JANUA LINGUARUM

STUDIA MEMORIAE

NICOLAI VAN WIJK DEDICATA

edenda curat

C. H. VAN SCHOONEVELD

Indiana University

Series Didactica, 7

INTRODUCTION TO THE STUDY OF LANGUAGE

by

TADEUSZ MILEWSKI

1973
MOUTON
THE HAGUE • PARIS
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSZAWA

This translation has been made from
Językoznawstwo
published in 1969 by Państwowe Wydawnictwo Naukowe
Warszawa
Translated by Marsha Brochwicz

Printed in Poland
(D.R.P)

CONTENTS

PART I

FOUNDATIONS OF LINGUISTICS

CHAPTER 1. THE DEFINITION OF LANGUAGE

Linguistics is the science of language; in order to understand the essence of linguistics, therefore, we must define what language is. When we listen to the speech of people in various countries, some we understand, while others we don't understand at all. We say that people whom we do not understand were raised in a social tradition different from ours, that they speak a language, passed down by their ancestors, different from ours. Thus, language manifests itself in speech, is an indispensable component of speech. Without a common language, there can be no comprehensible speech. We will define language as that which is simultaneously social, permanent, and abstract in speech. Speech is composed of other elements besides language.

Human speech, in its typical, complete form, is communication by sound between two individuals. Thus defined, speech involves several phases. Most indispensable is the process of speaking itself on the part of one of the persons in cummunication. Speaking can be defined as an individual process. It involves certain psychological processes in the speaker consisting primarily of a sequence of ideas represented by a definite set of sounds. Speaking involves the human speech organs as well—lungs, larynx, soft palate, tongue, lower jaw and lips.

Speaking is the primary, fundamental phase of speech. The second essential phase involves the receiver's, or adressee's, comprehension of the speaker's words. Sound waves set in motion by the person speaking reach the ears of the addressee and there, mediated by the ear-drum and other organs, evoke a sound impression in his mind. These sound impressions, in turn, lead to the formation in the adressee's mind of a sequence of ideas similar to those in the speaker's mind at the moment of speaking. Thus, we can define comprehension as a social process, for while speaking requires only one person, as in a monologue, comprehension assumes the presence of at least two conversing persons.

The third important phase of speech, after speaking and comprehension, is the text. That which the addressee comprehends constitutes a certain whole, and this product of speaking and comprehension—is the text. It is characteristic of human

products that they can be preserved, i.e., they can exist, to a certain extent at least, independent of the processes which produced them. A text possesses this property; it is, therefore, a product. Texts are preserved in the memory as a series of sentences together with their idea content and the circumstances accompanying them. There exist texts (e.g., ancient Indian hymns from the *Rig-Veda*) which have survived in human memory in unchanged form for hundreds of years. Writing is a more reliable form of preserving texts. At present, the oldest preserved texts, in Egyptian hieroglyphics on the Nile and in cuniform on the lower Euphrates, are over 5000 years old. A new means of preserving texts by recording them on records or magnetic tape has been introduced in the 20th century. A text is a concrete product of speech, because it preserves a particular concrete and unique thought, experience or event.

After speaking, comprehension, and text, the next phase of speech is language, the social and abstract product of speech. Texts are reproducible, but they cannot themselves serve as a direct basis for new texts. It is language, the system of words and rules abstracted from memorized texts, which constitutes the tool enabling us to create new texts. The ability to reproduce memorized texts does not by itself amount to proficiency in a given language. We have command of a language only when we are able to abstract from a large corpus of memorized sentences, individual words and grammatical rules, i.e., that which constitutes language. It is increasing capacity for constructing new intelligible sentences which are formally correct that indicates that we are making progress. By abstracting linguistic elements from memorized texts and placing them in a harmoniously constructed system, we arrive at language. Similar, if not identical, language systems exist in the minds of people of the same nationality, and existed in the minds of members of past generations, most important, of parents and teachers. Language, moreover, is embodied in memorized and written texts as an element which can be abstracted, isolated. This fact is exemplified by the Hittite language, in use between 2000 and 1000 B.C., unknown for the next three thousand years, and then, in the first half of the 20th century, abstracted from the cuniform texts by scholars. The Hittite language must have existed somehow in these texts, since it could be abstracted from them.

Since a given language exists, with only minor variations, in the minds of many people, and since it is embodied in texts, it is a supraindividual, i.e., a social product, a generalized system of norms of communication to which everyone must conform if incomprehensibility is to be avoided. Thus, language is an element of speech characterized by three aspects: it is social (in contradistinction to the individual process of speaking); it is a product, i.e., it is permanent (as opposed to non-permanent processes like speaking and comprehension); it is abstract (in contrast to the concrete processes of speaking and comprehension and to a concrete product, e.g., a text). Language is social, in that it involves a system of words and grammatical rules constituting norms of speaking which make communication among all the members of a given society possible. Language is a product in that it can survive ages without essentially changing, despite social upheavals. It is abstract in that

neither its words nor grammatical rules refer to specific concrete phenomena, but rather to abstract classes of phenomena (e.g., *horse* in general, *to go* in general) or to the general functions relating these classes to one another (e.g., the subject-predicate relation). The interrelation of the four phases of speech distinguished here can be represented schematically, as in Fig. 1.

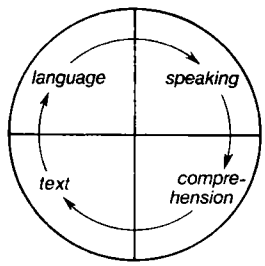


Fig. 1. The four phases of speech

The four phases of speech are interdependent in both origin and function. The abstract system of social norms embodied in language constitutes the basis for the process of speaking which, in turn, expresses individual, concrete thoughts and experiences. Because, however, content is expressed in speaking that conforms to the social norms of language, individual speaking constitutes the source of the social process of comprehension. The preserved product of comprehension constitutes the text, from which the norms of language are abstracted, and these norms constitute the basis for further speaking. The four phases of speech, following one after the other, form a closed circle in which they interact in such a way that it is difficult to precisely define the position of language in the total phenomenon of speech without first analyzing the distinctions amoung various types of signs.

Note: Terminology introduced in Chapter 1: Speech—communication by sound between two individuals, in which one tells something to the other. The four phases of speech: (1) speaking—an individual process, (2) comprehension—a social process, (3) text—a concrete product, (4) language—a social and abstract product. Linguistics—the science of language. How are the four phases of speech related to one another?

CHAPTER 2. THE RELATIONSHIP OF LANGUAGE TO OTHER TYPES OF SIGNS

A large role in human life is played by certain phenomena the meaning of which lies, not in the nature of the phenomena themselves, but in the fact that they call our attention to something else apart from themselves, often something involving a completely different field of reality. We call these phenomena signs. Their essence lies in the combination of two phenomena: the designating form, which directs our attention, and the designated content, to which our attention is directed,

Railroad signals, for example, are typical signs. As our train approaches a station at night, we notice a red light. It is a form of sign, to which a meaning content is attached: the track is occupied by another train—no through traffic. The form of the sign, a red light, is a matter of such indifference to us that it exists in our consciousness as a mere substitute for the message—track occupied. It is the form of the sign that enables us to recognize the message.

The form of a sign, like every other phenomenon, can be recognized by virtue of the fact that it can be distinguished from other phenomena. A sign can be distinguished, in the first place, from anything which is not a sign and which does not direct our attention. The red light of the railroad appears in the dark and is surrounded by darkness. The darkness is not a sign, but the contrast between the red light and the darkness enables us to recognize the signal. If everything were flooded with red light, the railroad signal would be unrecognizable. A sign, however, is usually contrasted in our minds, not only with things which are not signs, but also with other signs, which direct our attention to other message contents. Thus, the red light of the railroad signal, designating "track occupied", is distinguished from a green light, which designates "track free". The situation in reference to which these signals function—a train approaching the station at night—is the same for both, but the two signals give the opposite information, each directs our attention to a different message content. The existence of one of the signs makes that of the other indispensable, for if a signal "track occupied" exists, the opposite signal "track free" becomes necessary for the normal functioning of train traffic. Such signs, which set the conditions for each other's existence, constitute a system of signs. One cannot exist without the other, yet each can be distinguished from the other; we call this relationship between signs opposition. Red and green railroad signals, in constituting a system, stand, within the limits of this system, in opposition to each other.

In linguistic science, the word "code" has lately become common to denote a system of signs. This term is used in a very broad sense, and encompasses all systems of signs functioning in the realm of man, animal, or machine. It follows that every language, for example, Polish, is a certain type of code, and that all the languages of the world comprise a group of codes related to one another by certain shared characteristics which distinguish them from other, non-language codes. In order to isolate the properties characteristic of language codes, we must first present a general classification of all codes.

The numerous properties of codes proposed by various investigators as the bases for their classification may be divided into two main categories. One category of properties pertains to the information channel, i.e., the path by which the form of the sign reaches the receiver; the other category concerns the structure and function of codes. These two categories of properties will be discussed in turn.

Because the character of the information channel has a purely technical significance for communication, it determines neither the structure nor the function of

the codes which make use of it. In principle, we have as many channels of information as we have senses, and, it follows, as many kinds of signs. The signs perceived by the two senses most important to man—sight and hearing—can be further divided, on the basis of their form. Thus, visual signs can be divided into transitory signs, which appear and then disappear immediately, and permanent signs, which, having appeared, exist for a certain period of time; auditory signs can be divided into vocal-auditory and instrumental-auditory. These different types of signs deserve closer attention.

Transistory visual codes include various types of gesture and mimicry among animals and humans alike, from gestures of summoning, to the expression of feeling in dance, to directional gestures which orient us in space and which, in humans, involves pointing with the finger. In favorable conditions, such gestures develop into mimicry, best observed in deaf-mutes. In the bee dance, directional gestures have developed into an elaborate system of signs by means of which bees, upon returning to the hive, inform their hive-mates of the location of large quantities of pollen and nectar. This is perhaps the most highly perfected code in the animal world.

Tracks left on damp ground by animals and humans represent a primitive form of permanent visual signs. Roadsigns, directional arrows, colored tourist trail markers as well as various details of dress indicating that the wearer belongs to a particular social group (coats-of-arms, national emblems, distinctions, orders, etc.) represent a higher form of permanent visual signs. Railroad, ship and army signals, which can be prolonged at will, represent a still higher form. In this category we must include the plastic arts as well: sculpture, painting and graphics. In the course of the last six thousand years, writing developed from drawings presenting various objects or depicting the ownership of certain things by particular persons. As writing developed, it became more and more closely related to language. The oldest form of writing is hieroglyphic writing, combining depictions of objects and graphic equivalents of speech sounds, as found in Egyptian hieroglyphics. Sound equivalents played a greater and greater role in the evolution of writing up to the moment when writing took on the form of the syllabary (in which each sign designates a particular syllable). In Syria, c. 1300 B.C., one such syllabary developed into an alphabet (in which signs designate individual sounds of speech). All of the alphabets used in the world today originate from that Semitic alphabet. One realm of writing—that of numbers—proceeded along a different line of evolution and, enriched by signs used in the exposition of various sciences, formed, together with these signs, a complex code, difficult at times to translate into normal language.

