A. Kondratov

SOUNDS AND SIGNS

MIR PUBLISHERS

А. КОНДРАТОВ

ЗВУКИ И ЗНАКИ

ИЗДАТЕЛЬСТВО «ЗНАНИЕ»

A. KONDRATOV

Sounds and Signs

Translated from the Russian
by
GEORGE YANKOVSKY

MIR PUBLISHERS MOSCOW
1969

UDC 164+410 : 51+413.0+414 (023)=20

Revised from the 1966 Russian edition

Transliteration of the Russian Alphabet

Russian	Translitera-tion	Russian	Translitera-tion
А а	a	Р р	r
Б б	b	С с	s
В в	v	Т т	t
Г г	g	У у	u
Д д	d	Ф ф	f
Е е	e	Х х	kh
Ж ж	zh	Ц ц	ts
З з	z	Ч ч	ch
И и	i	Ш ш	sh
Й й	ĭ	Щ щ	shch
К к	k	Ъ ъ	”
Л л	l	Ы ы	y
М м	m	Ь ь	’
Н н	n	Э э	
О о	o	Ю ю	yu
П п	p	Я я	ya

На английском языке

CONTENTS

INTRODUCTION

We talk just as naturally as we eat, or walk, or sleep. But as a rule we do not give thought to the fact that the "naturalness" of our speech is only apparent. Human language is not at all inborn; it was taught to us, it was taught by the society we live in.

Over the ages man has wondered what language is, how it is constructed, how one's mother tongue differs from the languages of other peoples, and how the language of human beings differs from the signal cries of animals, and how our everyday speech differs from other media of communication in human society.

Thinking on these problems gave birth to linguistics, the science of the laws of language. Methods of investigating language improved with the development of linguistics and the accumulation of facts and knowledge. Today the latest tool is in the form of numbers and exact measures. Mathematical statistics and the theory of information, probability theory and mathematical logic, computers and sign theory are more and more coming to the aid of students of language.

That is what our story is about.

What Semiotics Has to Say

We begin our story with semiotics—the science of signs and symbols. It is semiotics that will tell us how language differs from other media of communication and how the human language differs from the language of animals or the so-called machine languages.

Do Trees Converse?

The language of the trees, the grass, the clouds and the forests, the language of the mountains and the waters. So say poets, but is there really a language of nature? Do the trees and grass and clouds talk to us in some way?

Primitive man thought so. In his view, nature spoke to man, warning or threatening, frightening or encouraging him. The Sun would give a friendly wink from behind the clouds, releasing a ray of light. Thunder would speak in terrible tones to anyone who failed to obey the gods.

All that has long since become poetical imagery, but once it was taken in its direct meaning and not at all figuratively. All the happenings of nature were in the language of a god or the gods. These primitive beliefs disappeared, and the naive picture of a "speaking nature" gave way to the knowledge that only living beings could speak to one another.

True, even nature can "speak" if by that we mean a transfer of information. Bending branches indicate a strong wind, dark clouds bespeak an approaching storm.

Of course, this language of nature is quite unlike any conversation, any exchange of information between human beings, or even between animals, for their signal cries are always directed to some addressee, while nature does not appeal to any one in particular: the clouds do not intend to give warning of a storm, the trees do not wish to say anything about the wind. Nature informs but does not converse.

The new science of semiotics, or the theory of signs and symbols (the word itself comes from the Greek 'semeion' meaning 'sign'), studies the language of animals, our own human language, and the numerous and diversified nonlanguage systems of signs and symbols like road signs, signal systems, show-window displays, maps, diagrams and the like.

Semiotics treats of any system of signs, any "language" used by "entities of any nature whatsoever"—human beings, animals, and, of late, man-made automatic devices, "thinking" machines. Semiotics is closely associated with another young science called cybernetics. Semiotically, we may regard man or animal or machine as a cybernetic device that performs operations upon diverse sign systems and texts.

The Alphabet of Semiotics

Signs, signals.... In ordinary speech we do not often distinguish between them. But the science of signs does. So we shall begin our discussion of semiotics with the distinguishing features of signal and sign.

All the multifarious phenomena of the world about us convey information, the material bearer of which is a signal. Pulses of current, the letters of this book, photographs in newspapers, the bioelectric currents of the brain are all signals. Signs are a variety of signals. They differ from the other signals in that they are conventional. The smoke of a fire is a signal (or a 'natural sign' or, as semioticians say, an 'index sign'); it carries information about the existence of a fire, though we do not see it. But the smoke of the fire becomes a sign if we have agreed with someone that it is to indicate that 'all is well' or that it is a 'warning' or that 'I am here'.

A signal bears information. Red, black, and white sails indicate the colours red, black and white, and that is all. But when Theseus, the Greek hero, set sail and agreed with his father, King Aegeus, that black sails on his ship would mean trouble and white sails would signify success, this elementary signal system was a system of signs; the colour indicated more than colour, it was a sign.

The black sails signified to Aegeus that his son had perished; for the sailors of the 16th and 17th centuries they signified pirates. A sign always has a sender of information and an addressee, the recipient. A signal does not necessarily have both: when we see thick smoke billowing up from a woods we conjecture that there is a fire. The smoke is a signal of this. But there is no sender. No one purposely sent up smoke to deliver information.

Semiotics distinguishes three types of signal. Index signals, also called 'natural signs', form the first type. Natural, because there is no previous agreement about the meaning of a sign. For instance, a deer smells a tiger. The smell is a signal that a tiger is in the vicinity, close by, though unseen by the deer.

The crash of glass says that a window has been broken, though we may not have seen anything. "Where there is smoke, there is fire," says the Russian proverb. Smoke is a natural sign of fire. Looking out into the street at passers-by wrapped in their coats, we conclude that it is cold out of doors. This is again a natural sign, or index sign.

Actually, the information we obtain from natural phenomena, from animals, is in the form of index signs. But not all of it is.

Let us take animal tracks. Are they index signals? It would appear to be so, for we have not come to any agreement with the wolf or rabbit that the tracks are to designate the passage of a wolf or rabbit. These are signals of a natural origin.

But tracks have a characteristic peculiarity that distinguishes them from natural signs. Tracks, signals that is, appear to designate the resemblance to a wolf or rabbit paw. The signal has meaning and external form. This second type of signal (called 'copy signals' or 'image signals') is peculiar in that the meaning (content) and the external form (expression) are similar. The tracks of animals and humans are an instance of this type of signal. Other examples are photographs, moulds, imprints, impressions.

Finally, the third type includes signals of communication or conventional signs. They are called signs in the narrow sense of the word. Most of the signals the people employ are in this category. The signal (!) has nothing in common with the concept of 'danger', yet we understand it as a sign of danger. The word 'elephant' has nothing in common with the African or Indian beast. A shake of the head is usually taken to mean 'no', but in Bulgaria it signifies agreement—added proof that such apparently natural signals as gestures are also conventional signs of society and are not an endowment of nature.

Conventional signs and communication signs are exhibited by animals too. The cry of the sacred baboon 'ack-ack-ack' is a warning signal, which alerts the herd. The brief 'ack' signal will put a whole herd to flight. Dogs and cats and a variety of birds have similar conventional signals.

All signs (and signals too) have something in common. Every sign (and signal) consists of that which is signified and that which signifies.

The pages of this book have been typeset; the set of black letters that form words, and the gaps that separate the words constitute the external or material expression; the meaning of the words forms the content. The sound vibrations emitted by the throat of an ape is the external expression of a warning signal (that which signifies), while the content of the sign (that which is signified) is in the order: 'warning!' or 'run!'.

A sign is meaningless without a system of signs. Let us take the most elementary case: what does '!' mean? The school child says 'exclamation mark', an automobile driver says 'caution', a chess player says 'excellent move', and the mathematician reads it as 'factorial'. And each one is right. That which signifies is the same (!), but it has four utterly different meanings in four different sign systems.

Polyglot Crows

We obtain information by observing the phenomena of nature. But nature does not converse with us, it does not talk to human beings. How about animals? How does their "language" differ from the language of human beings? And what do they have in common?

Primitive man endowed nature with a soul. The Middle Ages, on the contrary, only allowed man to be the "vessel of intelligence", the possessor of language and speech. But scientists have demonstrated beyond the shadow of a doubt that animals have a language too: very primitive and simple, naturally, when compared with the human languages. Signs are employed by chickens, dolphins, monkeys, cats, bees and elephants, birds and ants. Signs, not signals! Bees have an intricate "dance" system, grouse have a language of males and another of females, baboons have 17 word signs, and anthropoid apes have about thirty.

True, the signs of animals do not often form rigorous systems and are unrelated (the meowing and purring of cats). But in some animals these isolated signs may form a system and even produce combinations. To illustrate, the common 'warning' sign of chickens breaks up into four signs of warning: 'close danger', 'distant danger', 'man danger' and 'hawk danger'. Chickens have a language with some 10 elementary signs, various combinations of which form about a score of 'composite signs' (like the 'categorical order' sign which consists of two call signs repeated twice in a row).

Crows have still greater linguistic capabilities. Many years of research on the part of American scientists have shown that crows have a variety of languages: city crows do not understand rural crows, crows from Connecticut cannot converse with those from California. But then there are the wandering crows that go from town to countryside and from one state to another. They have their own specific language. What is more, they know the languages of the other city and rural crows and can talk (in crow fashion, naturally) with them. Human beings are not the only polyglots, it appears.

So how does the language of animals differ from that of people? Particularly since animals, like people, employ all types of signs: index signs, copy signs and communication signs.

There is a difference, of course, and it is this. The signs of animals are concrete. They are, as it were, attached to the event or situation. A cock cannot tell a chicken what happened yesterday or what will take place tomorrow. Neither can the loquacious chimpanzee. The sign exists only at a given time, in a given concrete situation; it has meaning only now, for the present.

Only human language separates signs from the situation, only man is capable of speaking about events of the future, the past, imaginary real and imaginary unreal. The polyglot crow, no matter how many crow languages it has learned, cannot tell its nestlings a fairy story or any kind of a story. A human being can say "I caught a crow" or even "A crow caught a man". A crow can't do that because its concrete language does not allow it to do so. The transformation of language into an independent system of signs has given man tremendous advantages over all other animals.

A bird emits a cry to attract the attention of the person who feeds it, but it is not aware of the cry being a sign. In the language of animals, the sign and the thing it designates are intimately bound together. That is why their language does not develop or change. Baboons, chimpanzees, chickens and cats "talk" in the very same way they did a hundred or a thousand years ago.

To animals, 'signs' can denote a state of joy, fear, hunger, or an appeal. Many of them can very easily be translated into the language of human beings, and not even in one word (an interjection, say) but as a whole sentence. Yet animals do not pronounce phrases; the sign, cry, word or phrase are all one and the same thing.

The language of animals differs from the word language of human beings both as to function and structure.

Incidentally, people do not always speak with words. On occasion this becomes very involved indeed.

How to Say It Without Words

One of the Scandinavian sagas sung by the skalds, legendary bards of the ancient Vikings, relates of a scholarly debate between a renowned sage of theology and a brave one-eyed Viking.

The sage put up one finger. The Viking countered with two fingers. The sage showed three, the Viking offered a fist. The sage took one cherry, ate it and spat out the stone. The Viking picked a gooseberry and swallowed it. The scholarly debate thus went on for a long time until the famous sage recognized defeat.

When asked why, the sage said (in human language this time and not in the language of signs): "My opponent is a true fount of wisdom. I showed him one finger meaning that 'God is one.' But he wisely countered with two fingers to indicate that besides God the father there is

God the son. Then I tried to trap my adversary by showing three fingers to indicate that perhaps there are three gods: the father, the son and the holy ghost. But he wisely avoided the trap by showing a fist to mean that God is one in three persons.

"Then I showed him a cherry, and I wanted to say that life is as sweet as the cherry. But he again put me to shame, for in eating the gooseberry he said that life is better than sweet fruit, it is streaked with acid and this makes it more dear and valuable. He is indeed the wisest of all churchmen in the world." Thus the wise man dejectedly concluded his tale.

Then the Viking was asked. In surprise he said: "I did not think of God at all. Simply his insolence was too much for me, for his one finger meant that I had one eye and so how could I compete with him. I showed him two fingers to indicate that my one is equal to his two. He then put up three to mean that the two of us together had three eyes. So what could I do but show him my fist so that he would see that not words but deeds could cure such impertinence. His answer was that he would eat me like a cherry and spit out the bones. So then I swallowed a gooseberry for him to realize that I did not intend even to leave any bones and would eat him up altogether!"

Sign language, as you see, is not a very exact conversational tool. The gesture signs were the same, yet one carried on a theological debate while the other urged to fight.

Why? The reason is that the debaters used different signs, though outwardly the signs found identical expression. This is possible in ordinary spoken language as well. English abounds in words which taken separately mean different things, like 'fair' meaning market and 'fair' meaning beautiful.

In sign language, a nod means agreement to some peoples and refusal to others. So the crux of the matter is not in the sign, that is, not in its external expression, but in the system of signs.

Our gestures are simple, a forward nod meaning agreement, a shake of the head, refusal, etc. The American Indians had more simple 'sign words' that designated 'tree' and 'leaf'. Giving first the sign of the tree and then of a leaf, and then of a leaf falling from the tree, the Indian was able to say to his companion 'fall' (autumn). Using such combinations of signs (word gestures), an Indian could describe a good deal: the beginning of war, the conclusion of peace; he could even translate intricate myths and legends into the language of signs.

The natives of Australia also have a rather developed sign language. It is used in a variety of cases: when the conversation is over large distances where the voice would not carry, when tribes meet that speak different languages; finally, when "to speak in words" is tabu. In Australia, the widow who had just buried her husband is not allowed to use words, neither is the youth who is being initiated into manhood, and so forth. (Even civilized nations have not always advanced much beyond these "speech tabus"; recall the taciturn monks of the Christians who never opened their mouths for years on end and conversed in sign language, for "the spoken word is a sin".)

The hand language of the Australian Aranda group has about 500 distinct gesture signs. These signs denote objects, actions, the qualities of objects, social terms and even whole questions, whole phrases.

True, in most cases, like in the case of the Indians, phrases are conveyed by combining the simpler word-signs. For instance, to say "brother has died" one takes three signs: 'brother', 'already', and 'to die'. For some word signs, Australians use only their fingers, which are of course visible only over short distances. For more distant conversations, one would employ movements of the whole arm, the head and even the upper part of the body.

The sign systems of the various European peoples are very much alike. The Englishman, the Russian, the Frenchman and the German will all shrug their shoulders to mean "I don't know". But there are certain differences. In Russia, the good-buy gesture by waving the hand with fingers together is interpreted in Brazil as "come here". For the "come here" sign, in Russia one would turn the palm upwards and move it to and fro. In many countries of the West, the farewell sign is a waving of the hand with the palm away from the person waving.

Or take the Russian sign for marvellous: the upright thumb. That same meaning is expressed differently by the Frenchman, who makes a circle with the index finger and thumb, brings them to his lips and makes the sound of a kiss. In Brazil, one would say 'great' or 'marvellous' by taking the lobe of the ear between one's fingers.

Among the modern peoples of Europe, hand signs and facial expressions supplement language rather than replace it. Gestures express our emotions, put stress on certain words, and bring out new nuances of meaning; sometimes, true, they help to convey the reverse meaning (ironical mimicry, winking). At times, one pays more attention to the intonation and facial expressions than to the words themselves.

The prominent French linguist J. Marouzeau had this to say about facial expressions (having in view the French, naturally): "Winking may mean conspiring, suspicion, or slyness, depending on the position of the lips. Wide open eyes indicate curiosity or amazement. Wrinkling one's mouth has the sense of dissatisfaction. A smile may express tenderness, doubt, or mockery. Wrinkles that run vertically down the forehead express deep thought or surprise; while a forehead with horizontal wrinkles signifies anger or a threat."

Hand gestures, he says, have the following meanings: "A flat wrist with palm up means honest agreement. But if the palm faces forward, it means refusal. Two palms together express a request. The raised forefinger signals a warning; pointing ahead, it aims and designates danger. The same finger put to one's forehead signifies deep thought, on the lips it means a call to silence. Hands on the hips is a challenge; hands crossed on the chest is bravado. . ."

The reader will see that in the main they coincide with signs familiar in Russia, England, the United States and elsewhere, though there are some differences. At times the impression is one of rather crude histrionics. This is not fortuitous because a word description is a very imperfect way of explaining sign language. For this reason, Marouzeau wrote only of the most primitive gestures that are within the grasp of even a mediocre actor. The reader himself has probably felt how imperfect sign language can be: folded hands—there are all kinds of ways of folding one's hands, go and figure them out.

If there were some simple and exact method of recording gestures (the 'metalanguage', as semioticians say; simple formulas like the language of chemical formulas or chess notation), we could compile a large and very interesting dictionary of the gestures (sign language) of every people. How useful to the actor or to the teacher inculcating habits of behaviour. Unfortunately, this is still only a dream, for we still lack a metalanguage of gesture and mimicry.

Two thousand years ago Cicero taught orators that "All the movements of the soul must be accompanied by gestures capable of explaining doings and thought: gestures of the wrist, the fingers, the whole hand stretched outwards, the foot striking the earth, and especially mimicry of the eyes; gestures are like the language of the body which is understood even by savages and barbarians." The famous Roman rhetorician Quintilian even compiled something in the nature of a lexicon of gestures. Since then two thousand years have passed and to our shame we know but little more about the language of gestures than did the ancient Romans. The whole trouble lies in the absence of a convenient and exact metalanguage of gesture notation. We hope that semioticians (in conjunction with linguists, physiologists, actors, and psychologists) will soon elaborate a good system for recording gestures and mimicry.

The Language of Signals, the Language of Whispers, the Language of Whistles...

Though gestures and mimicry are excellent accompaniment to spoken language they do not replace it.

Our spoken human language is the main medium of communication. But sounds are not audible over long distances, and if it were not for the telephone and radio, we would find it hard to communicate information.

The telephone and radio are fairly recent inventions, for thousands of years people lived without them. Still ways existed for conveying information over large distances.

The peoples of the Old World employed for this purpose "living witnesses"—messengers with letters or word-of-mouth messages of enemy attacks, etc.

The American Indians were much more resourceful. They worked out a whole system of signals, a 'signal language'. Messengers were not needed then and communications were delivered over great distances in an instant.

The most common signals were fire and smoke. 'Fire signals' were employed at night. Fires were lighted on high points that could be seen from a distance, or along the coastline. With their aid, Indians could report about the approach of strangers, about whales thrown up on the shore, etc.

Smoke served for daytime signalling. Wet grass or freshly cut branches were thrown into a fire to make for slow burning and heavy smoke that could be seen from afar. The source of the message' was the number of fires or 'smokes') and also the number of smoke flashes, which were made by throwing a leather blanket over the fire and then pulling it away and repeating the procedure as many times as was needed for a given signal.

The number, duration and sequence of flashes would indicate to other tribes, which were always on the alert, what exactly the neighbouring tribe wanted—a joint hunting expedition or aid, etc. When warriors returned from warfare, they signalled from afar about the number of lost warriors so that kinsfolk would learn about this beforehand.

But fire and smoke were not the only tools of the "signal language". A beautifully picturesque system of signalling with blankets was employed when moving on foot or on horseback. For example, the 'sign of the bison' was given in "blanket language" as follows: a blanket was held by the corners high above the head and then lowered to the earth. The 'coast is clear' was depicted with a stretched out blanket wayed rhythmically in front of oneself. A rapid waving of the blanket over one's head indicated 'approach of enemy', throwing it up into the air meant 'warning').

There was also a signal system generated by means of walking on foot or pacing a horse in circles, forward, backward or in zigzags. Some of the Indian tribes of North America during the age of colonization had already devised a system of mirror signalling. Thus, the American ethnographer Rigen related that when in 1902 he was descending a mountain path with Sibic Indians, the latter began to signal to their tribesmen with the aid of small mirrors reflecting sunlight. The mirror-reflected response from the valley below read: 'We are all in good health. We have sufficient supplies of food.' When the travellers finally reached the valley, the whole tribe was already assembled, for the mirrors had signalled their approach.

The "signal language" of the North American Indians is a visual, optical language. The peoples of Central and South America, tropical Africa, south-east Asia, and the islands of the Pacific Ocean created a special language of drums.

Drums are heard from quite a distance and so the Indians of Equador and Peru use them for "talking" to spirits and deceased ancestors, for the human voice could hardly be expected to reach distant "spirits" residing in the "other world".

But as a rule the loud sound of the drum did not serve for negotiating with the distant "other world" so much as for other urgent tasks. The beating of drums might announce the approach of an enemy, a coming celebration, a wedding or other ceremonial affair.

However, drums were mainly employed to

communicate the "latest news". Many tribes and peoples of South America, Africa, New Guinea developed a complicated signalling system that enabled a wide range of information to be communicated—from the arrival of foreign ships to a successful pig hunt. The Papuans of New Guinea developed a very complex system of signalization. If a hungry husband returns home in the evening and sees that his wife is away, he calls to her with the following signal:

. . . .

Three times five medium strong beats form the 'come home' signal; the last six dots represent the signature of the drummer, in this case the Papuan named Saiyam.

The inhabitants of New Guinea distinguish between "private conversations" and "general". There are special signals for each individual and

for the whole clan. And when a dweller of the village hears the drum signals, he can immediately tell whether the drum-telephone conversation is between two persons or is intented for the whole village. If all the women of the village are away, then the clan as a whole demands the return of the women with a special signal instead of each husband trying to reach his wife.

Researchers say that the drum signals of the Papuans are extremely diversified. The most important signals are: warning, a call for a meeting of the community, the betel-nut signal indicating that all men must come with betel nuts, the cocoa-nut signal, the pig (or dog) tooth signal (teeth are used by the Papuans as money), and the signal summoning the dwellers for trade.

Almost every home has its drum for "telephone conversations" with the neighbours and even with close-lying villages.

The drums may be heard over three, five and at times even ten kilometres. Sound covers these distances in seconds, whereas a messenger might spend a whole day making his way through the tropical jungles.

Curiously, the visual (optical) signalization system was devised by the Indians of North America, dwellers of the prairies and boundless plains; the sound (or drum) signalization was invented by the inhabitants of the jungles.

Perhaps the most ingenious mode of communication was contrived by the dwellers of the Canary Islands. They found a way of conveying human speech—yes, speech and not conventional signals—over distances up to five kilometres without any telephones, radio transmitters or other modern media of communication. By whistling, ordinary whistling, the way kids do.

True, not so ordinary. Each separate speech sound of the language is coded, so to speak, receiving its specific variant, like in the Morse code each letter is designated by a special set of dots and dashes. The Canary Islands were discovered by the Italians, but the Pope of Rome presented them to the King of Spain as the "State of Fortunia"; the Pope was considered the "deputy of God on earth" and had the right to do what he wished with all newly discovered lands! Colonization began, and as a result the natives of the islands—Guanche—were wiped out. The only thing left is their remarkable whistle language, and that only on one of the Canary Islands, La Gomera.

The Canary Islands are mountainous, cut by ravines and gorges, with precipices and steep cliffs. The whistle language enabled the Guanche to converse over distances up to five kilometres. The present-day Spanish dwellers of La Gomera continue to "whistle talk" when they have something to say at a distance.

The language of whistles is found only in one place on the globe. But the language of whispers is used by certain primitive tribes like the Veddas living in the jungles of Ceylon or the pigmies inhabiting the forests of Central Africa. This is not an ordinary whisper that conveys our speech, but a peculiar monotonic whisper which is pictorially compared to "a strange continually building up wave of sounds resembling the broken breathing of a pack of dogs".

The purpose of this language is obvious: to communicate during night hunting so that the keen ear of the elephant or antilope should hear no sound. The medium of communication here is the language of whispers.

Etiquette, or the Language of Behaviour

The language of animals and humans, the language of whistles and signal and sign systems are all media of communication studied by semiotics. But perhaps one of the most interesting sign systems is our own behaviour, more strictly speaking the rules of social conduct, or etiquette.

The basic purpose of etiquette is to serve as a means of communication, as semioticians are wont to say. The rules of conduct always imply a performer and a recipient, the former carrying out the rule, the latter (whether an individual or society at large) responding.

Languages and—more broadly speaking—sign systems serve as a means of communication. But etiquette? What does a person express in this specific language of behaviour?

The rules of conduct in society are a product of human civilization. Every person, to one degree or another, is familiar with the rules of etiquette and, depending on the situation, puts them to use. To give one's seat to an older person, to a woman or a guest (even though he may be younger) and so forth. We employ these rules and this special language where our actions take the place of words.

The rules of etiquette are divided into positive rules and prohibitive rules. There are many more of the latter, the idea being that the rules of conduct in society do not allow us to perform certain "natural" actions. No matter how much we want to demonstrate that we are right, we must not shout. No matter how bored we are, we must not yawn.

Etiquette is a peculiar system of signs, a language used in society and spoken via our behaviour. Quite naturally, such "languages" at times differ radically, depending on the country or the social stratum one revolves in.

In the Near East, a person on horseback meeting one on foot must be the first to greet, irrespective of difference in age, title or sex. An approaching person greets a standing person. A standing person is first to greet one seated. If a seated youth greets an aged man of prominence entering the room, this is considered to be tactless, though from our European viewpoint, the reverse situation would be considered lacking in tact—if the old man were to greet the young man first!

Tactlessness, or infringements of etiquette is what—though it appears paradoxical—makes etiquette a sign system. Indeed, when we observe all the rules of conduct in society, our "sign conduct" is of zero value. (Everything is as it should be.) Let us say that you shake hands when visiting friends, you take a seat after requesting permission, etc. But as soon as you refuse to extend your hand in greeting, or if you take a seat without first asking permission, the zero value vanishes at once, for your conduct immediately begins to "speak" of your lack of respect for your friends with whom you refused to shake hands and for your hosts when you slumped into a chair without asking permission.

Another sign system—our ordinary speech—is also subject to the rules of etiquette.

In Russian, as in German with its 'du' and 'Sie', there are two forms of addressing a person. English once had two forms, but now has one. In the language of Tibet certain notions are expressed by two different words depending on the person with whom one is speaking. For example, head in ordinary speech is 'go', in respectful speech, 'a'; thought in ordinary speech is 'sampa' but in respectful speech it becomes 'gongpa'.

The Japanese language has several different "forms of respect" or "degrees of politeness". There are specific grammatical phrases used in speaking with equals, with superiors and with persons of lower title or position in society, etc. At one time the Russian language also had a special particle (pronounced 's') of respect that was appended to verbs.

This brings us to yet another sign system, the spoken language.

The Most Important Language

From the viewpoint of semiotics, our ordinary language is a system of signs existing in society and for society, like all the other sign systems. But language has a peculiarity that distinguishes it from all other media of communication.

Our simple language is actually not so simple as it seems. It appears to be simple only because from an early age we absorbed it, mastering the laws and rules without realizing the process; full realization came later in school when we learned to read and write.

We speak freely and easily without realizing that the spoken language is a highly complex system of signs capable of conveying every imaginable idea, thought or concept.

Suppose that we did not have a language and that it was replaced by some other system of signs.

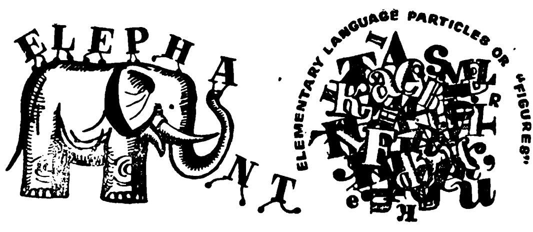
In Swift's *Gulliver's Travels* we read of the Laputan scholars who rejected human language and decided to converse with the aid of objects. When in need of bread, you point to actual bread, if you are thirsty, you indicate a bottle of water, and so on. Each scholar was accompanied by a servant carrying a bagful of the things to represent the 'words' of a conversation. But not every 'word-thing' can get into a bag. Take a bridge or a house. Then come the more abstract notions like 'conscience', 'absolute' and many many others that have no counterparts whatsoever in things.

Even if the 'word-things' were replaced by pictures, the situation would not be much better than what Swift's sages from Laputa encountered. Just imagine carrying around hundreds of thousands of 'word-pictures' or trying to locate the one you need. Then again that difficulty with abstract notions like 'conscience', 'absolute'—try and picture them!

Why is our language, the ordinary spoken tongue, the richest and at the same time the most economical sign system?

The reason, says semiotics, is that it is hierarchical. In all other sign systems we have the designating and that which is designated, the expression and the content. Language is constructed in a much more complicated fashion. Let us go back to the familiar case of the sign '!'. In traffic regulations it is a sign for caution. In the written language it is an exclamation mark. But what does it designate, a sign of what specific action, property, or object is this language sign '!'? The word 'elephant' denotes the animal, but what do the individual letters denote? Nothing at all except the letter. Then what makes them signs?

The system of signs, says the modern linguist. The significance or meaning of the letter



'a' is that it is the letter 'a' and not, say, 'b', or a combination of lines, but precisely that given letter, an element of system, which is the system of a given language. Something in the nature of the 'elementary particles' of a language, out of which we build up 'atoms'—syllables, words—which are then combined into more complex structures ('molecules') called sentences and texts.

Strictly speaking, in human language only words are signs. Letters (and the 'sounds of language' in speech, or phonemes) are not signs for the reason that they do not contain meaning. They are not correlated with anything in the actual world; letters are not signs but only parts that go to build up signs, like the sign-word 'elephant' with its component 8 meaningless letters. They are technically called 'figures'. But why are they needed, why not talk by means of simple 'pure words' not composed of separate figures? Why shouldn't language be like a system of traffic signals?

The answer to this question is given by cyberneticians, not linguists. They explain it as follows: the capacity of the memory system or the volume of the human memory is limited. A human being is not able to retain in his memory all sign words; in a traffic system there are several tens, in language there are tens of thousands! The English astrophysicist and science-fiction writer Hoyle describes a cosmic being in the form of a black cloud possessing a fantastic memory. For such a being, a language could be devised in the form of traffic signals, that is it could be made up of sign words. For the ordinary human beings that we are, this kind of language structure will not do, as is cogently demonstrated by experiments in teaching children to read.

One group was taught in the usual way: first individual letters ('m', 'a'), then syllables ('ma', 'ma'). Another group of children were given whole words at once: 'mama', 'cat'. At first things went more or less successfully, but then the children stopped: they could not remember more than about forty words. If after great effort another ten words were learnt, about that same number drifted out of their memory. Children are human beings and not the cosmic beings of Hoyle, capable of remembering limitless bodies of facts.

Suppose every sound distinguishable by the human ear were a separate 'word'. Misunderstanding and confusion would be rampant, for the slightest modification of tone, timber or loudness would generate new words. More, enormous numbers of sound-words would be beyond the capabilities of the vocal cords. Finally, there would not be enough sound waves for all the words, every language has tens and hundreds of thousands of words. The ear distinguishes far fewer sounds.

That is why our sound language is a multi-layer structure; first phonemes—the building blocks of the language—and then words and then phrases and sentences.

The number of phonemes, or building blocks of speech, varies in the different languages from 10-12 to 70-80. Even the least developed language has many thousands of words, and the number of sentences that may be constructed out of the words reaches astronomical figures. "Thus, language is so constructed that with the aid of a handful of figures and by means of endless combinations one can construct legions of signs," says the Danish scientist Louis Oehlmslev who bridged the gap between linguistics, the science of language, and semiotics, the science of signs. "As to purpose, languages are primarily sign systems, but as to their internal structure they are first of all something quite different, namely systems of figures that may be employed in constructing signs."

It is this property of language that enables us at any time to express any kind of thought or emotion. A few building blocks enable us to construct the extreme multiplicity and wealth of language, just like the relatively few chemical elements make possible the remarkably diversified world in which we live.

The Story of Semiotics

The Russian wit Kozma Prutkov was wont to say that the "specialist is like a swollen cheek from a toothache, it's one-sided". Bernard Shaw's aphorism was that a specialist is one who learns more and more about less and less until he knows everything about nothing.

Probably both were right, particularly for the time in which they lived. But towards the middle of this century there began a unification of the sciences along with increasing specialization. Cybernetics was the starting point. The electronics man and the psychologist, the biologist and the mathematician, the engineer and the logician, the linguist and the physicist all of a sudden found a common language. The problems being attacked by a wide range of the humanities and the engineering and natural sciences proved to have much in common.

Semiotics is just such a composite scientific discipline. Despite its youth, the subject matter that it studies—systems of signs used in human society—is as old as mankind itself.

Mathematicians were the first to approach the problem of sign theory. Mathematical logic was born in the middle of last century. It regarded mathematics and logic as specific systems of signs that are constructed in accordance with strict formal rules. Much later this 'pure theory' became the basis of cybernetics; cybernetic machines—computers—operate on the basis of rigorous formal rules.

Mathematicians were followed by linguists. To find out what a language is, say English, one has to compare it with the other languages of the world (German, Chinese, Bulgarian, Eskimo, etc.), to see what it has in common with them and what differences there are. To find out what human languages are in general, we have to learn about other, nonlanguage systems of signs: traffic signs and etiquette, games and gesture language.

Languages exist for society and in society, like all other media of communication. Human society could not exist without them. That is why, of late, problems of semiotics are beginning to attract the attention of sociologists, ethnographers, and economists.

Where are language programmes stored in the brain? Where are the cells located that are responsible for speech and understanding of speech, where is the speech centre? These problems of semiotics are solved by physiologists and psychologists. Psychiatrists and physicians are engaged in the treatment of deranged speech (more broadly, sign activities) of human beings, the abnormalities of the brain.

We learn signs from early childhood. But how do people achieve this who cannot see or hear?

The studies of the Soviet scientist Ivan Sokolyansky have demonstrated that a great deal may be attained in these extremely difficult conditions by constructing a special language for the deaf and dumb.

The senses of touch and smell are sufficient for a person without sight and hearing to become a full-fledged member of society. To illustrate, take the case of Olga Skorokhodova, a blind deaf-mute patient of Dr. Sokolyansky, who successfully defended her dissertation for the degree of Candidate of Pedagogical Sciences.

Semiotics is also an important aid to the science of teaching, because the process of instruction largely involves teaching people to learn signs, which may be the signs of the mother tongue or a foreign language, the signs of mathematics, physics, chemistry or other sciences.

The theory of signs is likewise exceedingly important in the field of cybernetics and computers. Semiotics is involved in the translation of human sign systems into the rigorous and unambiguous language of machines.

Incidentally, not all human systems of signs are amenable to translation. A semiotic analysis of art has shown that it cannot be formalized completely. The task of semiotics consists in finding the laws of a "language", which when followed generate a "text" (this may be a phrase in ordinary language or a "text" in behaviour, or a work of art). But the sign system of art differs from the other sign systems in that its "texts" cannot be completely generated by the laws of the "language", the laws of the system! Any pedantic adherence to rules in art does not yield a work of art but simply a pattern, a general something, a trite stereotype. (This peculiarity of art does not, of course, make it impossible to investigate art by emlpoying semiotic methods.)

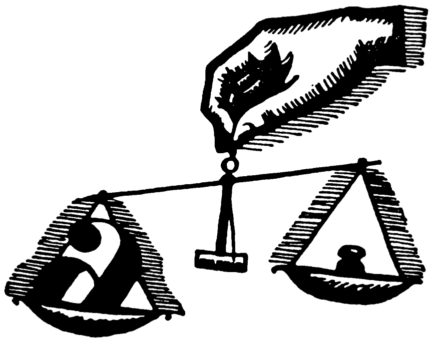
The role of semiotic methods in the humanities is comparable to the role of mathematical methods in the natural sciences. The theory of signs enables one to introduce into the descriptive sciences strict formulations and exact terms.

The problems of semiotics are intimately bound up with practical affairs; but aside from purely practical applications, the theory of signs has yet another no less important aspect. It helps man to identify himself and his place in the surrounding world.

What is man? What is his place in the world among the other living beings? How do we human beings differ from animals? From man-made automata? What will be the possible difference and similarity between us and intelligent beings from other worlds with which contact may be made in the twentieth century? And is contact with them a possibility?

