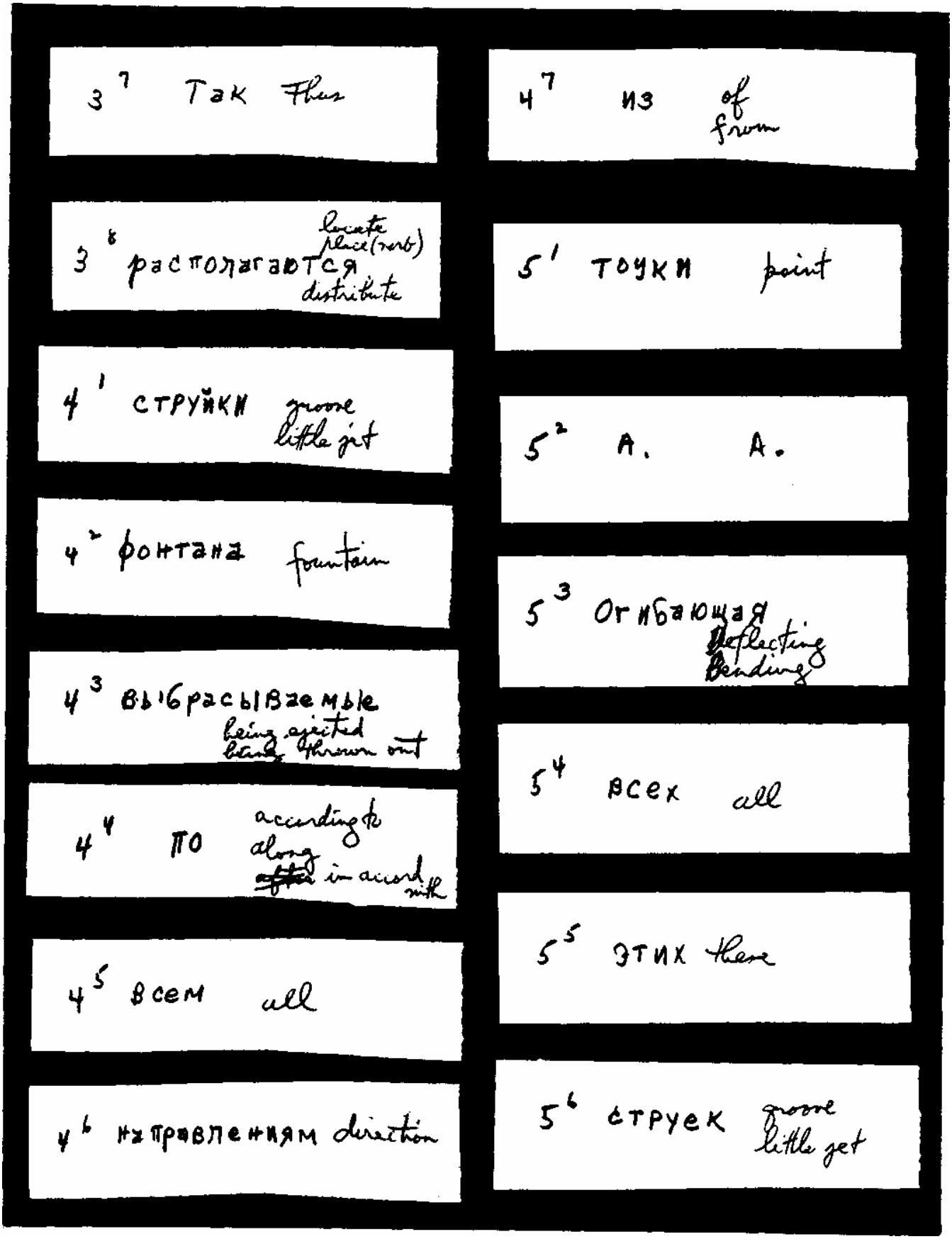
Preliminary experiments indicate that it is possible to apply machine methods advantageously to the problem of translating Russian scientific and technical material. Even the crude translation produced without systematic exploitation of the Russian grammar provide a surprising degree of insight into the subject matter of scientific and technical material. An important problem awaiting investigation is how best to exploit the possibilities inherent in the Russian grammar while still remaining within the realm of the economically feasible.