Vocal-auditory signs are produced by human and animal vocal organs, the respiratory organs in fact, which are simply used to perform the function of vocalizing. In practically all birds and animals, the receiving organ is the outer ear. The use made of this information apparatus in these two groups of vertebrates varies considerably. A high level of development has been reached by the system of vocal-auditory signs connected with mating habits in the songs of certain species of birds. Primates

demonstrate a particularly conspicuous development of this type of code. Among primates, the systems of cries of gibbons and other monkeys represent an earlier stage of development, while human language represents a later stage. In the 20th century, recording speech signs on phonograph records and magnetic tape has become popularized. This constituted the origin of auditory signs that are permanent, or at least reproducible.

Instrumental-auditory signs are produced by instruments and received by the ears. Such signs exist only in human society. Among them are included drum and trumpet signals used in the armies of various nations and epochs. Certain Negro tribes in the Sudan (e.g., Yaunde and Ewe) have highly elaborated drum signals by means of which messages are sent over various distances, as though by wireless telephone. The principal domain of instrumental-auditory signs, however, is that of instrumental music, which, in the 18th and 19th centuries, developed into an immeasurably complex and subtle code. Instrumental-auditory codes developed from vocal-auditory codes. The drum signals of Negroes arose from ordinary speech, while instrumental music developed from song. Reproductions of speech on records and tape, discussed above, stand on the border-line between the two types of signs—instrumental-auditory and vocal-auditory—for although, in such reproductions, the sign originates in the human vocal apparatus, the reproduction itself is produced by recording instruments.

Because they are received by dual organs—two eyes or two ears—both visual and auditory signs aid in space orientation. The reception of a sign at two points enables one to determine, not only the direction, but also the distance from the sign's point of origin. The perception is then specified by the information channel. Two limiting points are clear to the receiver, the point of origin of the visual or auditory sign, and the place where he himself is standing. These are the spatial boundaries of each act of information, and, therefore, of each act of speech. They are of fundamental significance for the structure of codes.

The differences, on the other hand, among various information channels are of secondary significance in that the same code may be realized by means of different channels. Such varying forms of a single code, executed through different channels, constitute its subcodes. The principal subcode is that which is most frequently used. The following five subcodes of language, executed through various information channels, can be listed: (1) the transistory visual subcode—e.g., the sign system of deaf-mutes, (2) the permanent visual subcode—e.g., writing, (3) the principal, vocal-auditory subcode—spoken language, (4) the instrumental-auditory subcode—e.g., Negro drum signals, (5) the tactile subcode—e.g., the Braille alphabet for the blind. Thus, we see that the information channel utilized by a given code is determined by the technical conditions in which it must function, that is, transmit information, as a result of which new subcodes continually arise as conditions change. New systems of writing are constantly being introduced (e.g., Morse code, various written ciphers) and even, in the case of pathological damage to the speech organs, the handicapped

individual creates for himself a new vocal-auditory subcode in which the sounds of the abnormal speech, although different from the sounds of normal speech, correspond to them functionally.

Having discussed the external and inessential features of codes, those related to their information channels, we will now discuss their essential features, those concerned with the structure and function of codes. The classification of sign systems based on these features is of fundamental significance.

In classifying signs in terms of their structure and function, we first divide them into two main types: symptoms and signals. Several pairs of contrasting features distinguishing these two types of signs from each other may be mentioned:

(1) Symptoms constitute an integral part of a complex of phenomena the basic function of which is usually biological; their sign function is incidental. Signals, on the other hand, fulfill no function aside from the sign function, in which they are specialized.

(2) Symptoms are never transmitted by anyone with the intention of informing a receiver; they are unintentional. They are simply phenomena arising from certain causes, and they become signs only in the presence of a receiver who associates them with a particular meaning content.

Symptoms are one-directional signs. Signals, on the other hand are produced by the sender with the intent of producing an effect upon the receiver. Because they are signs for both sender and receiver, who associate them with the same meaning content, they are two-directional.

(3) Symptoms are not intentionally emitted by a sender, and, since in the normal sense of the word, no sender exists, symptoms cannot be recognized by the sender as signs. Thus, symptoms are non-reciprocal. A person is capable of recognizing, to a certain extent, symptoms which manifest themselves in his person, but he will perceive only a part of what others perceive. Signals, however, are reciprocal. The sender who produces signals intentionally must himself receive them, for purposes of control. One cannot very well produce signals which one is incapable of receiving.

Here are a few examples. A crow hops about, crying pitifully—we see that it has a broken wing. A dog pants, his tongue hanging out—we say that he is hot. We see human footprints on the damp ground—we know that someone has passed by. A child has flushed cheeks; we feel the child's hand—it is hot—we understand that the child is ill. The patient sitting in the dentist's chair lets out a sudden cry—the dentist knows that his drill has touched a nerve. Upon analyzing all of these signs, we see that they are simultaneously complex, non-intentional and non-reciprocal, thus they are symptoms in the sense of the word outlined above. They are signs only for the receiver, Symptoms may constitute extensive codes, but only in the mind of a receiver, e.g., the system of disease symptoms in a doctor's memory, or the system of animal tracks in a hunter's mind.

While symptoms make up one-directional codes, existing only in the receiver, signals, which we will discuss next, make up two-directional codes, existing in both

the sender and the receiver. All other features of signals—specialization, intentionality and reciprocity—are determined by this fundamental property. Signals can be divided into two large categories: non-semantic appeal signals and semantic signals. The basic difference between them involves their relation to reality. The form of semantic signals directs the receiver's attention to some phenomenon in the surrounding world. The meaning content of the signal, that which it designates, is the conception of some external phenomenon, a conception which, because it is shared by all the members of a given society, and, most important, by the sender and receiver, has a supraindividual, objective character. Appeal signals, on the other hand, do not refer to the external world, but rather evoke certain defined changes in the mind and behavior of the receiver. Music is a typical code of appeal signals, especially the classical and Romantic music of the 18th and 19th centuries, the primary function of which was to play on the listener's emotions. Dance also belongs to this category, and, in the field of plastic arts, 20th century decorative and abstract art.

Semantic signals can be divided into two categories: motivated signals, or images, and unmotivated, or arbitrary signals. In images, the form embodies features which correspond to the features of the phenomenon referred to, to a great enough extent to enable the receiver, on the basis of these correspondences, to identify the image with its meaning content, i.e., the thing designated. The characteristics of the form are motivated, determined by their similarity to the thing designated. The characteristics of the form of arbitrary signals, on the other hand, have nothing to do with the characteristics of its meaning content, they are not motivated or determined by the thing designated.

The categories of signs so far discussed—symptoms, appeal signals and images are all motivated, i.e., the meaning content which they impart to the receiver is determined by the characteristics of their form; the category of signs which we are about to discuss—the various types of arbitrary signals—are unmotivated. The property of being unmotivated is related to three other distinguishing properties of arbitrary signs—they are conventional, diacritic and interchangeable.

(1) In motivated signs—symptoms, appeal signals and images—the form of the sign embodies the causal experience which constitutes its meaning content. The question then arises—what relates the form to the meaning content in arbitrary signs? In such signs, a certain correspondence, previously established between the sender and the receiver, must exist enabling the receiver to react to the signals of the sender. The animals of a given species face the same situations generation after generation, and, because they are identical in organic structure, they react to these situations similarly. The identical character of these situations (hunger, sexuality, hostile attack) and of the organisms and their reactions constitute the basis for the connection between the form and meaning content of signs which are identical in both the sender and the receiver—thus making communication possible. Something similar occurs in human beings where childish expressions of the type *mama* are involved. The first sounds produced by the child are the labials *m*, *b*, *p*, the lips

being especially active in sucking; the first person whom he sees is the mother. The constantly repetitive phenomenon of sucking, common to all children, and certain characteristics of the speech organs (lip articulation) are such, that in children of all nationalities, the image of mother is associated with the sound *mama*, and that of other relatives with the sounds *papa*, *baba*, etc. This case, in which the connection between form and meaning content is determined by natural cause, is exceptional among people, for generally the complicated details of language are dictated by social tradition. Younger generations learn the specifics of language from their elders, thereby entering into a certain social contract with them concerning the connection between the form and meaning content of arbitrary signs which, by virtue of this fact, are contractual, i.e., conventional.

(2) Motivated signs—symptoms, appeal signals and images—are non-diacritic. Their form makes use of a continuous scale of visual and auditory features generally blending into one another and evoking a particular reaction in the receiver. Arbitrary signals, on the other hand, are diacritic. The phenomena constituting the form of these signals possess two categories of features. The first category includes those features which designate or give meaning, features connected with the meaning content by convention. Having been chosen from the continuous scale of visual and auditory features of phenomena and narrowly defined by convention, they are relatively few in number; and, because they have been established by convention, their number cannot be increased in the course of communication. It is with these features alone that the meaning content of signs is connected, and only these features play a role in communication. The second category, that of non-designating features, on the other hand, exists quite apart from the conventional system of arbitrary signs and plays no part in communication.

(3) Motivated signs—symptoms, appeal signals and images—are non-recurring phenomena which evoke definite reactions in the receiver by virtue of their characteristic features. Signs may occur which are similar, but they will not be identical. Because they are non-recurring, motivated signs are one-directional. The addressee receives them, but he is unable to reproduce them. Arbitrary signs, on the other hand, may be reproduced indefinitely since each reproduction simply involves repeating a particular convention established between the sender and the receiver, or addressee. The addressee receives conventional signs and produces them in turn—thus, they are two-directional, interchangeable like money.

Certain works in the plastic arts—sculpture, painting and graphics—are typical images. Realistic portraits in particular, as well as landscapes and scenes taken from life are such that it is possible to precisely associate the form with the phenomenon represented on the sole basis of similarity. There is also a type of music, so-called program music, which operates in terms of sound images the acoustical features enable them to be associated with certain phenomena, e.g., the roar of a waterfall, the bombarding of cities during an air-raid, etc. Visual images can be contrasted to conventional writing, as acoustical images, to spoken language. The

distinction is clear. Visual or musical images are non-diacritic, i.e., they make use of a continuous range of shapes and colors or of pitches and volumes. Each feature of such an image is meaningful to the extent that it resembles the phenomenon which constitutes the meaning content of the image. Every image is a non-recurring phenomenon. Because it cannot be identically reproduced, it is one-directional and non-recurring. Writing and spoken language, on the other hand, make use of a narrowly limited number of conventional features—letters in the case of writing, vowels and consonants in the case of spoken language. These conventional signs may be reproduced indefinitely; they are totally interchangeable.