These problems cannot be solved without the help of the theory of signs. It is only semiotics, which deals with the similarities and differences between the languages of animals, machines and human beings, that can point out the similarities and differences between modes of communication of extraterrestrial intelligence and earth dwellers.

Such is the extraordinarily broad range of the theory of signs: from cosmic linguistics to the language of gestures, from the primitive cries of beasts to the remarkably complex intricacies of the languages of the arts: painting, music, poetry and sculpture.



Languages and Codes

Our ordinary human language.... All kinds of media of communication: signalling by flags, gestures and road signs.... The marvellous language of music, art and dancing.... The crisp rigorous language of commands and the numbers of computers.... All these are means of communication, means of conveying information. The information they convey may be studied and even measured with the accuracy and heartlessness of numbers. Our story is about how this is done.

The Theory of Information

Why do we speak? What is the aim of communication? What is the purpose of such long-range media of communication as television, radio and the telegraph?

Quite obviously for the transmission of information.

For a long time the concept of information was rather nebulous and indeterminate, and it appeared destined to remain that way for ever. Indeed, what is there in common between a brilliant discovery and a telephone conversation, between the answer to the question "How do I get to the Bolshoi Theatre?" and the readings of a barometer?

Yet a theory has been developed in recent years—and has now become an established mathematical discipline—that permits objectively assessing the amount of information contained in any message whether it is a memorandum or Pushkin's poetry, a telephone conversation or a Bach piano concerto, a weather forecast or the report of a discovery destined to revolutionize science. It is called the theory of information.

The theory grew out of purely practical problems: to find the most economical telegraph code, ensure reliable radio communications, eliminate interference in communication systems, and the like.

However, after the American Claude Shannon laid the foundation of the probabilistic theory of information in 1948, the theory was taken up by a wide range of researchers: biologists, linguists, philosophers, geneticists, arts specialists, mathematicians and psychologists. A code was defined as any system of signs designed for the transmission of messages. With a definition that broad, human language, nucleic acids (the carriers of genetic information in the body) and the arts began to be regarded—and even measured with numbers!—as specific codes.

How do we measure a quantity of information? The cornerstone of the modern mathematical theory of information is the concept of indeterminacy or entropy. When we toss a coin, it can only come down heads or tails. If we throw dice, the indeterminacy of the result increases, for there is an equal probability of any one of six faces of the die coming up. Information is what removes indeterminacy or, to put it bluntly, ignorance.

There are different kinds of ignorance, naturally. There are situations where only two answers are possible: yes and no. But there are also situations where the number of such answers is immeasurably great: the number of possible combinations of protein molecules produces the monstrosity 10^{1300}, which is one followed by a thousand three hundred zeros!

When does a message convey no information? When we already know what it contains. If I say that $2\times2=4$, you will hardly gain anything from my message. A ball is thrown into the air, and we always know that it will fall to the ground. A news item reporting that this is what occurred will not convey any information to us. The situation is different if we are trying to toss a ball into the basket in basket-ball. In basket-ball there is indeterminacy: the ball may go in or miss.

But then a lot depends on who is throwing the ball and from what distance. Players of the famous Harlem team of the United States can score five hundred times without missing once on penalty shots. We will hardly be getting much information if we learn that one of the Harlem 'Globe Trotters' scored on a penalty shot. When a beginner tosses, a report that he missed is also low on information content because we are pretty sure that he will anyway. But when we learn that a beginner has scored on the first shot—surprises like that happen—then we have much more information.

Why? Simply because it is an unlikely event. Credit goes to Shannon for introducing a quantitative measure for information that is contained in a choice of events out of a series of events occurring with different probabilities. Previously, only events of equal probability were taken into consideration.

As early as 1928 an American engineer, Hartley, introduced the concept of quantitative measure for information contained in a choice of events from a set of equally probable events (for instance, the faces of a die come up with equal probability, the highly desirable six just as often as one or two). Hartley proposed estimating this amount of information as the logarithm of the number of possible events. The accepted unit of measurement is the bit (acronym for 'binary digit') or yes-no unit. This explains why the logarithm to the base 2 is used instead of the logarithm to the base 10 (decimal logarithms) commonly used in school.

A statement that a tossed coin has come down heads supplies us with exactly one bit of information. The point is that $\log_2 2$ (heads or tails) $= 1$, which is one bit.

The fact that one gets a card of clubs, spades, or any other of the four suits of cards yields information to the extent of two bits because $\log_2 4 = 2$. A report about the outcome of a situation where there are eight possible (equally probable) variants will produce three bits of information ($\log_2 8 = 3$, or $2^3 = 8$, the number of bits being the exponent of the number 2).

However, this measure is convenient and valid only when the possibilities are equivalent, that is are equally probable. The suit of a playing card, the face of a cube, and the heads or tails of a coin all come out with an equal probability. But then suppose the probabilities are not equal? For instance, the probability that in July the temperature in Moscow will be above zero is very great and the probability of a frost is negligibly small. Yet according to Hartley, both cases of weather are considered to be equally probable. Therefore, a report that the temperature in Moscow on July 1 was above zero (which of course surprises no one at the height of summer) carries exactly as much information as the news of the temperature falling to zero and below zero (which, quite naturally, would amaze any Muscovite).

Hartley realized, of course, that the probability of an outcome exerts an effect on the amount of information carried in the message. One certainly cannot attach the same significance to an almost improbable outcome as to a highly likely one. But he believed that the difference between the outcomes could not be expressed in numbers. They are determined by psychological (when dealing with human beings), meteorological (if the weather is the subject) or other factors not amenable to mathematical procedures.

Claude Shannon demonstrated that this point of view is erroneous. No matter what the factor—psychological, meteorological or any other—it can be taken into account by appealing to the theory of probability. He proposed a formula (now called the Shannon formula) that may be used to measure the amount of information about events which occur with different probabilities.

This is Shannon's formula:

$$H_1 = -(P_1 \log_2 P_1 + P_2 \log_2 P_2 + \ldots + P_n \log_2 P_n).$$

H_1 is the amount of uncertainty that a message eliminates and, hence, it is a measure of the amount of information (for information destroys uncertainty!); n is the number of 'outcomes', and $P_1, P_2, \ldots, P_n$ are the probabilities of their occurrences.

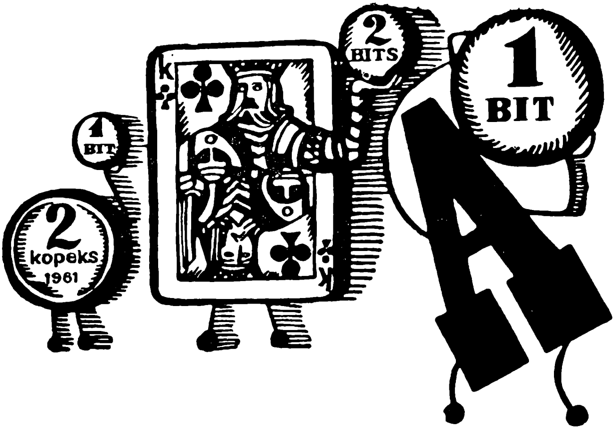
Thanks to this formula, scientists are now able to measure information contained in extremely diversified messages. The point is that every 'code sign'—we recall that a code is any of the extremely diversified sign systems—has a definite probability of appearing and, consequently, will carry a certain amount of information that we can measure. What is more, due to the fact that we take logarithms as a measure of information, we can combine the amounts of information contained in each code sign that comprises the news item; thus we can measure the amount of information of any statement.

Indeed, probability theory states that the probability of two events is equal to the product of the probabilities of the events. Hence, the logarithms are equal to the sum of the logarithms. And the sum of the items of information carried by code signs is equal to the information of the entire text made up of such signs. If it were not for logarithms, we would have to multiply the probabilities of appearance of these signs. The convenience of the 'logarithmic' Shannon formula lies in the fact that according to this formula the information of two pages of a book is the sum of the information of the first page and the information of the second page; the information of the whole book is the sum of the informations of all its pages.

But then we already pass out of mathematics into the field of another scientific discipline known as mathematical linguistics.

How Much Information Is There in a Letter?

How much information does a single letter of a language contain? This question arose right after the information theory was born. Let us try to answer it: How many bits of information are contained in a single letter of the alphabet (we take the Russian alphabet with its 33 letters). In addition we have the zero letter (or gap) between words. This brings the total to 34, but since the letter 'ё' and the letter 'е' are basically one and the hard sign and



soft sign may be considered as one, we get a total of 32 letters, or 32 code signs, which is a very convenient number for measuring by means of binary logarithms: $2^5=32$. Before getting information about some one letter of the Russian language we have an uncertainty with 32 outcomes, for we do not know which letter will be read to us. That means that one letter of the Russian alphabet carries an amount of information equal to $\log_2 32=5$ bits. That is the maximum quantity of information that one Russian letter could carry if it weren't for the fact that all languages (including Russian) have a property which in the theory of information is called redundancy.

Redundancy enables us to distinguish between the maximum information that a single code sign can carry and the actual information that it does carry. Redundancy is a measure of the uneconomicalness of the code language. Every natural language like Russian, English, and German, say, and not the artificially contrived code languages of specialists, possesses this property.

If all combinations of letters in Russian had meaning (and if in addition all letters occurred with equal probability), then we would have a maximally economical language with no redundancy. Actually, every language has a considerable amount of redundancy.

Why? Why is redundancy needed? Is it not possible to create a language in which any combination of letters would yield a meaningful word? In other words, couldn't we construct a language without redundancy?

Yes, but only theoretically. We can even calculate the saving obtained by a language devoid of redundancy. A calculation of this nature was carried out by the Soviet information-theory specialist A. Kharkevich. If the alphabet of our language contained 30 letters and all words consisted of one single letter, then there would be 30 different words, say, like 'я' ('I'). The number of two-letter words would come out to $30^2 = 900$— 'мы', 'он' ('we,' 'he'), etc. But we have no words like, say, 'вм' ('vm') or 'зы' ('zy'), 'фх' ('fh'), etc. Using 30 letters we could generate 27,000 three-letter words like 'нам', 'мир', 'май' (possible English combinations 'you', 'sit'), etc., but combinations like 'йма' ('ima') and 'сна' ('sna') do not exist. The possible combinations of four letters, using the same 30, come out to 30^4 or 810,000, five-letter combinations yield 30^5, or 24,300,000 words. Yet a language like English (with 26 letters) has something like a million meaningful words (the unabridged Webster dictionary contains close to half a million). Now returning to Russian, if we take it that the average number of letters in a Russian word is seven (statistical counts indicate that that is roughly so), then only about 0.0002 per cent, or two ten-thousandths of a per cent of all possible combinations of Russian letters form words!

Nevertheless, the redundancy of our language is not in the least an unnecessary frill. The reason is this. Try to imagine the situation if there were a mistake in one letter in a language devoid of redundancy. In the word 'impossiblk' the error in the last letter hardly at all offers any barrier to understanding, yet in a nonredundant language it would be a totally different word, meaning, perhaps "something quite possible".

Or imagine a doctor writing out a prescription and making one little mistake—one letter misplaced! The result might even be a poison instead of the needed drug. It is thus quite clear that redundancy is no superfluous embellishment but something quite useful and important. It arose thousands of years ago, but only now has been fully understood and realized, all thanks to the theory of information. It is a sort of "safety margin" in language.

It is possible to create an artificial language that has no redundancy. Actually, numbers represent such a language: any combination of numbers 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 yields a meaningful number.

Not a single combination of numbers is "lost", devoid of meaning. True, redundancy even here is not equal to zero because number combinations like 078, 0078, 00078 and 78 designate the same number! But the redundancy of the "language of numbers" is exceedingly small compared to that in human languages.

This is clear not only to scientists but to workers in more practical fields. Payment for typing numbers is two or three times higher than for ordinary straight texts. Translators experience great difficulties when they have to do with proper names and enumerations, particularly of numbers. The reason is quite obvious: the translator has no way of knowing what names and numbers follow, whereas in ordinary speech it is almost always possible, because of the redundancy of language, to see ahead.

All natural languages have redundancy. Recent investigations indicate that nearly all the languages of the world have approximately the same redundancy, roughly between 70 and 80 per cent though Warren Weaver in *Science and Imagination* (1967) writes: "It turns out that English is actually just about 50 per cent redundant." This means that in every 100 letters of text in any natural language of the world we can guess about 70 to 80 letters, if, of course, we know the language (though we may not necessarily know the contents of the text).

When speaking of language redundancy, we have in view language generally, the average literary Russian, English or any other one. This is precisely where the amount of redundancy fluctuates between 70 and 80 per cent. Now in any specialized language (say, technical, scientific, or juridical, or in some kind of jargon) the redundancy is as a rule higher than in an average literary text. This is because the specialized language has a smaller word stock, many more standardized collocations and constantly repeating terms; in a word, it is much poorer than the language "at large" (which by the way is one reason why we are able to learn to read foreign specialized literature without really knowing the language very well).

Experienced typists tell us that it is much more difficult to type poetry than specialized literature due to the unexpected words and combinations that appear in the text.

The Soviet researcher R. Piotrovsky has found that the redundancy of 'official style' in the Russian language is equal to about 85 per cent (scientific and political texts were analyzed). American scientists made a study of a still more highly specialized language—radio conversations between aircraft pilots and airport officers. Quite naturally, such conversations are limited to a few very narrow topics and are constantly repeated. There is also a high degree of standardization of form.

All these factors resulted in a redundancy of close to 96 per cent, which means that 96 letters out of every 100 are superfluous!

Incidentally, such a high degree of redundancy is in no way remarkable. A single error in reception—an error of one letter—may mean the difference between life and death. And errors are highly probable in conversations of that nature where there is constant noise and distracting disturbances. That is why the pilot and the duty ground officer talk 96 per cent extra so as to eliminate the slightest possibility of any error.

The extreme redundancy of specialized languages that has come to light thanks to information theory is beginning to find uses in such practical matters as the compiling of codes. Firms in the United States make use of special codes in their correspondence. Today, specialists in information theory are always invited to take part in the compilation of codes. Frequent repetition of words and phrases permits codification of whole words and word combinations instead of separate letters (like, say, the Morse code). For example, the salutation 'Dear Sir', which consists of 7 letters may be written as a single sign.

Unlike superfluous specialized texts, fiction literature has, as a rule, a lower level of redundancy than language generally. The language of good writers is not standardized but colourful, rich, and full of imagery, unexpected. Naturally, the redundancy of such writing will be lower than in ordinary language.

"Phrase Models"

What would a Russian text (or an English text, for that matter) look like if it were devoid of any redundancy and one letter of the language had the maximum information of five bits?

A Moscow mathematician R. Dobrushin has worked out a procedure for producing artificial phrases of the Russian language.

Suppose we have a bag filled with letters of the Russian language (the same could apply to English, naturally), all letters in equal number. We then extract them one after another at random. The result is something like this:

оухерроль дбщянхвщихйжтифнарфенвщт

Practically impossible to read. And to notice an error would be exceedingly difficult too. Incidentally, this illustration has appeared in scientific books from time to time and mistakes have occurred quite often. Yet they were never noticed by editors, proof readers or the authors, and all because the text has no regularities, no rules!

We have already said that information theory permits finding the amount of information contained in a message where all the code signs have not only equal but also unequal probabilites of occurring. Language (Russian or any other one) is just such a probabilistic code.

What probability is there of one and the same letter of the Russian language occurring? To find out, we have to take a sufficiently long Russian text and count the number of times a given letter occurs. Actually, letter frequency can be more precisely determined only on the basis of a number of texts, not one, because it often happens that the nature of the text determines the occurrence of a letter and this can vitiate the results. For example, if we take a text from a book on higher mathematics we will most likely get a rather higher frequency of 'ф' than is found in ordinary Russian language texts. This is due to the recurrence of such mathematical terms as 'дифференциал' ('differential'), 'коэффициент' ('coefficient'), 'функция' ('function'), and the like. A still greater deviation from the norm is found in poetry. Mayakovsky offers an excellent illustration in the Russian language. His poem entitled "150,000,000" gives us an extraordinary proportion of 'а' s and 'б' s:

Бей, барабан!
Барабан, барабань!
Были рабы!
Нет раба!
Баарбей,
Баарбань!
Баарабан!

But such deviations do not occur often, and practically every excerpt of any printed material is close to the average language as far as its statistical regularities go.

In an average Russian text, the letters of the alphabet (this includes the spaces between words as well) have the following frequency: out of a thousand letters of text, 'а' is encountered 62 times, 'б' 14 times, 'в' 38, 'г' 13, 'д' 25, 'е', (and 'ё' together) 72, 'ж' 17, 'з' 16, 'и' 62, (like 'а') 'й' 10, 'к' 28, 'л' 35, 'м' 26, 'н' 53, 'о' 90 (highest frequency), 'п' 23 times, 'р' 40 times, 'с' 45, 'т' 53, 'у' 21, 'ф' 2 times (lowest frequency), ‘х’ 9 times, ‘ц’ 4 times, ‘ч’ 12 times, ‘ш’ 6 times, ‘щ’ 3 times, ‘ь’ and ‘ъ’ 14, ‘э’ 3, ‘ю’ 6 and ‘я’ 18 times. The space between words occurs most frequently of all—175 times per thousand letters of text. (These frequencies are taken from *General Theory of Communication* by A. A. Kharkevich, 1956.)

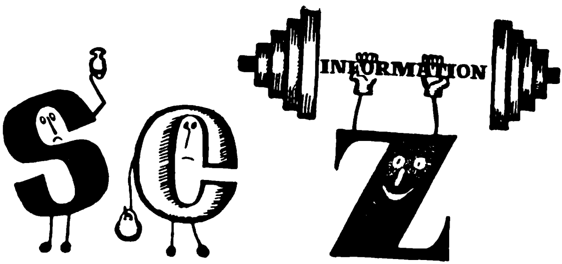
It is quite clear, as we can see, that the letters of the Russian language are not on an equal par as far as frequency of occurrence goes; the rare letters like ‘щ’ and ‘ф’ carry more information than the frequent letters like ‘о’ or ‘е’. On the average, one letter of Russian (taking into account the probability or frequency) will not carry the 5 starting bits of information, but 4.35.

Let us now attempt to construct a model of a phrase with allowance made for the probability of occurrence of each letter. We put into the bag cards with letters in quantities proportional to their frequency of occurrence (the spaces between words are included). There will be more ‘о’ s and fewer ‘щ’ s. We might get something like

т цыяь серв однг эбя энвтша буемлолйк

Now compared with the first phrase model we produced, this looks to be much closer to a real phrase of the language. Here, the word length is more like real Russian word lengths and quantitatively it even has a typically Russian distribution of vowels and consonants. Still and all, this sentence does not look like a piece of real Russian writing.

The point is that allowance was made only



for the frequency of individual letters of the Russian alphabet, all letters being considered independent of one another. Actually, that is not the case. Indeed, there is one letter, soft sign, that never appears after a vowel or at the beginning of a new word. There are other limitations to letter placement in words in Russian, all of which consequently reduce still more the amount of information that a single printed character of text carries. If a sufficiently large text is taken (this applies, quite naturally, to other languages, English, for instance, as well), we can compute the frequency of combinations of two letters and then we can use such data to calculate the quantity of information carried by a single letter of text and compile a 'phrase model'.

The quantity of information carried by one letter of text (for the Russian language) comes out to 3.52 bits instead of 4.35 bits (when making allowance for the frequency of two-letter combinations).

Now, with the number of paper slips proportional to the frequency of two-letter combinations, a random extraction of pairs might yield the following:

умароно как всванный рося ных ковкров недаре.

This is much closer to an actual Russian sentence.

Reasoning in similar fashion about triplets (combinations of three letters and their frequencies) we see a further reduction in the amount of information contained in a single letter of text. For Russian we have about 3 bits of information (more precisely, 3.01). The situation in English is similar. A sentence model might look something like the following:

покак пот дурноскака наконепно ане стволовил се твой обниль.

This is extremely close to actual Russian writing.

Finally, if we further take into consideration quadruple combinations of letters of text, we obtain a very close approximation:

весел враться не сухом и непо и добре.

Still closer approximations are obtained by taking into account combinations of five, six, etc., letters. The quantity of information is accordingly reduced each time. The question arises: is there a limit? If a quantity of 5 bits is the maximum amount of information that a single Russian letter is capable of carrying, then what is the amount of information carried by a single letter in a meaningful sentence of

Russian (the figures are probably roughly the same for other languages as well)?

Now, if to the restrictions imposed on combinations of two, three and more letters we add the restrictions of the rules of grammar and style and semantics (every sentence has to make sense—the models given earlier for purposes of illustration were all meaningless, and that is why they were called sentence models), we have a complex of very rigorous restrictions.

Means of communication serve to convey meaning and not simple combinations of letters or other conventional code symbols. The principal task of linguistic communication is to convey meaning. If we take a sentence like, say *"Coming tomorrow, John"*, our interest lies solely in the meaning and not in the least in the combinations of letters.

We are interested to know what lies behind the letters and what is designated by the code symbols. To the person who knows John everything is clear, but to anyone else who does not know him the telegram hardly conveys any meaning at all. Actually, the message does not transfer any information to anyone for whom it is not intended.

A sentence may be beautifully constructed and sound marvellous yet have no meaning whatsoever to anyone. That brings us to the question of the meaning of meaning. What is meaning? For instance, a modern physicist says that a photon is at the same time a particle and a wave; yet to a nineteenth-century scientist this sentence would be absurd.

Quite naturally, the only real verification of the sense of our communications is actuality. But even here there is quite a variety of subtleties as to what is to be considered meaningful and meaningless. In the languages of such sciences as physics, astronomy, chemistry, the meaning of sentences is verified by experiment, by the readings of instruments, in mathematics by the derivability of earlier accepted axioms. In ordinary life we simply use our common sense and dismiss meaningless concoctions like "Sugar is eaten with salt", though strictly speaking such a sentence is not without a certain amount of meaning for people with exotic tastes.

If we now turn to fiction (prose and poetry), the concept of meaning becomes still more involved. Some of the dialogues in Hemingway's stories are full of deep meaning yet are not so easy to grasp straightoff.

Despite the abundance of meanings in different types of writing—scientific, poetical, everyday speech—there is a commonly accepted criterion which ordinarily allows one to distinguish between meaningless sentences and sentences that carry real information. Any normal person seems to have a basically correct notion of what constitutes a meaningful sentence, even though there are no official instructions on this point. But how is one to determine meaning in a mathematical sense? In others words, how is it possible to calculate information that is contained in a single letter of a sentence with meaning?

An ingenious solution to this difficult and what appeared to be unresolvable problem was found by Claude Shannon. The method was then refined by A. N. Kolmogorov, who conducted experiments at Moscow University (at the chair of probability theory) to determine the amount of information contained in a single letter of a meaningful Russian text. A so-called 'guessing procedure' lay at the heart of these experiments.

The guesser was given a text to read through (to learn the style), and then the text was covered over up to a point; the next letter in the text had to be conjectured. Let us say the sentence ran as follows:

"Early one morning I was walki..."

The letters 'n' and 'g' are guessed, and the beginning letter of the next word is then provided. Let us say, it is 'a'. The guessing subject suggests 'l', for he immediately thinks of 'along'. At this point he will again need some information about the next letter. Taking into account the total number of letters and the percentage of letters guessed correctly and those missed (these include letters stated incorrectly), it is possible to determine how much information a person extracts from a single letter of the language when reading a meaningful text.

The experiments conducted at Moscow University suggest that the information for the Russian language is approximately equal to one bit. Only one miserable bit! The very same result was obtained in experiments carried out by the prominent Leningrad linguist R. G. Piotrovsky. There is a substantial difference between this single bit of information and the original five bits that a single letter of the Russian language could have carried, were it not for so much redundancy! The restrictions imposed by the rules for word formation and the rules of grammar, and, finally, the demands of meaningful speech all conspired to reduce the amount of information carried by a single letter by a factor of five.

In his book *Great Ideas in Information Theory, Language and Cybernetics* Jagjit Singh writes:

"In an actual [guessing] game played with a sentence of 124 letters, 89 were guessed correctly and 40 had to be intimated." This yielded "an average of 1.93 bits per letter [for English]".

"The real information content of the English alphabet thus seems to be about 2 bits per letter, if we take into account all the constraints due to redundancy."

The Shannon-Kolmogorov procedure permits determining the information carried by a single letter of any kind of writing: fiction or nonfiction. The amount of information in one letter varies with the writer as well. Quite naturally, pen hacks turning out reams of stereotyped writing will produce smaller quantities of information than skilled writers of high class and style.

On Paper and Out Loud

So far we have dealt only with the quantity of information in written material, but what about speech, which is the main vehicle of human communication? Is there the same amount of information in a single letter and in one phoneme of speech?

In the Russian language (the written code-language) there are 32 elementary code symbols, or 32 letters. Spoken Russian contains 40 code symbols, or phonemes. (By way of comparison English has 26 letters and more than forty phonemes.) Now if the written and spoken languages did not have any redundancy, the quantity of information per phoneme would naturally be somewhat greater than per letter. But, as we know, both modes of communication, written and spoken, are greatly restricted. One letter of a written message carries one bit of information. How much information does one phoneme carry?

The written language is somewhat more smoothed over than the spoken language. On the other hand, in writing one uses fewer inserts or extra words, which serve more to cover over pauses than to convey meaning (like, say, "now a", "as I was saying", "the point is that" etc.). In the spoken language we take liberties of grammar and style, inserting completely unneeded words and phrases, frequently just to keep the conversation going.

Our problem now is to find the quantity of information that is carried by a single sound of language, by one phoneme. Here we can resort to the "guessing procedure" proposed by Claude Shannon. However, in guessing the continuation of a written communication one can utilize his full knowledge of the language, whereas in guessing phonemes this is extremely difficult. The point is that from early childhood we read texts written with letters and not with conventional phonetical symbols.

Researchers found, however, that one and the same text may be recorded in phonemes and in letters. We are always in a position to read a conventionally recorded text and put an oral message into writing. In other words, both communications contain one and the same 'total of information', which is independent of the form in which it is recorded (coded), oral or written.

In short, the quantity of this total information is equal to the number of letters used to record the given text multiplied by the amount of information carried by a single letter of writing. (For Russian, it will be recalled, this is one bit.) And the quantity of total information will be equal to the number of phonemes multiplied by the amount of information carried by a single phoneme. Proceeding from this we can find the quantity of information contained in a single phoneme.

For the English language, it is roughly equal to the amount of information per letter of a written communication. Phonemes and letters turned out to be approximately equivalent.

But whereas speech and writing are equal partners as to the amount of information per letter, they differ in the distribution of information within a word, as has been shown by the recent studies of Piotrovsky. He writes: "Whereas in writing the peak informational load is carried by the first letters, in speech, the information is concentrated around the phonemes of the stressed syllable. This difference is obviously associated with different modes of perceiving linguistic units of oral and written forms of the language. A written word is perceived as a linear sequence of elementary symbols (letters). Orally, a word, or at least its syllables, is perceived as a whole."

An indication that information in written communications is concentrated at the beginning of a word stems not only from scientific experiments, but also from acronyms and abbreviations of words. Take such examples as MASER, LASER, NATO, or i.e., e.g., etc. All languages can supply illustrations of this fact. Even in ancient Rome, the first name of a citizen was abbreviated; for example, M. Tullius Cicero for Marcus Tullius Cicero.

In Spain in the old days, whole phrases of letters were contracted: SSS QBUM, which stood for "Su seguro servidor que besa a usted la mano"—your obedient servant who kisses your hand. Gentlemanly terse, to say the least.

Then of course we have the common abbreviations of countries: USSR, USA, etc. The Japanese name for Russia is abbreviated to 'Ro' and Vladivostock, which sounds like Urodsiosutokku, is cut down to 'Urodsio'.

The beginning of the word does seem to be the most important. At times, true, the end is also retained. But the middle is invariably dropped. Take 'Mr.' as an example for 'mister', the Spanish 'Sr' for 'Señor.' The Arabs appeal to Allah with abbreviations of lengthy messages that are repeated all the time. The god-fearing Mohammedan heaps blessings on his Allah and in a single letter will often mention him as many as five or more times, each time repeating the full list of the blessings. Quite naturally, one needs a system of abbreviation. And so Arabs take the first letters of the first three words and the last consonant of the last word of the phrase.

From the foregoing it is clear that the beginning (more rarely, the end) of a word contains a great deal, while the middle portion means hardly anything.

But this goes only for the written language! In speech, words are abbreviated differently. Here, very often the initial portions of words are dropped, precisely those that are so important in writing.

It will be noted that stressed syllables are always retained in speech, no matter how much else is dropped. The first letters of a word are important in writing (and those are the ones that are retained in abbreviations); in speech, the bulk of the information is concentrated in the stressed syllables of words.

Language and Cybernetics

The theory of information originally dealt with problems of coding, reliability, and the capacity of lines of communication. Later it was found, however, that the same concepts can be applied to 'lines of communication' used by living organisms for obtaining information via their sense organs. Automata and living beings obey the same laws of reception, processing and storage of information. This idea led to the birth of a new science called cybernetics, which was officially christened in 1948 with the publication of *Cybernetics, or Control and Communication in the Animal and the Machine* by Norbert Wiener, the American mathematician. The theory of information became the cornerstone of cybernetics because the latter deals with "the study of systems of any nature that are capable of receiving, storing and processing information and utilizing it for control." That is the definition given by Academician Kolmogorov, prominent Soviet mathematician.

"Systems of any nature" include human beings, animals and automata. They can all be described in unified terms of the theory of information.

How much information can be received and processed by the human brain? (Quite naturally, to find the volume of perceived information, we must take into account the time during which it is received. The rate of information intake per unit time is called the capacity of the system.) To answer this question we must first determine the capacity of the human sense organs.

The retina of the human eye has about 10 million nerve cells: cones and rods. About a million nerve fibres run from the eyes to the brain. Now if we take it that one 'nerve fibre of the eye' can perceive one bit of information in one second (let us say, it either reacts or fails to react to light), we then have one million bits per second. This figure is greatly reduced because the human eye is capable of reacting with a much greater rate than one response per second. That is why scientists estimate the capacity of the eye-brain system as of the order of 5,000,000 bits per second.

The number of nerve fibres in the ear is less than in the eye. For the eye there are one million fibres, for the ear about 30,000, or 30 times less. The capacity of the ear-brain system is about 50,000 bits per second. Incidentally by telephone we can convey somewhat less, about 30,000 bits per second; it is the 20,000 bits that we lack that make for distortions of the human voice in telephone communications.

But 50,000 bits is a tremendous quantity. What this means is that our ear-brain system can handle a choice from among $2^{500,000}$ possibilities (the bits are the exponent, if you recall Shannon's formula). This number exceeds one followed by ten thousand zeros. Compare a million, which has only six zeros. Or take a thousand million with nine zeros. To count up to a thousand million at the rate of one a second, working a ten-hour day, one would need nearly 90 years. Every single person on earth working through the history of mankind would never reach $2^{500,000}$.

50,000 bits per second is a tremendous quantity of information entering the brain. But how much information can be processed and consciously perceived?

Fifty bits per second, or 50 binary units of information, is the limit of perceivability of the human brain. This limit is reached in a fast conversation, very fast reading and in taking down speech in shorthand. Information coming in faster than 50 bits per second cannot be perceived by our brain, although the sense organs are capable of transmitting a thousand times more information (hearing) or even 100,000 times more (seeing). The capacity of the sense of touch lies somewhere between that of seeing and hearing.

Of course a great deal depends on the individual traits of the person, his physical and psychical state, the degree of training and other factors. A skilled typist or pianist can convey information at a rate of 25 bits per second, whereas an inexperienced person would hardly achieve 5 bits per second.

Yet when listening to a fast conversation, the human brain is working to capacity imbibing maximum quantities of information at rates of 40 to 50 bits per second. But that is only quantity. Information theory does not make allowance for the value of the information received: in idle talk it would probably be close to zero, though we would be receiving limiting quantities of information every second and the brain would be processing it all.

For Practical Matters and for Theory

Quite naturally, the task of information theory is not to figure out the amount of nonsense spoken every day. The capacity of the human brain is important not only to theoreticians, psychologists and computer experts, but to engineers and technologists on the job. The work of operators, dispatching officials, pilots and many others consists almost completely in keeping tabs on instrument readings or, to put it otherwise, in obtaining needed information. This is exactly where we have to measure the capacity of such information and compare it with the capacity of a living system such as the human brain and find the optimal number of instruments, their most effective arrangement, and so forth.

The modes of communication in society include the human being as the most important of all links. Hence, the practical demands for describing man and the technical facilities of communication in the same terms. This is done by information theory. To describe is not enough, it is necessary to improve such systems—telephone systems, to take only one illustration.

Telephone lines are overloaded during peak hours of service. But it is a fact that even existing telephone lines could transmit ten, a hundred, a thousand times more information than they do at present!

Indeed, we have already mentioned that in a telephone conversation about 30,000 bits of information are transmitted per second. Now in the fastest conversation the brain is capable of receiving only 40 to 50 bits of information per second. Which means that nearly 99.9 per cent of the information transmitted by phone is redundant, and only $^1/_{1000}$th part of telephone information carries useful facts. Now if the redundancy is eliminated (even a portion of it), then the capacity of telephone lines will be boosted tens and even hundreds of times.

Scientists of many countries are working on special devices that may be capable of analysing human speech at the input to eliminate redundant signals leaving only needed and useful information and then transmit it to the output, where other devices triggered by the signals received could synthesize or collect the sounds into meaningful speech.

Researchers hope to increase the capacity of other communication facilities as well that operate with a lower redundancy than the telephone. Phototelegraph and radio-relay lines, for example. But the record for waste appears to go to television.

Images on the TV screen are transmitted in the form of dots, or minute picture elements. Each one of these dots is a signal that can take on several meanings, depending on the degree of brightness.

How many such dots make up an image? In Soviet TV transmissions, about 400,000 to 500,000 elements, in American the number is 200,000 to 300,000, whereas in the transmissions of certain West European television centres an image is decomposed into nearly one million dots or elements!

Let us compare these dots with code elements, for example with the letters of the Russian language.

Just as the overwhelming majority of letter combinations do not make up meaningful phrases, so colossal numbers of element combinations (dot combinations) fail to produce something with meaning, in other words they do not produce an image.

We have already stated the calculations of mathematicians that only 0.0002 per cent of all possible combinations of Russian letters form Russian words. All the remaining 99.9998 per cent are only meaningless combinations of letters, nothing more. In exactly the same fashion, in television the majority of dot combinations have nothing to do with any kind of image. The redundancy here is much greater than in ordinary language.

The ear-brain system is capable of receiving 50,000 bits of information per second, while only a thousandth part of this information is processed. The eye-brain system perceives about five million bits of information per second. Television channels transmit fantastic quantities of information amounting to millions and tens of millions of bits! However, we consciously perceive not more than 40 to 50 bits per second, which is one one-hundred-thousand or even one millionth of the entire information that has been transmitted!

Scientists have been able to determine the quantity of redundant information that is transmitted in television broadcasts. Now the problem consists in reducing this redundancy of television.

In other communication facilities, the important thing is not so much economy as reliability. The point is that transmissions as a rule occur with interference and distortions (in electrical communications this is due to fluctuations in the loading in the network, in radio communications the causes are lighting discharges in the atmosphere and so forth). Pilots and spacemen are particularly interested in reliable communications. Information theory is helping communication experts, radio engineers, and designers to construct reliable long-range facilities for communication.

Our story has been devoted not so much to information theory proper as to its applications to the problems of linguistics. The study of language as a code is of use not only to linguists but also to all persons engaged in the study of codes in the direct meaning of the word—telegraph operators and communications men.

Even the oldest of modern telegraph codes, the Morse code, which was established many years before the birth of information theory, takes into account language laws. More frequently repeated letters (such as 't' and 'a' in Russian or 'c' and 's' in English) have shorter designations than others. Information theory permits establishing still more economical codes. For instance, the telegraph codes of the modern languages of India were constructed on the basis of up-to-date findings of language redundancy.