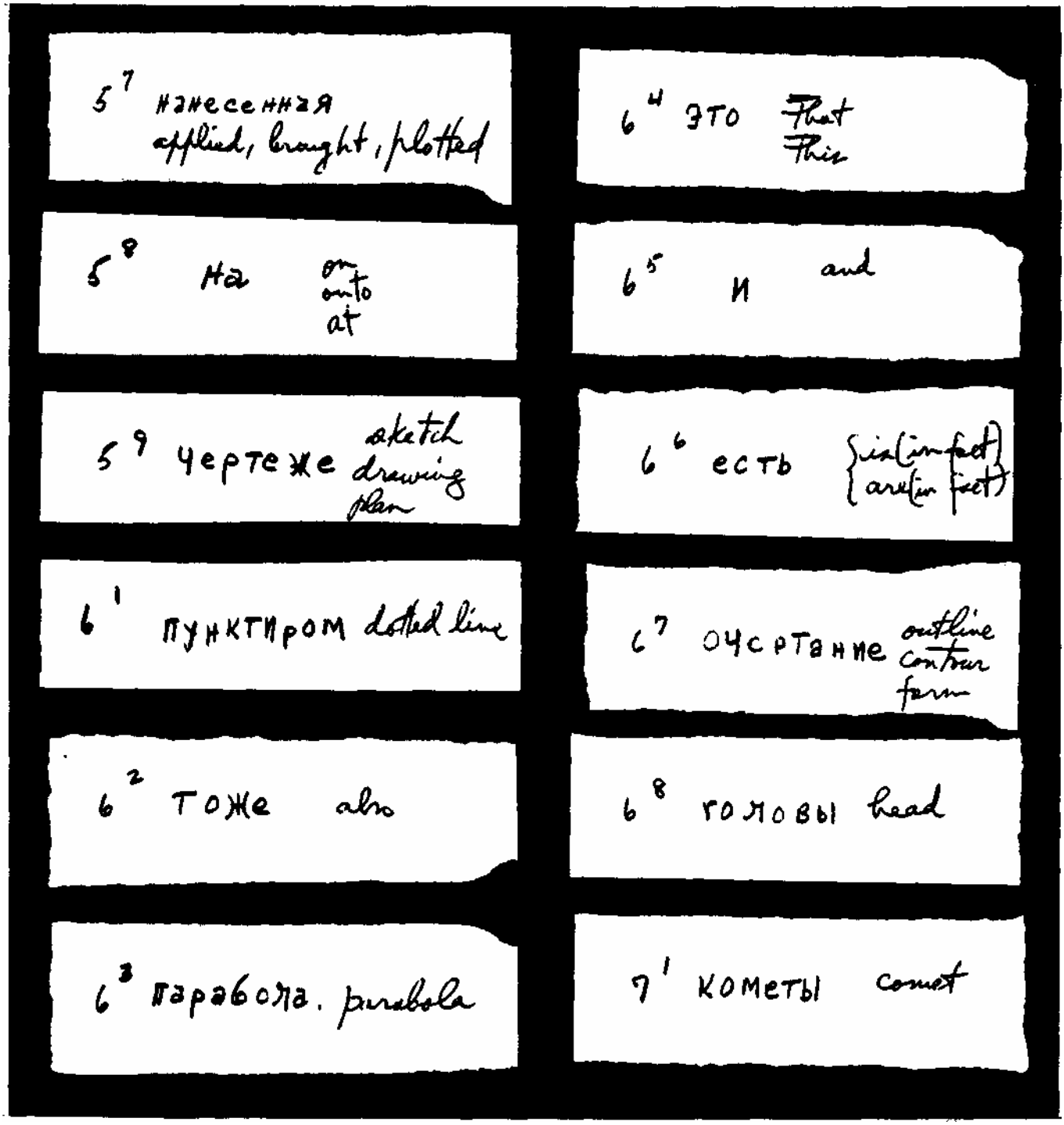
appendix—work slips from sample III

(The numbers refer to the arrangement on the original Russian page where the first line contained eight words and the last, only one.)




2 1 тела body
2 2 брошенные thrown deserted
2 3 со with from
2 4 скоростью velocity
2 5 10 10
2 6 м/сек м/sec.
2 7 под under below
2 8 углами angle
2 9 к to toward
2 10 вертикальной vertical
3 1 линии line
3 2 в into in at
3 3 15° 15°
3 4 30° 30°
3 5 45° 45°
3 6 60° 60°


3⁷ Так Thus
3⁶ располагаются locate place (verb) distribute
4¹ струйки groove little jet
4² фонтана fountain
4³ выбрасываемые being ejected thrown out
4⁴ по according to along in accord with
4⁵ всем all
4⁶ направлениям direction
4⁷ из of from
5¹ точки point
5² А. А.
5³ Отибающая Reflecting Bending
5⁴ всех all
5⁵ этих these
5⁶ струек groove little jet


5 7 нанесенная applied, brought, plotted
5 8 на on onto at
5 9 чертеже sketch drawing plan
6 1 пунктиром dotted line
6 2 тоже also
6 3 парабола. parabola
6 4 это That This
6 5 и and
6 6 есть {is (in fact) are (in fact)}
6 7 очертание outline contour form
6 8 головы head
7 1 кометы comet