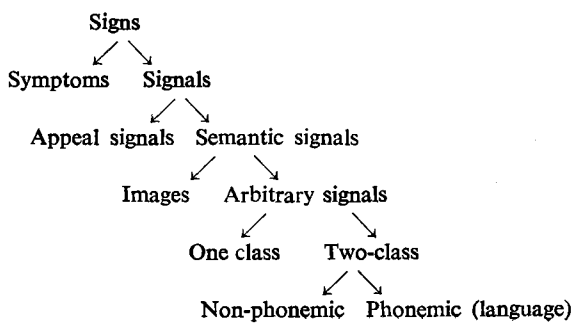
Arbitrary semantic signals, to which we will now turn our discussion, can be divided into two categories. Here, one-class signals are opposed to two-class signals. In one-class systems, i.e., closed, non-productive systems, as, for example, the system of gibbon cries or the system of railroad signals, the number of signs is strictly limited. Because, from the systemic point of view, these signs cannot be divided into smaller meaningful units, they are uniform in type, they belong to a single class of signs which constitutes the only class in the system. Every signal consisting of a definite set of formal features always bears the same information in reference to a given situation, e.g., the cry of a gibbon warning the troop of approaching danger, or a railroad signal (red light) informing us that the track is occupied. In this latter system—night-time railroad signals—we have only two signals, red and green, opposite in meaning. This system, therefore, is essentially closed. In two-class systems—open, productive systems—on the other hand, like the bee dance or human language, we have two types of signs—simple and complex—belonging to two distinct categories. In the first category, that of simple signs (e.g., the words of a language), the signs refer to a certain class of phenomena in the surrounding world; they may also be combined according to certain rules into an endless variety of complex signs of the second class (e.g., statements) which refer to definite, concrete and non-recurring phenomena in the world. This second class is essentially open and productive in that, through the combination of simple signs, a practically unlimited number of complex signs arises, which constitute the basic means of language communication.

Two-class codes include two types of elements. The first type includes simple signs, while the second includes the means of connecting simple signs of the first class to form complex signs of the second class. These second-class signs are not as a whole determined by the code; they are, rather, freely created, but meaningful nevertheless, because of the fact that the elements of which they are composed, i.e., simple signs and the means for combining them, are established by the conventional code.

The second distinction between one-class and two-class signals concerns their space-time range. One-class signals, strictly bound to a presently existing consituation in the immediate vicinity of the receiver and sender, in principle, refer only to things which are contemporary and near-by, which lie within the field of vision. Two-class codes, on the other hand, infinitely richer in signals and, at the same

time, not determined by the immediate consituation, make possible the formation of complex signs which designate phenomena outside the field of vision, distant in space and in time. The danger cry of the gibbon or other monkey, a cry belonging to a one-class system, always signifies that the enemy is near at the moment when the cry is produced. A gibbon is unable to report events that are distant in time or space. Bees, however, with their bee dance, or man, with his language can tell about things distant in time or space because of the greater richness of their two-class systems. Thus, there are two types of two-class codes in existence on our planet—the bee dance and human language. The comparison of these two codes might be of great interest, but, unfortunately, such a comparison is possible only in part, in that the bee dance is, and probably always will be, a partial mystery to us. Among the differences distinguishing language from the bee dance and unequivocally defining language, the most important in that language is phonemic, while the bee dance is non-phonemic. The forms of the immense number of words in a language which have defined semantic meanings are made up of a relatively small number of elements, i.e., sets of simultaneously occurring diacritic features functioning to distinguish and separate words from one another. These elements are called phonemes. Phonemes, as opposed to words, are absolutely devoid of permanent meaning. Various combinations of the three Polish phonemes *a*, *k*, *t*, for example, give us three individual words—*tak*, *kat*, *akt*. There is no equivalent of this aspect of language in the non-phonemic bee dance, which consists only of movements similar in function to our directing gestures. The bee dance involves no non-meaningful, conventional elements which, like phonemes, combine to form meaningful units. Language is the only known phonemic code and, as such, can be distinguished from all other types of signs. Only language codes are characterized by a hierarchical structure consisting of three levels of units gradually increasing in size—phonemes, words and statements. The smallest units are phonemes, devoid of meaning. Combinations of phonemes produce words which refer to particular classes of phenomena. Words combine to form true statements, complex language signs, which inform the receiver of concrete, actual ewents occurring in the surrounding world.

In summary, we can establish the following general classification of signs based on their structures and functions:



Thus, signs are divided into the following six groups: (1) symptoms, (2) appeal signals, (3) images, (4) one-class signals, (5) non-phonemic two-class signals, (6) phonemic two-class signals. All existing evidence indicates that the order proposed here represents the chronological order in which these types of signs came into use in the general evolution of organic life. Specialized appeal signals developed from symptoms, semantic images, from appeal signals, arbitrary one-class signals, from images, and finally, in human society, these arbitrary one-class signals developed into two-class phonological systems, i.e., languages.

Upon analysis, we can observe that all the categories of signs described here occur in speech. Moreover, in the development of the child, these categories of signs do not appear simultaneously, but, rather, in the same order in which they emerged in the course, covering millions of years, of the evolution of organic life. Here, in accordance with the well known developmental tendency, ontogeny, i.e., individual development, recapitulates philogeny, i.e., the evolution of forms of life on our planet.

By speech, we mean all the signs produced by the human speech apparatus. All of these categories of signs interact with one another in speech forming an infinitely complex set in which the role of particular sign categories varies from moment to moment. Sometimes one, at other times another of these categories of signs determines the basic tone of speech, while other categories are relegated, meanwhile, to a secondary position. Particular categories of signs make their appearance in the speech of children, and in such a way that the appearance of new categories limits but does not eliminate the functioning of older categories. As a result, adult human speech is the richest and most variegated set of signs known.

The first signs which appear in the development of the child's speech are symptoms. A child cries from the moment of birth and this crying slowly becomes differentiated. The principal function of the child's crying is not at all that of designating, which at this stage of development has not yet become a separate, specialized function. In relation to the total psycho-physical organism, the child's crying is an integral part of his experience. Affective stimuli make the child cry, which reduces tension, thus bringing the child relief. In the first phase of his development, the child is not aware of the effect which his cry has on his environment; he has no conception of its function as a sign. In this stage, therefore, there is no intentional production and control of signs. In the first months, the child's crying is merely a symptom informing the environment of his psychological and physical state of being. As with all symptoms, here we have only a receiver (a person in the child's environment); as yet, we have no conscious sender. The mother distinguishes various nuances in the child's crying, each of which informs her of a different state of his organism. These symptoms, therefore, form in her mind a certain one-directional code which does not yet exist in the child's mind.

This situation changes when the child, between the first and second years of life, notices that his crying has a certain effect on his surroundings. From that moment,

he begins to cry intentionally, in order to exert the desired influence on the behavior of his mother and other persons. Now, his cries constitute appeal signals, which begin to play a greater and greater role in the child's speech alongside the former symptoms, which of course remain. And so it may happen that the child falls down but, at the moment of falling, does not cry but, instead, looks around to see whether or not a familiar person is in the vicinity, and only upon seeing such a person, bursts into tears, which, of course, constitutes an appeal for help directed toward a particular person, who is expected to change his or her behavior—to run and pick up the child. Such appeal signals constitute a two-directional code—they involve a sender (child) and a receiver (person in the child's vicinity). These appeal signals are specialized, i.e., separated out from the total reaction of the organism, intentionally produced by the child, who produces them when he wishes and as he wishes, and, therefore, must himself receive them.

Toward the end of the second year of life, the development of the normal child takes a new turn. The child begins to grasp the semantic function of speech; he notices that certain elements of speech refer to particular phenomena in his surrounding world. Imitation constitutes the first means of producing semantic signs understood by the child. The child imitates movements in his surroundings, and, along with the general variety of movements imitated, appear movements of the speech organs from which sound imitating acoustical images, onomatopoeia, arise. The child imitates the sounds produced by humans and animals alike, and, in these sound images, a certain, as yet undefined, semantic function emerges.

In the further course of development, the child proceeds toward the mastery of arbitrary semantic signs in which the connection between the acoustical form and the objective meaning content is based, not on any similarity resulting from imitation, but on an association made at a particular moment of time. The first associations of this type are made independently of a social tradition. In the first sounds produced by the child, a large role is played by lip compression, developed in sucking. In this way, complexes arise which are composed of repeating labial sounds occurring at the beginning of syllables, e.g., *mama*, *papa*, *baba*. These complexes, originally symptoms, and later appeals for help, food, etc., acquire a semantic function, as previously described, in the further stage of the child's development. The child associates these complexes with those persons whom he most frequently sees, mother and father. Later on in the child's development, sounds appear which are associated with persons in the child's broader surroundings, like *niania* (nanny), *teta*, a childish form of *ciotka* (aunt). At this stage, all of these sound complexes are elements of a one-class system and, as such, cannot be compared with either the words or the sentences of adult language. They comprise a closed code of semantic signals which, within the context of certain typical consituations, become associated with parental functions and with the satisfaction of particular physiological needs of the child, such as hunger and other natural needs. They are at least partially independent of social tradition, as evidenced by the fact that they appear in nearly all the human

groups in the world. They resemble the systems of cries of gibbons and other monkeys, which also constitute one-class arbitrary codes which are not handed down by social tradition. In all of these cases, certain sounds arise in response to biological stimuli within the context of particular consituations, thereby becoming associated with these consituations.

A turning point comes in the child's life when he begins to adopt the traditional language of his environment. Heretofore, he has made use of the non-diacritic, continuous scale of sound features. His voice has been highly flexible and has produced with ease an infinite number of various transitional sounds fulfilling the function of symptoms, appeal signals and acoustical images. Once the child understands the essence of the semantic function, he begins to adopt the language signs of his social environment, which are based on social tradition, and are, therefore, conventional. The meaning content of these signs is related by convention to only certain features of their accoustical form, to those features which, by convention, are diacritic. The continuous scale of sounds now ceases to play a role, for only certain chosen features of the scale function to separate and distinguish conventional signs. Such features, which we call diacritic or phonological features, are different, at least in part, in different languages. In Polish, for example, voicing is one diacritic feature serving to distinguish words. In Polish, two words can be distinguished on the sole basis of whether a consonant is voiced or voiceless—e.g., *gość* (guest) (first consonant voiced) and *kość* (bone) (first consonant voiceless). In the Aztec language of Mexico, no distinction is made between voiced and voiceless consonants, and, therefore, words cannot be distinguished on the basis of this feature. In Latin, duration of vowels, i.e., the difference between long and short vowels, constitutes a diacritic feature related to differences in meaning—e.g., *venit* (he comes) (short *e*) and *vēnit* (he came) (long *ē*); in Polish this feature does not constitute a diacritic difference.