Compare that with the situation in the Morse code, where the letter 'o'—one of the most frequently occurring letters in the Russian alphabet—is denoted by three dashes, which is obviously undesirable. The reason for this wastefulness lies in the fact that Morse built his alphabet for the English language. In English, the letter 'e' has the highest frequency. Accordingly, the designation is a single dot (the shortest of all). In English, 'o' is a rarer letter and so received the longer code designation. When the Morse code was adopted to the Russian language, communications men did not reconstruct it, for the theory of information had not yet been devised.

Information theory is extremely important for applied sciences, yet it is no less important for purely theoretical scientific disciplines such as genetics, which studies the transfer of genetic information, for physiology, psychology and for other sciences. Thanks to the theory of information, scientists are now in a position to objectively compare the various languages: our conventional language, the language of animals, signalization of the nervous system and telegraph codes, machine languages and the chromosome code in the transfer of genetic information—in a word, all modes of transmitting information from man to man, from animal to animal, from man to machine, from machine to machine.

We have mentioned that the redundancy of human language is conditioned by its structure, by the fact that language has regularities and relationships. Which means that redundancy expressed in precise numbers can serve as an indicator or criterion of the organization of a language. More, it can even permit of conclusions about the effectiveness, the capacity of one or another language.

So far, only the first steps have been taken in this direction. For example, Ramakrishna and Subramannian of India have established that when translating from English to German one bit of the English text is equal to 1.22 bits of information contained in the German text. In a reverse translation from German into English, one German bit of information requires 1.06 bits of English information. The increase in text as the result of a translation is well known to all translators. But the important thing is that the increase is less when translating into English than when translating into German. Ramakrishna and Subramannian drew the proper conclusion that one bit of information contained in the English text is in meaning equivalent to roughly 1.15 bits of information of the German text, hence that the English language is more economical than the German.

Here is what the Soviet linguist E. V. Paducheva writes in her article in a collection entitled *On Precise Methods of Language Study:* "A comparison of different languages from the viewpoint of redundancy would contribute to a more exact content of the concept of 'progress in language'. It would be interesting to find out whether there exist any kinds of regularities in variation of redundancy in the development of various languages. A determination of such quantitative laws would permit predicting trends in language development by proceeding from a quantitative analysis of the present state of the language."

That would mean predicting the future of a language.

How Does a Language Differ from Technical Codes?

The word 'code' immediately brings to mind the Morse code, the Baudot code and other similar technical codes. In information theory, the term ‘code’ has a broader meaning (just like the word ‘language’ in semiotics).

A code is any mode of recording communications, any symbolic system that is used to transmit information as has been agreed upon by the sender and the receiver (that, roughly speaking, is the definition of the term ‘code’ given by the noted American linguist G. Miller).

This definition includes our conventional language as well. Of course, language as such has much in common with all other communication facilities and can be readily studied by the theory of information along with the technical codes. But still and all, how does our language differ from them?

The grammar of technical codes is very simple: it only contains rules for combining the elements of the code. A human language is complicated, multidimensional.

The structure of a technical code, like say the Morse code, may be described on two pages or less, but to describe the Russian language (or any other of the human languages) requires a dozen volumes or so. Such a full description would include the sounds of the language, the grammar, all the words, all set expressions and word combinations that we use in our daily speech.

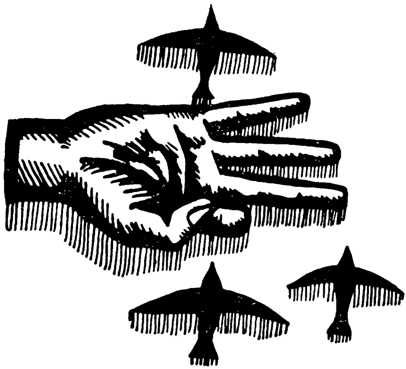
Technical codes are secondary and are constructed on the basis of our ordinary language or some other system of signs. Our ordinary language is primary and is not based on other systems of signs of other codes. To learn and speak another language to perfection, one has to spend many years of hard work. To learn the Morse code, one month or even a week suffices.

In technical codes, the rules for combining symbols are set up by the inventor. In language the situation is quite different, the whole nation is the creator of the language. A human language is the product of many thousands of years of development of a society.

The laws of codes are strictly defined in the Morse code and in all other technical codes. In language we have to do with texts, with phrases, expressions and words. And only by analysing such texts can we find the rules, the system of the language that make it possible to write. Now these laws are by far not so rigid as are the laws of technical codes.

It is important, to point out, however, that the rules of languages are frequently rather lenient; the broken speech of a foreigner can, after all, be understood, though practically every rule of grammar may be violated; the same goes for pronunciation—distortions may be extreme yet the meaning gets across. In technical codes, such transgressions would result in complete nonsense.

And, finally, the most significant difference between languages and technical codes. Language is not only a means of communication, a means for transmitting information, but also a vehicle for learning about the actuality that surrounds us, a facility for 'modelling' the world. It is precisely in this modelling capacity, the ability not only to convey information about the world, but to model the world, that human languages differ from technical modes of communication, from the signal systems of animals and from the language of machines.



People, Things, Words

Ethnolinguistics is a new scientific discipline that has come to life on the borderlines of linguistics, ethnography, history and culture. The subject matter of ethnolinguistics, elucidated in this chapter, is the interrelationships of language, culture and the surrounding world.

The Whorf Hypothesis

Benjamin Li Whorf, a fire insurance company official, was an observant man on his job. Fires flared up where people, reading 'Empty' on petrol tanks, grew careless and threw cigarette butts about, quite oblivious of the fact that 'empty' does not go so far as to mean that the remnant vapours in the tanks are not inflammable. Words gave rise to fires!

But Whorf was also a man of culture, and when he was not busy at the office he was studying the writings of the ancient Aztecs and Mayas; he later became interested in Indian languages, the languages of the aborigines of the Americas. Gradually his work gave birth to a theory that has since spread throughout the world. After Whorf's death an international congress was held in Chicago devoted to what has now come to be known as the Whorf hypothesis. The gist of this theory is that the thinking and behaviour of individual human beings is completely dependent on language. But was Whorf right in so thinking?

Things and Labels

How many colours can you name in the rainbow? Simple? Of course, there are six: red, orange, yellow, green, blue, violet...

That is true of English and German, but not of Russian where the blue is divided into two varieties, each with a special name, equivalents of light-blue and dark-blue. In the language of one of the peoples of Liberia, our six colours of the rainbow are covered by two words: one denotes what painters term the 'warm' colours (red, orange, yellow), the other the 'cold' colours (blue, violet and so forth).

That is not all. Some languages have different divisions of colours. Take a series like 'green', 'blue,' 'gray,' 'brown'. For this portion of the spectrum the Welsh have three words: 'gwyrdd', 'glas' and 'llwyd', the latter term denoting everything covered by brown and gray, or dark-gray. 'Glas' embraces light-gray, blue (light and dark) and green. The word 'gwyrdd' also takes care of that part of the spectrum called 'green'.

Thus, one and the same phenomenon in the world is called differently in different languages. Labels differ. The eyes of all persons in the world see the same colours, but different languages stress different hues, different portions of the spectrum.

Languages do not confine themselves to different interpretations of impressions of the outer world. The Russian language, like English, has words for 'sister' and 'brother', but the language of Hungary does not make that distinction. In Hungarian, one speaks of an elder sister or brother, or a younger one. In the Malayan language, brothers and sisters are combined under the generic term 'saudara'. Still more amazing divisions of the world (amazing from the point of view of certain European languages, naturally) are seen in the languages of the Indians, the inhabitants of Melanesia, and the peoples of Africa, America, Asia and Australia.

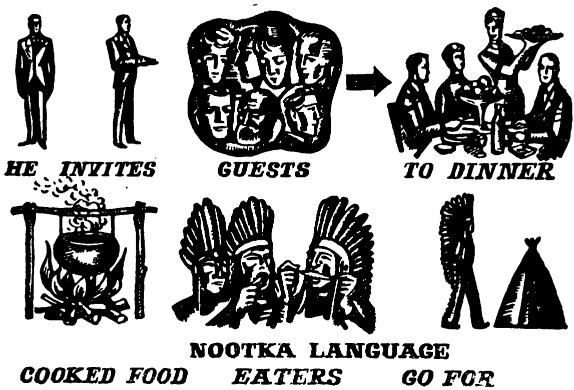
Nouns are generally taken to denote objects, verbs denote actions. This is common to Russian, French, German and English, and many other languages. Like most Europeans, we divide the world into two spheres: objects and processes. Actually, however, this is simply a property of our language and not of the surrounding world, which is in a state of eternal motion, becoming and variation.

'To strike', 'to run' are verbs for they designate processes in time, but why is an 'attack' a noun and not a verb? The phenomenon is the same—a process developing in time!

Why are such words as 'lightning', 'wave', 'pulsation' regarded as nouns denoting things and not as processes? Because that is the way our language has classified them. But another classification is possible. In the language of the Hopi Indians of the United States, these words are not nouns but verbs. And in the language of the Nootkas, island dwellers of Vancouver, Canada, all words would—from the standpoint of, say, English—be verbs. Indeed, that language knows no division into objects and actions: their view of nature is unitary and it has generated a single class of words.

In this unique language, a home stands or it simply 'homes', a fire occurs, or it 'fires'. Using suffixes or terminal units, the islanders form words that twist 'home' this way and that: "the house has long been standing", "a temporary house", "our future home", "the house that once was", "the house that is to become one", etc.

The Hopi Indians have a noun that may refer to any flying object or being, with the exception of birds. Birds are designated by a different noun. Whorf says that the first noun denotes the class F—B (flying minus bird). Indeed, the Hopis use one and the same word to denote insect, airplane, and the pilot, but not birds!



Different languages contain unequal temporal and spatial conceptions. "Present—past—future" appears to us to be the only conception of time possible.

But the Hopis have no tenses, they get along with moods only: the assertive mood ("I report about his arrival") may refer to all those processes that we denote as past or present ("he came" or "he comes"); the suppositional mood ("I express a supposition about his arrival") corresponds to our "he will come", or "he will probably come", and "he, apparently, has come").

We measure time in 'days' and 'years', that is, with the aid of nouns that do not designate any kinds of things at all. The Hopi Indians would never think of doing such a thing: their nouns denote only real things, actual physical bodies. In place of our "two days passed" the

Hopi would say, "this is the third time that it is light". They even dispense with the word 'time' in this context. They say something like "it is lighting up thrice". And a word-by-word translation of "two days have passed" into Hopi would call forth laughter: days have feet, they go in pairs.

Actually, a Hopi would probably not even grasp the meaning, for to him the second day is not a second object but the same process that was simply interrupted and has now been resumed. (We don't say that "a fourth Professor Thomas has come" to mean "Professor Thomas is delivering his fourth lecture today".)

Such unusual divisions of the world, so unlike our own, are found not only in exotic languages like Hopi, but even in Indo-European languages.

The Russian divide their 24-hour period from sunrise to sunrise into 'morning', 'day', 'evening', 'night'. The English have the subdivisions 'morning', 'forenoon', 'afternoon,' 'evening', and 'night'.

In German and English we distinguish between hand and arm. But in Russian both are ambraced by a single word 'рука'. We, in Russian or in English, say 'two eyes', but the Irish regard both eyes as a single organ, a single object. They use the singular to denote the 'organ of sight'. One eye is termed 'half the sight organ'.

The English, Germans and French distinguish between 'fingers' and 'toes', the Russians have one word for both.

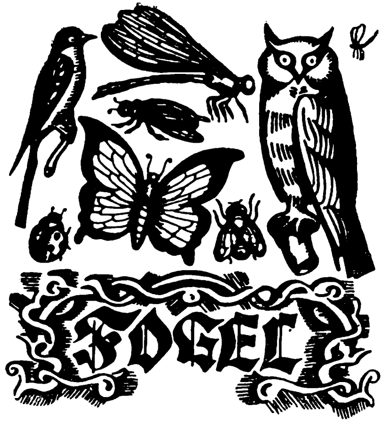
The Whorf Hypothesis Once Again

So the world about us is reflected differently in the various languages. Even such eternal and general conceptions as space and time have received different labels in the various languages. All the more justified are the differences in denoting things, colours, phenomena and properties. These are all irrefutable facts of linguistic science. But facts may be interpreted in a variety of ways.

A person learns his language in early childhood. At a very early age he begins to perceive the world within the framework of his mother tongue. And no matter how rich and diversified the world about him is, the individual will see and grasp only those phenomena for which word labels exist. Our mother tongue analyses the world in its own particular way and imposes such an analysis and world perception (or decomposition, if you like) on all of us. The Germans say, "Die Wortung der Welt" or "wording (word split-up) of the world".

Whorf's idea is this: people live not only in the world of things that surrounds us and in the world of social life but also in the world of one's mother tongue. The surrounding world is constructed in accordance with the "world of the language". And, as he puts it, every language includes, aside from terms, points of view and prejudices against other standpoints.

That is not all by far. Languages undergo change, and this involves changes in the world that surrounds the speakers of that language. To be more precise, the world remains physi-



cally the same, but in human consciousness it has become a different world.

Even a single language decomposes the world differently at different stages in its development. Let us take, by way of illustration, the designations of animals in old German and in modern German. The German word 'Tier' denotes all animals, but at one time it signified only the four-footed beasts in contrast to domesticated animals. The ancient German 'wurm' embraced worms, snakes and dragons and spiders. Modern German has a term for each one now. The ancient German 'fogel' covered all birds (now covered by Vogel), bees, butterflies and even flies. Thus, in the old days, the Germans divided animals into 'wild' and 'domesticated'. The wild animals—according to the mode of locomotion—were subdivided into 'tier' (running), 'fogel' (flying), 'wurm' (crawling) and 'fisch' (swimming).

This was a clear-cut classification, but it did not coincide either with the zoological classification or with the word-division of the animal world in the modern German language. The ancient language provided a very specific kind of picture of the world. And although the actual animal world is the same, the modern German regards it quite differently from his predecessors. This is all due to the structure of the language.

Whorf recalls that he came up against this problem in a field quite removed from linguistics—during his work in the fire insurance company. But the theory of "language, thought, and reality" that he constructed had nothing to do, of course, with fire-fighting, but resulted from an extremely careful study of a variety of languages that are built along lines quite different from the European languages; they analyse the world in utterly exotic ways.

To summarize, then, different languages reflect the world in unlike ways. A person acquires his mother tongue early in life and hence he perceives the world through the prism of his mother tongue from earliest childhood.

Whorf then suggested that we are prisoners of the word. We perceive the laws of our mother tongue in early childhood in an automatic fashion, unconsciously, so to speak. Together with this acquisition of the mother tongue, we just as unconsciously acquire a specific mode of thinking and a hidden "metaphysics".

Languages differ in their grammars; but perhaps a still more striking difference lies in the way languages split up the world into elements by means of words (a sort of lexical decomposition). The Aztecs have one term to cover 'snow', 'cold', 'ice'. The Russian and English languages have special words for each concept. Now the Eskimo language has dozens of designations for snow, describing it in every imaginable way: snow on the ground, snow in the air, snow that has frozen, snow carried by the wind, melting snow, icy snow, etc., etc.—all have specialized words!

There are numerous instances of this kind. Whorf made a great contribution to linguistics through his analysis of the languages of the Indians of North America. Whorf's services in the field of language study are unquestionable. But they are somewhat eclipsed by his intriguing and debatable hypothesis about the interrelations between language, culture and thought.

Language and Culture

Is Whorf's hypothesis true? Is he right when he says that each language has its metaphysics? And suppose Newton spoke and thought in Hopi instead of English, would the picture of the world be different in that physics? (In one of his papers, Whorf said just that!) What role does language play in perceiving and comprehending the surrounding world?

The significance of this problem is obvious. Also obvious is the philosophical stand that one must take to resolve it. For Soviet scholars, this is dialectical materialism, which states that the world precedes consciousness, and things come before words to designate them.

Does language exert any effect on thinking? Apparently it does, but only on the techniques of thought and not on the essence. The essence of thought is a reflection of reality. The purpose of language communication is the transmission of information about reality, the conveying of messages.

A Russian says the distance between Moscow and Leningrad is 640 kilometres. An Englishman would more often say the distance is 400 miles. The approach is somewhat different, the concepts do not coincide (a mile differs from a kilometre). But the reality of the reflection is exactly the same. And the reflection is valid.

Despite the differences in the technical facilities of a language (hence, variations in the mode of thinking), every language is capable of giving a true picture of the surrounding world.

The majority of scientists throughout the world recognize this fact no matter what the philosophical stand they may take. They are forced to conclusions by irrefutable facts.

A few years ago a symposium was held in Chicago to discuss problems associated with Whorf's theory. The conferees included linguists, logicians, psychologists and anthropologists, philosophers and ethnographers. "Language and Culture" was the chief problem discussed at the symposium.

Does language determine the world outlook of human beings? Whorf said it does. Most modern scientists think otherwise.

Psychologists point to the following facts. A child begins to perceive the surrounding world even before it has learned to speak. Man decomposes the surrounding world before there is any "linguistic thinking" going on at all. Later, when the child has learned to speak, he begins to use his language to attach verball labels to his acquired experience. Things come before words and not vice versa.

Here is what the noted American linguist Joseph Greenberg had to say at the Chicago symposium. Suppose two human beings, speaking two different languages, land on the moon. They are in a new environment because the moon is quite unlike the earth. The two spacemen return to the earth and then relate their impressions of our satellite. According to Whorf's view , we should have two utterly different pictures, two different distinct moons (the Russian moon and the English moon, let us say).

Incidentally, to disprove Whorf's theory we do not need to go so far into space. Human history abounds in cases similar to the lunar trip. A thousand years ago, Arab travellers visited northern lands. The nature, customs and habits of the Norman Vikings were quite foreign to the Arabian voyagers, in fact just about as different as the lunar world is to us earthdwellers. What is more, the Arabic language is totally unlike the language of the Normans. Still, the descriptions of the Arab travellers coincide with those given in the ancient language of the Northmen. We can see the same phenomena, events, cities and mountains. Each language paints the world in its own way, but ultimately the message about reality is conveyed properly and correctly.

Modern linguists compare language to a system of geometrical coordinates. To pass from one language to another is like going from one geometrical system of relations to another. The surrounding world depicted in the coordinates of the different languages is the same, but its reflections in the languages differ.

Whorf was right in contending that under certain circumstances language can exert its effect on our thinking (on the mode of thinking, not on the essence of it, we shall add), hence, on occasion, on the very behaviour of people. Recall the fires caused by "empty" petrol tanks. But Whorf forgot about a different and much more important fact, that thinking is affected by reality, by the practical experience of human beings, by life. In the final analysis, it is life and reality that play the dominant role and not language.

Again recall the case of the empty petrol tanks. Such things could happen in any country because 'empty' means 'empty' whether it is the French 'vides', the Russian 'пустые', the German 'lehr' or the Japanese 'kara'. But things like that would hardly occur in an oil-producing town where everyone is keenly aware of the dangers of petrol. The important thing is not the difference between English and Russian or between English and Japanese, but between the language of people who are knowledgeable in a specific field and the language of people who are ignorant of the facts. Fresh experiences expand one's knowledge of the world, enrich thought and generate new concepts and words to correspond to them.

Coming back to the light spectrum that was mentioned earlier, we see that the Russian language distinguishes two shades of blue using separate word ('голубой' and 'синий'). English and German, for instance, have to use word combinations in this case: light-blue, dark-blue, hellblau and dunkelblau.

The world is infinitely diversified and the number of words in a language is limtied. Quite naturally, when reducing the infinite diversity of things in the world (say the tremendous variety of colour shades) to a small number of concepts (in our instance, a small number of adjectives describing colour), we are able to disregard vast segments of inessentials and concentrate on certain dominant features. The painter dealing with colours finds it necessary to distinguish between very fine shades and so gives them diversified names, but in ordinary life we usually find that just a few adjectives suffice to describe our daily needs. For a quick picture without details the term 'dog' is enough, while the dog lover prefers to specify whether it is a bulldog, setter, pointer or any number of other types. We tend to disregard the finer and subtler points of any matter as long as they are not required. But as soon as the need arises for more detail, language is capable of finding means to express the extra shades of meaning. The experience of our life leads to refinements in thinking and in our language.

However, it may happen that a demand arises before any experience has been acquired, before life has had time to teach us to classify a definite set of objects. For some time we remain ignorant of the subject. That is when mistakes are made: faulty thinking due to ignorance of the matter at hand may lead to accidents. Again we are reminded of the pertrol-tank fires. If our thinking—either because of faulty language tools or for any other reason—gives a distorted picture of the world, the practical affairs of life will sooner or later rectify the picture.

There are, of course, fields of reality where an experimental verification of our conceptions within a short period of time is quite out of the question. For thousands of years man has been trying to get a proper picture of the structure of the universe. Or take the question of the rational and purposeful arrangement of affairs in living nature. What lies behind inheritance? Why are children like their parents? What underlies the psychical workings of the human brain?

It is only just recently that science has actually approached a solution to these problems. These areas have always been a fertile ground for imaginings and fantasy, and that is why mythology, superstition, mysticism and religion still have firm roots in the minds of men. Here it is, perhaps, that we find language exerting its effects on the thinking of peoples. We say 'perhaps' because the problem does not appear to have been investigated concretely in any way as yet. But as soon as science takes over from religion and the free play of fantasy, the effects of language in this sphere cease to act. Language still retains phrases like "the sun rises", "I was beside myself with rage", though we know that the sun does not move round the earth and that no soul leaves our body.

The effects of language on thinking are more apparent in folklore and literature where imagery is often built up via words with specific gender. A story in one language based on words with contrasting feminine and masculine gender mean nothing in another language. This is a very interesting field that is awaiting its investigators.

Whorf was right in thinking that language can exert effects on thinking, but he was wrong when he said that it occurs at all times. Actually it occurs only in cases where the practical affairs of life have not yet become dominant. The effects of language are limited and are corrected by life experience. That is why Whorf's investigations are highly valued though the hypothesis as a whole is rejected.

A Model of the World

Through the medium of philosophy, science and art we model the world.

But one of the most important 'modelling' systems is the ordinary language of our daily life. Language is the main vehicle for conveying information in human society. But language is capable of more than that, it can model the world or, so to say, pass the world through its specific prism. Therein lies the principal difference between language and technical codes that transmit information impassionately, without modelling it or in any way evaluating it.

Whorf's service lies in the fact that he drew attention to the concrete facts of this "reality of thought" (as Karl Marx termed it) and pointed to the actual role of verbal modelling in a variety of languages.

True, Whorf overemphasized its role as against other tools of inquiry and regarded language as the basic tool. But as often occurs in the history of thought, the most debatable views are the most fruitful, and, as one of the critics of Whorf's theory said, Whorf's mistakes are far more interesting than the trite banalities of more cautious scholars.

In the modern world, language is not the sole channel of inquiry, but language has enormous advantages over other modelling systems.

We master the natural sciences, mathematics and art at an age when we are fully aware of the world about us, while language is acquired in early childhood. Language is common to all. Thanks to language we can discuss phenomena that science has not yet deciphered, we can talk about the fantastic, the imagined, and the totally impossible. Language allows us to transmit information and, what is more, to express our attitude towards the message.

Einstein is reported to have said that no scientist thinks in terms of formulas. Mainly, we think with the aid of language. The other sign systems are auxiliary and insufficient. They are too rigid, too conventional and too immobile compared to our ordinary human language.

Language has yet another tremendous advantage over all other systems of signs. Language is accessible to every person, with its aid we can model the world; this distinguishes it from science and art which require special training.

An outstanding artist can create his own style, his own "language" of art, and a scientist can develop a new system of scientific symbolism or notation in physics, chemistry, mathematics or logic. But not one of them, no matter how brilliant, can substitute for the mother tongue of a people another language, foreign to it or artificially created. In early childhood we acquire language unconsciously, only later in school do we begin to become aware of the laws of language when we learn to read and write.

This universality of language may be utilized—and so put to use—by scientists to solve problems that at first glance appear to be quite out of this world. We are able to penetrate into the depth of the ages to find out how our ancestors lived many centuries and millenia ago by examining and deciphering historical writings and archaeological findings.

But there are periods not covered by any extant writings, never mentioned by Herodotus or Tacitus.

Still there is a vehicle that helps us to learn about our distant ancestors, and not simply reconstruct the types of production and material culture. This vehicle is far more powerful, for it permits us to penetrate into the spiritual life of our forebears, the inner being of people that would seem to have been lost irretrievably. This marvellous tool for resurrecting the past is language.

Reconstructing a Model of the World

As a tool of inquiry and a vehicle for modelling the world, language belongs to all members of society. The peasant and the king, the rich and the poor all employ one and the same language. In a later chapter entitled "The Linguistic Clock" we shall show how changes in language may be utilized as a special peculiar kind of clock to describe events that occurred in the distant past. This is possible only because language is universal and cannot be changed arbitrarily. Vocabulary and the meanings of words undergo the same extremely slow process of change.

Popular magazines often have articles devoted to the "ways of words"—fascinating lexical biographies. The science that deals in word origins is called etymology. It not only helps us to learn when a given word appeared in the language and for what reason, it also gives clues to earlier meanings.

But etymology abounds in problems that are far more exciting and important. The origin of a word, the history of a name characterizes a person just as much as the object he has named.

The science of the biographies of words is capable of revealing the inner world of people that died thousands of years ago, the inner life of man of distant epochs!

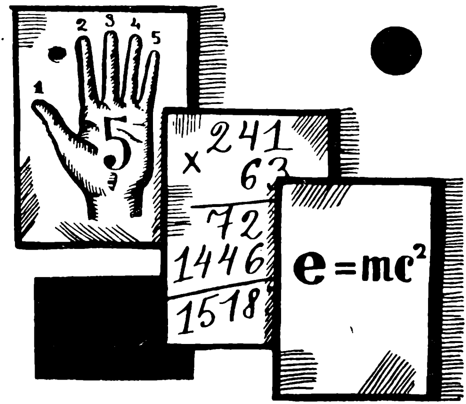
Let us take the word 'bear'. The Greek is 'arctos'. In Latin, bear is 'ursus' (recall Arctic and the constellation Ursus Minor). In the language of ancient India it is 'rkshas'. All these words come from the ancient Indo-European name for bear which sounded like 'rktos'. This Indo-European word is not found in Russian or the other Slavic languages.

Why have such words as 'water', 'nose', 'two', 'three', 'I', 'mother' and others retained their ancient Indo-European forms, and why has the earlier name for bear disappeared?

Linguists give the following explanation. The superstitious ancestors of the Slavs feared calling the bear by its own name. And the terrifying word was replaced by hints, by allegories. The ancient Slavs began to speak of the 'honey-eater', the ancient Germans spoke of it as brown (whence the German name 'Bär' and the English 'bear'). This is how language helps us to learn about the inner world of our distant ancestors, even including their superstitions and fears!

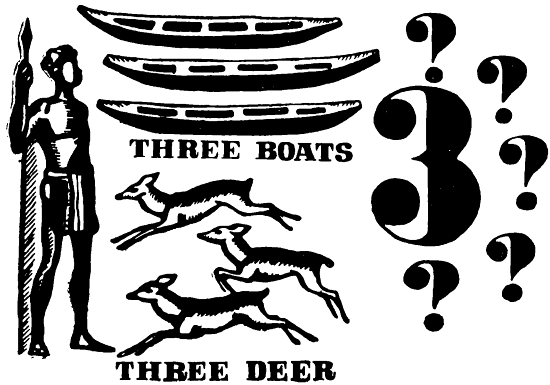
Words and Numbers

The peculiarities of primitive thinking are perhaps most vividly seen in the names of numbers. And not only in the names but in the very mode of counting, the formation of numerals and their usage.



Mathematics is the oldest of the sciences. But this sign system is young indeed compared to ordinary language. Certain researchers claim that many peoples are not capable of counting beyond three. That is so, yetn ot completely so. The languages of certain Australian tribes have only three numerals, 'one', 'two' and 'three'; but this does not mean that the Australian aborigine cannot distinguish three kangaroos from four. It is simply that their language does not have the necessary language labels for numbers greater than three. This is bound up with the economic way of life of the Australians.

Incidentally, there are not so many languages of this kind. Most languages, even among the most primitive, have more labels. The number 'four' is denoted by 'two and two'; the number 'five' by 'two, two and one'.



Dwellers of the Andaman islands count in the following manner: they use words up to four (and then they use their fingers saying "and also"; when the Andaman runs out of fingers he touches his nose.

In the beginning, people counted on their fingers, or used sticks or notches on a tree, and so forth. Thousands of years were needed for man to realize that numbers exist independently of the properties of things. Many centuries passed before numbers were isolated and their properties and functions were set up into an independent sign system.

It is extraordinarily difficult to reconstruct the inner world of people who lived thousands of years ago. But gradually the picture is becoming fuller. Using a few bones, scientists have been able to reconstruct the whole beast of ancient epochs; in language studies too, science is able to reconstruct the world of our prehistoric ancestors.

Ethnolinguistics

The science of language is being more and more correlated with the social sciences in which and for which language exists and develops. To illustrate, the history of the period of Peter the Great gives a clue to the influx of Dutch words into Russian marine terminology; if one disregards the history of geographical discoveries, one cannot understand why the language of the inhabitants of Madagascar has so many Arabic and French words and why the language itself—Malagasy—is related to the languages of Indonesia and Polynesia.

Certain purely linguistic facts cannot be comprehended without involving ethnography and even zoology. For example, the languages of the aborigines of Brazil (Indians) and of certain other tropical countries have no word for 'parrot', though there are many words for the different species of parrots. Why is this? For the simple reason that the word 'parrot' embraces quite different and rather distant zoological species. The word 'parrot' is found only in the languages of peoples who have never had much to do with parrots!

Still more important are the findings of ethnography and sociology. The point is that the vocabulary of a language is in strict accord with the social needs of the people using the language. Inhabitants of oases in the Sahara desert have 60 different words to designate species of palm trees but not a single word to denote snow—quite naturally, for they have never seen snow.

Looking at this matter from the north, we find the Nantsi with 40 different words describing all kinds of snow, hard, soft, loose, sticky, fresh, and so on and on. Snow is all important to these people, for it greatly affects the hunting and pasturing of their reindeer.

The languages of peoples at a low level of social development have few abstract concepts. It would be very hard to write about physics in the language of the Bushmen or in the languages of the aborigines of Australia. It is even practically impossible to set forth the fundamentals of nuclear physics (or even arithmetic) in ordinary Russian or English, for we have to introduce such notions and terms as 'quantum', 'multiply', 'minus', and so forth. The same applies to the languages of the Bushmen or of Australian aborigines. These languages are capable of handling all sciences, provided the new concepts and terms are introduced. Take the Chukchi people who study in Soviet schools and acquire the fundamentals of every science, though their language does not have any roots indicating 'seven', 'eight' or 'nine'!

Languages cannot be divided into primitive and refined. As soon as there is a demand for new concepts, the words are produced. The Mongols had a very intricate system of abstract terminology expressing highly sophisticated concepts for Buddhist religion and philosophy (self-refinement, nirvana, the transmigration of souls, and the like). None of this existed prior to Buddhism. In the eighteenth century, when the genius of Lomonosov found Russian science lacking a number of scientific notions, he boldly introduced the needed abstract terms.

The languages of peoples living in a primitive social structure have few terms of culture. But that only goes to show that primitive culture is poor! On the other hand, all the vitally important aspects of life are fully designated in the language. This refers in particular to terms of kinship, which are very important in the social life of Australian tribes. In the language of the Kurnai tribe there are 20 terms to denote kinship, the Wati-Wati tribe has at least 22, and the Color-Kuridit all of 50 terms!

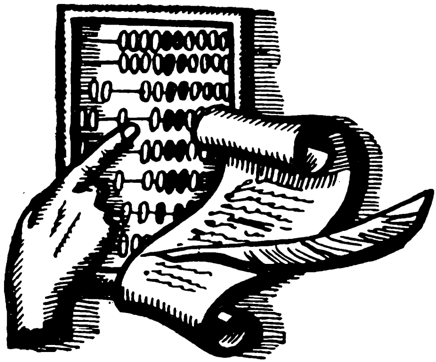
The same applies to designations of age (which again is particularly significant to the Australians); such words far outnumber "age terms" in any European language. Compare our words 'child', 'youth', 'adult', 'old man' with the 15 words found in the Aranda language used to designate various ages of the human male.

If we examine the situation in which the Bushmen or the aborigines of Australia live, we will find that their languages are no less adapted to their specific roles than are our civilized languages. The Aranda language does not have any concepts like 'mountain', 'hill', or 'river'. It would be a hard job to teach geography in this language. But on the other hand, it has a name for every individual mountain, even the tiniest hill. Using this language it would be easier to give directions (and more reliably too) than in Russian or English, say.

The Aranda language has 28 names of snakes and also a generic name for all snakes: 'apma'. But the Australian hardly ever uses it because he always wants to know what kind of snake, whether poisonous or not, edible or not, and so forth.

Different social conditions and pathways of development generate different vocabularies and disparate language structures. But it is meaningless to speak of the "best" or the "worst" languages, in fact just as senseless as stating which is better, a palm tree or a pine tree, Africa or Europe.

The population of our planet has reached 3 thousand million. And there are several thousands of languages. All people are equal, irrespective of sex, culture and race. Likewise, all languages of the globe are of equal value and have equal rights.



Numbers and Language

Karl Marx said that a science attains perfection only when it succeeds in making use of mathematics. Numbers and precise measures are today being applied in biology, archaeology, economics and psychology. Even linguistics has begun to utilize mathematical methods. The science of language is becoming an exact science, as this chapter will show.

Words, Words, Words

Basic to any knowledge of a foreign language are words, the vocabulary of the language.

Developed languages of the world have vocabularies ranging from thousands to hundreds of thousands of words. No one ever learns them all, it is simply impossible. In studying a foreign language the main task is to learn the most important vocabulary, a minimum word stock of greatest usability.

Let us examine vocabulary lists in 16 different textbooks of French used in American schools. One would think that the essential vocabularies of the different books should coincide at least to 70-80 per cent if not completely. (Quite obviously the basic core of the vocabulary words like 'I', 'you', 'mother', 'eat', 'sleep' have to coincide in all textbooks.)

The remarkable thing, however, is that only 2 per cent (!) of the words were found to be common to all books. Sixteen textbooks containing 6,000 different French words had only 134 words in common.

Now suppose two people who had studied the language by two different books wanted to converse in French. Would they be able to understand each other? Hardly, 134 common words are too few to form a basis for conversation.

But perhaps this is simply a curio. Unfortunately it is not. Ten textbooks of Spanish containing a vocabulary of 4,500 words had only 249 lexical items in common. Twenty-six textbooks of the Spanish language were found to contain 13,000 different words. Some basic word list indeed for anyone undertaking the study of Spanish!

Then how is one to find a list of the most needed words and compile a basic vocabulary free from the whim of the compiler?

The first ten or so common words are easy enough to find (the numerals from one to ten, the pronouns, and nouns like 'mother', 'fire', verbs like 'to sleep', 'to be'). But as soon as the list reaches several hundred words—and this is obviously necessary—different compilers bring in different words. One believes that the verb 'to create' should be included, while another one claims that 'do' will be quite sufficient, and in place of 'create' decides to include the verb 'hunt', which was missed by the compiler of the first vocabulary list. Hundreds of examples could be cited to illustrate this point. And the result is that 16 textbooks have only 134 words in common, and 26 books contain the appalling total of 13,000 different words, which is no longer a basic word list but a real and active vocabulary of any language. To demand that beginners master such a mass of words is obviously absurd.

Where can we find an objective criterion for the inclusion of words? What words are to be considered of high frequency and hence to be included in a minimum vocabulary list and which must be discarded?

Mathematics is able to help linguistics in this respect. To be precise, mathematical linguistics and precision methods of language study.

Frequency Dictionaries

What is a high-frequency word? A word which is frequently encountered in speech and writings. The more often it is encountered, the higher the frequency of occurrence.