In adopting conventional semantic signs, i.e., in learning the language of his environment, the child must, first of all, learn to distinguish and produce the diacritic features of the adult language which he is learning. This does not, of course, mean that he ceases to use the continuous, non-diacritic scale of sounds. A certain state arises, persisting throughout the individual's lifetime, in which both scales—diacritic and non-diacritic—are simultaneously made use of on different planes. The signs of a three-year-old child's speech, just as those of adult speech, continue to function as symptoms and appeal signals; there also exist, to a certain extent, accoustical images, or onomatopoeia, which function semantically by virtue of the similarity between their form and their meaning content. The continuous, non-diacritic scale, which in everyday speech we refer to as the individual tone of voice of particular persons at a given moment, continues to function within the realm of these three categories of signs. At the same time, the child, from the beginning of the third year of life, begins to use conventional language signs, which make use of the non-continuous, diacritic scale of accoustical features. This brings about

a great change in his speech. The hitherto existing infinite ease with which he produced sounds disappears. Diacritic phonological features which function in conventional signs and which, therefore, must be constantly reproduced, now constitute the framework of speech. These features, however, represent only certain aspects of the capabilities of our speech organs. Other aspects, for example, pitch level, are not included among phonological features and function within the continuous, non-diacritic scale. The diacritic scale limits, but does not replace, the functioning of the non-diacritic scale. One scale operates where certain features of sounds are concerned, and the other scale operates where other features are concerned. The child gradually learns to use both scales and to coordinate them.

In mastering the phonological features of sounds as well as the entire phonological system of a language, the child adopts the conventional signs of adults, composed, as we have seen, of a relatively small number of phonemes arranged in various orders. By the same token, he enters into communication with his environment, for conventional signs are two-directional, interchangeable. For the child, however, at the beginning, these are one-class, indivisible signs. He knows several language complexes and he knows that in reference to a certain situation each of these complexes refers to a certain objective state of things. Another great task now awaits him—to analyse these complexes. The first step in this task involves comparing language complexes with one another. When the child has memorized a sizeable number of language complexes, he compares them with one another and notices that a certain part of these complexes are identical, while other parts are different—e.g., *mama idzie* (mother is coming) and *mama siedzi* (mother is sitting) where *mama* is the part which these two complexes have in common and *idzie* and *siedzi* are the parts in which they differ. In this way, each language complex can be divided into two parts. Such analysis concerns both the form and the meaning content of speech signs, and, as a result of such analysis, each isolated element of form can be associated with its appropriate meaning content. The child notices, moreover, that the individual elements of the language complex remain in a particular formal and semantic relationship to one another, in this way, the child proceeds from a one-class to a two-class system in which he is able to distinguish the class of simple signs—words and the grammatical agents by means of which words are combined to form complex signs, i.e., sentences. Consequently, the hitherto closed system becomes open, i.e., absolutely unlimited in range. The consituation, indispensable to the functioning of the closed, one-class system, now ceases to be a prerequisite for communication, thus enabling signs to refer to phenomena distant in time and space. At this stage, the child has mastered language, such that in the following years, it remains only for him to supplement his knowledge with a large number of details in order to finally mature into a normal partner in conversation.

From the point of view of both species development, or philogeny, and individual development, or ontogeny, language—the two-class semantic conventional code—is the latest form to appear. Its development rests upon that of the series

of older codes—symptoms, appeal signals, accoustical images, and one-class signals—which functioned earlier. These old sign systems do not cease to exist, but, instead, participate in language in the speech of adults, which constitutes an orchestra of all types of signs.

Note: Terminology introduced in Chapter 2: Sign—the combination of two phenomena: the designating form and the designated content. Signs which set the conditions for one another's existence constitute a system of signs, i.e., a code. Signs belonging to the same system but differing from each other, create an opposition within the system. The information channel is the path by which the form of the sign reaches the receiver. Varying forms of a single code, executed through different information channels, are called its subcodes. Types of signs: (1) symptoms, which are signs only for the receiver, (2) signals, which are signs for both the receiver and the sender. Types of signals: (1) appeal signals, which refer to the emotional content of the receiver, (2) semantic signals, which refer to phenomena in the surrounding world. Semantic signals can be divided into: (1) images, which are related to the phenomena designated by virtue of features which they have in common with these phenomena, (2) arbitrary signals, which are connected with their meaning content on the basis of a previous association with them. Arbitrary signals can be divided into: (1) one-class arbitrary signals, the number of which in the system is limited, (2) two-class, complex arbitrary signals, the number of which is unlimited. Two-class systems can be divided into: (1) non-phonemic, (2) phonemic. Phoneme—a set of simultaneously occurring diacritic features functioning to distinguish and separate signs from one another. Codes of semantic, arbitrary, two-class phonemic signals are called languages. What is the relationship among the various types of signs in speech?

Chapter 3. THE HISTORY OF LINGUISTICS

We have defined language as a system of social norms, permanent and abstract, regulating that category of speech signs which are characterized as semantic, conventional, two-class phonemic signals. The science which is concerned with codes possessing these properties is called linguistics. Language is an aspect of speech, and at the same time it is a type of code. In order, therefore, to clearly define the position of linguistics, it is necessary to discuss those sciences which deal with various types of signs and various phases of speech.

At the beginning of the 20th century Ferdinand de Saussure proposed the creation of an independent science—semiology (from the Greek *sēmeîon*, 'sign')—which would investigate "the life of signs in the life of society". This proposal was realized in the period 1945–1960 owing to the appearance of information theory, formulated primarily by Claude Shannon and Norbert Wiener. Information theory is the general science of signs which function in the realm of animals, men and machines, a science concerned with all forms of information transmission. At its present stage of development, information theory utilizes primarily methods borrowed from mathematics, (probability calculus, mathematical statistics, algebra and mathematical logic), thus representing a field of applied mathematics encompassing the domain

of signs and information. From this point of view, linguistics represents that part of information theory limited to the study of language codes. This conception of linguistics has led, in recent years, to development along the lines of mathematical linguistics, which constitutes a practical application of information theory to language research. Although, at its present level of development as regards strictly deductive methods, mathematical linguistics represents only one of many fields of linguistics, it is possible that, in the future, information theory, suitably expanded, will become the general foundation for all research on language.

Connected, in a certain way, with the general science of signs is a group of sciences concerned with various phases of speech. Phonetics and certain fields of psychology analyse the processes of speech and understanding. Phonetics studies the sounds of speech as natural phenomena. It analyses the phonetic possibilities of the human speech apparatus, describing the behavior of the speech organs during the articulation of various sounds and classifying them. For linguists, phonetics is an auxiliary science, the results of which are indispensable to their work. There are three fields of psychology that are likewise significant for the linguist: (1) animal psychology, which analyses, among other things, animal communication, (2) psychological research on speech development in children, (3) the psychopathology of speech, involving studies of disturbances in speech and understanding resulting from damage to the cerebral cortex, a phenomenon referred to by the general term aphasia.

As opposed to phonetics and psychology, which investigate the processes of speech, stylistics and linguistics analyse the products of speech. Stylistics investigates the extent to which texts represent more than the mere reproduction of language norms and constitute the result of a choice of forms and their careful arrangement on the part of the person speaking or writing. Stylistics defines the principles of choice and arrangement of forms in the texts analysed, principles not included in the set of language norms and referring exclusively to a certain text or group of texts. Linguistics, on the other hand, which is our sole concern here, takes as its subject the system of language norms itself appearing in any given text, but having broader implications as norms obligatory in the whole society.

Linguistics can be divided into descriptive and comparative linguistics. Descriptive linguistics describes each system of language individually, while comparative linguistics compares these systems with one another and classifies them. Comparative linguistics, so defined can be divided into two fields: historical linguistics, concerned with the historical classification of languages according to the degree to which they are of common origin, and typological linguistics, which classifies languages according to their structural similarities. These three fields of linguistics, i.e., descriptive, historical and typological, will be discussed individually, but first, we will present a short history of these sciences.

The earliest systematic investigation of language developed in India around 500 B.C. There, it was believed that the recitation of sacred hymns from the *Vedas* fulfilled their religious-magical function only when performed with absolute linguistic

accuracy. Thus, the need arose for defining the rules of correct speech, and this became a stimulus to phonetic investigations. Hindu phoneticists made use of visual and tactile observation of the speech organs during the articulation of sounds and, on this basis, arrived at sophisticated descriptions of the articulation of particular sounds as well as a consistent system for classifying them.

In time, phonetic observations led to the theoretical formulation of language problems. As early as the 5th century B.C., PATANJALI introduced the term *sphoṭa* meaning a sound form functioning as a sign. The smallest element of this designating form is defined as a *varṇa-sphoṭa-*, 'letter-sound' and is clearly distinguished from *dhvani-*, 'speech-sound'—which is equivalent to the present-day distinction between phonemes and sounds. Hindu grammarians described *varṇa-sphoṭa* as an element devoid of independent meaning but nevertheless possessing a certain designating function in so far as its substitution by another element of the same type sometimes produces a completely different word, and failure on the part of the listener to hear this element, may prevent him from understanding the meaning of the word.

In the 5th century B.C. in India, the sacred language, Sanskrit, disappeared from daily use, at the same time remaining as the vehicle for all intellectual life. Sanskrit had to be specially learned and, therefore, in order to facilitate such study, descriptive-normative grammars began to be compiled, describing not only how the language was in fact spoken, but also, how it should be spoken. A synthesis of these various compilations was worked out by PĀṆINI, who probably lived in the 4th century B.C. and who included in his grammar nearly four thousand short rules to be memorized. In Sanskrit, the word for grammar is *vyākaraṇa-*, 'division', 'analysis'—which fact alone indicates that the main concern of grammar was the morphological analysis of the word, breaking it down into the smallest units possessing an independent meaning value, that we term morphemes. In his investigations into the nature of morphemes and their designating function, Pāṇini outdistanced all his predecessors. The further development of grammar in India involved primarily popularizing Pāṇini's work.

Another center of language investigations in antiquity was Greece. In the 4th and 5th centuries B.C., Greek philosophers were concerned with the question of the relation of words to things. Two theories existed. One theory considered this relationship a natural one (*phýsei*), while the other theory considered it conventional (*thései*). The first group were followers of HERACLITUS of Ephesus (540–480 B.C.), and the second, of DEMOCRITUS of Abdera (460–370 B.C.) Followers of the theory of natural connection between words and things understood words as necessary reactions of human nature to feelings and sense impressions, as in coughing, screaming and moaning. Followers of the theory of conventional connection, on the other hand, claimed that there is no necessary connection between the form of the word and the thing to which it refers, but that chance (*týkhē*) alone gave a given thing its name, and that a contract or convention (*thésis*) agreed upon by the members of society established these meanings as permanent. The dialogue of PLATO (428–

348 B.C.), *Cratylus*, in which the author confronts both traditional theories, represented a step forward in the development of notions about language.