Quite naturally, much will depend on the text that we take. For example, in Pushkin's story *The Captain's Daughter* the word's крепость' ('fortress') is encountered 98 times, since the scene is laid in a fortress. If we took any other text of comparable length (roughly 30,000 words), the word 'fortress' would occur once or twice at the most.

There is a problem: to avoid words characteristic only of a given book and hardly at all encountered in other writings. The best way, naturally, is to take a large number of sources and then count the number of times different words come up and compare the results. If they coincide, the word is a high-frequency element (or, contrariwise, low-frequency). If there is low coincidence, this will mean that one of the texts had a higher supply due to the type of plot or to some penchant of the author for this particular word.

Thus, 'fortress' which we find so often in *The Captain's Daughter* occurs in only 14 sources out of 133. Which all goes to show that the word is a low-frequency one and should not be included in any list of frequently used words of the Russian language. In a frequency word list based on 133 texts of different authors it is found only in the fifth thòusand, whereas if we were to confine ourselves to Pushkin's story alone we would find it among the first twenty most-used words of the language!

In another story *(The History of Pugachov)* it has an even higher frequency—134 times! Again natural enough if we recall that Pugachov was attacking fortresses almost all the time. Now if we examine Pushkin's writings outside of these two stories, we find that he used 'fortress' only once (in *Dubrovsky)* and only twice in all his poetry!

However, it is not only the quantity of different texts that can rid our calculations of the fortuitous. No less important is it to take large texts. The longer the texts, the more assurance we have of eliminating accidental factors that might affect the precision of our computations.

A large number of frequency dictionaries have been compiled in English, German, Russian, Polish, Czech, Portugese, Spanish, and other languages.

The Spanish dictionary compiled by García Hoz was based on computations of texts containing 400,000 words. The Czech dictionary was based on 200,000 words; the Polish dictionary, on 7,000,000 words; the French, on 1,500,000 words, the German, on 11,000,000 words, and, finally, one of the English dictionaries is based on a treatment of texts containing a total of 18,000,000 words.

Frequency dictionaries are usually in the form of lists of words arranged with words of highest frequency placed first. The largest of all the frequency dictionaries is *The Dictionary of the English Language* by Thorndike and Lodge, which gives data on the usage of 30,000 different words encountered four or more times in diverse texts totalling 18,000,000 words.

A frequency dictionary of the Russian language has been compiled on the basis of an analysis of a million words taken from 133 different texts. It contains 5,230 different Russian words that are encountered 13 or more times per million words.

Text-Forming Capacity

We see that statistics enables us to choose the most needed and most frequently occurring words.

How many words do we need in our minimum vocabulary? The teacher has a frequency dictionary of the English language with 30,000 of the most used words. It is quite clear that not all these words need be included in the basic list. But how many should be included? One thousand, two thousand or perhaps five thousand?

Statistics will again help us to escape an arbitrary decision and objectively determine the most needed words of the language. But arbitrary rulings again come to the fore when we attempt to determine the total number of such words!

Again mathematics can save the situation.

In Pushkin's *The Captain's Daughter* the pronoun 'я' ('I') is encountered 1,160 times per 29,000 words of text, which is an average of once every 25 words. The Russian preposition 'в' 724 times, or once every forty words. Generally, form-words make up about 35.3 per cent of an author's usage and 45.9 per cent of the characters' speech (that is, speech and writing), according to the findings of the Soviet linguist Yu. Markov.

There are also monopoly words in the notional vocabulary and not only among form-words. Tens of thousands of rare words make up only an insignificant portion of any text; in the main any text in any language is composed of a small number of words of extremely high frequency. Which means that different words have different text-forming capacities.

Seven hundred thirty-six high-frequency English words make up 75 per cent of any text. This means that three-fourths of all the words of a text will be familiar if we begin reading with a knowledge of only 736 common words.

A thousand of the most common words cover 80.5 per cent of an English text, 83.5 per cent of a French text, 81 per cent of a Spanish text. The data are close, as we can see. Two thousand of the most common English words build 86 per cent of a text, 3,000 about 90 per cent, 5,000, 93.5 per cent. What this signifies is that if we know 5,000 of the most common words, we will know 281 out of every 300 words of text (about one page!) and will fail to recognize only 19 words! A foreigner studying English would then be able to read rather freely any English text.

True, there still remains 6.5 per cent not covered by high-frequency words. Perhaps we should extend the list and attempt to embrace the remaining 6.5 per cent.

It is easy to calculate (and the calculations have already been made) that it would not be wise to expand the vocabulary. Even if we doubled the list and learned 10,000 words instead of 5,000 (always keeping to words of highest frequency!), then in place of 93.5 per cent we would embrace only 96.4 per cent of the text. We double the vocabulary but gain a mere 2.9 per cent.

This is not a peculiar property of the English language alone. All languages exhibit the same pattern. Three hundred of the most commonly used words make up 65 per cent of all the words of any text; 500 words make up about 70 per cent, a thousand words cover 80 per cent. An increase of another thousand words yields only a 6 per cent increase in the understanding of a text (English, 6.1 per cent, French, 5.9 per cent, and German, 5.6 per cent). Adding another thousand words (the third thousand) covers only about 3 per cent more, a fourth thousand only 2 per cent more, and the fifth thousand permits adding a bit over 1 per cent of the text!

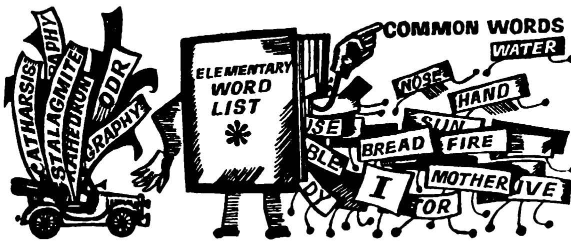
To summarize, then, the mathematician gives the foreign language teacher not only a list of words of highest frequency, but also determines the extent of the minimum word list. It then remains for language-teaching experts to reappraise the findings and determine just how many of the most frequently used words should be included in any textbook for the first month of study, and also the rate at which they should be absorbed, the type of text best suited for the initial stage of reading and other relevant problems that lie outside the sphere of mathematics. Incidentally, mathematics can be of help in purely methodological questions, say, in the teaching of speech habits.

A Basic Word List for Speaking

Knowing a language means above all the ability to speak. To achieve free-flowing correct speech in a foreign language requires overcoming numerous obstacles: getting away from habitual patterns, mastering the phonemes or building-blocks of foreign-language sounds, mastering grammatical constructions and building proper sentence structures, to mention but a few. But for speech, as in reading, the first thing is to acquire a vocabulary. Speaking even more insistently demands this because the words have to be at the tip of our tongue, whereas in reading there is always the possibility of looking up an unknown word in a dictionary.

Is it possible to compile a dictionary of speech? Definitely. And we can assume from the very start that most of the bookish words will remain in this new list. Actually, it may readily be conjectured that all one needs in order to compile a basic list of words for speaking is to eliminate the obviously bookish words from the starting list of frequency dictionaries.

This idea was verified in the United States in an experiment in which 607 speeches of 274 college students were taped (a total of about 300,000 words). The recording was done on the sly so that the students were not prepared and their pronouncements did not contain any bookish, oratorical or other elements alien to everyday speech. Here is how the book words scored.



Only five such words remained in the first five hundred of the most common words; 35 in the first thousand, 178 in the second thousand, 1,500 in 5,000. All this goes to show that even in a most rigorous sifting of bookish words, a minimum word-list will still contain about two-thirds of the reading vocabulary.

Is this enough? Can a person using 3,500 high-frequency words carry on a free-flowing conversation in English?

Calculations show that even 2,000 words are quite sufficient. The text-forming ability of the most common words of speech is greater than in written texts. A thousand of the first category of words covers nearly 90 per cent (in writing it was only 80 per cent, as you will recall). Two thousand common words of speech embrace over 95 per cent of any text, which is more than 5,000 words could do for written material! This means that the first 2,000 words of speech function better than the first 5,000 of the most common words of literature. Michael West came to the conclusion, after much experimentation, that for English a speaking vocabulary can be built up consisting of the following numbers of words.

A total of 450 words is needed for primitive story-telling at the fairy-story level. A detailed recounting of such stories requires 750 words: adventure stories demand a knowledge of 1,400 words, and with 3,000 words it is possible to describe the events of any piece of literature in great detail.

A Writer's Vocabulary

Dictionaries of writers' vocabularies began to appear a long time ago. They include all the different words that a given writer used in his works. The more words he used, the richer his vocabulary.

The English poet Milton used about 8,000 different words in his works. Dante in *Divine Comedy* used 5,860 words; Ariosto used 8,474 different words in *Orlando Furioso*. Horace, the Roman poet, 6,084 words in his works; Homer's poems contain about 9,000 words, Shakespeare used 15,000 different words (other investigators even put the figure at 24,000!). Pushkin's works contain 21,000 different words.

It is interesting to compare the word usage of ordinary people with the vocabularies of great writers. Psychologists claim that a child uses 3,600 different words; a 14-year-old has a vocabulary of 9,000 words, and an adult uses anywhere from 11,700 to 13,500 words.

True, when we say that Dante used 5,860 different words (of which 1,615 were proper nouns, including geographical names) in his

Divine Comedy, this does not in the least mean that Dante did not know any more words than that. Actually the great Florentine knew tens of thousands of words and most likely used them, but in this particular poem he made use of only a portion of his vocabulary; this portion is called the "vocabulary of Dante".

The vocabulary of a writer can easily be converted into a frequency dictionary that indicates how often a given word is used by the writer. True the work is much more involved than when compiling a simple vocabulary list of an author.

Soviet publishers recently put out a four-volume *Dictionary of Pushkin's Language,* which is at the same time a frequency dictionary as well. The information here includes the frequency of usage of every single word in Pushkin's *Complete Works,* the place it occurs, the sense in which the word is used and its grammatical forms.

The *Complete Works* of Pushkin contain about 600,000 words, 21,200 different words. Only 720 words occur over 100 times.

Of the 600,000 words of Pushkin's texts, 6,440 different words occur once, 2,880 words occur twice, 1,800 occur three times. These findings are an irrefutable and exact indication of the rich and diverse vocabulary Pushkin made use of.

The compilation of this monumental work took a large team of specialists a very long time—after all, over half a million words had to be investigated!

At the present time, electronic computers have come to the aid of language experts executing in hours and even minutes the computational work of what formerly took years. Computing machines are also useful in the compilation of frequency dictionaries. For instance, the dictionary of the Russian language that we mentioned earlier was compiled with the aid of a computer.

The vocabulary of a sufficiently long text, whether a novel or play or scientific book or official document, has a specific structure. We intuitively realize that the style and vocabulary of *The Captain's Daughter* is quite different from *An Introduction to Mathematical Analysis.* Frequency dictionaries and dictionaries of a writer's language enable us to express these differences numerically, as elements of objective proof.

Such exact statements are made not only by frequency dictionaries. Counts of the average number of words in a sentence likewise permit characterizing the style of a writer by means of numbers.

"For instance, we can say that Alexei Tolstoi had a preference for long sentences, while Kuprin liked shorter ones. We can even go further and say that in Tolstoi's *Sisters* the mean word counter per sentence is equal to 11.9, while in Kuprin's *Duel* the average number of words per sentence was 9.5. Exact figures are always more cogent. It is quite obvious that the latter statement bears a greater load of proof," says the Soviet mathematician R. L. Dobrushin.

Numbers at the Service of Linguistics

Originally, numbers and numerical methods found more applications in the solution of problems in engineering, communications, psychiatry. For example, what is the best arrangement of letters on the keyboard of a typewriter? Naturally, high-frequency letters should be in more convenient locations and the rarer ones on the periphery. Only calculations can determine the frequency of letters.

Let us take another instance of "applied linguistics"—the transmission of speech or writing (letters, telegrams, telephone conversations). As the Soviet mathematician R. L. Dobrushin, writing in his article "Mathematical Methods in Linguistics", says: "The rapid refinements of communications technology, the expansion of information transmission, the 'crisis of the ether' in which information transmitted electromagnetically has nearly reached the saturation point—all these factors pose acutely the problem of creating more economical methods of transmitting information."

What is the most convenient way of coding words and letters into electrical signals? In what way can we shorten texts by eliminating various portions that do not convey information and are actually redundant? Only exact numerical methods of language study can give the answers we need.

High-frequency words are still more needed by teachers of foreign languages. We have already related the role of frequency dictionaries as an aid in compiling basic word lists and the role of statistics in determining the most rational volume of such a basic vocabulary.

In recent years, fresh applications of linguistics have called for exact measures and numbers for describing language factors. With the advent of cybernetics and computerization, we have come up against the problem of man-machine relations; machines understand only the rigorously formalized, unambiguous language of numbers and logical commands. That is precisely the language into which we have to translate our ordinary human language.

Machine translation from one language into another, the machine storage of information, machine translation from speech to writing, and finally the man-machine conversational link (spoken input and output of information in computers) are all problems which a few decades ago belonged to science fiction and which today are realistic and extremely important problems of science. And the solutions lie in the marriage of linguistics of mathematics.

Numbers and precise measures help not only in practical problems of language. They are also needed in handling linguistic theory. One problem—to illustrate—is determining the degree of borrowings of one language from the vocabulary of another one.

Calculations, for example, have shown that in Albanian only 430 words out of 5,140 come from original stock, the rest are borrowings from other languages. The Korean language has borrowed up to 75 per cent of its vocabulary from Chinese. Modern English has borrowed from the French, Latin and other Romance languages up to 55-75 per cent of all its words.

Of course, a large number of borrowings in a language does not in the least mean that the language is in any respect inferior. All languages of the world are equal, and any one of them is capable of expressing everything found in the life of the people. The most common borrowings have to do with elements of the everyday life, culture, and work that had previously been lacking in the life of a given people. Naturally, the new object comes into the language with its name.

Mathematics, as we see, is a great aid to the study of lexicology—the science of word study. But precise methods have helped in many other ways, too, such as will be described in the next section.

Numbers and Cases

The noted Danish linguist Louis Oehlsmlev has calculated that theoretically a language can have up to 216 different cases. Of the living languages in the world today, one in Daghestan—Tabasaransk—has 52 cases!

Russian, English, Latin, ancient Greek, French, Persian, the Indian languages and many others have a common origin: an extremely ancient family of Indo-European languages that had eight cases: nominative, genitive, dative, accusative, instrumental, prepositional (locative), ablative, and vocative.

Not one of the modern languages of Europe or Asia has retained all the eight ancient cases. Their number has invariably decreased through the process of merging. In ancient Greek, the instrumental, prepositional and dative merged into a single (dative) case. In Latin, the ablative, instrumental and prepositional cases merged into a single ablative case. The Spanish, Italian and French languages—all descendants of Latin—gradually gave up modifications as to case altogether.

The Old Slav language had seven cases (out of the original eight): the genitive and ablative merged into a single genitive case. Modern Russian has dropped the vocative case.

Is there a process of disintegration of the case system at the present time? Are some cases more frequently used than others (say, in Russian)? And is there any preference in the use of the various cases in different spheres of writing: scientific, fiction, social and economic?

Answers to these questions are found in an interesting study carried out by the Soviet linguist V. A. Nikonov in *The Statistics of Cases of the Russian Language*. Here are some of the conclusions that he has come to.

In modern scientific, official and political usage, the genitive case holds first place: it handles from 36 to 46 per cent of the work of all the cases. Second place goes to the nominative case (from 20 to 25 per cent), the other four cases taken together provide 35 to 40 per cent, which is less than the genitive case alone.

Quite a different picture of case usage is found in fiction. Here the nominative is dominant (over 33 per cent), second place going to the accusative (21-23 per cent); and genitive is in third place with from 16 to 18 per cent. In the spoken language, the nominative case is used about as often as all the other cases together (close to 50 per cent). Second place is firmly occupied by the accusative case, while the genitive hardly gets beyond the modest 11-to-16 per cent bracket (recall the constant 36-to-46 per cent usage in official prose). Nikonov concludes that the various spheres of the language are by no means indifferent to cases. Scientific prose prefers the genitive and neglects the nominative, while the spoken language does just the opposite. These preferences—and this is most important—are not whims but regularities, which apparently are very significant, seeing that they are so stable. We have what might be called "case spectra" of different types of speech and writing.

Now the question arises: Were these case spectra always as we find them today? Again, numerical methods will help solve the problem.

The most pronounced changes in the case spectrum are found in political and scientific writing. During the 17th and the beginning of the 18th centuries, roughly one-fifth of all case usage (20 to 23 per cent) centred on the accusative case.

During the reign of Peter the Great, a radical change occurred. In the 18th century the usage curve of the accusative case exhibited a series of ups and downs. During the 19th century the level stabilized at about 13 to 18 per cent, and hasn't changed since.

Changes, though not so fundamental, took place in the spectrum of the dative case. In modern scientific writing it accounts for 3 to 6 per cent of all case usage, occupying last place. A hundred years ago it was functioning twice as intensively (5 to 7 per cent), two hundred years ago it was at 10 per cent, and three hundred years back, one-fourth of all case usage fell to the dative case! Imagine, in this linguistically insignificant span of time (300 years) the use of two cases (accusative and dative) has fallen off drastically. And just as steadily and triumphantly has the genitive case taken over, as will be evident from the following data.

During the second half of the 17th century, the genitive case was doing 19 per cent of all the work. In Lomonosov's writings it functioned at just under 30 per cent. Last century, it was operating at about 35 per cent. This century it is doing 46 per cent (nearly one half) of all the work of all the cases! Nikonov states that the genitive case has risen fantastically in recent times in scientific and political writing, one of the reasons being the precise terse style that may be fashioned with its aid. Without extensive use of the genitive case, the Russian language would be overloaded with circumlocutions and heavy, unwieldy constructions.

It is interesting to note that these shifts in the case spectra of scientific language have hardly at all touched on the ordinary everyday language of the people, which fails to reveal either a steady fall in the use of the dative and accusative cases or any such increase in the use of the genitive case.

Karl Marx in his time noted the peculiar stability of grammatical forms of the spoken language, and the statistics of case usage corroborates this fact completely.

In the spoken language we can rely on intonation and gestures and thus dispense with the precision and unambiguousness that the genitive case imparts to scientific writing. That explains why the case spectrum of the spoken language has not changed since the time of Peter the Great though scientific writing has exhibited fundamental changes in the case usage.

Mathematical Linguistics

Science is sometimes compared with a river which has two sources, theory and practice—

the purely internal problems of the sciences and the practical applications of theory to life and society.

Every science strives towards precision, exact measurements and numbers. During the first stages of its development a science accumulates facts and takes pains to describe them accurately. Such facts are then accounted for by a theory, which progressively develops proofs with the aid of measurements and numbers. The descriptive science changes into an exact science. Marx, the founder of the science of human society, and Pavlov, the creator of the science of human thought processes, and many other great minds in the past dreamed of a time when the humanities and social sciences would become exact disciplines.

It is not merely the development of the science itself that makes for precision. The tasks of society and practical affairs demand exact measurements and numerical methods. This is particularly evident in computer technology where calculating machines handle in hours and even minutes intricate scientific problems and calculations that formerly took years of the most painstaking work. Today logic and psychology, economics and physiology are becoming exact sciences. So is the science of language, linguistics.

The reason is twofold: theoretical (internal) and practical (external). Over many long years, linguistics has amassed a fantastic quantity of material. As the celebrated French scholar Marcel Cohen wrote in 1949, to continue to ignore numbers when considering linguistic phenomena would be tantamount to putting a brake on the development of linguistics.

But aside from this inner need for exactitude, numerous practical problems are clamouring for solution. These include such old and traditional aspects as the teaching of foreign languages, the development of economical codes, and such new and almost science-fiction problems as conversing with machines in human language and machine translation from one language to another. Without resorting to numbers we cannot solve such problems.

That is why mathematics is penetrating ever deeper into language studies, and mathematical linguistics is reaching out into new realms in the field of language research.

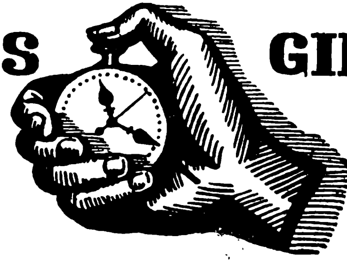
True, the term "mathematical linguistics" cannot be compared with the allied term "mathematical physics", which has been in circulation for a very long time. The latter is a fundamental division of mathematics created specially for solving theoretical and applied problems of physics. From the standpoint of methods, mathematical physics is no less complicated than any other division of mathematics.

In mathematical linguistics only the first steps have been taken. There is no special mathematical apparatus here, conventional procedures are used.

Elementary mathematics has found applications not only in linguistics but in the other sciences (biology, psychology, aesthetics) as well. But that is characteristic of the first stage only. Later, there will be set up a special mathematical apparatus to serve specifically the needs of the given branch of science.

In physics this was achieved by the genius of Newton, Boltzmann and other scientists many years ago. In economics, this is occurring before our very eyes, a special mathematical machinery is in the making. Linear programming, queueing theory and game theory are making possible the solution of specifically economic problems with the aid of numbers. And so the time will come when linguistics will have developed a special mathematical apparatus for the solution of its problems, and this is because mathematics and linguistics are advancing hand in hand with the development of modern science at large. There was a time when mathematics found its source of inspiration in the sciences of dead nature, astronomy and physics. Today its problem is much more involved: to create special tools for the solution of problems that biology, linguistics and psychology—to name but a few new disciplines—have brought to the fore.

LASS



GIRL

The Linguistic Clock

Language not only as a peculiar system of signs, a code studied by means of information theory and mathematical statistics, not only as a tool of inquiry and a reflection of the surrounding world, language as a clock to measure the rate of change of words. Language the timepiece is what we will discuss in this chapter.

In the Search for Exact Time

For the European who lived some two hundred years back, the history of mankind began with Homer and the legends of the Bible. Only in the 19th century did a revolution of thinking take place in the natural sciences and the various sciences of man, including anthropology, linguistics, ethnography, archaeology and the study of the arts. Man's history was broadened beyond comprehension. The few thousands of years the Bible allotted to mankind gave way to 70,000, 700,000 and then a million. And the most recent findings of the English anthropologist Louis S. B. Leakey have extended the time span of humankind on this earth to one and a half million years!

However, the scientific revolution of the 19th century did much more than lengthen the human span of life. Facts about languages, peoples and cultures and events ceased to be merely amusing and instructive tidbits of knowledge. They became scientific facts.

In this twentieth century, knowledge about the human sciences is becoming exact. Linguistics, history, psychology, ethnography, the study of arts, archaeology, are striving to become just as provable and free from the personal opinions of the investigator as physics, astronomy and mathematics. Quite natural, then, that they turn for help to the older members of the family, the exact sciences. Mathematics has come to the aid of psychology and linguistics, astronomy is helping history, and physics and chemistry have teamed up with archaeology.

One of the most important problems of the sciences of man is that of time. It is just as important to know when a given event took place as it is to know where and why it occurred. Many subtle, ingenious methods have been developed in recent years that permit dating past events with great exactitude.

How do we know that such and such a thing happened in the year 1781 B. C. or that something else occurred about five thousand years ago? Such statements are often made, you know. Who can prove that the event actually did take place in 1781 B. C. and not before or after? There are no living witnesses, naturally, but there are numerous 'dead' ones! First of all, we have written documents. The chronicler has written that in the year 6351 after the creation of the world, a fire occurred in the town.... With this coordinate of time (only some 7,000 years ago) we can compare it with the modern time reference system and convert the Biblical clock into a modern timepiece.

But what if the reference time point is not known? What can we do in that case? Here too scientists have found a way out. Nearly all ancient manuscripts (though here the word "manuscript" is hardly exact, for we have "books of stone", inscriptions hewn on the walls of Egyptian temples, and "books of clay"—the Sumerian, Babylonian and Assyrian clay tablets, and many other written monuments of history) mention mysterious and magnificent "celestial signs" which indicate solar and lunar eclipses, the appearance of bright comets and showers of falling stars. Astronomers readily calculate the time of such "signs" (their apparition) thousands of years into the future and just as many thousands of years into the past. The 'signs of the skies' recorded by ancient chroniclers serve historians as excellent coordinates for proving historical time to within a single year. Astronomers are able to date events that occurred two, three, four, five and six thousand years ago with amazing exactitude.

But what if there are no written documents? Which is the case more often than not because the history of mankind began many many thousands of years before the invention of writing. In fact some peoples acquired their written languages only in this century. And some peoples in Africa, Australia, and South America still have no written language. How do we date events that occurred in the pre-alphabet epoch?

If events are not too distant, a simple procedure may be used (one that was applied by a scientist in order to determine the date of appearance of an African tribe on a given territory). The researcher interrogated a large number of people that were well acquainted with the mythology of their people and found that they all stated the same exact number of tribal chieftains from the time of the first mythical leader. Taking the average reign at 25 years and multiplying it by the number of chieftains, the investigator obtained an approximate date —the number of years that had passed since the tribe settled in the given area.

However, this method is not very exact and is applicable only to brief (historically speaking) spans of time of one hundred to three hundred years.

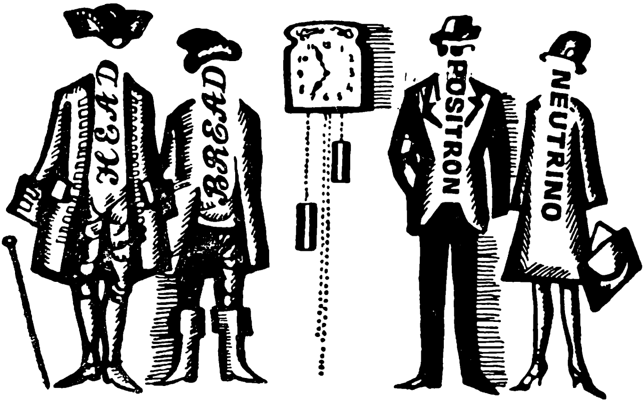
Where myths and legends served earlier scholars as a shaky foundation for conjectures of time spans, research workers today have elaborated a number of objective techniques to find out the time of occurrence of prehistoric happenings. The most reliable method is that of radiocarbon dating.

Every animal and plant contains at all times and places the very same percentage of radioactive carbon. After the death of the organism this unstable isotope begins to decay, and the disintegration proceeds at a constant rate. Thus, the remnants of bone, tree or any other substance containing radioactive carbon becomes a kind of natural clock tirelessly ticking away the seconds of time. Now when archaeologists find organic remains in their diggings, they can use the radioactive clock to give a precise measurement of the date when the event took place. To give an example, researchers have established that a certain bonfire lighted by the hand of man flamed up on the territory of America not 5 to 6 thousand years ago as had earlier been presumed but 30,000 years back. Quite a significant correction, to say the least!

Archaeologists and historians have found a very reliable clock to date events in years long past.

The Method of Glottochronology

But how about events that are not amenable to the method of radioactive dating? Either no organic remains were found or there couldn't be any (say, language changes cannot, quite naturally, be traced with the aid of radioactive carbon). Does that mean that such events cannot be dated? Perhaps there is some natural clock, like archaeological or astronomical timepieces (dates of eclipses, the appearance of comets, and so forth) or the radioactive clock?



A way to objectively and impassionately indicate the time.

In recent years, scientists have found something of this nature. True, 'found' is not the right word, for we have been using it every day and all the time. This timepiece is the vocabulary of our language.

That our language is constantly undergoing change is a familiar fact. True, we do not perceive such changes the way we do the movement of the second hand of a watch, for variations are slow. The change of language with time has been used as the basis of the linguistic clock. And the method of dating by means of language is known as glottochronology (from the Greek 'glotta' meaning tongue and 'chronos' meaning time) or 'lexicostatistics' (the statistics of words).

So language changes with time. The changes however are not uniform. Some portions of the vocabulary of a language change more rapidly than others. Suffice it to recall the new words of the space age, few of which had been invented before the year 1957 when the Soviet Sputnik ushered in the cosmic era.

Very often new words are closely associated with social changes, with science, culture and life. 'Lunik' came in with the first Soviet lunar spacecraft that landed on the moon. Probing into the atom brought forth such terms as 'electron', 'positron', 'neutrino', 'nucleon', 'atomic reactor', 'proton synchrotron', 'electron-volt', and so on and on.

Such language changes occur in step with alterations in the structure of the cultural and social life of the people involved, old words become obsolescent, new words are generated to serve the changing world.

Basic Vocabulary

Still and all, despite the rapid changes in language, fathers understand their children and the latter converse freely with their grandfathers and even great-grandfathers. Why? Because within the ocean of words in any language, there are a set that form a kind of core. A person may not know 'bit', 'nucleon', 'perception' and still be generally educated. But every single person will in his mother tongue know such words as 'water', 'bread', 'land', 'eight', 'work', 'house', 'head', etc.

These absolutely necessary words that every person must know go to form what we may call the core of the language. This core also serves as a source for the generation of new words and lives for a very long time in the language, over many centuries. Variations take place in the core, but with extreme slowness. Hundreds of years pass before any significant change occurs in the core vocabulary of a language.

Why do some words change very quickly and others remain stable for centuries?

Any language as a whole is constantly undergoing change, but the rate is slow, and this particularly concerns the more important words of the language, otherwise people would not be able to understand one another.

We have compared language with a clock in which the motion of the hour hand is not noticeable. Continuing the comparison, we might say that language has two hands: one, the minute hand, that moves in jerks and jumps reacting to the slightest variations in culture and everyday life. The movement of this hand is quite evident to the users of the language. The other hand has to do with the basic vocabulary of the language and moves very slowly. Reforms, wars, scientific discoveries and changes in our day-to-day life, which so rapidly renovate the cultural stratum of our vocabulary, hardly at all affect the basic word stock.

Such ancient words as 'mother', 'sky', 'two', 'fire', have served the English-speaking peoples for centuries, although reforms and revolutions and numberless wars and endless changes have taken place; on the other hand, the development of aircraft, computing machines and space travel have brought thousands of new words into the language.

Still and all, the basic word stock of any language does undergo change; this is because language is the most flexible and subtle instrument of human thought and cannot remain static. Even the most needed words, the very basic vocabulary slowly changes.

This has long been a familiar fact: any observant person can see that his language is in a state of flux. However, the idea of a linguistic clock appeared only in the 1940s, giving rise to such questions as "Does the hour hand of a language move at a uniformly slow rate? Is the rate of change of the basic word stock of a language constant? Do we not discern within the basic vocabulary of a language a peculiar type of 'radioactive decay'? Can we consider the rate of change of the basic stock to be a clock impassionately marking the march of time?"

That was what the American linguist Morris Swadesh suggested, who wrote that the newly devised technique of radiocarbon dating spurred him to investigate the rate of change of the vocabulary of languages.

Following the discovery of the radioactive clock, scientists started the search for a linguistic clock.

The Coefficient of Retainability

The linguistic clock is by no means such a precise and universal instrument as the radioactive clock—language is a product of society and not nature. What is more, the rate of change is very slow and compels one to take rather large segments of time—centuries and millenia.

To find the speed of the hour hand of language, investigators took a sufficiently long period of time: one thousand years, and then compared Modern English with the Anglo-Saxon of the year 950, say. How many basic-stock words has the English language retained from the vocabulary of its Anglo-Saxon progenitor? Statistical computations give the answer.

A list was drawn up of 215 commonly used words comprising the core of the language; calculations then showed that about 190 English words (or 85 per cent) have not undergone any change during one millenium.

Then followed the most interesting thing: Is this figure (85 per cent) indicative of all other languages or does it lack universal significance and is it capable of describing the rate of change of the English language alone? It might be that the Russian language has retained all 215 words of that list, the Chinese language only 10 words and German, none at all. Careful calculations again give us the answer. Scholars possess written monuments that are precisely dated, and they are in possession of "historical" timepieces which may be used to correlate linguistic time. The modern French, Portuguese, Italian, Spanish and Rumanian languages all stem from Latin. What changes have the 215 words of the irreplaceable basic list undergone in these languages? Rumanian has retained 77 per cent, French, 79 per cent, Portuguese, 82 per cent, Italian, 85 per cent and Spanish, 85 per cent.

Amazingly close agreement! Comparisons of early German and modern German show a retention of 78 per cent of the basic word stock. Again, a striking similarity!

But perhaps this constant rate of change is typical only of European languages and other languages have different rates. Other languages were studied, and it was found that 79 per cent of classical Chinese (of 950 A. D.) has been retained in modern Chinese. The Ancient Egyptian language of the Middle Kingdom (2100-1700 B. C.) held on in the Copt language (a descendant of ancient Egyptian) to the extent of 76 per cent in one thousand years. A remarkable coincidence!

Actually, it is not a coincidence because miracles don't happen.

If the most diverse languages change at a constant rate over thousand-year periods—and the results of statistical findings show this to be the case—then one can speak of regularities and not coincidences. The rate of the linguistic clock is a constant quantity!

From the foregoing examples it is evident that the rate of change of the basic vocabulary of different languages (or as scientists say, the "coefficient of retainability" of languages) fluctuates between 76 and 85 per cent per millenium, and the considerable fluctuation (9 per cent) is largely due to the inexact historical dating of certain control monuments. For example, ancient Egyptian which was compared with the Copt language is dated between 2100 and 1700 B.C. which is a time interval of four centuries. Quite natural then that the invariability during one thousand years turned out to be the most extreme instance—76 per cent of the words. Thus, certain fluctuations are due to historical timepieces and not the linguistic clock.

Statistical treatments of large numbers of languages have led researchers to the conclusion that the "mean constant of the rate of change", the mean time of change of languages, is approximately equal to 81 ± 2 per cent per millenium. That is the rate of the hour hand of the linguistic clock!

A Check on the Time

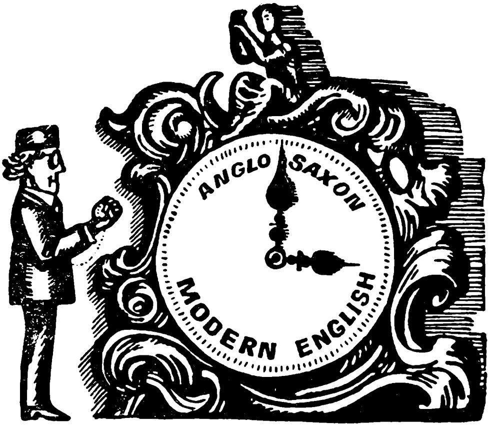
The best way to verify the truth of a theory is to use it in predictions. Recall the discovery of Neptune on the basis of the theory of gravitation, or the discovery of a new island in the Arctic by a researcher sitting in his study.

The linguistic clock was set by the historical clock. Would the timepiece show the exact time without the aid of historical material? That is important, because when we possess only linguistic material and know the rate of change of languages, we can attain extremely precise dating. We have only to compare the progenitor language with the descendant language and calculate the percentage of basic words retained in the latter. If 81 per cent has been retained, that means a period of one thousand years has elapsed; if the percentage is greater or less, it is easy to calculate (using simple mathematical formulas) the time that elapsed during the given period.

The linguistic clock is an important tool for linguists, for with its aid it is possible to date the time of emergence of one language from another (for example, Rumanian, Spanish, and French from the Latin mother tongue). The language expert does not always have historical evidence for determining the date of separation or, as scholars term it, divergence of languages. We know when the Russian, Byelorussian, and Ukrainian languages separated, for scientists have masses of historical and literary monuments.

However, it is much more difficult to figure out when the eastern Slavic languages (Russian, Ukrainian, Byelorussian) separated from the western Slavic languages (Czech, Slovak, Polish) and the southern Slavic tongues (Bulgarian, Serbo-Croatian, and Slovenian). Language scholars can only surmise rather approximately that the separation of the pre-Slavic language into the eastern, western and southern groups occurred somewhere about the sixth century A.D. This is of course an approximate date because the Slavs did not have a written language at that time. That is when the linguistic clock is most needed—when there are no written documents.

Its range of utility is not confined to purely linguistic problems, it is much broader. The linguistic clock can help to restore many facts of the prehistory of mankind. By correlating the readings of the linguistic clock with the readings of archaeological, astronomical and



radioactive timepieces, it is possible to attain a very reliable and exact dating of events that would seem to have left no traces at all.

In conjunction with the findings of history, archaeology, ethnography and linguistic geography, which deals with the dissemination of languages on a global scale, this method enables us to establish the prehistoric settling of peoples and their paths of migration in times immemorial.

Immediately after the discovery of the linguistic clock, it was used—and verified—to illuminate the prehistoric obscurity of the earliest periods of human existence.

We are almost completely ignorant of the early settling of the American continents. Man was born in the Old World; the Americas were settled—so say most scholars—by people from

Asia. Asiatic tribes moved to the new continent one after another. Some believe this migration to have continued for millenia. The last to come to the American continent were the ancestors of the modern Eskimos and Aleuts.

The Eskimos settled the whole northernmost territory of North America from Alaska to the island of Greenland (only a few remained in Asia; on the territory of Chukotka live about one thousand Eskimos). The Aleuts inhabited the islands now called the Aleutian islands and in part settled in Alaska. An Eskimo from Alaska understands an Eskimo from distant Greenland but cannot understand his neighbour Aleut. But ethnographers prove that Eskimos and Aleuts were once a single people. And the science of linguistics states that despite the fact that the Eskimo and Aleut languages now differ, they were once a single language—Eskaleut (from Eskimo+Aleut).

When did the languages separate? Taking a change of 81 per cent of the basic work stock per millenium and allowing for the fact that both languages changed at the same time, they should have diverged by 81 per cent of 81 per cent, or 66 per cent. Researchers have found that the languages diverged (hence, the peoples as well) roughly 2,900 years ago. The linguistic clock was soon double checked by the radioactive clock. Radioactive carbon was found in the most ancient sites of the Aleuts, and careful analyses have demonstrated their age to be equal to three thousand years. The readings of the radioactive clock and the linguistic timepiece coincide!

Having correlated the data of glottochronology with the findings of other sciences, such as ethnography, history, archaeology, scientists now advance the hypothesis of possible pathways of the prehistoric migrations of peoples from Asia to the North American continent. The Eskaleuts were the last to settle on the territory of America; hence, it is logical to seek tribes having similar languages not in the New World but in the Old World.

The time of migration—three thousand years ago—inspires hope that we may be able to find some related languages among peoples still living today. Some believe that the Chukchi, Finns and also certain Indo-European peoples, which include the Slavs, the Germans, Indians, and Persians, may be thus related.

In the Search for a Universal Vocabulary

To repeat: the hour hand of the linguistic clock is steadily and slowly moving with time. Scientists have established the words that are most resistant to change, the most stable ones.

Some concepts change with extreme slowness. The pronouns are an instance ('I', 'you', 'who', 'what'), the numerals from 2 to 10, the names of parts of the body ('ear', 'nose', 'eye', 'tongue', etc.), notions such as 'water', 'sun', 'die', 'name' and many others. The Soviet linguist A. B. Dolgopolsky has calculated that in the history of 150 languages of Europe and Asia the concepts 'I', 'two' have never changed their designations; the notion 'you' changed in only two

of the 150 languages, 'ear' in only 10 languages, and so forth. On the other hand, there are concepts which are much less stable: 'foot' has changed in 35 languages, 'river' in 33 languages and the concept 'boy' has changed its designation a number of times in the history of nearly every language.

To summarize, then, different words have unlike degrees of stability. Now what does this lead to? Let us suppose that some parent language broke up 6,000 years ago into several branch languages. During the first millenium or two, these tongues sloughed off the most unstable words in our check list, leaving as common words only the most stable ones. These will then change very slowly indeed! Thus the constancy of the coefficient of retainability is violated!

But this is not all. The list of 215 words has, it turns out, absorbed a number that depend on the culture of the people, say 'rope', 'spear', 'salt' and the like. Figuratively speaking, the objective course of the hour hand of the linguistic clock is thus affected by the "subjective" motion of the minute hand—words belonging to culture. But that we cannot allow! And so eliminating such words, the American linguist Swadesh cut the list of 215 down to about 100.

Linguists are confronted with the still unresolved problem: what words are to be left in the universal list? What words are essential to all people: Hottentot, Chinese, Russian, Eskimo, German, Australian? And is such a universal vocabulary possible, one applicable to all languages of the world?

Quite obviously, the list of all-human words cannot be allowed to include those associated with geography or climatic zones, with animals or plants.

Such words would be present in some languages and absent in others: besides, they are easily borrowed. For example, 'giraffe' is a modified Arabic word 'zirafah', which means 'nice'. Many European languages have the Australian word 'kangaroo' and the American Indian 'opossum'.

Neither are the numerals fit for our list, for they refer to the sphere of culture. We have already mentioned (in the chapter "People, Things, Words") that many peoples have no numbers for counting beyond two objects.

There is yet another reason why the universal word list cannot include numerals: they can be borrowed and passed from one language to another. In Mexico there are Indian languages that have borrowed numerals from Spanish; the Japanese language has borrowed numerals from the Chinese.

Also to be eliminated from the universal words are those concepts that lack names in certain languages. By this criterion, the notion 'wife' must be excluded for the reason that many languages express it with the word 'woman.'

Finally, even such generally human words as those denoting the parts of the body may be absent in some languages. One of the tribal languages of Western Australia has words describing the upper part of the arm, the forearm, the right hand, the left hand, but there is no generic term for that part of the human body. English has two words ('hand' and 'arm') to convey the Russian 'рука'.

One should not forget yet another important peculiarity that is closely associated with the problem of compiling a universal vocabulary. For some peoples and cultures a word may be vitally important, for others it may be totally absent, obviously of no significance whatsoever. Indians of the Amazon river valley lack such terms as 'clock', 'train', 'station'. The English and Russian languages lack words that are important to Yakutians: 'tuut', skis with a leather sliding surface, 'soboo', meat that is tasteless because the animal was exhausted; 'oloo', which means to winter over on the growing grass (of a horse); 'kharys', the distance between the tips of the extended thumb and middle finger.

Then there are notions which in one tongue are described by a single word, yet require a complex of words in another. For instance, the Shilluk language (Republic of Sudan) has no word for 'apology', but it has a phrase that goes like 'to spit on the ground in front of a man' (which signifies 'to apologize'). The Uduk of Sudan does not say 'to become reconciled' but 'to want to unite grasping fingers'. In the language of the descendants of the Incas—the Quechuas of Peru—the word for 'year' is 'tying up the sun', which is connected with sun worship in the ancient state of the Incas.

So you see it is extremely difficult to separate words that belong to a human being as such from words that are a part of a specific culture, society, nation. Some researchers are doubtful whether it is even possible to build up a universal list of words of humanity. They believe that people live not only in the objective world of things but are also under the effect of their mother tongue, which is the vehicle of communication for the given society. The actual world is thought to be perceived through the prism of one language.

But most scholars believe that language barriers are surmountable inasmuch as the world in which the Russians, Australians, the English, Germans, Indians, Africans and Eskimos live is a unitary world despite all kinds of geographical, linguistic, cultural and historical differences. The surrounding world remains the same in any language, for any person, though a wide variety of linguistic replicas may be obtained from the world of reality.

An Almost Exact Instrument

All the reasons and arguments given above have compelled language experts to cut the universal word list from 215 to 100. But even this reduced list does not satisfy scientists. It may be a long time yet before a final list of universal words is compiled.

The hour hand of the linguistic clock will then indicate rather exact time, and the method of glottochronology will be almost as precise as the technique of radiocarbon dating, and the linguistic clock will indicate time that is practically as accurate as what our natural clocks show.

But all this demands a thorough and detailed investigation of the speed of the minute hand, which will perhaps also be found to be moving uniformly. Or maybe a more sensitive and mobile second hand of our vocabulary will be found. All these problems require further study and investigation by linguists, and also by psychologists, historians, ethnographers, in short by all students of man.

The whole period of the "childhood of man" is still an obscure abyss. It is into this darkness of prehistory that workers of a whole gamut of professions have entered: historians, archaeologists, ethnographers, anthropologists, linguists and students of culture. The combined efforts of linguists and geologists, historians and mathematicians, biologists and chemists have lifted the curtain of millenia and are showing us the life of our distant ancestors.

The senior sciences like physics, astronomy and mathematics, which became exact sciences thousands of years ago, are today helping the junior sciences, the humanities, to take the pathway of objective knowledge. Even such an "inexact" science as linguistics is capable of helping exact methods of research. The linguistic clock, wedded to the radioactive clock, the archaeological clock and the historical clock, is helping scientists to fix prehistoric dates, the dates of the "childhood of mankind", because man wants to know about his past, no matter how long ago it was, no matter how obscurely it is shrouded in the mist of that heroic era when our progenitors fought their fierce battles with nature defending their right to be called MEN.

DWcM LOcT MAch TURn

The Universal Code of Science

In this day and age, highly theoretical and extremely practical problems are very often found to be tightly linked together. This bond of theory and practice is clearly evident in the science of language....

Two Times Two Is Four

Scientists say that our human language is a sign system, the elements of which may be used to express the content of any other system of signs. For example, the signs of traffic signalization can always be translated into ordinary language. A still simpler system is the multiplication table: $2\times2=4$, $2\times3=6$, etc. When learning the multiplication table in school we preferred using words: two times two is four.

A curious fact it is that there are no "favourite" languages among the natural spoken tongues of the earth; none is better or worse than another. That which is said in one language can always be expressed in another.

At an international linguistic conference, a young scientist maintained that the Eskimo language was inferior to English on the ground that the conjunctions 'and' and 'or' were fused in a single word in the Eskimo language. This was countered by the noted linguist Jakobson, who stated that modern American business and scientific writing very often made use of 'and/or', which is exactly what is done in Eskimo! Incidentally, in cases of real necessity, the Eskimo uses 'naligmyng', a conjunction with the meaning of 'either'. True, there are very few words in the Eskimo language capable of expressing concepts of quantum mechanics or cybernetics. But there was a time when English, Russian and French didn't have any of those terms. The Eskimo language is just as capable of generating such terms as any of the "cultivated" tongues.

But though any idea or concept may be expressed in any natural language, it is not always convenient to do so. Let us take the operation with numbers: $2 \times 2 = 4$. In French, Russian, Yakutian, Malagasy this simple truth will sound differently, while the language of numbers is comprehensible to all. What is more, it is extremely convenient. Just try solving problems in which all the numbers are given in words, say add two hundred eighty thousand four hundred fifty-two to four million six thousand fourteen. No easy job in the absence of numerals.

It is still more difficult—using words in place of numbers—to derive or even prove the simplest laws of mathematics. To illustrate, take the formula of the cube of the difference of two numbers: $(a-b)^3=a^3-3a^2b+3ab^2-b^3$. A common school formula. Put into words, it reads: the cube of the difference of two numbers is equal to the difference of the cube of the first number and the tripled product of the square of the first number by the second, to which is added the tripled product of the first number by the square of the second minus the cube of the second number. That is how it reads. But the proof would take up quite a bit of space indeed.

By replacing words with numbers, linguistic signs by the signs of another system, we have at our disposal a universal language, the language of mathematics that surmounts all linguistic barriers and difficulties. More, it is precisely mathematics that has been suggested for the basis of a cosmic language designed for communication with unknown "brothers of reason". Numbers are unambiguous, they express one and only one concept. Compare the number 1 and the word 'one'. The number has a single meaning, the word has several, one of which is "will see you again one day".

Thanks to the system of numbers and mathematical symbols we save enormously in space and time and in thought processes; besides, we have the possibility of reasoning (calculations, proofs and deductions) which would be practically impossible with only the ordinary language at our disposal. Try—without using mathematical symbols—to solve a set of equations (especially with parentheses and brackets)! The language of numbers—perfectly unambiguous and disallowing for any inexactitudes or vacillation—contributed unbelievably to the development of mathematics as a science.

Other sciences followed suit. Logicians, physicists, chemists set up special systems of signs used to express the complex concepts, ideas and facts of science. Engineers use engineering drawings. Geographers use the language of maps. Compare the convenience and exactitude of the map with any verbal description of locations and layouts. Compare the drawings of housing construction with a verbal description of how to build a structure.

The symbols O, R, +, ≡ are familiar from school. Of course, they can be expressed in words: oxygen, resistance, the plus sign, a triple chemical bond. Strictly speaking, they are not simple words but scientific terms. At times, especially in mathematics, scientific symbols do not require any verbal expression. Examples are all manner of computations.

Special scientific symbols first appeared in ancient Egypt, Babylonia, and China. The development of mathematics, astronomy, surveying, navigation, and trade demanded symbols. As knowledge grew and society developed, the number of such signs steadily increased. At the present time there are about eight thousand (!) special scientific symbols. There are scientific papers which consist almost entirely of signs and symbols with hardly a human word in between.

It is impossible to think that anyone without a mathematical background will be able to comprehend a monograph in topology, functional analysis or metamathematics. The contents might be expressed in simple, accessible language, but how terribly combersome and time-consuming.

In every branch of science, symbols have to be systematized and classified. Also, a universal system of scientific symbols is necessary to unify the sign systems of the different sciences. Yet the number of signs increases with every year and makes understanding more difficult even for specialists in a single branch of science.

That is not the only trouble that lies ahead.

A Niagara Falls of Books

By the end of the fifteenth century, half a century after the first printing press appeared, roughly 40,000 books of different titles had been published. This avalanche continued through the centuries increasing in nearly a geometric progression until in the twentieth century the libraries of the world hold in the vicinity of 30,000,000 different books. Thirty million! How many books can a person read in a lifetime? A million? A thousand, a hundred thousand?

Let us try to figure it out. Suppose a person devotes exactly half of every day to reading. We take the average book to be 300 pages long and every page to contain 2,500 letters. Our book-lover will, let us say, read every day for 50 years at the rate of 20 letters a second. We take it he has a tough head and can stand 12 hours a day, retaining almost everything and reading at one and the same level of concentration.

How many books would this miracle reader be able to plough through in half a century? Only 24 thousand, it turns out. What a miserable number compared with the millions of books that have been written. The ordinary person is of course a far cry from our ideal book devourer—he gets through only two or three thousand books.

Nearly half of the books of the Lenin Library in Moscow have never once been ordered or dipped into. Remarkable? Yes, and no, too. The catalogues of the library have 25 million cards with book titles. About 200,000 new books are registered every year.

Easier to Invent It Again

Is there any way out? Of course there is no need to read everything. There are many who live a lifetime reading a dozen books, a hundred books. But scientists have to keep abreast of events in their field and so do a great deal of reading, otherwise they would lag behind.

In the United States, five years were spent in developing a certain device at a cost of 200,000 dollars and a great deal of work. Soon afterwards, it transpired that this very same device had been made in the Soviet Union some years before the first studies had begun in the United States. The Soviet investigation had been published, but American scientists did not know about it.

But how can anyone be expected to read all the publications devoted to chemistry when two suitcases of chemical literature come out every single day of the year. It has been calculated that chemists spend only a third of their time on experimentation, the rest being taken up with locating material and reading the literature.

So that scientific workers should not suffocate in this welter of books and articles, special reference guidebooks, bibliographical surveys, abstract journals, and, finally, whole specialized institutes are put at their disposal. In the Soviet Union, the information service is headed by the All-Union Institute of Scientific and Technical Information (VINITI), which employs a staff of over 2,000 specialists and about 20,000 translators. Over 10,000 scientific workers and engineers write the abstracts (the journal of chemical abstracts alone handles about 100,000 articles a year!).

Even that does not resolve the problem completely. A great deal of time and effort still goes into the search for materials. American researchers have even calculated that if a study is expected to cost less than 10,000 dollars it is cheaper to carry it through than to search the literature. A very unpleasant situation, to say the least.

And the stream of journals, newspapers,

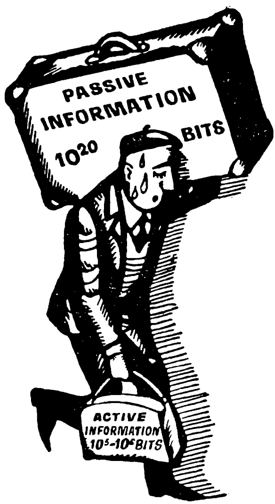
books continues unabated. Libraries are hardly able to keep up with the torrent of publications. Bibliographies have proliferated to the point where we need bibliographies of bibliographies. The present-day terror of scientific thought is the knowledge explosion and the unbridled growth of the literature.

Something had to be devised to harness this ocean of books and see that not a single grain of knowledge that man has been able to acquire is lost. And it was: a logico-information machine, an invention comparable to that of printing.

A Mechanical Memory

Taking advantage of the mathematical theory of information, it is possible to assess a magnitude that would appear to be unmeasurable—knowledge. Not only are we able to estimate this magnitude but we can measure it as well. The unit of measurement is the bit. One bit signifies the amount of information that states which of two equally probable events has occurred. An example is the heads or tails of coin tossing. More about this can be found in "*Languages and Codes*" (p. 41). This information can be transmitted, received, stored and utilized for control of all kinds of processes and machines. You will recall that all these are problems dealt with in information theory.

Just as mathematics is not concerned with the material of the cones, cubes and cylinders it studies (the geometrical properties of these bodies are obviousliy independent of their composi-



tion), so information theory cares nothing for the communication channel used to transmit information (telegraph wires, sound waves, the nervous system and so forth). The recipient of the message is also totally irrelevant; it may be a human being, an animal, or an automatic device; likewise, the information may be stared in the human brain or in the memory department of a machine. For information theory, which is a mathematical discipline, the only thing of importance is the amount of information, the number of bits stored in the human memory, in the memory of an animal or a machine.

How many bits of information can a human being assimilate?

Man's memory is limited, not everything can be retained. During one's lifetime, an enormous quantity of information is examined by means of sight, hearing and touch, but only an insignificant portion of it remains in the memory and is recorded in the brain. Hundreds and thousands of people pass by in a lifetime's acquaintance, and yet how many names remain fixed for instant recall?

Scientists have calculated the number of nerve cells of the brain, called neurons, to be in the vicinity of 10,000 million. If we take it that every neuron stores one bit of information, the capacity of our brain can be put at 10^{10} (10,000 million) bits of information. This is a great deal indeed, roughly the information contained in a library of 10 to 15 thousand books.

Many believe, however, that this is an underestimation. True, the active, operating memory of the human brain stores up only about 10^5 to 10^6 bits of information. This is the material of instant recall. At the other extreme, the passive or subconscious information is estimated at about 10^{20} bits!

The storage components of modern computers are far indeed from such a monstrous capacity. But subconscious information is not something that we can make use of whenever we want to (we very often labour to extract from the recesses of our memory some stubborn fact and only recall it later on and of a sudden). But modern computing machines are quite capable of storing the 10^5 to 10^6 bits that we keep handy for instant recall, and in the near future this range of information will be surpassed by a new generation of machines.

The memory of a machine has enormous advantages. Firstly, it can assimilate information round the clock without in the least tiring. Secondly, it has a high rate of assimilation. Our wonder-reader, you will recall, imbibed information at the rate of 20 letters per second, which is the equivalent of 20 bits of information. This is not the limit, however; psychologists and cybernetics men claim that the human being is capable of receiving and processing up to 40-50 bits of information per second (take another look at the chapter "Languages and Codes"). But not more than 50 bits. This is the maximum human speed and it quickly brings about extreme fatigue. Now the rate of machine assimilation is hundreds of times what the human can do.

The human being can remember information incorrectly; the information can decay in the brain, because the brain is a living structure, while the "iron brain" of the machine records facts reliably and firmly.

And finally—the most important thing—we can always add more memory units to our computer, thus mechanically building up new layers of knowledge, or we can wipe out the memory in toto and obtain a tabula rasa. Nothing of the sort can be done with the human memory.

In Search of a Language

Cybernetics has thus brought to man a reliable and true helper in the form of the computer. The human memory is now supplemented by an electronic memory, the capacity of which can be increased without bounds as it engorges fresh facts and whole books that it can be made to read and digest.

The language of the machine is the language of numbers and precise unambiguous statements. Try translating *Hamlet* or modern poetry into the machine language and you get nothing. The language of poetry is vague and of many meanings, while the computer demands absolute precision and single-valuedness.

Researchers then asked: What is a translation? What can be translated into the machine language and what cannot? The answer was given not by engineers teamed up with mathematicians but by linguists and semioticians, specialists in the theory of signs.

First of all, they said, it is necessary to distinguish between translation and coding. All the letters of a language can be encoded by means of binary digits—zero and unity. Say, 'a' will be 01, 'b', 10, 'c', 11, etc. Any text can be transferred into this code, but it will never be termed a translation. Just as the Morse code—which does just that—does not produce a translation.

A translation—unlike coding—is primarily a conversion of meaning from one medium to another. It is a complete transformation of the system of signs. The only thing that remains is meaning. Coding, on the other hand, leaves intact not only the meaning but also all the words, their order and so on.

There are three kinds of translation. One is a conversion within a single language. Essentially this is simply a replacement of synonymous words and constructions and definitions. Defining dictionaries afford many instances of this nature.

The second type is translation proper—between two languages. The words (or signs) of one language are replaced by words (or signs) of another. Say, the French 'grand' might be translated into English as 'great', 'large', 'significant'.

Finally, the third type of translation is a kind of transmutation or interpretation. The words of a language are replaced by nonverbal signs. Say, the word 'caution' may be depicted by the traffic sign '!', or the phrase "I want to eat" and its meaning may be conveyed by gestures.

Everything expressible in one language can be translated into another. All accessible knowledge is amenable to expression in any language. Physical laws may be expressed in the form of mathematical equations, the description of physical experiments in the form of verbal statements or geometrical drawings.

However, there are many things that do not submit to the unambiguous, rigorous language of science. In fact, the very language of certain sciences is still all too vague to be understandable to machines or even to the specialists themselves.

That is why translating human knowledge into machine language is so far only possible in the field of the exact sciences such as mathematics, logic, physics, chemistry and engineering.

The Language of Words and the Language of Concepts

The basic means of human communication is via the spoken language. But with every year we are obtaining more and more information from other systems of signs. Illustrations in books, advertisements and signs, diagrams, maps, formulas, drawings, road signs, window displays and the like. Scientists make particular use of such nonlinguistic languages for they are more vivid, succinct and unambiguous.

Machines are not yet capable of understanding our human, polysemous language. They require a dry, formalized language, into which scientific knowledge has to be translated. If we succeed in doing this, then logico-information machines will indeed become a well of wisdom capable of digesting scientific articles, books—all kinds of publications—at the rate of a thousand characters a second, picking out new information and recording it firmly for all time.

We have already said that the language of science is the language of concepts. The Russian word 'бензин' is 'gasoline' in the United States, 'petrol' in England, 'essence' in France, and 'Benzine' in Germany, but everywhere the meaning remains unchanged; in fact it would probably best be expressed not verbally but as a structural formula. Thus, machine language has to be a language of concepts too. To every concept there must correspond one and only one symbol, just like in mathematics every magnitude is expressed by one number.

There are infinitely many numbers, yet this boundless multiplicity can be expressed with the aid of ten figures (in the binary number system only two—unity and zero—are needed). The recording of scientific information also demands its figures and its basic, elementary meanings. By combining them we shall be able to express even the most complex concept in the same way that we can write any arbitrarily large number simply by combining numerals. Is the problem solvable?

Long before the advent of information machines, physicists had developed the theory of dimensions where any physical quantity can be expressed by means of certain basic quantities. Take mechanics. Length L, mass M and time T are elementary "numerals" which combine to form expressions such as LMT^{-2} for force, LT^{-1} for velocity, ML^{-3} for density, and so forth. As you see, formulas can be built up out of physical quantities as well as mathematical ones, which means there is some hope for other sciences too! If, naturally, they attain the same level of precision and rigour as physics has.

Computing machines handle mathematical symbols and formulas with ease. Logico-information machines will have to deal with the formulas of sentences consisting of separate "words" or scientific notions.

A Semantic Code

Thus, a logico-information machine requires a special language, the language of meaning: basic concepts are to be used to derive others. Of course, the smaller the number of initial quantities, the better (machines operate best with the binary system of numbers). Just as the basic quantities in physics proliferate, so should the starting concepts multiply here too. This principle underlies the machine languages under development today in many countries.

One of the best known information languages is that developed by the Americans Perry and Kent. Here the basic concepts are defined as semantic factors (by analogy with basic physical quantities, though no multiplication is implied, since concepts do not have numerical values). Among the basic concepts are 'instrument', 'information', 'transmission', 'temperature', 'pressure', and so on. These are the atoms of meaning, the construction material for building up other more complicated concepts. The semantic factors are designated by special code signs. For instance, 'instrument' is coded as 'M-ch' in the machine language of Perry and Kent, 'transmission' as 'T-Rn' and so forth. Now let us take the word 'telephone'. How can it be expressed in the machine 'language of meaning'?

Telephone is a special instrument designed for the transmission of information by means of electricity. Its definition embraces four semantic factors: 'instrument', 'transmission', 'electricity', 'information'. They are respectively denoted by 'M-ch', 'T-Rn' 'L-cT', 'D-cM'. Now, to make the definition of our telephone complete in this language of meaning we have to indicate in what relation these four factors comprising the word 'telephone' stand to each other. You most likely have already noticed the gaps in each of the code symbols. These gaps are filled with what are called relation signs. The telephone is one of a multitude of familiar instruments denoted by the semantic factor 'M-ch'; by convention, the letter A (which means 'there exists', 'is') is used to denote that an entity belongs to the class of instruments. We insert it in the gap and get 'MAch' (there exists an instrument of the class 'M-ch'). The telephone is utilized for the purpose of transmission; we insert the letter U in the semantic factor 'T-Rn' to indicate that the given instrument ('MACh') produces an action; 'TURn' signifies 'action-transmission'. In the third factor 'D-cM' ('information') we insert W, which states that the concept that we have coded operates on what is designated by this semantic factor (in our case it is information).

The telephone operates by means of electricity. 'LQcT' denotes this fact ('L-cT' means 'electricity', the letter Q signifies 'with the aid of').

The result is 'DWcM LQcT MAch TURn'—'an instrument for the transmission of information with the aid of electricity'. That is how the sentence reads in the language of meaning.

This language is no simple theoretical construct. It is the basis for the operation of information machines. First, an ordinary abstract of an article is compiled. Its contents are analysed. The basic elements, the basic semantic factors are then extracted. They are written in the conventional code language and fed in this shape to the memory store of the machine.

The Future

Information machines are not only reservoirs of facts, dispassionate bibliographers that supply researchers with material at lightning speed. More exciting and creative problems are already being attacked. The machine can determine whether the information being fed to it is new or merely a compilation of familiar facts.

This requires comparing the incoming material with the memory stock.

The machine can generate abstracts of books and articles (experiments of this nature have been carried out in the Soviet Union and other countries). When scientists construct a unified terminological code and all the concepts and definitions of the various sciences are rigorously and uniquely defined (like the unitary system of measures, etc.), then information machines will become encyclopedic. The memory stores will have all known theorems, formulas, definitions—in a word, the entire range of human knowledge that can be formulated in a rigorous and unambiguous machine language. A machine of this kind will accept only such information as conveys something new. All rehashing of old ideas, imitations and simple compilations will be straightway rejected by the "all-knowing" machine.

What is more, the information machine is not only able to store up information passively and merely reject overlapping and duplicating material, it can draw fresh conclusions, produce new pieces of information, diagrams, and it can

derive laws by processing information stored in its memory department.

Fantastic! The robot scientist come true! Yet there is nothing out of the ordinary at all. Problems of a creative nature have already been attacked. Recently, the Chinese logician Hao Wang, working in the United States, carried out the following experiment. A computer was fed with the basic axioms of mathematical logic, and in a matter of seconds the machine independently derived and proved a series of theorems that had already been proved in the fundamental work of Russel and Writehead *Principia Mathematica*. But that is not all, the machine formulated a whole range of fresh theorems!

In England and the Soviet Union (at the Institute of Cybernetics, Ukrainian Academy of Sciences), researchers have worked out a programme for a machine that was to prove Euclidean geometry on the basis of the axioms. The machine not only coped with the task but pointed to new lines of proof that had not occurred to mathematicians in the course of two thousand years!

"Mechanization of mental work" is the slogan of this age aimed first at relieving bookkeepers, statisticians, and mathematicians of extended and tiring computations, and then taking over the work of bibliographers, librarians and translators, and finally coming to the aid of creative efforts in general.

Computers Have Their Limits Too

Everything that is amenable to translation into the language of formulas and numbers, that can be formalized, may be handed over to cybernetic devices. And the machines will handle such formalized problems much faster than humans. But is everything translatable into machine language?

Time and again devotees of cybernetics come out with statements that the capacity of the computer is unlimited. Which would seem to be very close to the truth, for the history of cybernetics from its very inception has been a series of negative predictions that have been overcome one after another, starting with

Wiener's conjecture that the computer would probably never be able to translate from one language into another. But we can be sure of one thing: machines will never be able to completely replace human beings, particularly with regard to creativity.

We have already mentioned the fact that the human brain has an information store of the order of 10^5-10^6 bits. For modern technology that is not very much, in fact machines can be designed that operate on a still greater quantity of information. But the machine will never become a human being because simply storing information is not enough, it has to be put to use.

It is quite possible that the cultural facts we carry in our heads may be estimated at something like a million bits of information. That is not the point, however. The gist of the matter lies in the programme with which the machine operates on the information, the rules of operation, the algorithms that are utilized both consciously and subconsciously. So far, a computer operates solely by the method of scanning the whole range of its memory. In Swift's *Gulliver's Travels* there is a description of the Laputan Academy. A professor of this academy invented a method by which the most ignorant person could write books on philosophy, poetry, politics, law, mathematics and theology with a minimum of effort and expense and with no erudition or talent in the least.

The secret of the invention was simplicity itself. The surface of a huge frame was covered with a multitude of wooden boards connected by fine wires and having words pasted on in a variety of tenses, moods and cases. Upon command, forty men turned forty handles thus changing the combinations of words on the frame. As soon as three or four randomly chosen words produced a meaningful combination, scribes immediately recorded them. The procedure was repeated until a manuscript was produced.

That, roughly, is how a computer "creates". By running through all possible combinations of words, the machine will indeed hit upon a meaningful group that might correspond to the words of Byron's "I would I were a careless child". But despite the lightning speed with which a computing machine works, it would take hundreds, thousands of years, perhaps millions upon millions of years!

But that is not all: there are problems which are fundamentally beyond the limits of computing machines. One of them is the problem of formalizing translations of works of art. We have already said that the information language of a machine is a language of concepts. Naturally, translating a piece of poetry results in standardized, stereotyped phrases, generalizations of a sort, and that is all.

Translators into French complain that the beauty of some of Pushkin's poetry when translated into French is totally lost. One can easily imagine what nonsense will result from transferring musical works to a machine language. Translation implies a transference of meaning. The meaning of scientific writing, our ordinary spoken language, or a 'machine language' does not depend on the specific form in which it is written, that is, the code in which it is recorded. But this does hold in the case of works of art. Form and content, the code and the communication, the what and the how are inextricably linked into an integral whole; they cannot be separated or put into another code without distorting the communication. In this case, a translation of meaning is impossible. The only possibility is a creative transposition, a variation on the theme of the original, that is all.

Tchaikovsky did not translate Pushkin's poem *Eugene Onegin* into the language of music, he created an independent piece of music. The drawings of a painter. the stories of a writer are not simply a rehashing of a certain content in the language of graphic art of written characters but a re-creation.

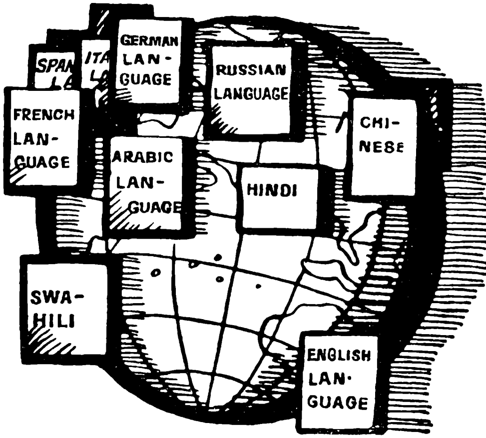
Poets Need Not Worry

The language of science an technology is rigorous, unambiguous. That is the sole reason why it is 'understood' by computers. Colloquial language is complicated, polysemous. It is precisely this polysemy and nonformalized nature that permits us to speak about and even explain events that are unaccountable in terms of science, to speak of happenings of the imagination, fairy-tales and science-fiction doings.

Language is flexible and variable for the simple reason that the real world about us is constantly undergoing change and one "cannot get into the same stream twice". The problem of creating a machine language capable of imbibing the entire wealth of the living languages is many times more complicated than that of establishing an information language. An information language is indifferent to the style of description, the emotional undertones and linguistic subtleties. To take an example, it reacts identically to such a synonymous spectrum as: woman, female, dame, petticoat, skirt. All these words would be recorded by a single character of the meaning-code, which describes an individual of Homo sapiens, female sex,—and that is all!

Machine translation is being investigated by researchers in the United States, the Soviet Union and other countries. They hope to overcome the numerous linguistic and stylistic difficulties and create a machine that will make it possible to translate from one natural language into another just as easily as it handles translations from the language of one science to that of another.

Human society, however, has media that do not submit to translation into the machine language—poetry, music, painting. These will always remain in the domain of man proper.



The Tower of Babel

Can a universal language be created? One that would permit a Russian to converse with a Japanese, a Turk with an inhabitant of Polynesia?

The history of a mediating language goes back thousands of years. What does modern linguistics have to say about a common tongue?

Tales Are False But the Hint Is There

The Bible says that at one time everyone on the earth spoke one language, one tongue. Mankind had a single medium of speech. But the people challenged God: they decided to build a tower to Heaven. And God was alarmed: with one human tongue, man would be able to do what he had set out to. And so he confused the tongues of the builders so that they could not understand each other any more and the tower of Babel was not completed.

Naive though the story is, there is truth in it. When people are united, they are capable of great deeds, and language is a tool with the aid of which unity and understanding can be attained.

Man has—after all—reached the sky without building a tower. Spacemen of the human race have penetrated deep into the cosmos and have ambitious plans for still greater space ventures. But the world still lacks a common language. To reach the skies was easier than to set up a universal language for the inhabitants of the earth.

What can be the reason? As the great philosopher of France Voltaire said, "the difference of languages is one of the greatest misfortunes of being". A resolution of the congress of the First International contains the following words: "A universal language would be a universal boon and would greatly contribute to the unity and brotherhood of the peoples." Why hasn't this problem been solved?

First of all, let us agree on what is to be understood by a 'universal language'. What is it, an obligatory world language or simply an intermediary coexisting with all the national languages?

Our mother tongue that we learn in early childhood is our real home. But we have to leave home in order to speak with foreigners. Is it possible to construct an artificial language for all peoples of the world? And what kind would it be, a supertongue that would swallow up all the little native languages, or would it be of necessity a delicate structure that could never replace our real mother tongue?

Soviet scientists believe that the mother tongue is an inalienable medium of communication with the outside world and with other people. It is intimately bound up with the history of the people and the national culture. For this reason it cannot be supplanted by a 'superlanguage' or by any kind of artificial universal tongue.

Thus, when speaking of a world language, we can only have in view a supplementary intermediate language and no superior structure that would dissolve the mother tongues of the world. It would be an auxiliary language that would not in the least displace the national tongues and would be needed only for communication with foreigners.

Living Lingua Francas

Lingua francas (mediating tongues common to a number of peoples) have been in existence for ages.

Primitive tribes, both those that have long since died out and those still alive today (aborigine Australians, Bushmen, Papuas, etc.) have used sign language (the language of gestures) as a common mode of communication.

Sign language was particularly developed among the Indians of the prairies of North America. Numerous tribes of Sioux, Caddoes, Algonquian and other Indians were in constant hostile or friendly contact and had to have a common mode of expression since their tongues were quite different. And so the sign language developed. It was understood by all Indian tribes from Canada to Texas. And it was capable of describing almost anything from simple questions and answers to negotiations dealing with alliances between tribes, to tales of hunting, and involved mythology and legends and stories.