The formulations of EPICURUS (341–270 B.C.) constitute the most important attempt in antiquity to reconcile these two theories. He distinguished two periods in the development of language—the first natural, and the second conventional. In the first period, words arose as automatic reactions of the human speech organs to certain feelings and impressions. Only in the next period did causal factors in social and intellectual life begin to play a role. In order to avoid ambiguity, each tribe established conventional meanings for the particular words which had arisen as spontaneous reactions and, in addition, introduced names for abstract notions existing only in the mind. The theory of Epicurus was not generally accepted. In antiquity, the conventional theory triumphed.

In the 5th and 4th centuries B.C., alongside the problem of the relation of words to things, emerged the problem of the structure of words themselves. Democritus, Plato and Aristotle (384–322 B.C.) developed the view that language is composed of indivisible sound units, devoid of meaning of their own, but capable of comprising meaningful series, i.e., words and sentences. Such a unit was called a *stoikheîon*, 'prime element', equivalent in meaning to our phoneme. According to Plato, human speech cannot be understood unless a definite number of individual *stoikheîa* is distinguished in the infinitely indivisible stream of sounds produced by the human voice, and, at the same time, we cannot recognize a single *stoikheîon* without recognizing them all. In this notion can be detected, in embryonic form, the concept of the phonological system.

In the second period of development of Greek linguistics, in the 3rd and 2nd centuries B.C., problems of language were the concern primarily of Alexandrian philologists—who confronted such problems in connection with the editing of classical Greek texts—and philosophers of the school of Stoicism in Athens, on Rhodes, and in Pergamum—who studied language within the framework of their broad notion of logic. Two schools of thought arose, the analogists and the anomalists. The Alexandrian philologists were adherents of the first school, while the Stoics were anomalists. The terms *analogía*, 'proportionality', and *anōmalía*, 'irregularity', come from mathematics and were introduced by the principal founder of the school of Stoicism, CHRYSIPPUS of Soli (282–208 B.C.) for solving grammatical-logical problems. In the following generations, however, these two terms became the focal points of two different concepts of language. The Alexandrian analogists held the view that regularity prevails in language, as attested by the total agreement between logical and grammatical categories. According to the anomalists, on the other hand, options obtain in language, and in many cases grammatical and logical categories do not correspond to one another. The analogists understood language as a system of proportional relations, a notion which constitutes an embryonic form of the modern concept of the language system. The anomalists, on the other hand, occupied themselves primarily with criticism of the oversimplified concept

of this system as presented by the analogists. The analogists claimed that given word endings always refer to certain conceptual categories. The anomalists contested this view by pointing out various exceptions. This in turn forced the analogists to define those classes of words in which given endings are exponents of given categories. In this way, they proceeded to establish categories of noun and verb inflection and to specify the words which are inflected in the different ways represented by these categories. Thus modified, the analogists' theory gained general recognition.

In Greece, those views which treated language as contractual, conventional, proportional and systematic triumphed. Only at this stage could grammar develop. The first formulation of grammar was accomplished by DIONYSIUS of Thrace (170–90 B.C.) around 100 B.C. in a work entitled *The Art of Grammar* (*Tékhnē Grammatikḗ*). His views were developed and supplemented by the greatest of Greek grammarians, APOLLONIUS Dyscolos ('The Crabbed'), who lived in the first half of the 2nd century A.D. In his principal work, *On Syntax* (*Perì Syntákseōs*), he presented the foundations of Greek syntax.

The true concern of Greek grammatical investigations was the function of words in a sentence, a subject most thoroughly embodied in their theory of the parts of speech. We can follow the gradual elaboration of this theory in the writings of Plato, Aristotle, Dionysius of Thrace and Apollonius, the ultimate form of which has been preserved, with slight changes only, to the present day. The Greek theory of parts of speech is based on three criteria. First, Apollonius presented the criterion of inflection, dividing words into non-inflected and inflected, and further dividing the latter category into declension according to case and conjugation according to person. The remaining parts of speech were defined on the basis of a syntactic criterion, i.e., the rules governing their use in the sentence, (adverb, preposition, conjunction), and on the basis of a semantic criterion, i.e., their meaning (e.g., *ónoma* 'name').

The tradition of the Greek school of grammar was carried on in Byzantium up until the middle of the 15th century; two other schools of grammar as well, the Latin school—which developed from the 2nd century B.C.—and the Judeo-Arabic school—which flourished between the 7th and the 12th centuries A.D.—owe their origin to the influence of the Greek school. Of these three medieval schools of grammar, the most vital was the Latin school. In the 13th century, on the basis of the tradition of these schools, the development of grammars for individual modern European languages began. The first work dedicated to Polish orthography was written by Jakub PARKOSZOWIC around 1440, and the first Polish grammar was published by Piotr STATORIUS-STOJEŃSKI in 1568. *Russian Grammar*, published by LOMONOSOV in 1755, was of critical significance to the development of literary Russian.

In the 16th century, the European school of linguistics absorbed the traditions of the Latin, Greek and Hebrew schools and, in the 17th century, with this tradition as a foundation, *General Grammar* was compiled, published in Paris in 1660 and

based on materials from French, Latin, Greek and Hebrew. This grammar confronts the similarities existing among these languages, analysing them by means of logical schemes.

Despite its many faults, *General Grammar* must be considered a manifestation of concern with certain problems reaching back into antiquity and leading, in time, to the creation of comparative linguistics. Quintilian, a Roman orator of the 1st century A.D., compared the linguistic properties of Latin and Greek. Jewish grammarians of the 10th and 11th centuries confirmed the affinity of the Hebrew, Arabic and Arameic languages, which constitute the nucleus of the Semitic language family. In the 16th century, the notion of the Romance, Celtic, Germanic and Slavic language families was established. In 1556, in the *Polish Courtier* (*Dworzanin Polski*), Łukasz Górnicki correctly determined the composition of the Slavic language family. In the 17th century, the affinity of the Ural-Altaic languages was hypothesized. In the 18th century, interest in linguistics developed in the entire world, a phenomenon which to a great extent can be credited to G. W. LEIBNIZ (1646–1716), who encouraged the study of living languages. In 1786 William JONES discovered the affinity among Sanskrit, Greek and Latin, at the same time suggesting that the Gothic, Celtic and Old Persian languages also belong to this group. Thus, in skeletal form, the concept of the Indo-European family of languages originated. In 1800–1805 in Madrid, L. HERVÁS published his *Catalogue of Languages and Nationalities* (*Catálogo de las lenguas de las naziones conocidas*), in which he compares the vocabulary of three hundred languages. He was the first to establish the affinity of the languages of the large Malayo-Polynesian family. In the beginning of the 19th century, the Bantu family in Africa, the Sino-Tibetan family in East Asia and the Dravidian family in India were discovered. The horizons of linguistics were immeasurably broadened; but this fact in itself is not what is most important to us. The mere claim that a relationship exists in a certain group of languages does not constitute comparative linguistics. Comparative linguistics makes its debut as a science only when the process of comparison takes on a methodological form, which did not occur until the beginning of the 19th century.

Many factors contributed to this development. The most important was the development, in the course of the 18th and 19th centuries, of the concept of evolution, which transformed both nature and society. The second factor in the development of comparative linguistics was the introduction into Europe, at the end of the 18th and beginning of the 19th century, of Sanskrit and Indian literature on linguistics. Europeans were quick to adopt the significantly superior methods of linguistic description used by Indian grammarians. Moreover, they procured superbly prepared Sanskrit linguistic materials, which could be compared with historically related Greek and Latin materials. Franz BOPP (1791–1867) took a decisive step in this field when, in 1816, he published a book entitled *On the System of Sanskrit Conjugation as Compared with Greek, Latin, Persian and Germanic Systems* (*Über das Conjugationssystem der Sanskritsprache in Vergleichung mit jenen der griechieschen, lateinischen,*

persischen und germanischen). The essence of Bopp's accomplishment lies in the fact that he compared whole systems of language forms, i.e., all the conjugations of verbs, and not single words lifted from the language as a whole, as had been practiced to this time.

While Bopp introduced the notion of system to historical linguistics, it was Wilhelm von HUMBOLDT (1767–1835) who applied this notion in his study of types of language structures. In the years 1836–1840, his principal work, *On the Kavi Language on the Island of Java* (*Über die Kawisprache auf der Insel Jawa*), appeared, the introduction to which was the superb treatise, *On the Variety of Language Structure and its Influence on the Spiritual Development of Man* (*Über die Verschiedenheit des menschlichen Sprachbaues und ihren Einfluss auf die geistige Entwicklung des Menschengeschlechts*). Humboldt introduced the notion of internal language structure (*innere Sprachform*), which is the equivalent of our concept of language system. This structure appears, in the first place, in the grammatical categories of language, the relation of which to lexical elements varies from language to language. On the basis of these investigations, Humboldt worked out a typology of languages which, with only minor changes, maintained its authority throughout the 19th century. Humboldt distinguished the following types of languages: isolating languages (e.g., Chinese), in which lexical and grammatical elements are completely separate, agglutinative languages (e.g., Turkic), in which lexical and grammatical elements are combined mechanically only, incorporating languages (e.g., American Indian languages), in which the verb form contains the exponent of person of the direct object as well as of the subject, and inflected languages (e.g., Indo-European), in which grammatical categories are expressed by word endings.

The further development of linguistics proceeded along the lines of historical studies of particular language groups. Germanic philology took the lead. In a work entitled *Studies of the Origin of the Old Nordic or Islandic Language* (*Undersøgelse om det gamle nordiske eller islandske sprogs oprindelse*), published in 1818, Rasmus Kristian RASK (1787–1832), a Dane, compares the lexical materials and grammatical system of Old Islandic with other Germanic languages, with Slavic, Lithuanian and Latvian, and with Greek and Latin, establishing the primary phonetic equivalence and the affinity of all these languages. His work was carried on by Jacob GRIMM (1785–1863), author of *German Grammar* (*Deutsche Grammatik*), essentially a comparative grammar of Germanic languages. In the second edition of the first volume of this work, in 1822, "Grimm's Law" appears, the first law of phonetics, which deals with the evolution of the Germanic consonant system. This law became a model for later investigators and, in particular, for August POTT (1802–1887), author of the first etymological dictionary of Indo-European languages (1830–1836), the true founder of the comparative phonetics of these languages.

The studies of August SCHLEICHER (1821–1868) represent the peak of the first period of development of comparative linguistics (1816–1870). Bopp had already laid the groundwork for the comparative grammar of Indo-European languages,

his work being primarily dedicated to the problem of the genesis of the inflection system. Another work, entitled *Compendium of the Comparative Grammar of Indo-European Languages* (*Compendium der vergleichenden Grammatik der indogermanischen Sprachen*) was published in 1861 by Schleicher. The most important part of this work is a phonetic system which makes use of a large number of sound rules. Schleicher introduced the method of reconstructing the sounds and forms of the proto-language, a method accepted, with minor modifications, by his successors.