Quite naturally, this common sign language did not displace the sound languages of the Indians, Australians or other peoples. It served solely as an auxiliary means of communication. At higher stages of development we see the advent of sound languages of an auxiliary nature.

Primitive man surrounded by hostile neighbours did not find much need for a common language. But the development of slave societies and the state generated a need for a common language understood by merchants and travellers, by statesmen, rulers and kings, particularly since dozens, hundreds of different tribes and peoples were conquered by powerful monarchs.

The victors dispensed with ceremony when it concerned vanquished peoples. They imposed their language. In many monarchies of antiquity, documents of state, the laws, and orders were announced and recorded in the languages of the rulers.

It was not only the languages of great empires that became 'international languages'. More often these were the languages of great cultures or religions. Thus, Sanskrit in India was a common language not only for the numerous peoples of northern India that spoke tongues allied to Sanskrit, but also for the inhabitants of southern India, the Dravidians, whose tongue was in no way related to Sanskrit, and even the island dwellers of Java who took up the ancient culture of India.

Another Indic language, Pali, that was used to write Buddhist books became the common language of Birma, Thailand, Cambodia (but not Tibet, Mongolia and countries of the Far East, for northern Buddhism differed from southern Buddhism). Arabic was the international language in all countries of the Moslem world, even where no Arabic conquests had been made (in Central and western Africa, and in Indonesia). Old Slavic was a common language for Russians, Bulgarians, Serbs who spoke Slavic languages and also for Rumanians, whose language is of Latin origin.

Empires, religions and cultures vanished, but there were cases when the language outlived the state and culture and religion, and for a long time served as an international medium of communication used by merchants, travellers and even scholars and writers of a variety of countries. That is what happened to Latin, which outlived the Roman Empire by more than a thousand years.

In the Roman Empire Latin was the language of the state and of culture, later, of the church as well. The upper classes of peoples subjugated by Rome assimilated the language of the victors. The Roman Empire crumbled, however, under the onslaught of German and other tribes. New states were in the making and the languages merged and disintegrated. Latin disappeared as the language of a people. But it remained as a medium of cultural relations and was studied by Charles the Great (Charlemagne) and other emperors and kings of the Middle Ages. Church services were conducted in Latin; Latin was studied in the monasteries; Latin was used in the writing of treatises by theologians, scholars and philosophers, by every cultured person of medieval Europe.

But when the Middle Ages came to an end, so also did the dominance of Latin. It ceased to be the "language of culture". Dante, Rabelais, Servantes and Shakespeare wrote in the living tongues of their peoples and not in dead Latin. In a short time Latin ceased to be the international language of scholars.

In the eighteenth century—13 centuries after the fall of the Roman Empire—the Latin language died out completely.

It was about this time that Arabic ceased to be an international language. The Arabic language had been used by scholars to write learned treatises, by poets and cultured people of many nations; it had been widely used in drawing up treaties and in trade. Over vast areas, from the Philippines to the Atlantic Ocean and from Central Asia to Central Africa.

Arabic had faithfully served poets and scholars, warriors and merchants. It was the "Latin" of western Asia and Africa.

The growth of national consciousness and the birth of nations helped to demote Arabic from its position of an international language. The Latin of Europe and the "Latin" of Africa and Asia became defunct at the same time and for the same reasons, which were social and not linguistic.

Until quite recently, English was one of the official state languages (alongside Hindi) in the Republic of India. English still remains a convenient common language for communication among the many peoples of India with their tens of languages and hundreds of dialects; note also and particularly that the tongues of southern India—Dravidian languages—are not related in any way to the languages of northern India—Indo-Aryan. Although in 1965 Hindi was announced the official state language, English will continue for a long time to serve as a mode of communication among the many peoples of India.

Incidentally, India has yet another language that had a life-time extending for at least two and a half thousand years—Sanskrit.

Semiartificial Languages

In the slave societies of remote antiquity, priests were the bearers of culture and of the traditions of language. That was the case in ancient Egypt and Babylonia and in Assyria, and also in ancient India.

Four thousand years ago, a people that went by the name of Indo-Aryans entered the valley of the Indus River. The language they spoke was ancient Indic. Gradually these Indo-Aryans conquered the valley of the Ganges, the second largest river in India, and the language spread throughout all of northern Hindustan. The single language began to decay into a multitude of dialects.

Out of the one ancient Indic language there grew up a host of Prakrits, as the new (middle-Indic) dialects of various parts of India came to be known.

In the south-west, near present-day Bombay, a language called Maharashtri was used (whence came the present Indic language of Marathi), in the east it was Magadhi (or today's Bengali), and in the centre of Hindustan, the language of Sauraseni.

That was not all. In the south of India lived the Dravidians, in the centre the Munda, both of which were indigenous dwellers of the country who were there when the Indo-Aryans arrived. They spoke (and still do today) numerous languages quite unlike that of the Indo-Aryan invaders.

There were very many languages in the country and the need for a single literary language was evident. A language was needed to unite into a whole the linguistic confusion of the country.

And such a language was created. It received the name Sanskrit (from the ancient Indic 'samskrta' which literally meant 'perfected'). It is apparent that this language was in the making for many decades with the participation of large numbers of ancient scholars. However, we know of only one, Brahman Panini, who lived 25 centuries ago, in the fifth century B.C.

Panini's grammar was a strange concoction, that is, when compared with the grammar books of today's schools. But any modern mathematical logician or mathematical linguist would find many familiar items in Panini's work, because it is the world's first (in the history of linguistics) instance of the structural study of a language.

Panini's work consists of about four thousand short phrases. The great Indian scholar gave a description of the structure of the Sanskrit language in small precise doses very much like those of modern mathematical logic.

The ancient Indic language that was used to write the sacred books of the Hindus—the Vedas—lay at the heart of Sanskrit. However, like any living language, the language of the Vedas teemed with competing synonyms taken from the various dialects of ancient India. Archaic forms coexisted with later ones. One might say that the Veda writings represented an uncultivated language. It lacked the order and regularity of a literary language. Ancient scholars proceeded to weed the irregularities out of the Veda writings and attempted to normalize it.

In a short time Sanskrit became the language of culture in India. And in the course of two thousand years it was used as a literary vehicle by great Indian poets, dramatists, philosophers and logicians! For two thousand years!

The ancient language of the Indo-Aryans continued to disintegrate into dialects and separate branch languages, which existed independently and developed and even had a literature. Sanskrit existed alongside these living languages as an auxiliary or international language, as the language of culture, very much like Latin and Arabic.

To this very day, Sanskrit continues to be the language of literature and science like it was two and a half thousand years ago. Literature of all kinds, from ancient epic writings to modern detective novels, comes out in the Sanskrit language. Sanskrit is used in scientific articles, newspapers, magazines, and is the mode of expression for scientific, philosophical and religious debates.

In 1951, the young republic of India was in the midst of a census. To the question: "Your mother tongue", 555 Indians gave the anwer: "Sanskrit". For all the numerous languages of India, Sanskrit is the source of modern terminology in the political, scientific and technological spheres. Even the constitution of the republic speaks, in a special article, of the importance of this language which was created 25 centuries ago.

Panini and other grammarians created, or rather organized, the language. Practically the same thing was done in the 9th century by Cyril and Methodius. Linguistic talent enabled them to create the Slavic alphabet and translate into the Slav language the religious books of the christians. (It will be recalled that these writings were replete with abstract words denoting concepts unknown to the Slavs.)

For a long time this Old Slavonic or church Slavonic language was the language of the culture of Kiev Rus, Bulgaria and Serbia. It played an important role in the birth of the Russian State. It was only after the development of the various national languages, Russian, Byelorussian, Ukrainian, Serbian, Bulgarian, etc., that the Old Slavonic language ceased to be the language of literature and culture of the Slavs. Today it is used only by the Orthodox Church, but even here it is giving way to the vernacular. The same thing has happened in the Catholic Church where Latin is the official language: the clergy more and more find it advisable to speak to the common people in the living language.

Both Sanskrit and the Old Slavonic literary languages functioned in culturally unified areas. The Indo-Aryans of northern India and the Dravidians of the south differed in language but were related by a common Indian culture: they had the same Vedas, the same legends and the same customs. No wonder then that Sanskrit was able to operate as a common medium for intercourse between an Indo-Aryan and a Dravidian. The Old Slavonic language was used by the people of ancient Rus, the Bulgarians, Serbs and Rumanians. They were all united by such an important (for that period) cultural force as the Orthodox Church.

But how were matters handled in the case of people from quite different civilizations: Arabs and Italians, English and Chinese, Europeans and Africans, where not only the languages stood far apart, but also the customs, faiths, cultural traditions and ways of life. There was no common language of culture, yet trade had to go on and communications had to be established. One could not always rely on interpreters. What language could be used? Gradually the idea arose of creating an artificial language on an international scale.

Me You. We Buy Sell

Long before the idea arose of building artificial auxiliary languages, such tongues sprang up on their own without the aid of science. When two or more peoples came into close contact, a common jargon evolved that helped individuals to communicate, to trade and exchange goods. These tongues came to be known as pidgin languages.

In the Middle Ages, sailors, merchants, travellers and artisans used a common jargon called lingua franca, which consisted mainly of Romance-language (Italian, Spanish and partly French) words, but without the Romance grammar. Each people on the shores of the Mediterranean made its contribution.

A very interesting case is presented by the commercial language Chinook created by the Indians of a vast territory from Alaska to California along the north-western seaboard of North America. The basis of this lingua franca was the language of the Chinook Indian tribe that inhabited the lower reaches of the Columbia River. The grammar of the language however was simplified.

In the beginning this was a purely Indian tongue spoken by the Indians of North America, but at the start of the nineteenth century Europeans began to use it when the fur trade sprang up in those areas. That was when this commercial lingua franca began to absorb distorted English, French and even Russian words. Most of the words were of course English. By the middle of last century the pidgin language Chinook already had roughly 50 English words; forty years later, in the nineties of last century, there were 570! In the twentieth century, the Chinook jargon fell into disuse and today has almost completely been supplanted by the English language.

English has served as the basis for two hybrid languages: Kru English, or Broken English—a franca lingua of western Africa, and pidgin English.

Pidgin English is the jargon of sailors in Japan, the southern seas, Hongkong and even in California to some extent.

In pidgin English the words are English, but there is no English morphology; grammatical relations are expressed by the word order like in Chinese or by syntactic words constructed out of English words.

This should come as no surprise since pidgin English originated in Hongkong and elsewhere with Chinese forming the bulk of the traders. Pidgin English represents an alloy of English vocabulary and Chinese grammar.

In the islands of the Pacific Ocean is a lingua franca that goes by the name bêche-de-mer. The reader may recall Jack London's *South Sea Tales* and other short stories and novels where the action takes place in Oceania.

In the Far East, before the Russian Revolution of 1917, there was a fairly common lingua franca based on a mixture of Russian and Chinese. Despite its apparently absurd nature, the language had specific rules and a curious grammar.

This 'Russo-Chinese tongue' disappeared, but what of other pidgin languages? What is their fate? They too vanish with the passing of colonialism, together with the upsurge in education round the world and with the growth of national consciousness.

But by far not all of them disappear. Some lingua francas become full-fledged national languages! Thus, bêche-de-mer has become the mother tongue of thousands in the northeast of New Guinea and today goes by the name

Neomelanesian. It is taught in school and used in newspapers.

The pidgin language Kru English served as the basis for the Krio language used by the inhabitants of Free Town, the capital of the African state of Sierra Leone. Among the pidgin languages that have developed into national languages are the Surinam language of Guinea, the Papiamento jargon of Curaçao, and the language of the Solomon Islands.

A number of fresh languages have thus been added to the existing family of independent languages. So the question of a common tongue on an international scale still remains. Is it possible to create such a medium of communication?

The Algebra of Thought

The Latin of the Middle Ages died with the Middle Ages. A large portion of Newton's writings is in Latin but Lomonosov, Lavoisier and Franklin wrote their scientific work in their mother tongues and not in Latin. Scholars and philosophers in those days were occupied with the dilemma of what language was to replace Latin. Would it be possible to devise a universal language scientifically, a more refined language than any of the existing natural languages of the world?

Scholars of the 17th century, who were wide-ranging thinkers encompassing technology and mathematics, physics, philosophy and linguistics dreamed of even more—not merely a simple language accessible to all for purposes of communication. They pictured the universal language as a rationalized and refined sophisticated medium for the transmission of thought, a vehicle of the thinking process. The first to express this idea was the great French scholar and philosopher Rene Descartes.

He wrote in 1629: "An artificial language is possible and it is possible to establish the science upon which it depends. With its aid the peasant will find it easier to reason about the essence of things than do philosophers today." In Descartes' view, the universal language should be a peculiar logical key to all scientific and human concepts. He believed it possible to establish an orderly arrangement of all ideas in the manner of the sequence of natural numbers. Just as it is possible in one day to learn—in some unknown language—to name and write all the numbers to infinity, so a way should be found to devise all the words needed to express everything that originates (and can originate) in the human mind. All depends on finding the simple ideas which are peculiar to the conceptions of every person and out of which is composed everything thought of by human beings.

Descartes advanced the general idea of an artificial language, 'a tool of reason'. Another great scholar of the 17th century, Leibniz, pointed out ways of constructing such a language. In his view, this common world language would become the 'alphabet of human thought'. The words in it would not only convey ideas but also make more vivid their interrelations. Calculations replace reasoning with language becoming a kind of algebra of thought, a logical algebra of concepts. All complex ideas are combinations of simple ideas and notions. Just as we obtain all divisible numbers from the multiplication of indivisibles ($9 = 3 \times 3$, $121 = 11 \times 11$, $15 = 3 \times 5$, etc.), so we can obtain all complex concepts from the elementary 'atoms' of thought. Logical truths may be obtained with the aid of arithmetical truths.

But what would speaking be like in such a language? Leibniz suggested investing the 'alphabet of human thought' in a sonic form. The first 9 numerals would correspond to the first nine constants of the Latin alphabet: 1—b, 2—c, 3—d, 4—f, 5—g, 6—h, 7—l, 8—m, 9—n. The tens, hundreds, etc., would correspond to the five-vowels: 10—a, 100—e, 1,000—i, 10,000—o, and 100,000—u. Higher orders could be designated by means of combinations of two vowels (say, millions=au, etc.). That would make it possible not only to write but also to speak the language of thought. To illustrate, the number 873,740 would be pronounced 'mulodilefa'.

Leibniz' ideas were extremely valuable; true, not so much to linguistics as to mathematical logic and cybernetics. The 'algebra of thought' proved possible when applied in a very narrow and rigorous ("formalized", as mathematicians put it) field of human knowledge. But it is quite unfit for any living natural language, for who can compile a 'table of concepts and ideas'? How are such concepts to be classified and in accordance with what kind of rules? What is more, no one knows whether it is possible to enumerate all human concepts, perhaps their number is infinite. At any rate, the number is so great as to differ hardly at all from infinity, practically speaking.

And note that this is only the sphere of notions and ideas. The realm of human emotions, intentions, desires, and so forth cannot at all be classified by means of the 'alphabet of thought'. All too multifarious, complex and often contradictory is the sphere of human consciousness termed 'emotional life'.

Leibniz came up against these difficulties at the very start of his attempts to develop a 'world language'. He never completed the work and left only some notes about his 'language of thought'. All subsequent attempts to devise such a language came to nought. Such artificial languages are too far removed from practical everyday life. As early as last century, the development of a 'language of meaning' moved from linguistics to logic and the more abstract divisions of mathematics. In place of the barren 'algebra of language' scientists got to work on the 'algebra of thought', which culminated in cybernetics, thinking machines and other wonders that are beyond the scope of this book.

However, attempts to construct a universal language continued on a purely linguistic, not logical, basis. The first artificial language of this nature was devised in the latter part of last century and went by the bizarre name of 'Volapük'.

Schleyer Invents Volapük

In th middle of the 1870's a fifty-year-old catholic prelate by the name of Johann Martin Schleyer living in a remote Bavarian village got the idea of a unified written language, supposedly after the letter of a neighbour to his son in America was returned as being totally unintelligible—how could a simple peasant be expected to know English.

Schleyer's idea was to compose a 'world alphabet' capable of transcribing the sounds of all the languages of the world. Schleyer was a devout catholic and in publishing his project of an 'international alphabet' substantiated it as follows: "Christian Europe needs a unified alphabet as it needs a single faith."

But the pater did not stop there. He was not deterred by the fact that Europe continued to use the ordinary alphabet and not his world alphabet. In the year 1879 he produced a language called Volapük.

The grammar of this artificial language knew no exceptions. The words were pronounced exactly as they were written. The accent (stress) invariably fell on the last syllable. For the most part, the vocabulary was made up of the roots of English and, in part, German, Latin, and French words, but the words were distorted beyond all recognition. World was called 'vol' (from the English 'world'), language—'pük' (from the English 'speak'), whence the name Volapük, or world language. Even proper names were twisted: Portugal became 'Bodugan'.

The Catholic Church eagerly supported the

new language. A society was established for the study and spread of Volapük, later a congress was convened. The Church helped the spread of Volapük in many countries of Europe and America. The society of Volapükists turned into an unwieldy order of knights at the head of which stood Schleyer.

The society of Volapükists, called the 'Volapük Academy', soon came into conflict with the views of Schleyer (the society had a large number of serious linguists who were at first enthusiastic about the idea of an international language). Schleyer announced the dissolution of the Academy and formed a new one. He established a whole hierarchy of Volapükists: senators, rulers of continents, countries towns and so no down to the smallest Volapükist organizations. In Rome the Pope conferred on Schleyer the title of 'Personal Gentleman in Attendance of His Holiness'. By 1899 there were 283 different societies engaged in the study of the Volapük language, and 25 journals were published. Over a thousand teachers with degrees sedulously propagandized the new language, in which hundreds of books and pamphlets were coming out. Then the epidemic passed.

Many were discontent with this absurd language that did not have anything in common with any of the living languages of the world. By the start of the new century the movement had come to a virtual standstill.

Schleyer considered as personal offenders those who had betrayed his brain child. But still the number of betrayers increased. The fashionable passion of "educated people of the world" (for whom, as Schleyer himself had said, the language was designated) died out. In 1912 the creator of Volapük died. Immediately, a book was published describing his life and work—the last publication in the language Volapük.

Languages and Needs

All the words of Volapük were concocted by Schleyer himself. To study the language, one had to learn by heart totally unfamiliar roots. "Perhaps that is the reason this international language did not take root," reasoned those who were fascinated by the idea of building an international artificial language and getting people to learn and use it.

The way out would appear to have been found in 1887 by Dr. L. Zamenhof of Warsaw, who for his international language, called Esperanto, used the familiar roots of international words. The grammar of Esperanto was logical and simple and the language was easy to learn. There were numerous enthusiasts and the language spread to very many countries. Still and all, this language did not become a vehicle of international communication. Why?

Some thought that it might be because the language was not refined enough. That gave rise to a spate of fresh international languages —Ido, Novial, Occidental, and others. Among the inventors were prominent and talented linguists, such as the famous Danish scholar Otto Jespersen. Yet nothing came of all this effort, the languages were fine specimens, but did not spread.

But why? The matter, it turns out, does not lie in the structure of the language, but in the fact that a language must be a necessity, something that people cannot get along without in their normal lives and in their work.

A physician will undertake to study a foreign language when he finds that a great deal of medical literature is published in it that is not to be found in any other language. Tourists study the languages of the countries in which they travel and which are a must in those countries. Hence, the reasons are more social than linguistic.

Let us take an instance from the history of languages in the twentieth century.

In 1920, after a long struggle, Ireland won its independence. The population had already forgotten its ancient and beautiful Celtic language and nearly everywhere English was spoken. Only in the remotest fishing villages along the western coast was real Irish still spoken; there were even some people who did not know English. But these villages played no role whatsoever in the life of the country and the language fell into desuetude. However, Irish nationalists called upon the Irish people to speak their native tongue. The language was encouraged and taught in all schools. The results? Today, every Irishman who finishes school can read Irish after a manner but practically never speaks it!

Could this be because of a lack of love for one's motherland? Of course not! The matter is simple indeed: English is familiar to all in Ireland, whereas Irish is understood by only a few. Obviously, it is easier to converse in English. Now if there were large numbers of people in Dublin and other places in Ireland that did not know English at all, then they would willy-nilly learn to speak Irish, and the beautiful ancient language would come to life.

The situation is much the same in India today where English is still used, notwithstanding the fact that the language is associated in the mind of every native with colonialism. But what is to be done if a Tamil from southern India cannot understand a Bengali from Calcutta unless he speaks English? The point is that ancient Sanskrit, which was once used as a common vehicle of communication among the educated people of all parts of India, has fallen into desuetude and English is far more familiar.

Yet here is another illustration. In Israel the official language is the newly introduced old Hebrew (Ivrit), a language that has not been spoken for thousands of years, and one that is extremely complicated. And still this is the language in practical use in the everyday life of the country today. The secret is very simple. The population of Israel is a mix of people from many different countries speaking different languages. Those from Yemen spoke Arabic, and did not understand a single European language. The Jews that escaped from Hitlerite Germany of course knew no Arabic. The Sephardim from Greece (descendants of Spanish Jews) spoke Greek and their Spanish dialect and quite naturally could not understand the Yemenites or those from Germany. Quite obviously, the only way out was to speak Ivrit, which was taught to all new citizens of the country (some had known it because it was the language of religion). The result was that without a knowledge of Ivrit, one could not communicate with one's comrades at work and even ask directions in town, or buy bread or study in school. Thus, very practical considerations compelled the population of a whole country to go over to a new language.

These examples are highly instructive. A language only gains currency when it is indispensable.

Now let us go back to the artificial languages.

It would, of course, be very nice to have a relatively easy international language for scientific intercourse, even Esperanto, say. It would be much easier for a Hungarian, a Japanese and a Somali to learn to write Esperanto than to spend many years of hard work learning a living foreign language. The advantages of the artificial language are obvious, it would seem. But there is one great disadvantage: all the scientific literature is written in Russian, English, French, and not at all in Esperanto! If, say, a Japanese chemist knows Esperanto but does not know English or Russian or German, he will not be able to follow events in his field of research. For a chemist that is a great danger! And so he prefers to study the difficult living languages.

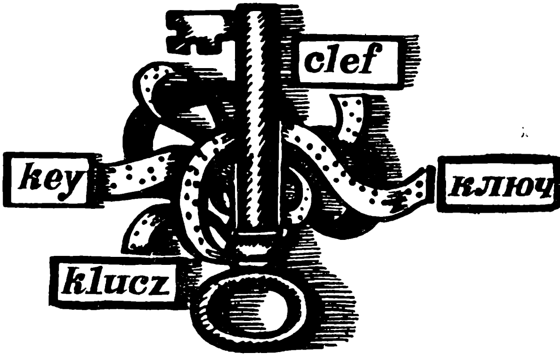
An artificial language would become widespread if by some miracle the bulk of scientific and engineering journals and books were written in that language. But miracles don't happen.

Thus, the iron laws of the science called linguistic sociology are a barrier to the introduction of artificial languages for the purpose of international intercourse. However, the work devoted to their construction over the past century and more has not been wasted. The search for a world language has led to curious and frequently very valuable finds.

The idea of an 'algebra of thought' inspired Leibniz to lay the foundations of mathematical logic, which in turn served as the basis of the "language" of computing machines. In 1897, a project was advanced for a decimal classification of ideas and concepts. True, it found application not in speech but in the library.

The classifications of books as to field of knowledge that we find in libraries are the result of this decimal classification (for instance, division 6 represents applied knowledge, 61 medicine, 616.8 diseases of the nervous system, and so forth).

The 'language of meaning' has also found applications. True, not in the language of human beings, but in machine languages. Today scientists round the world are actively working on a special intermediate language that will indeed become a universal, international language, but for machines, not human beings!



Machine Translation

For a language to be international, it is not enough merely to have that label. This is true of all existing artificial lingua francas. Scientists are in earnest search of a universal language, but not for humans, for machines.

Machines Do Miracles

We are no longer surprised to read about machines that play chess, compose music and do translations. Today electronic computers have taken over all manner of professions. Yet a decade or so ago such talk could be found only in science fiction.

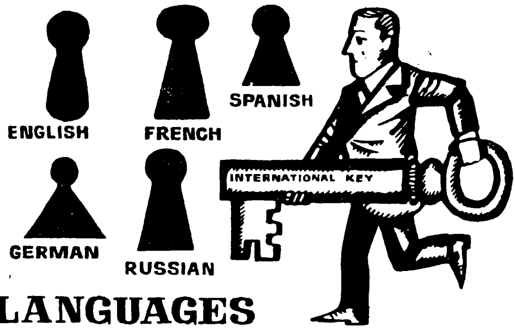
Take machine translation, computers that translate from one language into another. But how can a machine translate? One has to know languages to do translation. Surely that cannot be.

Incidentally, such views on the impossibiity of machine translation were expressed by people closely connected with the computer world. In fact, the creator of cybernetics, Norbert Wiener, said in 1947: "...as to the problem of mechanical translation, I frankly am afraid the boundaries of words in different languages are too vague and the emotional and international connotations are too extensive to make any quasi mechanical translation scheme very hopeful."

This skeptical picture, however, did not discourage investigators. Work on programmes for translation continued. In 1948 and 1949 scientists were still discussing the possibility of machine translation; in the fifties, practical work had already begun. True, even then many saw little hope for the future of machine translation.

Despite setbacks and difficulties that studded the way, on January 7, 1954, International Business Machines in New York publicly demonstrated for the first time a translation of a mathematical text from Russian into English, the first one ever done by a machine—an electronic computer called IBM-701.

At the end of 1955, a Soviet electronic computing machine did the first machine translation job in the USSR.



LANGUAGES

Today, machine translation is a matter of prime economic importance. The problem is being attacked by logicians, engineers, linguists and mathematicians.

The first attempts at machine translation were only experimental and did not represent serious works of translation. The texts were rather primitive and were based on a small vocabulary with highly restricted rules of grammar. At the present time, researchers are attempting to make machine translation a going concern capable of turning out large quantities of material.

The problem is by no means an easy one and the way is tortuous with snags and pitfalls at the most unexpected places. Human language is much too polysemous, flexible and rich; it is unbelievably difficult to reduce such an instrument to the rigid language of electronic computers. This chapter is devoted to the exciting problems of machine translation.

Man as a Machine

The idea of machine translation originated long before the birth of cybernetics and electronic calculating machines. Back in 1933, the Soviet engineer P. P. Troyanovsky was granted a certificate for inventing a "machine for selecting and printing words in the process of translation from one language into another or into several other languages at the same time". True, the invention was not put into practice, which was not surprising if we recall the primitive state of automatics in the thirties. And the translation was of the word-for-word kind (without grammar) and very slow. It was only the electronic computer that could put machine translation on a firm basis.

The idea of automatic translation goes back much farther than 1933, however, if one disregards the use of machines. Automatic does not invariably signify machine operation. Automatic could be applied to human translation by a person that has no knowledge of the foreign language he is translating from and hence does not expend any mental energy in the process.

The idea is to translate by means of special manuals reminiscent of the phrase-books used by tourists containing lists of phrases for everyday use. On the left-hand side of the page are words in the mother tongue, on the right, their equivalents in the foreign language (written in the alphabet of the mother language). Say, 'good-bye' equals the Russian 'do svidaniya'.

Everything would seem to be ideally simple.

Actually, however, it is not by a long shot. Such conversation books usually give one expression, an unambiguous translation from the language of the user into a foreign language. Yet there may be several words in the foreign language all synonymous to the original word. The question of synonyms, in which all languages abound, is not a simple one. Still and all, despite their complexity, languages obey certain rules, which may be taken into consideration, otherwise learning a foreign language would be practically impossible. In fact, people speaking the same language would not understand each other if there were no rules!

Is it possible to compile dictionaries capable of accounting for all the multifarious meanings of words? And also to list the rules embracing all the grammatical niceties of a foreign language? Definitely, but it is no easy job on the practical side.

Now if it is possible to account for all rules, they can be automatized. A person speaking his mother tongue has actually done just that. We do not stop to think about the rules of combining words into sentences (grammar). We do it automatically for the simple reason that the rules of the language have been so thoroughly drilled through our years of speaking that we can entrust them to our subconscious sphere and not to the conscious area of thought.

We learn our mother tongue in early childhood. To learn a foreign tongue is far more difficult. Still, we master foreign languages because we learn the rules and even gradually drill them to a certain state of automation much like we did the rules of our mother tongue. This, of course, takes years to do.

So far we have been speaking of the automatics of subconsciousness. The rules of a language are so thoroughly mastered and have become so firmly rooted in our brain that we do not think about them. But a "conscious automatism" is also possible. A person with absolutely no knowledge of a foreign language is only able to read the letters. Let us suppose he has a list of the grammar rules of an unfamiliar language and a complete list of all words and their meanings. Would he be able to translate from that language? Definitely.

Such translating would probably take a long time, but it could be done, provided of course that the vocabulary gave every meaning of every word, and that goes for expressions that cannot be translated literally (idiomatic expressions, linguists call them).

The first experiments in this kind of "human automatization" have already been carried out. The Hungarian linguist Mihaly Gabor published some time ago a book entitled *An International Key to Translation,* designed for six European languages: English, German, French, Spanish, Hungarian, and Russian. Using this international key, it is possible to translate from one language into another without knowing a single rule of the foreign tongue, in a purely mechanical fashion.

Gabor's translation key is meant for human beings only. And it is within the grasp of anyone. Gabor reasoned that if the most complicated mathematical operations can be accomplished with the aid of machines, surely it is possible to automate and mechanize the use of grammar and thus raise the efficiency of translators. The rules of grammar can be handled mechanically, something like putting together a jig-saw puzzle and juggling root words into their proper places, and also suffixes, endings and parts of sentences on the basis of appropriate methods and instructions.

Though Gabor's key was designed solely for use by humans, scientists working in the field of machine translation became interested immediately, for electronic computers are capable of following strict clear-cut rules much faster than any human. The purpose of machines is precisely to carry out work that does not require any creativity, only a mechanical observance of preset rules.

True, machine automatization, in contrast to the human version, gives rise to a whole series of additional difficulties.

The Machine and Meaning

The point is that a person who translates automatically from a foreign language knows his mother tongue, whereas the machine has no mother tongue. A person translating from an unfamiliar language chooses expressions that best fit the overall meaning, whereas a machine cares nothing for meaning.

Even in dictionaries of narrow specialized fields of knowledge, practically one-fifth of all the words have more than one meaning, are polysemous words. In choosing the proper meaning, we work from the general meaning of the surrounding words and the overall context. Now a machine does not understand anything and cannot for this reason work on the basis of meaning, like a human being does.

The first investigations in machine translation were conducted by mathematicians and engineers. During the Second World War electronic calculating machines were successfully used to decode secret communications of the enemy. The methods of cryptography set scientists to thinking of the possibility of applying these very techniques for translating by means of machines. Thus, one of the pioneers in the field of machine translation, Warren Weaver, of the United States, asked whether one could not consider the problem of translation to be a problem in cryptography. He said that when he looks at a piece of writing in Russian he says that it is written in English but is coded in unknown symbols, which he will then undertake to decode.

If that were so, the problem of machine translation would be solved in a purely mathematical fashion. Unfortunately, subsequent studies demonstrated that the problem of translation and the problem of decoding are not the same by far. Here is what Soviet scientist D. Yu. Panov, who headed the first Soviet efforts in machine translation, has to say on that score: "In coding and decoding, we do not handle the language as such, we change only the external form of the words by recording them in a specific code. It is quite natural that such a problem is fully amenable to formal methods. Now in translating we change the language, that is to say the entire exceptionally complicated and subtle system of expression that for centuries has been elaborated by a people and is most intimately associated with that people's thought process, history, customs, way of life, and so forth."

This means then that the principal problem of machine translation is the problem of meaning because a translation is ultimately a conversion of meaning. As the Americans R. Richens and A. Booth, prominent specialists in the field of machine translation, put it, in the general sense translation is a substitution of one language for another to express the same set of ideas.

At present, special laboratories are investigating problems of machine translation in the Soviet Union and in other countries. Even machines are being pressed into service in the elaboration of the rules of machine translation. They compile certain rules of action, the algorithm of translation of the machine, and then verify them in operation. In the process, all the linguistic facts missed by the programmers are collected and taken into account by the machine. After that, attempts are made to modify the translation algorithm in such a manner as to account for the machine-detected inaccuracies. During the first stage of research, the programme of machine translation was frequently based on a simple adjustment of the most suitable word equivalents. The translation of the English word 'solution' into Russian can serve as an example. The word has two meanings: the solution of a problem and a chemical solution. How does the computer choose between them? At first this was handled as follows: if the word 'problem' came in the same sentence or was found in the general vicinity, the machine translated into the Russian word 'решение' (mathematical solution), otherwise the translation was 'раствор' which means chemical solution.

This purely practical approach could yield results only within a close range. As soon as the computer came up against a more complicated text, with involved linguistic problems, it came to a halt. The reason for this was the haphazard collection of diverse facts devoid of any kind of rigorous system. The number of such facts that had to be accounted for increased to astronomical proportions.

That is why machine translation is mor and more becoming a problem for linguists and not mathematicians and engineers. Only language experts can precisely define the system of a language that will take into account colossal quantities of linguistic facts. Only a linguist can indicate points of convergence and divergence of the systems of two languages and spot features that are common to all languages of the world.

During the first stages, researchers worked out the rules for machine translation from English into Russian, from Russian into English, from Russian into German, and so forth. Today, a more general and a truly fantastic problem has been advanced, that of creating a common language, or a machine version of a lingua franca (an intermediate language), into which and from which any language in the world can be translated.

An intermediate language has a great advantage over the compilation of programmes for every language. Let us say we have to translate from three languages, English, Hungarian and German into Russian. We then compile the rules (algorithms) for English-Russian, Hungarian-Russian, and German-Russian translation. But a Russian-English translation requires a new set of rules—the Russian-English algorithm of translation. To translate from German into English again demands a new algorithm—the English-German set of instructions. Four languages require 12 algorithms for translation from any one into any other. Five languages require 20 algorithms, 20 languages, 380 algorithms. But there are several thousand languages in the world! A great deal of time would be needed to prepare the algorithms for that enormous number.

It would seem to be much simpler to create a single intermediate language into which the computer would translate from some foreign tongue and from which the translation would then be made into any desired language. Press the Mongolian language button at the input and the Hungarian at the output, and the machine automatically translates the text from Mongolian into Hungarian.

The problem is to construct such a marvellous intermediate language. Is it possible?

Weaver gives a vivid picture of what the intermediate language amounts to. He imagines people living in tall open towers all with a common base. When the people want to converse, they shout from tower to tower. Communication in this fashion is of course inconvenient and unreliable. Now when a person comes down from his tower to the great open basement, he can talk freely to any other person who has likewise come out of his tower. True, the view is not so beautiful here as it was aloft, but communication is both reliable and convenient.

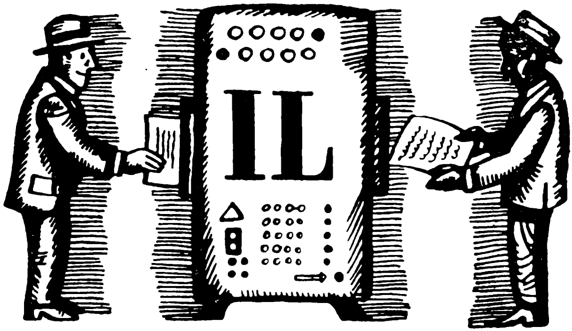
The idea is to come down from each language to some not yet devised universal language of humanity and then to go aloft by the most convenient route.

How do we picture this universal intermediate language?

The Intermediate Language Debate

The structure of an intermediate language is constantly being discussed. Experts propose all manner of versions. Some say that the best solution is to take one of the existing natural "living" languages, say, Russian or English, and then draw up the rules of translation of all other languages of the world into this language. Simple as that.