In the second stage of development of comparative linguistics (1870–1914), the most significant school was that of the Neo-grammarians, which originated in Leipzig and gradually embraced all the countries of the civilized world. In Germany, this school was represented by Karl BRUGMANN (1849–1919) and Berthold DELBRÜCK (1842–1922)—who, following Bopp and Schleicher, produced the third elaboration of *Comparative Grammar of Indo-European Languages* (*Grundriss der vergleichenden Grammatik der indogermanischen Sprachen*; 1893–1900)—and Hermann PAUL, author of *Principles of Historical Linguistics* (*Prinzipien der Sprachgeschichte*; 1880), in which he formulated the school's program of investigation. In France, work similar to that of the Neo-grammarians was carried on by Antoine MEILLET (1866–1936), author of *Introduction to the Comparative Study of Indo-European Languages* (*Introduction à l'étude comparative des langues indoeuropéennes*). The principal Neo-grammarians in Poland were Jan ŁOŚ (1860–1928) and Jan ROZWADOWSKI (1867–1935), and in Russia, F. F. FORTUNATOV, A. A. ŠACHMATOV and A. I. SOBOLEVSKIJ, active at the end of the 19th and beginning of the 20th centuries.

The work of the Neo-grammarians was characterized by psychologism and historicism. Like Paul, they considered language norms a fiction; for them, true language exists only in individuals. They believed that there does exist in the subconsciousness of individuals, however, a permanent set of memorized language forms related by a system of associations enabling it to function and constantly creating, by analogy, new forms and sentences. This subconscious language system, because it is to a great extent independent of consciousness, functions automatically, undergoing a gradual evolution which the Neo-grammarians attributed to three main phenomena: the fact that phonological laws function without exceptions, the functioning of analogy, and borrowings. The Neo-grammarians held the view that the designating function of language has no influence on its evolution, which proceeds mechanically, in accordance with laws which, like natural laws, are without exception, but which, unlike natural laws, are not universal. In a given region, at a given period, one sound, having a given position within a word, becomes transformed into another sound in all words, without exception. After the functioning of such a phonological law has expired, in the next period of the language's evolution, new forms may arise through analogy or borrowing, which do not agree with the expired phonological law. These exceptions are apparent only, since they did not exist in the period when the law was in function.

Two scholars, Jan BAUDOUIN de COURTENAY (1845–1929), the greatest Polish

linguist, and Ferdinand de SAUSSURE (1857–1913), the greatest linguist of France and Switzerland, developed a perspective on language problems completely differing from that of the Neo-grammarians, their contemporaries. While the Neo-grammarians were solely concerned with the evolution of language, these two scholars confronted the question—what is language and what is its relation to the total phenomenon of speech. They were influenced by each other's research in this area. In his treatise entitled *Reflections on the Original Vowel System in Indo-European Languages* (*Mémoires sur le système primitif de voyelles dans les langues indoeuropéennes*), published in 1879, de Saussure presents the problem of the language system and defines the phoneme (from the Greek *phōnēma*, 'sound') as an element of the phonological system of the proto-Indo-European language capable of being distinguished from all other elements of the system. This concept was accepted by Baudouin and his student at the University in Kazan, Mikołaj KRUSZEWSKI (1851–1887). Baudouin introduced the distinction between two aspects of language—its static aspect, encompassing the relationship among simultaneously existing elements of language, and its dynamic, evolutionary aspect. Baudouin included phonemes among the static elements of language, at the same time emphasizing the difference between the physical nature of sounds and their role in the mechanism of language. A sound is a physical sound; a phoneme is a sound which is connected with the meaning of words, an indivisible element of a system. Baudouin presented a synthesis of his ideas in a treatise entitled *An Hypothesis Concerning Phonetic Alternation* (*Próba teorji alternacji fonetycznych*, Kraków, 1894).

The views of Baudouin and Kruszewski, in turn, influenced de Saussure, who formulated his new perspective in lectures at the University in Geneva in the years 1906–1911. These lectures, entitled *A Course in General Linguistics* (*Cours de linguistique générale*) were published posthumously in 1916. De Saussure distinguishes language (*langue*) from speaking (*parole*) as two separate components of speech (*langage*). According to him, language is "the social part of speech, external in relation to the individual, who is able neither to create it nor to change it. Its existence is solely dependent upon a kind of contract agreed upon by the members of society." In contradistinction to language, which is the social and essential part of speech, speaking is the individual and more or less accidental part. A science of speech, therefore, must be divided into two completely different disciplines: a linguistics of language, which investigates language, a social creation, and a linguistics of speaking, which analyses the individual aspects of speech, i.e., speaking and phonation.

"Language—according to de Saussure—is a system of signs which express ideas and, as such, can be compared with writing, with the deaf-mute alphabet, with military signs, etc. Language is simply the most important of these systems. Thus, we can consider the creation of a science which will investigate the life of signs in the life of society". De Saussure called this science semiology from the Greek *sēmeîon*, 'sign'. The purpose of this science was to define the essence of signs and the rules

governing them. Linguistics would constitute one aspect of this science. As stated above, de Saussure's proposals have been realized in present-day information theory.

In laying the foundation of linguistics, de Saussure divided linguistics into two parts, synchronic and diachronic. "Synchronic linguistics is to be concerned with the logical and psychological relations which connect the simultaneously existing elements of language constituting the system, such elements as those presented to the collective consciousness. Diachronic linguistics, in contrast, is to concern itself with the investigation of relations connecting the elements of language that succeed one another in time and are not perceived by the collective consciousness, elements which, when substituted for one another, do not constitute a system."

The idea of value constitutes the basis of de Saussure's concept of the language system as described by synchronic linguistics. De Saussure underscores the fact that the range of a word's use, or its value, is determined by the limits imposed on it by the ranges of use of neighboring words. In this way, the value of each word is determined by its opposition to other words. It is this interdependence among the values of words which transforms them all into a uniform language system, and that which pertains to the content of words, pertains to their form as well. "It is not sounds in themselves which give words their meaning, but phonetic differences enabling us to distinguish a given word from all others—for it is with these phonetic differences that meaning is connected."

The publication of de Saussure's *Course in General Linguistics* (1916) marks the beginning of the third stage in the development of comparative linguistics, which reaches up to the present day. The two trends of investigation continue: historical linguistics as represented by the Neo-grammarians and descriptive linguistics as formulated by de Saussure and Baudouin, but their relative importance has changed, becoming opposite to what it was in the second period. Now, historical studies retreat to a subordinate position and, beginning with the end of the First World War, the central problem of linguistics becomes the question of the structure of language, thus the name structuralism. Several phases can be distinguished in its development.

In the period between the wars, the greatest activity was manifested by the Prague school, the principal representatives of which were N. S. TRUBECKOJ (1890–1939) and R. JAKOBSON (born 1896). This school, having achieved a synthesis of the views of de Baudouin de Courtenay and de Saussure, concentrated on the development of phonology and related fields. Trubeckoj's *Foundations of Phonology* (*Grundzüge der Phonologie*), published posthumously in 1939, embodied a synthesis of the Prague school's achievements. Trubeckoj began his investigations with a distinction between two basic functions of speech sounds in intellectual communication: the distinguishing function, which involves the differentiation of words from one another, and the delimiting function, which involves separating words from one another in the stream of speech. Analysing the first of these functions, the author claims that any phenomenon may be distinguished from another only when it constitutes, by virtue of

some feature, a contrast or opposition to that other phenomenon. The distinguishing function, therefore, can be fulfilled by an acoustical feature only when this feature is in contrast to some other acoustical feature, i.e., when it is a member of an acoustical opposition. Trubeckoj called such acoustical oppositions, which in a given language distinguish two words having different semantic contents phonological oppositions. Each member of such an opposition can be a short phonological unit (e.g., *a*:*i*) or a long one (e.g., *kość*:*gość*). The shortest phonological unit, i.e., that which in a given language cannot be further divided into shorter phonological units—is a phoneme.

Any feature of a phoneme is determined by the existence of another phoneme distinguishable from the first by virtue of that feature alone, and capable of filling the same position in relation to other phonemes. Thus, for example, the distinguishing quality of voiced *b* in Polish is determined by the existence in that language of the phoneme *p* differing from *b* by virtue of the single fact that it is voiceless and capable, like *b*, of appearing before the vowel in words having different meanings—compare the following oppositions of these phonemes: *był*:*pył*, *basy*:*pasy*. A phoneme is a set of features each of which is determined by its opposition to a contrasting feature in some other phoneme. In this way, phonemes form a coherent system of acoustical oppositions, which we call the phonological system.

The features differentiating the members of an opposition can also be characterized by their number in the opposition. From this point of view, three types of opposition may be distinguished: privative, gradual, and equipollent.

Privative oppositions are based on the presence or lack of a single feature. To this category belong oppositions between Polish voiced and voiceless consonants *b*:*p*, *d*:*t*, *g*:*k*, etc., in which the voiced consonant, which possesses the distinguishing feature is the marked member of the opposition, while the voiceless consonant, which does not possess this feature, is the unmarked member. Gradual oppositions are those in which the members are opposed to one another on the basis of differing degrees of intensity of a particular feature. They form a chain of several oppositions in the members of which a certain feature appears in greater and greater intensity. The opposition of the vowels *u*:*o*:*a*, for example, constitutes such a chain in respect to the degree of volume, from the minimum degree of intensity of this feature in the vowel *u*, through the medium member *o*, to the maximum member *a*. Equipollent opposition is based on the contrast between two logically corresponding features the first of which appears in one and the second in the other member of the opposition. We have, for example, the opposition of Polish labial and laminal consonants of the type *p*:*t*, *b*:*d*, etc.

The second field intensively studied by Trubeckoj and the entire Prague school is morphology, the study of vowel and consonant alternations within a given element of a word—e.g., Polish *rękᵃ*:*ręcz-ny*:*ręc-e*, alternations *ę*:*ą* and *k*:*cz*:*c*.

The work of Karl BÜHLER, professor of the University in Vienna and author of *Theory of Language* (*Sprachtheorie. Die Darstellungsfunktion der Sprache*; 1934), was

close to that of the Prague school. The author begins his exposition by establishing four axioms of linguistics. The first axiom states that the most complete form of the concrete phenomenon of speech is communication, in which one person informs another person about something. This act presupposes the existence of four elements—a sender, a receiver, a phenomenon constituting the meaning content of the sign and phenomena, normally acoustical, perceivable by our senses, i.e., the form of the sign. Next, the author describes the inter-relationship among these four elements, which will be presented at the beginning of Chapter 4. Axiom two presents a definition of the sign. According to Bühler, a sign is a physical phenomenon which stands in place of something else; not all of a sign's features, however, serve this function, but only certain of them, abstracted from the whole. Consequently, each sign may be considered in its entirety as a physical phenomenon (from the point of view of phonetics) or as a set of abstracted features pertinent to the semantic function (from the point of view of phonology). Axiom three distinguishes four aspects of the phenomenon of speech, which, in somewhat revised form, was presented in Chapter 1. Finally, axiom four describes one-class and two-class sign systems (cf. Chapter 2). Bühler divides words into referential and denoting words, which will be more fully discussed in Chapter 6.

The tradition of the Prague school was carried on and developed by Roman JAKOBSON in the Scandinavian countries during the Second World War and, after the war, in the United States (professor at Columbia University, presently at Harvard University). Among his enormous number of works—collected in *Selected Writings* (Vol. I, including papers in the field of phonology, was published in the Hague in 1962)—those treating the problem of phonological opposition, in which Jakobson modified and further developed Trubeckoj's concept, were most influential. In his book entitled *The Language of Children, Aphasia and General Phonological Laws* (*Kindersprache, Aphasie und allgemeine Lautgesetze*, Uppsala, 1942) Jakobson distinguished between primary and secondary elements of the phonological system. Primary phonemes are those which are clearly differentiated from one another, as a result of which, the oppositions obtaining between them are easier to discern (*i*:*a*, *p*:*t*, *t*:*n*). Consequently, primary phonemes and the primary oppositions connected with them are the first which the child adopts when learning language and the last to disappear in the progressive disintegration of an individual's language in the course of the speech disturbance known as aphasia (aphasia is caused by damage to certain centers in the cerebral cortex); these primary phonemes, moreover, appear in nearly all the languages of the world. Secondary phonemes and their oppositions—e.g., Polish (*a*:*ą*, *s*:*sz*), appear late in the child's speech, disappear in the first stages of aphasia, and are found in only certain languages. These concepts are of great significance to typological linguistics. Certain other ideas of Jakobson, presented in *Preliminaries to Speech Analysis* (1952, in cooperation with G. Fant and M. Halle) and *Fundamentals of Language* (1956, in cooperation with M. Halle), are also of significance for investigations in this field. According to Jakobson, all phonological

oppositions occurring in all the languages of the world are binary, i.e., they are the result of a contrast between features. The first of these features are sound, or acoustical, features; they must be described, therefore, not only in terms of articulation, as practiced in the past, but also, and above all, in terms of acoustics. Thus, we must describe not only the movements of the speech organs during the articulation of a given feature of sound, but also the properties of that feature of sound itself, the properties of the sound waves transmitted from the sender to the receiver. Jakobson divides all binary oppositions, comprising the phonological systems of all the languages of the world, into twelve types, each of which involves the contrast between two mutually exclusive acoustical features, e.g., vowel : non-vowel, consonant : non-consonant, voiced : voiceless, nasal : non-nasal, etc. Each phoneme is a set of several such features, which contrast with the features of other phonemes. The set of types of acoustical opposition is different in each language. Thus, the variety among languages.

Toward the end of the period between the First and Second World Wars, Copenhagen became the second center, along with Prague, of structural linguistics. This school is represented by Louis HJELMSLEV (1899–1965), professor at the University of Copenhagen since 1937, editor of the periodical *Acta Linguistica*, and author of *Principles of General Grammar* (*Principes de grammaire générale*, 1928), *Category of Cases* (*La catégorie de cas*, I, 1935; II, 1938), *Prolegomena to the Theory of Language* (*Omkring sprogteoriens grundlaeggelse*, 1943) and *Linguistics* (*Sproget*, 1963). Hjelmslev's theory is based on a completely new notion of the proper subject of linguistics perhaps most clearly presented in an article entitled "Language and Speaking" ("Langue et Parole": *Cahiers Ferdinand de Saussure*, II, 1942, pp. 29–44). In analysing the meaning of the term "language" (*langue*) as used by de Saussure, Hjelmslev concludes that this term includes three different concepts: (1) the language scheme, i.e., the pure language form defined independently of its social realization and physical manifestation, (2) the language norm, i.e., the material form defined by its social realization but independent of particular manifestations, (3) the language custom, i.e., a set of customs accepted by a particular society and defined by observable manifestations.

Thus, the French *r*, for example, may be defined in three ways, depending on whether we look at it from the perspective of the scheme, the norm, or the custom of the language. From the first point of view, we can say no more than that *r* is a consonant, i.e., a non-syllable-forming element which must accompany a syllable-forming vowel. In addition, we may say that, in French, *r* is a consonant which stands at the beginning and at the end of words and, in consonant groups, always next to a vowel. Here, we have defined *r* only in respect to its position in broader structures, i.e., in syllables and words, in terms of its distribution in these words and syllables; we have not, however, attributed any positive features to the French *r*. In this treatment, *r* is merely an abstract element of the communication system,

element which may be expressed by a sound, by a written letter, by a gesture, etc., and which remains the same abstract element of the language scheme.

From the perspective of the language norm, the French *r* may be defined as a rolled sonant. The mouth is half open and the air stream escaping from it is interrupted at infinitely short intervals. Here, we have to do with positive features, but we take into account only those distinctive features serving to differentiate words.

From the point of view of the language custom of the French *r*, it is defined as a voiced rolled sonant and is articulated by an interrupted stream of air from the vibration of the uvula, which constitutes the end of the soft palate in the back part of the oral cavity or—more rarely—by the vibration of the tip of the tongue in contact with the gums. In both cases, the vibration of the speech organs alternately close and open the oral cavity, interrupting, in closing, the air stream. Here, we have all of the features which are found in customary French pronunciation.

Apart from these three notions—language scheme, language norm and language custom—which are encompassed by de Saussure's concept of language (*langue*), Hjelmslev introduced the notion of the individual act, corresponding to de Saussure's concept of speaking (*parole*). Hjelmslev points out that there is a close connection between a concrete individual act of articulation and the language custom, which together form the material side of language, as opposed to the non-material side, as represented by the language scheme. Hjelmslev considers the notion of norm, however, to be, in practise, dispensable.

Hjelmslev's formulation facilitates our understanding of the differences between two important currents in contemporary linguistics. The first is represented by the Prague school (Trubeckoj, Jakobson and their followers in various academic centers), the second, by the structuralists, inspired by Hjelmslev. Both of these linguistic schools are based on the concept of language system. They define this language system somewhat differently, however. The Prague school conceives of language as a system of social norms which are realized in individual processes of speaking. These norms are material to the extent that they define those sound features that are indispensable to communication, to the differentiation and separation of words, e.g., the particular quality of voiced *z* which is necessary in Polish for distinguishing words of the type *koza:kosa*. For structuralists of the Copenhagen school and related schools, on the other hand, the language system is a non-material scheme, capable of being realized in various media: in sounds, as in speech, and in graphic form, as in writing, etc. The difference between these two currents in language is not a fundamental one. It does not prevent scientific collaboration, and many linguists take into consideration in their research both the language scheme and the language norm.

Another position, to a certain extent intermediary between those of the Prague and Copenhagen schools, is held by certain investigators in other centers: in Poland, Jerzy KURYŁOWICZ—cf. papers included in the collection *Linguistic Sketches* (*Esquisses linguistiques*, 1960)—and in France, André MARTINET—author of *The*

Economy of Phonetic Changes (*L'économie des changements phonétiques*, 1955) and *Elements of General Linguistics* (*Eléments de linguistique générale*, 1960).

In recent years, the American school, created in the period between the First and Second World Wars by E. SAPIR (principal work: *Language*, 1921) and L. BLOOMFIELD (*Language*, 1933), has became significant. In the post World War II period, American structuralism is represented by Z. S. HARRIS, G. TRAGER, K. L. PIKE, Ch. MORRIS, and, in the last decade, by U. WEINREICH (*Languages in Contact: findings and programs*, 1953), N. CHOMSKY (*Syntatic Structures*, 1957) and Ch. F. HOCKETT (*A Course in Modern Linguistics*, 1958).

The newest trend in linguistic investigations, already mentioned at the beginning of the previous chapter, is mathematical linguistics, carried on primarily in the United States, France, and the Soviet Union. Two factors were of primary importance in its development. The first was Hjelmslev's concept of the abstract language scheme. The elements of this scheme, completely lacking in material features, remain in quantitative relationships to one another, as a result of their frequency in the anguage system, or in a text, and these relationships may be determined statistically. The second factor was the development of cybernetics (from the Greek *kybernetes* 'steerman'), a field of technology dedicated to the construction of automatons which perform complicated functions. The construction of translating machines, e.g., from Russian into English, or *vice versa*, is of the greatest significance for linguistics. Mathematical linguistics proceeds on the assumption that these translating machines are constructed and function in a way somewhat similar to that of the human brain, and that, therefore, experimentation with them may constitute a basis for conclusions concerning the processes of speaking and the structure of language.

Mathematical linguistics has come to be divided into two trends—statistical and algebraic—both combining purely linguistic questions with problems of logic, mathematics, physics and technology. The main effort of linguists consists in exploiting technical and mathematical achievements in order to enhance our knowledge of purely linguistic phenomena. Today, mathematical statistics is widely used in dealing with such language problems as describing the distribution of language elements in a text or dictionary, studying the individual language features of various authors, and dealing with certain questions concerning the typology of languages. A great role in the evolution of language has been played by the tendency of speech phenomena to be realized by the least possible effort; this tendency is presently being investigated by mathematical methods. The principal representatives of mathematical linguistics are: G. K. ZIPF (*Human Behavior and the Principle of Least Effort*, 1949), B. MANDELBROT (*Logic, Language, and Information Theory.—Logique, langage et théorie de l'information*, 1957, in cooperation with L. Apostel and A. Morf), P. GUIRAUD (*Problems and Methods in Linguistic Statistics.—Problèmes et méthodes de la statistique linguistique*, 1960) and G. HERDAN (*The Calculus of Linguistic Observations*, 1962). Certain problems which have been elaborated by mathematical linguistics will be described in Chapters 5, 14, and 15.