Actually, however, it is not so simple as that. Every living language has its specific rules. And also its specific exceptions to these rules. If Russian is taken as the intermediate language, trouble will arise with the numerous case endings and their large number of exceptions, which will have to be taken into account even



when translating from Spanish into Japanese. Any other language used as the intermediate link will offer special rules and considerable exceptions to them. If a living language is taken, we will invariably distort all the other languages of the world in the act of double translation.

Say we are translating from Russian into Turkish. Neither, thank god, has any articles. But if the intermediate language is English, we will first have to stick in all the articles (which is ever so hard for the machine to do) and then we will have to throw them all out again in the final version.

Let us again try Turkish or Russian as the intermediate language. Inconvenient, no question of it. Try translating from English into French: again an enormous amount of useless and complicated work. In the first operation, translating into the intermediate language (Russian or Turkish), we have to cross out the articles in the English text that is undergoing translation. Then when translating from the intermediate language into French we have the excruciatingly difficult task of restoring articles. Besides all this, every living language is extremely complex and much too polysemous for it to be the language for electronic translators.

But then, say the experts, why not try one of the international languages like Esperanto or Volapük? The rules of grammar are simple in the extreme in these artificial languages, not a single exception. Why spend time and energy on elaborating a special language if we have access to already existing artificial languages?

Unfortunately, not one of the artificial languages is suitable for the future intermediate language. The natural living languages are far too complicated, and any one of the artificial languages, whether Esperanto, Interlingua, Volapük or any other international medium is patterned after the natural languages. In this respect, the artificial ones are only a bit better than the natural ones.

Esperanto has an article, which means that in translations from Russian into Turkish (if Esperanto is the intermediate language) a lot of machine time will be lost in seeking out articles that are not needed either in Russian or Turkish or any of a large number of other languages.

That is why most researchers believe that the intermediate language should differ from both natural and international languages. It will have to be constructed on the basis of some kind of special principles. The Leningrad linguist N. D. Andreyev suggests the following structure for an intermediate language. It will be a language with a definite vocabulary and grammar but it will be impossible to speak such a language (although a phonetic system could be worked out—but, then, electronic computers never need to converse). The phrases of the intermediate language will have to be combinations of symbols expressing specific concepts.

The closest thing to an intermediate language of this kind is our customary number system, where each expression has its unique precise meaning.

The grammar and vocabulary of the intermediate language would be built up on the basis of the grammars and vocabularies of the different living languages. There would occur a sort of "averaging" of all the real languages of the world, with the new intermediate one imbibing only the most typical and high-frequency grammatical rules and words common to most human tongues.

Many Soviet experts in this field do not agree, however. They think that the intermediate language should contain all categories of all languages if they designate something in the real world. It would have to be rich to the extent that it could express any phrase in any language. Then there would be no fear of losing valuable information in the transfer act (recall the ubiquitous article in translations from English into French).

Says Soviet mathematician V. A. Uspensky, "If for example most languages lacked a category like the future tense of the verb, Andreyev would reject that category for his intermediate language. I believe we should proceed from a different angle. If even one language has the future tense category, that in itself is sufficient grounds for including it in the intermediate language. On the other hand, a category like grammatical gender does not need to be included even if it is characteristic of most languages."

Uspensky takes that view of gender because the gender of inanimate nouns is meaningless. To illustrate, take the Russian noun 'стол' which means 'table'. What is there to be gained when we learn that it is masculine? If it were feminine, the translation would be exactly the same anyway.

I. A. Melchuk, a young Soviet researcher, suggests constructing the future intermediate language on the principle of universality. His view is that it should be a system of correspondences between all the languages of the world.

Actually, the language should not exist in a material form. It should simply be a system of correlationships between the various languages of the world, correspondences between words and word combinations that are equivalent. For example, the Russian 'разряд', the English 'class', the German 'Klasse', the Chinese 'teng', the Japanese 'i' have the same meaning. It is this semantic correspondence that will form a 'word' in the intermediate language. Correlationships of other words in the most diversified languages of the world will go to form new words. The intermediate language will be built up out of series of word correlations of actual languages and will not have any materially embodied words of its own. Incidentally there might be a material embodiment in the form of a label. Say, a number could be assigned to each series.

But how about morphology, the relationships of cases, numbers, tenses, moods, voices?

Russian has gender, number and case for the noun, Hungarian, Turkish and Chinese nouns do not have any gender. French and Chinese have no case endings for their nouns. Chinese nouns do not even have number: the singular and the plural sound alike and are written alike, other words (numerals) being used to indicate plurality. How then can grammatical relationships be established?

Melchuk maintains that the intermediate language should be devoid of any morphology. All the morphological categories (gender, number, case for nouns, say) will have to be reflected in the vocabulary of the language and not in the grammar. Thus, the Turkish word 'dash' ('stone') would be given in the intermediate language as two words: one with the meaning of 'stone' and the other with the meaning of 'singular'. The Russian word 'камень' ('stone') would have four words in the intermediate language: one with the meaning 'stone', another signifying 'singular' (as in Turkish), a third stating that it is of masculine gender, and a fourth conveying the nominative case. The Chinese word 'shihtou' ('stone') would yield only one word with the meaning 'stone' because the Chinese language does not use grammatical means to designate gender, case or number.

This is not only a theoretical consideration, but also a practical expedient. The machine language is unambiguous and mathematically precise. It is based on the formal language of mathematical logic. Now the language of mathematical logic does not have any morphology. It consists of a set of simple symbols, the alphabet (this corresponds to the vocabulary) and of rules for connecting the symbols into sentences (which corresponds to the syntax of ordinary languages).

Melchuk says: "An intermediate language constructed in this fashion satisfies best the demand of universality: on the one hand, it includes only what is common to all possible languages—vocabulary and syntax, and on the other hand, in any translation from any language it expresses everything that is expressed in that language."

In the Search for Universals

What the future intermediate language will be like is not yet clear, for scientists have not yet solved the most difficult of all problems, that of meaning, and without meaning nothing will be meaningful.

"Machine translation is not a simple translation by means of machines but is a transformation of text that retains only the 'meaning' of the text," write the Soviet scientists N. D. Andreyev, V. V. Ivanov and I. A. Melchuk. Hence, the intermediate language will have to be supplied with such words as express all the multiplicity of meanings of our speech. In other words, it will have to be a language of meaning.

This requires finding general semantic categories, some sort of 'units of meaning' that are peculiar to all languages of the world, and then isolating the specific 'units of meaning' of each concrete language. This problem is closely associated with another highly important one, that of creating a special information language for electronic logico-information machines (see the chapter "The Universal Code of Science").

Science deals with terms that are unisemantic, clear-cut and definitely isolated as to meaning from other terms. Languages exhibit quite a different picture. A word in an ordinary language can often have several meanings. The word 'solution' that we analysed earlier in this chapter is an instance.

International terminological commissions, federations of standards and other organizations of scientists and technologists can establish the exact meanings of terms, but how is one to figure out the meanings of ordinary words? Here, commissions are of no help, for the living language cannot be curbed by decrees and fiat.

To all this we have to add the difficulty that one and the same notion may be expressed by several words. In the different languages there may be one or two or five such concepts. Let us take the Russian word 'знак' ('sign'): the notion is expressed by three synonyms: 'символ', 'знак', 'обозначение'; in English there are at least six words ('sign', 'symbol', 'note', 'mark', 'notation', 'index') expressing the same meaning. Japanese, like Russian, has three words of the same meaning: 'go', 'fugo' and 'nigo'. It will thus be necessary to take into account that one and the same content, meaning, can be expressed in a variety of ways in any one language.

Still, scientists hope to develop an intermediate language with the aid of which it will be possible to translate from any language into any other language. In fact they are already working on the problem in a number of countries. The Cambridge group in England are attacking the problem via the thesaurus method.

The Thesaurus Method

A thesaurus is a dictionary of a special kind. The words here are grouped according to categories, classes, ideas. The best-known work of this kind is Roget's thesaurus of the English language, first compiled in the middle of last century and since republished in numerous editions with amendments and additions. The classes in this dictionary are divided into sections, and the latter are further divided into categories. Altogether there are 6 classes, 24 sections and over a thousand categories. For example, the class 'SPACE' includes the sections 'Space in General', 'Dimensions', 'Form', 'Motion'.

The section 'Motion' is subdivided into the categories: Motion in General, Degrees of Motion, Motion Conjoined with Force, and Motion with Reference to Direction. Further subdivisions include Travel, Navigation, Aeronautics, Traveller, Mariner, etc. All the words of the language are thus arranged in categories.

There is also an alphabetical list of all words for convenient reference with indicated class, section and category to which the word belongs.

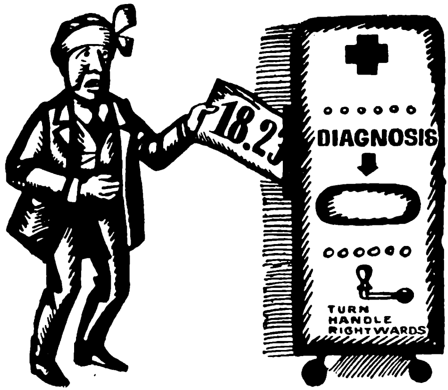
When a word has many meanings, it is included in several groups. Say, 'flat' in Roget's list is included in such groups as No. 172 inert, 191 suite, 207 low, 213 horizontal. Thus, a word of many meanings is split up into various categories in accordance with the various nuances of signification.

Roget's dictionary contains only words of English. Today, thesauruses are being devised for many languages. General classes, sections and categories contain the words of different languages with closely associated meanings. Each naturally has a common category number, which numbers together go to make up the vocabulary of the intermediate language.

Translating with the aid of a thesaurus is a rather unusual procedure. It is a process of translating whole paragraphs instead of sentences. A single 'semantic formula' is first constructed for a paragraph; the meanings of the sentences that make up the paragraph are expressed by numbers of the thesaurus. Then a search is made for semantic formulas in the language into which the translation is to be made that correspond to this semantic formula.

Semantic Factors

A 'semantic language' of this kind is under development in the Soviet Union in the laboratory of machine translation at the Moscow Peda-



gogical Institute of Foreign Languages. Here the thesaurus method is not employed but a more promising and interesting one—the method of semantic factors.

Every language has hundreds and thousands of different words, each with its own meaning. On closer scrutiny, however, one notices that very many meanings consist of more elementary meanings.

To illustrate this point, let us take eight words: 'father', 'mother', 'son', 'daughter', 'uncle', 'aunt', 'nephew', 'niece'. Each has a distinct meaning. But if we take 'father', 'son', 'uncle' and 'nephew' we find a common element of signification, or a common seme—they are all masculine. Thus, we have one seme, sex: masculine denoted by A, feminine by $\bar{A}$, which means 'not A'.

A further analysis shows that mother and father are associated with daughter and son by direct kinship; daughter is associated with aunt or son with uncle via indirect kinship. Hence we get another semantic factor: kinship (the symbol B designates direct kinship, $\bar{B}$, indirect kinship).

Let us finally compare mother, father, aunt and uncle on the one hand and son, daughter, nephew, niece, on the other hand. The seme is determined via 'generation', which may be preceding (symbolized by C) and subsequent (symbolized by $\bar{C}$).

Now the meaning of any one of the eight words can be expressed in the form of a combination of three semantic factors. Father is ABC (maschuline sex, direct kinship, preceding generation) daughter is $\bar{A}B\bar{C}$ (feminine sex, direct kinship, succeeding generation), nephew is $A\bar{B}\bar{C}$ (masculine, indirect kinship, succeeding generation), etc.

Thus, every word can be represented as a combination of simpler semantic units. Just like the meaning of a sentence stems from the combined meanings of the words that comprise it, so the meaning of a word is made up of combinations of elementary semantic factors, or elementary units of meaning. The basic task is to find these simple 'meanings'.

It is not so difficult to find semantic factors in the language of science as it is in everyday speech. Thus, Perry and Kent, of the United States, developed a special semantic language used in logico-information machines. It is confined to a single field of technology—metallurgy.

The laboratory of machine translation of Moscow's First Institute of Foreign Languages is working out a semantic language for our everyday speech. The following semantic factors have been found (the conventional symbol is indicated in parentheses): negation (symbol '—'), positivity (symbol '18'), sensation (symbol '23'), greatness (symbol '10').

The meaning of the word 'inadequate' is written $\overline{18}$ in the semantic language, which is the same as the word 'bad', which is a synonym. The word 'unwell' is written as $\overline{18}.23$ (bad + sensation), the word 'small' as $\overline{10}$ (negation of greatness), the word 'remarkable' as 10.18 (greatness + positivity).

Work on isolating elementary meanings or 'atoms of signification' is still in the early stages. Yet it holds out great promise: the point is that in translating by means of semantic factors, the machine models comprehension of the text by the translator. To begin with, the machine analyses the text semantically (which is exactly what the human translator does), and then reproduces it in its own words in the target language (again like the experienced human translator does). Applying such an analysis, we are able to produce the following semantic translation: from the English 'this is not new' into the Russian 'это уже было раньше' ('This has already been').

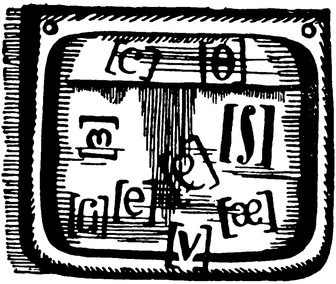
The Language of Humans and the Language of Machines

Thus, using semantic factors researchers hope to teach the machine to 'understand' the text it is translating and then to state the comprehension in another language. It is hard to overestimate the value of such a machine capable of grasping the meaning of a text. True, machine comprehension demands extremely precise rules. American linguists found, for example, that in a superficial semantic analysis (by the thesaurus method) the machine translated the proverb 'out of sight, out of mind' as follows: 'negation of sight', 'negation of mind' = 'blind idiot'.

The rigorous and exact language of machines is already beginning to exert its effect on scientific language. The point is that the machine cannot grasp vague or inexact statements, and so every problem which we wish to present to the machine has to be formulated with extreme clarity. Perhaps the machine translator operating with semantic factors will even help workers in the humanities to achieve greater exactitude of expression. What is more, we may even gain something in the way of more precise wording and style in our everyday affairs and rid much of our writing of meaningless and clumsy words and turns of phrase.

But will that mean that the human language has become something in the nature of a machine language? By no means. The living languages have elements which would appear meaningless if expressed by semes. A feeling for style, emotional colouring and similar elements of human intercourse are quite beyond the scope of the electronic machine. And of course translations of works of art are actually far removed from the machine, for in fiction one does not simply translate the text but creates a new work of art, through a rewritten text that conveys imagery, poetical models and not merely literal meaning. One can easily picture the machine translator conveying 'the youthful maid trembled' as 'the young girl shook'.

Most likely, there will be two languages: one for business that is rigorous, unambiguous and accessible both to humans and machines (this will be used exclusively in scientific and official spheres) and a second one, inaccessible to machines, will serve as a vehicle for conveying emotions with multifarious nuances that can never be formalized.



Sounds and Meaning

How do people speak? How is it we undestand each other? It is only recently that the science of phonology—which is to the humanities what nuclear physics is to natural science—was able to penetrate into one of the greatest of human marvels, speech.

Inborn or a Product of Society?

The child emits sounds from the very moment of birth. Emits sounds, which is not to say "speaks". "Cries of discomfort" and "gurgles of satisfaction"—such are the responses of the newborn to events in his environment. These are purely biological utterances, not speech. The child was never taught them, they come out spontaneously like the cries of animals.

But the child grows, his world expands, and the number of sounds increases. At first, the scream of irritation and the squeal of comfort are vowel sounds, if matched to the sounds of ordinary human speech. The sound range of the child gradually increases and consonants appear.

Why does the child first pronounce vowel sounds? Simply because the human throat finds them the easiest to emit. The child does not speak but merely emits sounds just like the young of any other animal that possesses vocal chords.

Later, there sets in a transformation of instinctual physiological, inner, speech into the true speech of human beings. The conversion does not take place of itself but only under the impact of surrounding people: mother, father and others.

There are no inborn sounds in any child anywhere in the world: the little Bushman, the German tot, the Japanese toddler, and the tiny being on Terra del Fuego all emit the same sounds. If a Bushman aged one to three months is put into an English family, he will grow up speaking English as his native tongue. Conversely, a little English-born child will master the Bantu language to perfection if brought up in their environment.

Yet there was a time when many believed in a 'congenital language' claimed to be peculiar to all earth dwellers.

According to an ancient legend, the Egyptian pharaoh Psammetichus once wished to find out which of the languages was first on this earth.

He ordered two newly born children to be taken to an old speechless man so that they could not learn to speak from anyone. Let their speech be natural, the pharaoh decided, and when the children begin to speak, that will then be considered the "very first" language of humans.

As the legend goes, they began speaking Phrygian. We are not sure the experiment ever took place; at any rate, if it did, the pharaoh was definitely fooled. Somebody had obviously taught the children to speak Phrygian, for otherwise they would never have learned to speak any human tongue. So say the facts. Of course, such a harsh inhuman experiment could never be repeated, but nature on its own has carried out similar experiments.

Science knows of cases when children were reared by wild beasts: wolves, leopards, monkeys, bears and even by a sheep. These wild children did not speak Phrygian or Russian, but a sort of "beastian". Their vocal chords emitted the signal cries of the animals that brought them up; they howled like wolves, squealed like monkeys, and bleated like sheep. Later on, it was with great difficulty that they were taught a human language.

Now we come to the problem of how a normal child learns to speak, one brought up by human beings, not beasts.

During the very first weeks of his life, the child begins to react to sounds. It is not a simple response but a differentiated one in which the tot distinguishes between unpleasant, sharp and loud sounds and pleasant, melodious and gentle ones. To the little being, the pleasantest sound is that of its mother's voice (this is no trite phrase but an experimentally established fact). He isolates it readily from the background of noise, and beams when he hears it.

Through this faculty not only to speak, emit sounds, but also to hear and distinguish one sound from another, the child begins to learn human speech in a real way. It comes about by imitation: he imitates the sounds he hears (true, this is common not only to man but to birds, say parrots or starling) and he babbles and plays with sounds like only humans can.

Baby-talk when applied to adults is derogatory but in the life of a child it plays a very great role, for these are no instinctual cries of discomfort or shouts of joy; they represent a certain mass of sounds that is destined to crystallize into a harmonious system of language.

The child begins to prattle at the age of three to four months. The sounds are the same in all children of the world, irrespective of the different languages of the adults. Researchers studying baby-talk have found a fantastic diversity and complexity of sounds including sibilants and even the clicking sounds found in the languages of the Bushmen and Hottentots.

Then why are the languages of people different if baby-talk throughout the world is all the same? Why doesn't a unified adult tongue originate from the prattle of infants?

The reason is that human speech does not develop; it is imparted, inculcated by society. It does not appear of itself but in the process of the development of the child.

Through the influence of adults, the babble of babies turns into the speech of the child. The sounds absent in the mother tongue are forgotten and vanish for the simple reason that they are not used by grown-ups who repeat only the sounds to be found in the native tongue.

The system of a language is like a sieve that sifts out the unneeded and retains the "native" sounds. Just what is this sieve? Why do we speak differently? What is the 'system of a language' and why is it needed?

Atoms of Speech

Take any dictionary and you will see that changing a single letter changes the word completely: hot, hop, hip. A single letter, a single sound, and the word means something quite different. This property of sounds in changing the meaning of words is called the semantic differentiating function of sounds. Every language has similar series of words differentiated by a single sound each time. The differentiating sound has no meaning in itself but is capable of converting one meaningful word into another one, as we have seen in the example given above.

But does every alteration of sound lead to a change in meaning? Take any word spoken in the high-pitched voice of a little boy and in the low-pitched bass of a grown-up man. The overall sound is different but the meaning remains unchanged.

In some languages a long vowel sound or a short one plays no part at all—the meaning of

the words remains the same. Say in Russian the word 'стул' whether pronounced short or long, as in the German 'Stuhl' or the English 'few', remains unchanged in meaning. Now in German such shortening or lengthening can modify the meaning of a word: 'stehlen' (long) means 'to steal', 'stellen' (short) means 'to put'.

Short or long vowels mean nothing to the Russian, but Germans, English-speaking people and Czechs find them extremely important in building new words.

From the foregoing it is clear that not all sound differences are important in language processes. Sounds whose variations are of no significance in a given language go to make up a single unit of sound called the phoneme.

Phonemes can be pronounced differently depending on the peculiarities of the voice of the speaker and for a number of other reasons. 'Mama' can be pronounced in a variety of ways by changing the position of the lips for 'm', yet the meaning will remain unchanged, though if we replace 'm' by 'p' we get quite a different word: 'papa'. Consequently, 'm' and 'p' are different phonemes because they generate different words.

It frequently happens that in one language certain sounds are variants of a single phoneme, while in another language they are different 'atoms of speech'.

In English 'k' has one basic pronunciation and is thus one phoneme, but in Hindi it has two pronunciations and is capable of generating different words. In the language of the Avars of Daghestan in the Caucasus 'k' is uttered in 14 (!) different ways, hence has 14 different phonemes that must not be confused if one wants to be understood properly.

A child and an adult can pronounce the same sounds, but the ability to speak begins with the ability to distinguish sounds and not merely to utter them. And not simply sounds as such but those that differentiate words in the language.

To summarize, then, we can say that not all differences in sound have significance but only the differences inherent in phonemes. Is this assertion true? It is and yet it is not. True in the sense that only phonemes are important as far as the meanings of words go, but when we listen to someone speaking we extract information not only from the meaning of the words. From a person's delivery we can learn much about his mood, something that will not be revealed in the words he uses.

One and the same phrase may be uttered in a variety of ways depending on the mood of the speaker; his intonation in a single word like 'hello' can run through a whole gamut of emotions from cheerfulness and elation to servility and dejection. The peculiarities of intonation and mode of delivery or utterance can pinpoint a person we have long forgotten.

Finally, we can learn a great deal about the age of the speaker, whether child, grown-up or old man, about the sex, and even about the place of birth (from the accent).

Nevertheless, English to a Londoner, a person from New York or a New Zealander will be the same. Investigations of acousticians and linguists have demonstrated that no two persons speak in exactly the same way. There is always some small difference in the way they utter words.

Even one and the same person speaks differently upon different occasions. In ordinary speech we swallow sounds and whole syllables, but this is hardly permissible in the speech of a radio or television announcer.

Speech sounds are infinitely diversified. The important thing is to distinguish only specific sounds (phonemes) that form the basis of the language—words and the meanings of words. Even in intonation, not all variants are needed, rather only the typical differences that are commonly understood; they are called intonemes, or intonation patterns. Their study has only just begun, but their existence is hardly to be doubted. Phonemes and intonation patterns comprise a separate science called phonology.

Why Some People Speak with a Foreign Accent

"His accent gave him away. He was immediately recognized for his foreign accent." And true enough, an accent, actually an incorrect pronunciation, stamps a person as having a different native tongue.

But why do people speak another tongue with a foreign accent? Everything would seem to be in order: a good knowledge of the vocabulary and the rules of grammar and the basic sounds, yet one immediately feels the difference. Why? The question appears to be simple, but only phonology has been able to give the answer.

Every language breaks up the continuous spectrum of speech sounds in its own specific way. Whence the difference in pronunciation that produces what we call a 'foreign accent'. A person speaking a foreign language frequently introduces elements of his own native tongue into the phonological system of that language. This is particularly evident when the native tongue and the foreign language would appear to have similar sounds. To illustrate, the Russian 'пот' and the English 'pot' are much the same in pronunciation yet reveal a perceptible difference in the utterance of the 'p' sound. Differences are still more evident when we deal with vowel sounds. Eward Sapir says that the vowel sounds of English and Russian are totally different and that hardly any two could be considered as coinciding.

A person begins to master the phonological system of his mother tongue in early childhood; once learned and firmly rooted in one's mind, it is used throughout the rest of one's life automatically. But what about mastering a foreign tongue? Quite naturally, an attempt has to be made to learn the phonological system, the range of phonemes of a foreign language. And of course in the process we introduce the customary habits of our native tongue into the new speech pattern and the result is an accent. Russians, for instance, make English and French voiced consonants voiceless at the ends of words, in accord with the typical pattern of the Russian language. Besides distorting phonemes, a foreigner places his native stamp on the intonation of his utterances as well, and quite unintentionally. It is no easy job to overcome speech habits that were so thoroughly drilled in early childhood.

Until recently, there was only one way out: to repeat the sounds under the guidance of an experienced teacher, best of all a native, until the sound is mastered to a point where no accent is perceivable. The trouble has always been, however, that there are few good teachers, and the process is so complicated and time-consuming that not many had the opportunity to engage a teacher for long periods of time.

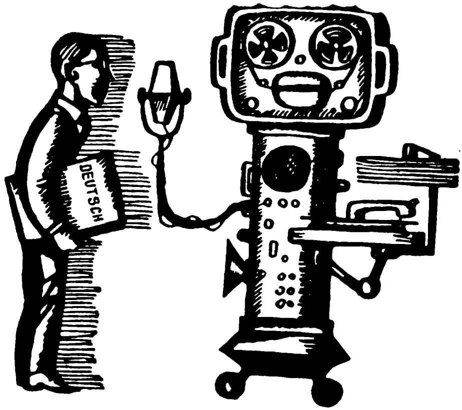
Engineering developments have suggested a marvellous way out of this impasse. The traditional teacher with chalk and blackboard has given way to phonograph records and tape recorders. Let us take a look at a modern language laboratory at the Moscow Pedagogical Institute of Foreign Languages. No chalk here, only microphones, tape recorders, and earphones. The pride of the laboratory is a tremendous "collection of sounds", a whole library of recordings—six thousand in eight languages. One can listen to the voice of Thomas Mann reading selections from his novel *Bekenntnisse des Hochstaplers Felix Krull*, the voices of

Louis Aragon, Paul Eluard, Erich Weinert, the recordings of Goethe's *Faust* and Shakespeare's *Hamlet,* plays, short stories and poetry read by Gerard Philippe and other celebrities. The library also has a recording of the famous accusatory speech of Georgi Dimitrov at the trial of the burning of the Reichstag. Too, there are recordings of 500 talks, interviews, lectures and more. Besides such "classics" we have whole textbooks recorded together with exercises, pronunciation drills, vocabulary and grammar studies.

Rows of desks equipped with inbuilt tape-recording devices, and earphones. The student puts in his tape and listens. Or he can record his own voice in the form of answers to questions or translations from other recordings. The sounds of the language under study are mastered by listening to recordings and following the printed text.

The advantage of recordings is that they may be repeated as many times as one desires. Imitating the speech of excellent native speakers is one of the best techniques for acquiring a perfect pronunciation. The student can then record his rendition of a selection and see the weak spots for himself.

One's native tongue appears simplicity itself, almost as natural as breathing. We never think about inhaling and exhaling and the movements of our chest, for they are all done automatically. Handling one's native tongue is much like that, for we never have to think about phonemes or even know that they exist; yet we have complete mastery of our language. The



same is true of a foreign tongue that has been mastered to perfection.

To achieve automatization of a foreign language, one has to drill tens, hundreds, thousands of times the linguistic patterns of the language; in the process, we first consciously and then subconsciously master the new phonological system, the atoms of the foreign tongue, which are quite different from those of our native tongue (though at times very much akin to them). Today, with such teaching and learning aids as cinematography, tape recorders and phonograph records—whole language laboratories—one can master a foreign language in record time.

A language is the most subtle instrument we have for conveying human emotions and thoughts. A knowledge of other languages helps us to learn about other peoples and thus to get a better understanding of ourselves, for as the ancient saying goes, truth is revealed in comparisons.

Style Through Sound

Phonemes are the 'atoms of language'. They are what converts a flow of sounds into human speech. We have already said that speech can reveal the age and the sex of the speaker, the mood he or she is in, and much more.

At times it is even possible to tell, without seeing the speaker, whether he or she is thin or stout; ailments like asthma or dyspnoea even show up in one's speech.

However, this is due not to the language but to the system of signs. Someone in a bad mood imparts a special tone to his voice not because the language (English, say) has any particular vehicle for denoting such a mood. And illness, age and mood are all perfectly non-linguistic, they are devoid of any relation whatsoever to language as such.

Still and all, language has elements of a system of signs that enable one to distinguish more than simple meaning.

Language is above all a social phenomenon, and the sound agents that can express the social class to which one belongs, the educational level of the person, his background, and so forth are conventional or 'sign' elements.

The Russian scientist N. S. Trubetskoi, one of the founders of phonology, writes that certain peoples located at a low stage in the development of the clan system exhibited specific differences as to groups based on sex and on age with a corresponding difference in 'female' pronunciation and 'male' pronunciation of the same phonemes. The pronunciation was thus governed by social and not physiological factors.

For example, one of the sounds of the Chukchi language is pronounced 'r' by the men and 'ts' by the women. In Tikhon Syomushkin's novel *Chukotka* one of the girls tells the Russian teacher, who asked her to pronounce the letter 'r', that "it is not nice."

In the language of the Yukagirs (in Kolyma, Siberia) there are sounds pronounced differently by men, women, children, and old people.

In class societies there are distinct differences in the pronunciation of representatives of the various groups and classes. Trubetskoi goes on to say that in Vienna the everyday speech of an official of a ministry differs from that of a salesman. The Russian literary language distinguished between a nobleman's and a merchant's pronunciation. Probably in every language one finds differences in the pronunciation of townsfolk and peasants, the educated and uneducated. One often comes across a specific "fashionable" pronunciation in high society with a typically careless articulation peculiar to the dandy.

The Tamils of southern India had a strict division into castes. The upper caste were the Brahmans (priests), below which stood the Kshatriyas (governing and military occupations); lower still came the Vaisyas (commercial and agricultural occupations), and at the very bottom of the hierarchical ladder were the Sudras (traditionally assigned to menial occupations). This caste system was reflected in the Tamil language. The Brahmans pronounced one of the phonemes like 'sh', the Vaisyas and Kshatriyas uttered it like 'ch' and the Sudras like 's'!

In Russian, the letters 'a' and 'o' are pronounced practically in the same way when unstressed, but in old Russian the clergy made the 'o' in 'отец' ('father') sound like a stressed 'o' to emphasize their ecclesiastic position.

In every language of course there are differences in pronunciation based on the locality from which the speaker comes.

Here we wish to emphasize the fact that not only the intonation, the loudness, and the timbre of the voice but also the very system of the language (its phonemes) can express origins, education, upbringing and the sex of the speaker. This, however, is not the sphere of phonology, but that of the linguistic science called sound stylistics. Here we are interested in phonology, the science of systems of all languages of the world, the science of the sounds of human language.

Language and System

Each language has its own subdivision of the world of sounds into phonemes. Each one sifts out the phonemes it wants for the given native tongue.

How do we count the number of phonemes

in a language? Actually, anyone who knows a given language can figure out the number of 'atoms of the language' by comparing words and isolating the factors (phonemes) that distinguish words semantically.

This book is written in English, and so you can find the phonemes of the English language by collecting pairs of words where a single sound (phoneme) substituted for another changes the meaning of the word. This is most conveniently seen in words of one syllable: 'top', 'tap', 'tip', 'bit', 'bat', 'but', 'sun', 'son', 'sin', and so on.

The phonological systems of different languages naturally differ. In Russian there is one phoneme 'a'. In Latin, there are two (one long and one short): 'malum' meaning 'bad' if the 'a' is short, and 'apple' if the 'a' is long.

In the same way, Latin has two 'i' phonemes, two 'o' phonemes and two 'u' and 'e' phonemes. This gives us a total of ten phonemes based on five letters. In the ancient Indian language Maya there were fifteen! In addition to the long and short vowels there was a special kind which may be depicted by an apostrophe in writing. It was produced by a break in the voice brought about by a closing of the vocal chords.

On the other hand, the Mayas did not distinguish such obvious (to the average European) differences as 's' and 'z', 'v' and 'f', and 'd' and 't'. To these people, 'dan' and 'tan' were not separated as distinct words.

In some languages, meanings are differentiated by means of stress in addition to phonemes;

'subject' (noun) and 'subject' (verb), for example in English. This cannot be done in French, Finnish, Czech, Georgian, or Hungarian.

Languages differ also in the kind of combinations of sounds which they permit or prohibit. In Russian no word can begin with the letter 'ы', whereas in Estonian, Chukchi, Mari and many other languages, there are numerous words that begin with 'ы'.

The Polynesian languages of the South Pacific islands do not allow for two consonants in succession. What is more, every word has to end in a vowel sound. How difficult for them to pronounce 'Landknecht' or 'information'. But these same languages allow for long words without a single consonant! The Hawaiian word 'oiaio' means 'truth'—five vowels and not a single consonant.

Then there are languages, the sounds of which cannot be conveyed at all by English or Russian sounds: the clicking sounds of the Bushmen, Hottentots and Zulus.

Every language extracts from the continuous spectrum of sounds the ones it needs to build its system of phonemes. The smallest number of phonemes is found in the Polynesian languages and in the languages of the natives of Australia—from 12 to 15 phonemes. The Hawaiian language has only 7 consonants and 5 vowels (Russian has 5 vowels and 35 consonants). Some of the Caucasian languages have a record number of over 70 phonemes! For instance, the Abbasi language has over 70 consonants and only two vowels, some researchers even claim only one vowel, the other vowel phoneme being considered a variant of the first.

Twelve Distinguishing Factors or Universals

All told, there are several thousand different languages in the world. In America alone there are about 2,000. This means that we have a few thousand distinct phonological systems. Even more, if one takes into account the varieties of dialects of a single language, each dialect with its own phonological system. Even closely related languages are different.

But perhaps these thousands of phonological systems have certain common features. Isn't it possible to isolate a few universal factors to be found in all languages of the world, no matter how divergent they may be?

Linguists were startled when after a careful study of the languages of Australia and the Americas, Asia and Oceania, Europe and Africa they were able to establish what the linguist Eward Sapir of the United States characterized as the remarkable fact that distinctive phonetic features are spread over great areas irrespective of vocabulary and structure of the languages.

The difference between 'a' and 'b' is that the former is a vowel and the latter a consonant. Thus we have found the first factor in the classification of phonemes: vocal or nonvocal, say phonologists. Most languages of the world exhibit this same distinction: phonemes are either vowel or nonvowel.

Like Russian, many other languages have a number of vowels and a series of consonants. What distinguishes them?

Let us take 'b' and 'p'. The difference here is that the former is voiced and the latter voiceless. Thus, the next universal factor is the 'voice-voiceless' factor. Other pairs of sounds are possible, like 'd' and 't' and 's' and 'z'. Very many languages exhibit this distinction.

Another distinguishing feature among sounds is that the sound may be either continuous or interrupted. The vibrating (rolling) Russian 'р' ('r') is an instance of a continuous sound, in contrast, say, to the Russian 'л' ('l') which is an interrupted one.

By analysing the phonemes of a single language, we can establish their distinguishing features. We can compare a phoneme to a circle, and the distinguishing factors that separate it out from the other sounds of the language, to coloured squares that are superimposed on the circle. If, say, the phoneme is a vowel, a dark blue square is placed on the circle, if a nonvowel, a light-blue square is superimposed; if the phoneme is voiced, a dark red square is used, if voiceless, the square is light-red, etc. Then each phoneme will be represented in the form of combinations of squares of different colours, or in the form of a set of distinguishing factors. Such a colour system is described by one of the prominent Soviet phonologists Sebastyan K. Shaumyan in a collection of articles entitled *Problems of Structural Linguistics*.

The natural question arises: Will there be enough colours to cover the variety of factors needed? In other words, will there not be too large a number of distinguishing factors among the phonemes of a single language? All the more so if we take all the languages of the world—several thousand. The average number of phonemes in a single language is about 30. Simple arithmetic then shows us that the number of phonemes will reach a hundred thousand: 30 times three or four thousand. How many distinguishing factors do we need for this great number?

Researchers in linguistics, acoustics, speech, physiology and communications have produced the answer. And a highly unexpected one it was. It was found that a mere dozen distinguishing factors suffice to differentiate all 100,000 phonemes!

Spoken language may be viewed from two angles. On the one hand, it takes the form of wave vibrations or a continuous stream of sounds. On the other, it involves the movements of the vocal chords, tongue and lips.

Sound waves may be recorded on tape, on an oscillograph. The operation of the speech apparatus can also be analysed. In this fashion we can obtain descriptions of phonemes and distinguishing factors from two sides: sonic or acoustic and speech or articulation.

Now if we take the acoustic features, there will be found to be only a dozen paired differentiations (or distinguishing factors) for all the languages in the world.

Why are the distinguishing factors constructed in the form of pairs, on the binary principle? Vowel-nonvowel, etc., why not triplets or quadruplets?

The question might at first glance appear to be quite senseless, something like "why is a wheel round". But actually this is not so. The reason may be rooted in the human psyche. Recall the infant in his efforts to learn his mother tongue and the work that he performs in sifting out the needed phonemes from the multiplicity of speech sounds. For the infant mind, the binary principle (yes-no, white-black) is much more comprehensible than other more complicated systems. Of course, the infant of 6 to 10 months of age does not give any thought to the distinguishing features of phonemes (actually, linguists themselves learned about them only in the 20th century). But the child subconsciously discerns the difference between the voiced and the voiceless consonants.

Incidentally, the binary (or paired) yes-no principle is not only best for children—it is the ideal for the computing machines of cybernetics.

"Hallo, Robot"

A machine can be outfitted with vision—the electronic eye of the photoelectric device. A machine can be made to feel by means of punched cards, tape, magnetic discs, tapes and drums. Even the sense of hearing is possible—in the form of a microphone. How convenient it would be to speak with a machine in the normal human fashion, dispensing with the machine language of programmes and algorithms.

But how can we teach the calculating machine to comprehend human speech? Imagine how efficient and convenient it would be to give oral instructions to banks of computers handling huge production complexes. Or imagine translating machines converting the speech of a lecturer into other languages.

But if it is possible to teach a machine to analyse spoken language, it naturally seems possible to make it comprehend commands given in the human voice, and even answer like a human being. This is important not only for machine translators but for machines for the blind and for numerous instruments that are in contact with human operators. It has even been suggested that in aircraft an automatic device might be able to keep the pilot informed about the readings of a large number of instruments in the form of brief oral phrases. Such "speaking instruments" could be very useful at control consoles in the spacecraft cabin (so far only science-fiction stories have them).

In principle, then, communication between humans and machines may be accomplished not only in machine language but in human language as well. But how?

The first attempts to construct a "speaking machine" were made long before the birth of cybernetics—in the 18th century. In 1780 the Russian Academy of Sciences offered for solution the following problems: "I. What are the properties and characters of the pronunciation of such divergent vowel letters as a, e, i, o, u? II. Is it not possible to devise tools similar to the organic pipes known as the human voice, such as would be able to pronounce the vowel letters a, e, i, o, u?"

The first "automatic speaker" was constructed by the marvellous engineer and mechanician of the end of the 18th century, Farkas Kempelen of Hungary. However, all the "automatic speakers" of the 18th, 19th and first half of the 20th centuries were only toys. It was only after the advent of the computer and cybernetics that problems of communication between human beings and machines were under serious discussion. What is the present situation in this sphere?

The first approach viewed the word as the basic unit of speech. In a recent experiment in the United States, a computer was connected with a spectroscope that analysed streams of sounds. The findings of the spectroscope were then converted into numbers and fed to a computer.

The speaker working with this set-up pronounces a definite word, say, 'two' several times. On the basis of a series of repetitions, the machine develops in its memory a sample or pattern of the word. Then the word 'two' in printing is fed to the machine. The machine goes through a process much like learning to read.

After this training, the machine can readily identify the "learned" word pronounced by the teacher and compare it with the standard it has in its memory store. Seven women instructors and six men participated in the experiments in teaching the machine. The automaton not only learned to identify precisely the word that was spoken but even identified the speaker! The women's voices were guessed without a single mistake, the men's in 93 cases out of a hundred.

Following this training session, it was decided to go on to "self-instruction". By listening to new speakers, the machine was able to identify the earlier learned words and to modify the standard of these words when they were not pronounced exactly like the "teachers" did.

But such experiments are possible only when the vocabulary is very limited. The set-up we have described could identify only 83 English words at a rate of indentification of one word every second and a half. To increase the number of words in the vocabulary of the machine, it would be necessary to expand its memory.

Identification difficulties increase in proportion to the number of words. When there are only a few words, like 'two' and 'three', it is rather simple to teach the machine to distinguish them. But when close-lying pronunciations like, say, 'three' and 'triple', 'two' and 'stew' are introduced the problem becomes enormously involved.

The identification time also increases with increasing complexity, which means that the machine speed has to be increased drastically if we want to keep our computer from being a dunce. But the greater the machine memory store, the more time is required to search out the needed word. And the bigger the vocabulary, the more extensive must the memory be.

More, as Soviet linguist V. V. Ivanov says: "If the model used is one in which the underlying identifiable unit is the spoken word, the problem will become unsolvable for a large vocabulary."

The storage capacities of modern computers are too small for them to remember one thousand of the most used words of a language. But. as we know, the vocabulary of any language is many many times greater. Which means that the machine memory store has to be expanded that many times—a complete impossibility at the present stage of computer development.

Writes Ivanov: "Insuperable technical barriers include not only the volume of the machine memory but the exceedingly small time allowed for extracting from the vocabulary the needed word that has to be identified prior to the identification of the word that follows it."

And so, though a machine can be taught to understand a few human words (and some machines are already doing that), the 'word' method is not suitable for any kind of real conversation between humans and machine.

But then why not teach the machine to distinguish phonemes instead of words? Particularly since they do not number more than 70 or 80 in any language of the world. This would make the machine memory smaller than that required to remember a thousand words.

During the past fifteen years, scientists the world over have been hard at work trying to teach machines to learn phonemes. Sad to say, they—the machines—are not yet up to the task. The flow of speech sounds is a continuous flow of waves, and machines are not yet capable of picking out the separate phoneme-elements from this stream.

Again, the pronunciation of phonemes depends largely upon the age, sex and origin of the speaker. We intuitively make allowances for such factors and can identify even distorted phonemes by comparing them with the standards stored in our memories. But how can a machine be taught to do this?

Perhaps there is another solution. In fact, there is. It has to do with the 12 universal distinguishing factors of all languages. When the machine hears a sound of human speech, its first job is to divide it up into the twelve universals and then compare it with the standard of a whole word that is stored in the machine memory as a sequence of the distinguishing factors.

Modern computing machines have two types of memory: one small-size high-speed and the other large-size and relatively slow. The high-speed unit may be used to process speech sounds and test them for the distinguishing factors. The large-size memory can be used to store the entire vocabulary.

Here is what Ivanov has to say about this system. "In this kind of identification of speech the volume of the memory is made to approach the capabilities of existing machines, and the time of extraction will no longer depend on the duration of word pronunciation since identification of acoustic signals and vocabulary search will take place at different times and will be done by different parts of the machine. A machine of this type could simultaneously accomplish automatic identification of sounds and automatic analysis of sentences, which would make it possible to solve such complicated problems as the separation of a sentence into its component words."

This model was proposed on the basis of purely linguistic reasoning, and recent studies in the USSR headed by L. A. Chistovich have demonstrated that identification of human speech takes place according to a similar principle.

The machine has learned its lesson: when among men, do as they do.

The Language of Men and the Language of Beasts

Human languages have an average of 30 to 40 phonemes each. Studies of zoopsychologists have shown that there are roughly the same number of ultimate sound units in the signal systems of animals, such as the chimpanzee, and the dolphin.

Is this an accidental coincidence? Most likely not. Language differences among humans consist in the fact that the elementary units, or phonemes, combine to form syllables, words and sentences. This is not so among animals. Their signal cries are devoid of grammar—rules for combining phonemes into more complex linguistic units.

Why is it that animals are not capable of this? Why is it that only man can construct words out of elementary sound units? The number of such units is about the same in all the higher organized inhabitants of our planet, whether man, dolphin or monkey!

It is not linguistics that will give the answer but other sciences such as history, psychology, and the science of society.

Dolphins and chimps find a few dozen sounds quite sufficient for their signals. They do not need any more; they do not feel the need to express complicated ideas, concepts, intentions, and so why combine elementary signs into more complex units?

The reason for this is that they have no society, only groups of animals, where there is hardly any need to communicate. All animals actually require are signs of warning, satisfaction and calling and similar elementary signals. For such purposes a few tens of signals suffice.

An analysis of the skulls of our ancient progenitors such as the Pithecanthropes, man-apes and cave dwellers, the Neanderthal men, show that they did not possess speech in the way we understand it. Articulate human speech was a thing unknown to them. Apparently, they had just as primitive a signal system as monkeys.

Activities involving work brought about the necessity of new words and new signals. Not one or two, but tens, hundreds and thousands of them. And to solve this problem the indissoluble animal signals began to be separated into elementary phonemes and sign words consisting of them.

Scientists in the future may be able to demonstrate exastly how the human language was born out of the signal system of anthropoid apes, man's forefathers, how work and its activities called forth new signals, and how the increasing vocabulary demanded an expansion of the memory system of primitive man, how labour habits and the expanded vocabulary in turn contributed to the evolution of the brain; how the development of the brain paralleled improvements in the speech apparatus, and how the laws of language were consolidated in society, in the minds of the members of society—in a word, how the human language was born.

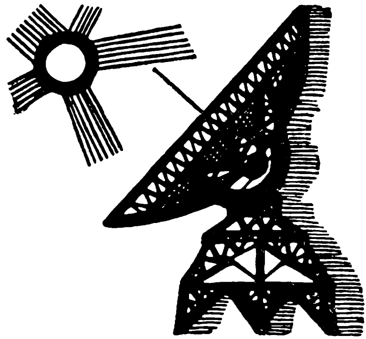
Phonology has helped greatly in the study of human speech, and also in exact studies of the 'languages of animals', the signal systems of dolphins, apes, bees, and ants.

Just a short time back, the Soviet researcher N. I. Zhinkin, who is studying the physiology and psychology of speech, analysed the sound signal system of hamadryad baboons.

The study was done according to all the rules of modern linguistics. The baboon cries were measured with the aid of an oscillograph. Spectrograms enabled Zhinkin to make a "microscopic analysis" of the sounds. A roentgenoscope was used to measure the "speech", to record exactly the articulation movements made by the throat of the baboon during "talking". Finally, the data obtained were analysed in accordance with the theory of phonological distinguishing factors. The 12 universals of the human language proved applicable to the language of baboons.

Studies of animal language are important for linguistics (how does human language differ from the signal systems of animals?) and for semiotics (which deals with all manner of sign systems irrespective of their artificial or natural origin and whether they are used by humans, animals or machines). Cybernetics is also interested in the development of more refined machine languages.

Finally, they are important in tackling the problem of what may appear to be in the realm of science fiction—the language of cosmic communication, the linguistics of space designed for conversing with unknown intelligent beings of other worlds.



Space Linguistics

Primitive man could communicate only with the immediate members of his tribe, his group. In time, man came to understand other languages of other tribes. Modern man is searching for ways to communicate with intelligent beings of other worlds. Today scientists have undertaken the first attempts to communicate via a cosmic language with intelligence outside this world.

On the Threshold

Scientific workers of today are tacking problems that just a short while ago were in the sphere of science fiction only. Man has begun the conquest of space and is on the threshold of flights to other planets. Will he find intelligent beings in these worlds, and what about the problem of communication with them? Astrobiologists, cyberneticians, astronomers, mathematicians, logicians, communications experts and zoopsychologists are all engaged in this problem.

What will we encounter in these foreign worlds? Is man alone in the universe or has he thinking brethren elsewhere? Will he be able to understand them, communicate with them, exchange the attainments of culture of the different planets?

Perhaps there are no stellar brothers at all.

No one has ever seen a Martian, an inhabitant of Venus, or any other planet of the solar system or of other stars of our Galaxy.

But scientists knew about atoms and molecules long before they detected them. Astronomers have been discovering planets and asteroids with pencil and paper.

Our vast array of laws, physical and biological and social suggest that man is not the sole inhabitant and intelligent being in the universe.

Indeed, can it be that only our sun, a modest star in a plain portion of the Galaxy, has a planetary system? And if planets exist round other stars, then why should not life and intelligence originate there as it did on our earth?

The number of stars in the explored portion of the universe is measured by the incredibly enormous number 100,000,000,000,000,000,000. The American astronomer Harlow Shapley believes that there must be in the vicinity of 100,000,000,000,000 (one hundred million million!) planets with conditions quite suitable for the life of highly organized organisms. Among them would definitely be planets with intelligent beings.

Man is no prisoner incarcerated in the cell of his planet. More likely the Galaxy—our stellar world—is a densely populated city where the earth is but a tiny structure on the outskirts somewhere.

Tsiolkovsky once wrote: "Is it possible that Europe is populated and the other parts of the world are not? Can there be one island with dwellers and multitudes of other islands with none? Is it likely that only one apple-tree in the endless gardens of the universe is covered with apples and all the infinitely many others, with greenery only?! Spectral analyses have shown that the substance of the universe is the same as the material of the earth. Life is everywhere in the universe, and this life is endlessly diversified."

One day this theoretical problem may become practical—the era of spaceflight has begun and the conquest of distant space is on.

What awaits us in these worlds, what manifestations of intelligence? Human-like beings? Or creatures endowed with intelligence but quite different physically? At what stage of development will they be? At the stone-age level? Or perhaps our human civilization will in their eyes be a stone-age achievement. How far behind? A thousand years behind, a hundred thousand or a million? Perhaps, indeed, there is an enormous circle like that described by Efremov in his book *Andromeda* and we will be able to establish contact with them. Only the future will give the answers.

Talking with the Stars

Tsiolkovsky's prophetic words that "mankind will not remain on the earth forever, but in the pursuit of light and space will at first timidly penetrate beyond the limits of the atmosphere, and then will conquer all the space around the sun", are coming true.

But man is not satisfied with exploring the solar system; he strives beyond, to the stars. The distances then become millions of times greater than the distances to the planets of our solar system. It takes light travelling from the earth one second to reach the moon, 8 minutes to get to the sun, but to the nearest star—Proxima Centauri—it takes 4.27 years! Diminish our solar system to the size of a postcard—the nearest star will be half a kilometre away, and our Galaxy will occupy the territory of the Soviet Union! At present speeds, trips to the stars would last thousands and even tens of thousands of years.

But it is not necessary to send human beings on such interstellar voyages. We can communicate with intelligent beings via cosmic signals. Radio is the most reliable long-range communication facility for conversations with the stars.

Astronomers have been receiving radio signals from outer space for the past ten years or so, but the source is nature and not intelligent beings.

Radio signals arrive from all points of our Galaxy and from stars in other galaxies distant from us 5, 8 and 10 thousand million light years. Some of them are very weak, others exhibit fantastic strength. There is one note that stands out in this interstellar radio noise background. Atoms of hydrogen colliding in empty space send signals on a wavelength of 21 centimetres. These collisions are accidental, and the signals of the hydrogen wavelength arrive from corners of space that are as yet beyond the reach of the most powerful optical telescopes in the world. Now if we begin to perceive in such fortuitous radio outpourings a semblance of order and regularity, it might mean that the source is not nature but human-like intelligence!

The first such experiment was carried out in 1960. On the night of April 5 and 6, the Green Bank Observatory in the United States turned its huge 25-metre radio dish towards stars that scientists thought most likely had planets similar to the earth: Epsilon Eridani and Tau Ceti. The radio noise from these stars was caught by the giant ear of the radio telescope, then amplified and recorded on tape and finally analysed by electronic machines.

Radio signals from deep space on the 21 centimetre wave band coming from these stars, candidates for planetary life similar to that on the earth, were found to contain certain regularities. Could they be...

Alas, the highly ordered signals were not coming from human-like intelligence of another world but were simply due to malfunctioning apparatus. The hopes of scientists were dashed. No signals were found to be emanating from the planets of Tau Ceti and Epsilon Eridani. Perhaps some other time.

Engineers and scientists are intensively developing another mode of communication with cosmic inhabitants—by means of heavy beams of light. Lasers are capable of generating beams of such high strength that they will cover the immense distance of interstellar space to unknown planets. Many scientific workers take the view that we will succeed even in this century in establishing two-way cosmic communication with the inhabitants of other worlds in our Galaxy, at any rate with the closest neighbours. It should occur sooner or later, but how will the cosmic conversation take place?

The noted physicist and historian Ralph Lapp says that the simplest way for us to give the inhabitants of planet X to understand that their message has been received is to repeat a signal corresponding to the repetition of an unknown word spoken by a foreigner as he points to an object. We would have to change somewhat the sequence of the signal, he adds, in order that society X should not think that they are dealing with some kind of radio echo.

But this is merely a cosmic "hallo" that we exchange with our interlocutor from space to let him know that we are listening to him. But from then on how will the conversation proceed?

" 'Let us introduce ourselves,' whispered Pyotr Mikhailovich. He rose to his feet with dignity and said distinctly:

" 'We are dwellers of the earth. That is what we call our planet, which is located at the end of the third spiral arm of the Galaxy.'

"...Suddenly, a series of exotic signs and lines appeared in the cupola of the amphitheatre.

" 'Aha,' said Pyotr Mikhailovich with satisfaction, 'they want to speak to us, it seems.'

"He peered intently into the strange signs for a few minutes and then a smile broke out on his face.

" 'They have written some kind of mathematical formula or equation. Judging by its structure, there seems to be something reminiscent of the law relating mass and energy, that universal law of nature.' "

Such is the description A. Kolpakov gives in his space fantasy *Griada* of a conversation of earthlings with the dwellers of a planet located in the centre of the Galaxy.

From then on, information is exchanged in the following manner: the earthlings write the alphabet, and our academician "pronounces each letter in a loud distinct voice". In turn, the Griadites write the letters of their alphabet and the sounds are pronounced "by what seems to be a mechanical throat". The only difference between the two alphabets lies in the number of letters: the Griada alphabet has at least a hundred.

This is followed by a study of the language, during which Academician Pyotr Mikhailovich expresses his dissatisfaction that they were being taught "like little children". The chief difficulty lay in the "hardly perceivable regularities" of the grammar of the language of the Griadites.

We leave on the conscience of the writer the ability of Pyotr Mikhailovich to grasp at a glance mathematical formulas whose structure resembles the relations in the mass and energy law, and the no less remarkable perspicacity of the Griadites, who straight off realize that they are being shown the "earth alphabet", and immediately write their own.

The point is that we are not able to do this even with many peoples on our own planet, simply because a vast number of notions and words which would seem to be almost self-apparent in modern Indo-European languages are lacking in the languages of the Papuas, Eskimos, Australian natives, Indian tribes of the Amazon River and other peoples at a low level of cultural development.

Specialists translating ancient Indian philosophical writings into European languages often complain of the untranslatability of many abstract concepts. Yet all these difficulties which arise in the linguistic intercourse of peoples here on the earth, will appear infantile when compared to the problem of establishing cosmic intercourse with beings in other worlds, with the problem of communication, exchange of information with intelligent creatures of other planets. People, despite the many differences in behaviour, language, culture, upbringing are beings of the same kind, children of a single unitary cycle of evolution of life on our planet.

Our world, the world of the Australian native, the Aleut, the Bushman, the "earth" world, is something comprehensible to all us earthlings. By virtue of our planetary kinship we find a common language. But creatures of another planet? Things may be quite different from ours: external appearance and internal world, conduct, culture and language. Everything may differ substantially.

Let us suppose that there are people which, by virtue of some kind of physiological defect, are capable of seeing only blue. They would hardly be able to formulate the notion that they see only that one colour. The term 'blue' would then be devoid of meaning. In their language there would be no names for colours and the words used to denote shades of blue would correspond to our 'light', 'dark', 'white', 'black', etc., but not to the word 'blue'. To realize that they are looking at the colour blue, these people would have to have perceived other colours at some time. Now these are only terrestrial pitfalls. Any cosmic conversation would most likely involve difficulties hundreds of times greater.

That is why some scientists believe that it is practically impossible to get to understand intelligent beings in other worlds. Their psyche, conduct, culture will differ from ours to such a degree that neither human beings, nor 'thinking' cybernetic complexes will be able to establish contact with our cosmic brethren: earthlings will not be able to comprehend "non-people" elsewhere in the universe. Still, most researchers claim that cosmic talks will become possible. The reason is that the world in which we live is unitary: the same atoms and elementary particles comprise it; the same physical laws govern it both in the vicinity of our sun and in the region of the Andromeda nebula. It may be argued—and researchers are debating the issue—whether protein life is the only form of life and intelligence in the universe or whether a nonprotein type of life is possible. But the laws of the material world and, hence, the principles of processing information are of the same nature throughout that portion of the universe in which we live. This unity instils confidence that no matter what the difficulties that may be encountered in cosmic intercourse with human-like creatures of other planets, they are surmountable.

Despite numerous linguistic barriers, one man understands another. Even without any language at all, a certain amount of meaning can be transmitted by means of gestures, drawings, formulas and diagrams. One has only to recall the performances of that remarkable French mime Marcel Marceau.

Communication and mutual comprehension are possible for the reason that all people are children of the earth. Communication between intelligent beings and mutual comprehension, irrespective of the number of hands, noses, heads, tentacles, will be possible for the reason that we all—creatures of reason—are children of a single universe.

What form would a cosmic language take? In what language would we converse with our human-like brothers in other worlds?

Mathematics as the Basis of Communication

Mathematics is called the language of the universe. Its laws are common to the whole world, including Mars, Venus and the Andromeda nebula—a minus times a minus will always result in a plus. That is why the majority of scientists believe that mathematics will be the vehicle for a cosmic conversation with intelligent beings.

Not all agree, however. Professor Kolman, of Czechoslovakia, speaking at a discussion devoted to problems of cybernetics, had this to say: "If it is assumed that somewhere in the Andromeda nebula there are highly organized fluid inhabitants living in a fluid medium, they can have no geometry or arithmetic in our meaning of the words and, consequently, notions taken from these sciences could not be employed to communicate with such creatures."

Indeed, our terrestrial Euclidean geometry would not be comprehended by "fluid dwellers". But haven't mathematicians constructed a whole series of nonterrestrial, non-Euclidean geometries? Modern mathematics (for example, topology otherwise called rubber-sheet geometry) is sufficiently abstract and rich to represent and comprehend any geometry, any arithmetic, and any kind of logic (for the logic of unknown reasoning beings may not be two-valued—true or false—but three-valued, four-valued and multivalued).

Many years ago when radio and other long-range systems of communication were not yet known, scientists proposed constructing, on the vast plains of Siberia, a gigantic luminescent figure of Pythagoras' theorem for communication with Martians, who were then supposed to realize that the earth is inhabited by intelligent beings.

Tsiolkovsky too suggested conversing with intelligent creatures of other worlds via mathematical concepts. In the article "Can the Earth Communicate to the Inhabitants of Other Planets the Existence of Intelligent Creatures Here" that he published in 1896 he wrote that the first candidates for cosmic conversations would be the dwellers of Mars (the possibility of finding intelligent beings on Mars is now regarded with great skepticism, although the debate is still on).

Writes Tsiolkovsky: "The shields convince the Martians that we are capable of counting. They are flashed on once, twice, three times, etc., with intervals of 10 seconds or so between each group of scintillations. In this manner we could show off to our neighbours all our extensive arithmetical knowledge: for example, we could demonstrate our ability to multiply, divide, extract roots, and more, we could demonstrate our astronomical knowledge by giving, for instance, the relationships of the volumes of the planets. . . . It would be best to start with things known to the Martians, such as astronomical and physical data. A series of numbers could also convey to the Martians any desired shape, that of a dog, a person, a machine and so forth. Indeed, if like human beings, they are to any extent acquainted with analytical geometry, they would find no difficulty in guessing the meaning of these numbers."

Many researchers today suggest using powerful radio transmitters to send mathematical concepts or symbols, such as the natural integers 1, 2, 3, 4, 5, etc., into space. Or the number 'pi', the ratio of the length of the circumference of a circle to its diameter, and similar mathematical truths.

Scientists are attempting to use the language of numbers and formulas to tell about mathematics and other sciences as well, even human culture at large. The following is a story of the first attempt of this kind.

Lincos (Lingua Cosmica)

The Dutch mathematician Dr. Hans Freudenthal is a prominent specialist in mathematical logic, topology and other extremely abstract fields of mathematics. In 1960 in the series *Studies in Logic and the Foundations of Mathematics,* Freudenthal published his latest work entitled "Lincos, Design of a Language for Cosmic Intercourse".

With extreme pains and care the author of Lincos (lingua cosmica) develops his cosmic language designed as a vehicle for communication with intelligent beings of other planets. The elaboration is so thorough that one gets the impression that the cosmic conversation is going to be established in the near future. There is something portentous in the fact that Freudenthal completed the first volume (the second is not yet written) of his study in December 1957, the year of the conquest of space by the world's first artificial earth satellite.

The logical foundation of Lincos is the idea that on the basis of the unitary laws of the universe, people of the earth can establish comprehensible contact with human-like beings elsewhere. Freudenthal immediately introduces the significant reservation: Lincos is an abstract scheme of a language and not its concrete or physical manifestation. However, in the author's opinion, the first "letters" of Lincos must be graphical signs (we have already discussed them): their meanings must be closely associated with physical embodiment, something like the interjections of our speech, say, 'sh' or 'cock-a-doodle-doo'.

Most workers in this field believe that mathematical notions are easiest to convey to human-like beings. And so, after the Introduction, in which the author elucidates the principles of his method, he takes up an exposition of the foundations of mathematics. While the Introduction is written in ordinary human language, the other chapters of the book require of the human reader almost as much attention as it will of the intelligent being it is aimed at H. Freudenthal proposes to begin communication with intelligent beings by transmitting a set of initial Lincos signs a large number of times and in different sequences. Basic mathematics is started in the following manner. First, the cosmic graphical sign is transmitted in the form of a brief flash of light or a radio beep, which is the graphical sign of a dot. Then, following three such dots, the sign > (greater than) is introduced and then two more dots: ... > ..

The very same sign is given between five and two dots, between ten and five dots, etc.:

..... > ..
.... > .....

The sign < (less than) is introduced in similar fashion:

.. < ...
.... < ....
. < ..

Then the equal sign (=) is introduced:

... = ...
.. = ..
...... = ...
.... = ...

Having given an explanation of 'greater than', 'less than', and 'equals', Freudenthal passes on to the binary (base 2) system of numbers (binary is employed because it is the simplest, and the human-like being out in deep space may have three or eight fingers instead of our ten, or he might not have any fingers at all). One graphical sign of a dot—the equal sign—the sign 1, two dots—the equal sign—the sign

10 (and not 2 as in our familiar decimal system of counting), three dots—the equal sign—the sign 11, four dots—100, etc.

$$. = 1;\ .. = 10;\ ... = 11;\ .... = 100.$$

Then follow the rules of addition, subtraction, multiplication, division (it will be noted that these rules are very simple in binary notation). For example, the entire multiplication table in binary looks like this:

$$1 \times 0 = 0;\ 0 \times 1 = 0;\ 1 \times 1 = 1;\ 0 \times 0 = 0.$$

After the brief course in basic arithmetic, the author of Lincos takes up algebra. To do this, he introduces the concept of an abstract number:

$$a + 100 > a + 10$$
$$a + 11 < a + 101.$$

Then, using algebra, Freudenthal introduces the human element into mathematics: the notion of a question. This is followed by a rigorous and systematic exposition of the basic notions of algebra and the chapter of Mathematics ends with an introduction to mathematical analysis, which belongs to higher mathematics. Many sections of analysis, the functions of many variables for instance, are so nonrigorously explained in our terrestrial language that to translate them into the rigorous language of Lincos was impossible.

So you see that the foundations of mathematics are rather satisfactorily explained with the aid of Lincos. But mathematics is an abstract science. Will we be able—even if we acquaint our human-like brothers abroad with our advances in mathematics—to tell them about ourselves, our way of life, ethics and conduct? Such things are immeasurably more difficult and complicated than mathematical formulas and laws, for such laws are common to the whole universe, while the rules of behaviour, morality and culture differ greatly even on our planet, even among the different classes of a small country!

Signs, Language and Behaviour

"Someone cautiously and lightly tapped on the window. Half a minute later, the knocking was repeated. A yellow row of eyes moved back a few metres.

" 'They are inviting us to come out,' said Knyazev."

That is how writer G. Martynov, in his trilogy *Astronauts* describes how earthlings get acquainted with the dwellers of Venus. Without even knowing what the inhabitants of the planet look like, Martynov's heroes are confident that a tap on the window signifies a polite request to come out—the earthlings found food near their spaceship. And so the author, together with his hero, says: "By bringing their 'bread' to the spaceship, the Venusians demonstrated that they want peace. That was the only interpretation that could be given."

Just like human beings, although the inhabitants of Venus turn out to be "strange beings that appear to have just stepped out of a fairy-tale, with three black eyes and thin flat mouths." Yet that does not stop them from performing a "heathen-like worship of fire" and from bringing "gifts"—a cup. Writes Martynov: "With a movement of his hand, just like humans do, the Venusian invited the two astronauts to follow him."

So simple! Men arrive on the planet Venus, the local inhabitants of this foreign world bring "gifts", remove their hats and reverently bow, and then with a friendly wink of their third eye and a come-along-with-us movement of the hand invite them home—just like normal human beings.

That may be Martynov's idea of communication but what if our "human-like intelligent beings" have no hands, are spherical and live in water, are like spiders or trees, or are absolutely different from anything we know both as to appearance, habits and conceptions? Or suppose that in place of a lot of creatures, there is only one gigantic organism on the whole planet? How then do we explain to our strange space creatures our principles of behaviour, even if we have somehow explained the rules of mathematics?

Freudenthal, and that is one of the great merits of his book, demonstrates that with the aid of the universal language of mathematics it is possible to explain even the rules of conduct.

In Lincos a great deal of attention is paid to behaviour. Using abstract mathematical concepts and the abstract language of Lincos, Freudenthal was able to set forth many of the regulations of human behaviour, including even the rules of etiquette.

He begins the principles of behaviour with the notion of an acting person—an abstract being carrying on a conversation with other abstract dramatis personae on mathematical topics. Actor A sets a problem, another one—B—solves it correctly, while a third one—C—solves it incorrectly. In this way, to each actor is ascribed specific rules of behaviour.

Freudenthal stresses the fact that dozens of diversified examples have to be given so that an intelligent being of another world should be able to grasp the system of human notions. The rules of morality start with explanations of the notions of 'good' (Ben) and 'bad' (Mal), which are the human evaluations that underlie any moral judgment. Here is how it is done by means of the lingua cosmica. Actor A poses a problem, actor B solves it correctly, C incorrectly, D solves it correctly, but in too long a fashion. This solution is evaluated as 'Mal', but not as false, for the solution is correct, though long.

The short and correct solution done by B is evaluated as 'Ben' (good). Thanks to such appraisals, Lincos is able to state a large number of human notions and conceptions. The following is an instance of how the cosmic language states a rule of etiquette: Acting person A puts a problem to B; acting person D solves the problem correctly. A says 'Mal' (bad) because he asked B and not D. Hence the conclusion: do not answer (even if you answer correctly) questions that are not asked of you!

In the chapter Behaviour, Freudenthal introduces the notions 'prove', 'more or less exact assertion' and gives examples of 'misunderstanding' and 'mispronouncing' by the same ingenious method, a mathematical conversation of acting persons, as in the instances given above. Explanations in Lincos are at times extremely subtle, like explaining to our stellar cousins that there may be situations in which "the receiver can conclude that somebody might refuse to answer a question though able to do so".

In the chapter Behaviour, much attention is devoted to the fact that the number of people moving, wishing and perceiving exceeds the number of speakers. Animal is something that differs from human; it wishes but cannot speak. Human is he who can speak. On the earth there are about three thousand million speakers.

Beginning the next chapter, Freudenthal writes: "So far the members of the class 'Hom'* might be ghosts. The only extension we needed, was time." He then undertakes to explain the basic physical laws that govern the actions of humankind. "In too short a time a person cannot go from one place to another." "If an object is too large it cannot be carried by a single person, though perhaps by more persons together," etc. The notions of mass and movement are introduced on the one hand "practically", from the viewpoint of human behaviour and, on the other hand, as physical laws, axiomatically.

* Humans —*Tr.*

The author of Lincos explains the advantages of the group as compared to a single person ("if an object is too large it cannot be carried by a single person, though perhaps by more persons together").

'Mother' (Mat) and 'father' (Pat) are explained as follows: "The existence of a human body begins some time earlier than that of the human itself. The same is true for some animals."

"Before the individual existence of a human, its body is part of the body of its mother." Freudenthal gives the biological laws and also the basic physical laws that humans have learned.

At the very end of his study Freudenthal gives the basic principles of the theory of relativity, the famous formula relating mass and energy: $E = mc^2$. And that is the end of the first volume of Lincos.

In the second volume, still uncompleted, Freudenthal intends to use Lincos to discuss matter, life and human behaviour, giving more subtle aspects in these spheres than was covered in the first volume in the chapter Behaviour.

Greetings, Stellar Brother!

Lincos is the first attempt to create a cosmic language that human beings could use to exchange information with intelligent creatures of other worlds; it is a fundamental study devoted to the problem of extraterrestrial communication. It might very well be that in the future, when we receive signals from human-like beings in deep space, our terrestrial radio stations will begin transmitting in Lincos or some more refined version of lingua cosmica.

In the introduction to his book, Freudenthal emphasizes that communication via Lincos is only possible with beings that have at least reached the level of development of present-day man. Communication with beings endowed with intelligence of a lower level than ours will require some other language.

There can be no doubt that its creator will be a mathematical logician, a mathematician, or a specialist in semiotics, the science of sign systems. The most abstract of our sciences, those far removed from practical everyday affairs, would seem to be coming to the aid of man both in his terrestrial activities and his cosmic ventures.

Difficult though the problems of communication with beings in other worlds are, they are solvable, just as the problems of communication among the peoples of the earth are. The Galaxy is our stellar home. Just as the peoples of the earth are children of one planet, and the great multitude of languages they speak are comprehensible, so the dwellers of all the worlds in our Galaxy should be able to understand each other in some way, for they are all children of one great universe.

* *
*

The vehicles of communication in human society include gestures, the beating of drums, whistles and whispers, traffic lights, and road signs, the expression on one's face, and the language of human behaviour, the language of machines, the language of numbers, programmes, algorithms.

But not one of these modes of communication can compare in flexibility, universality and efficiency to our common everyday human language. That has been the topic of this book, a story of our rich linguistic heritage, of the precise methods of investigation that have helped to uncover the secrets of this 'wonder of wonders'.

But though the mysteries have been revealed, we still stand in amazement of language. It compels us to love it still more, though its secrets be disclosed, to admire and strive to penetrate into the finest subtleties of one's mother tongue and to learn the languages of other peoples.

SUBJECT INDEX

NAME INDEX

TO THE READER

Mir Publishers would be grateful for your comments on the content, translation and design of this book. We would also be pleased to receive any other suggestions you may wish to make. Our address is: 2, Pervy Rizhsky Pereulok, Moscow, USSR.

Printed in the Union of Soviet Socialist Republics

ABOUT THE BOOK

Over the ages man has wondered what lan guage is, how it is constructed, how one's moth er tongue differs from the languages of other peoples, and how the language of human beings differs from the signal cries of animals, and how our everyday speech differs from other media of communication in human society.

Thinking on these problems gave birth to linguistics, the science of the laws of language. Methods of investigating language improved with the development of linguistics and the ac cumulation of facts and knowledge. Today the latest tool is in the form of numbers and exact measures. Mathematical statistics and the theory of information, probability theory and mathematical logic, computers and sign theory are more and more coming to the aid of stu dents of language.

That is what our story is about.

MIR PUBLISHERS MOSCOW