Note: The position of linguistics as a science: Information theory (semiology)—a science concerned with all types of signs and forms of information transmission. One branch of information theory is the theory of speech, which is limited to the study of signs occurring in vocal communication among people. Speech includes four phases, which are analysed by the various sciences representing branches of the theory of speech. The psychology of speech, one of the branches of general psychology, investigates those processes occurring in the brain and nervous system which are involved in speaking and understanding. Today, cybernetics, the theory of automatons, is also concerned with these processes. The psychological aspects of the speaking process, the movements of the speech organs and the sounds resulting from these movements, is analysed by phonetics, a field known as experimental phonetics when it makes use of mechanical equipment. Acoustics, a branch of physics, studies the structure of sound waves transmitted from the sender to the receiver. A whole series of sciences study the third phase of speech, the text. Philology is the theory of defining and interpreting texts. The purpose of philology is to recreate the original form of a text to define the function of its elements in the process of communication, and to establish the text's history. Stylistics investigates the language structure of a text by determining the principles of choice and arrangement of forms appearing in it, principles not included in the set of language norms. All of the texts studied by philology and, in practise, by stylistics, and which form the basis for linguistics, are written texts. Forms of writing and their distribution in time and space—the history and evolution of writing—constitute the subject matter of paleography. Graphology, on the other hand, analyses individual characteristics of handwriting, relating them to psychological characteristics of individuals. The fourth and final phase of speech, i.e., language—a system of permanent, social and abstract norms—is the realm of linguistics, which is divided into descriptive and comparative linguistics. Descriptive linguistics describes each language system individually, just as it functions in a given social group. Normative grammar presents the language system as a model for imitation on the part of speakers. There are two branches of comparative linguistics, historical and typological. Historical linguistics compares languages in respect to the degree to which they are of common origin, while typological linguistics compares languages in respect to their structural similarities. Through what developmental stages did these various branches of linguistics pass?

PART II

DESCRIPTIVE LINGUISTICS

CHAPTER 4. COMPONENTS OF THE LANGUAGE SYSTEM

Human speech fulfills three basic functions: expressive, impressive, and semantic.

The expressive function of speech characterizes the sender, informs us about him. We recognize the sex, age, emotional state, social class and education of the person speaking by his tone of voice and manner of speaking. Someone is speaking in the adjoining room—I recognize by the voice that it is my father. I open a book without looking at the title page—I recognize my favorite author by the style.

The impressive function of speech is the effect which it has upon the emotional state and behavior of the receiver. There is a well known army saying: "As the command is given, so it is carried out." The same words of a command evoke completely different reactions on the part of the soldiers depending on how they are spoken. A difference in intonation may cause the command to be carried out sloppily in one case, and efficiently in the other. Equally well known are cases in which the rhythmic repetition of orders coordinates and increases the efficiency of a group of people working. It also happens that a novel in the original fascinates the reader, while the same book in translation bores him. Since the content is the same in both cases, the difference in the reader's reaction is connected solely with the difference in form.

The semantic function of speech involves both calling attention to phenomena in the surrounding world, lying within the speaker's field of vision, and presenting phenomena that are distant, that cannot be seen. When we hear a story or report, we imagine that we can see the events being described.

The following scheme is helpful for understanding the three functions of speech distinguished in Fig. 2.

In this scheme, point *A* stands for the person speaking, point *B* for the listener, point *C* for the acoustical phenomena constituting the form of speech, and point *D* for the meaning content, existing either in reality (e.g., *thunder*) or in the social imagination (e.g., *the god Poseidon*). The acoustical form of speech (*C*) is connected with the person speaking, whom it characterizes (*C-A*), thus fulfilling the expressive function. The acoustical form also influences the psychological state and behavior

of the listener (*C-B*), fulfilling the impressive function. Finally, through the mediation of the sender and the receiver, the acoustical form is connected with the objective phenomenon being designated (broken line *C-D*), and so fulfills the semantic function.

The signs of speech are, first of all, symptoms as regards their relationship to the sender, whom they characterize; secondly, they are appeal signals for the listener, whose behavior they direct; thirdly, they are semantic signals which present reality as a result of the fact that their form is connected with objectively existing things and the relationships existing among them. These three functions are to a certain degree independent of one another, and they are of equal importance, for each predominates in different circumstances. The first appears primarily in complaints, the second in commands and the third in stories and reports.

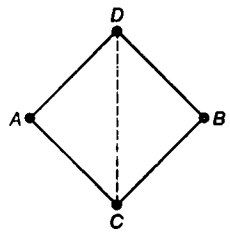


Fig. 2. The three functions of speech

In onomatopoeic utterances, the semantic signs of speech acquire the form of acoustical images. Particles, of the type *yes* and *no* are one-class semantic signals. The huge majority of speech elements, however, belong to the two-class phonemic system which we call language.

In speech, there is no sharp boundary between these types of signs. To a certain extent, signs belonging to the language system also fulfill in speech the expressive function of symptoms, the impressive function of appeal signals and, at times, function as acoustical images. One-class particles are constructed from the phonemes of the language. There exists, therefore, a peripheral sphere of intermediary forms which to a certain extent are part of language, but which simultaneously fulfill the function of extra-linguistic speech signs: symptoms, appeal signals, images or one-class signals. This sphere of different types of intermediary forms constitutes a separate field in relation to the true language system and in so far as it is the concern of linguistics, it constitutes the subject matter of a separate branch of linguistics. Charles BALLY, one of de Saussure's students, calls this branch linguistic stylistics, in contradistinction to stylistics, referred to above, which he calls literary stylistics and which is concerned with the study of texts.

Having set aside this peripheral branch of linguistics, we will proceed to the language system proper. Speech is an infinitely complicated set of all types of signs. In this set, semantic, conventional, two-class signs, which we call language signs, can be distinguished, not on the basis of the means of transmitting them, but on

the basis of the structure of the code to which they belong. In the flow of speech, language signs pass along the same information channel as do all other kinds of signs and are so closely linked with them that they can be isolated only by a careful analysis which considers the nature of signs.

Non-conventional signs—symptoms, appeal signals and images—may be differentiated from language signs on the basis of the fact that they are non-reproducible and non-interchangeable. The expressive cries of a child—which are symptoms, appeals and commands—and rhythm and onomatopoeia in poetry make a different impression on us each time we come into contact with them, and they cannot be identically reproduced. The semantic function of all non-conventional signs lies in their potential, rather than in their constant, function. The sounds of speech may or may not fulfill this function. A poetic work, which one day impresses us with its onomatopoeic qualities, may make no impression at all the next day. The function of conventional signs, on the other hand, is constant. Conventional signs are reproducible and interchangeable because they represent reproductions of the elements of a system which has a permanent existence in society. By language signs, we mean the permanent, reproducible and interchangeable component of speech, which we distinguish from the non-permanent, non-reproducible and non-interchangeable, extra-linguistic component.

Conventional signs may be divided into one-class and two-class signs. The first are indivisible and totally bound to the consituation; the second appear primarily in the form of complex signs, or utterances, which may be broken up into simple signs, i.e., words. Because of their unlimited variety, complex signs may exist independently of the constitution. We consider as true language systems only these made up of semantic, conventional, two-class signs, while we define one-class signals and all composite forms as constituting the periphery of such systems. For the present, we will not concern ourselves with this peripheral sphere, but will concentrate our interest on the true language system, abstracted from speech and existing permanently in society.

From the point of view of information theory, the languages of the world represent a group of codes possessing certain features in common but differing from one another in other features. In the study of languages, therefore, two problems are of primary importance. The first involves the analysis of the universal features of language systems; the second involves the description and systematization of those features on the basis of which various languages are differentiated. Such a description of the common features of all systems and of the features characterizing particular languages is the realm of descriptive linguistics, which naturally divides linguistic phenomena into general and particular phenomena. Descriptive linguistics describes the common features of all languages of the world, elaborating a kind of model, or simplified general diagram which, to a certain extent, refers to all of them. In particular linguistic studies, however, each language is described individually, as though it were the only language in the world. Such a description consists of thousands

of individual descriptions having in common only the method by which they are arrived at. In the present work, which is of a general nature, our primary concern will be to present the common features of all languages, thereby establishing a certain universal language model. The specifics of this general model will be illustrated primarily by material from Polish and other Indo-European languages. At the same time, problems concerning modern methods of language description will be considered. This will constitute the subject matter of Part II, dedicated to descriptive linguistics. The analysis of differences among languages will be presented in Part IV, in which typological linguistics is discussed.

The definition of language presented here assumes that it is a system of conventional, semantic, two-class signs. These three properties of language correspond to the three functions of language signs: the diacritic function of differentiating the forms of signs on the basis of their conventional features, the semantic function of simple signs of the first class, i.e., words, which designate phenomena in the surrounding world, and the syntactic function, which involves combining simple signs into complex signs of the second class, i.e., sentences. The set of means for fulfilling any of these three functions is called a component language system, i.e., a separate component of the total language system.

Three such component systems can be distinguished in language. We have the phonological system fulfilling the diacritic function of differentiating signs, the semantic system fulfilling the function of designating reality, and the syntactic system including the norms for combining simple signs into complex ones. These three component systems are hierarchically arranged. The phonological system constitutes the base; the semantic system comprises the next level, which is composed of designating phoneme sets, i.e., words; at the highest level we have the syntactic system consisting of means for generating sentences out of words, sentences which constitute independent complex signs functioning in speech.

The relative independence of these three systems of language is a fact supported by observations of aphasia, a speech disorder. Aphasia refers to a disturbance in speaking and understanding of speech caused by disease and consequent damage to the brain tissue. Cases of aphasia exist in which deterioration occurs in only one of the three component systems of language, while the two remaining systems continue to function, at least partially. Thus, in aphonia, the patient only loses control of the phonological system, he loses the ability to produce the learned phonemes of a given language, while his capacity for thinking in language is unimpaired. In the case of alexicalism the patient loses control of the semantic system of language, and is unable to properly name objects presented to him, while the other components of language continue to function. In the case of agrammaticism the system of syntax ceases to function. The patient is unable to express in language form the conceptual relationships among words, although he is still able to produce sounds and words, and even properly uses words for signifying particular objects. These

facts prove that linguistic elements have a particular order in our minds, in principle corresponding to the division into three component systems presented here.

Consequently, the description of language must include four parts: (1) phonology, (2) linguistic semantics, (3) syntax, (4) linguistic stylistics, which discusses the peripheral sphere of language. These four parts will be discussed in succession (Chapters 5–8). First, however, we must direct our attention to the fact of language productivity.

Language is a system of norms regulating individual processes of speaking. Speaking, however, is a living process, constantly changing its form and constantly satisfying new individual and social needs. For this reason, it cannot be a closed system, made up of a strictly limited number of elements; it is an open system capable of increasing the number of its elements without disturbing the system of language itself. This extension of the system and its adaptation to new needs is what we call the productivity of language. All living languages, like English or Polish, function to increase the number of words and other elements. If a language ceases to be productive, if the number of its elements becomes strictly limited, it becomes a dead language, as happened in the case of Latin. In living languages, all three component systems, i.e., the phonological, semantic, and syntactic systems, function to increase their number of elements, although each in a partially different way. Because of the variety of these types of productivity, they will be discussed in detail through the analysis of particular component systems of language. Here, however, for purposes of illustration, we will point out one of the most important types of such productivity, one which involves the capacity for filling in empty slots, or gaps, in the system.

The elements of language form oppositions, the members of which are in part identical and in part different. We call the identical part of these members the common base of the opposition, and we call the differing parts distinguishing features. If, for example, we compare the Polish words *domek* and *domisko*, we see that in the opposition which they compose, the element *dom-* is the common base, while the elements *-ek* and *-isko* are the distinguishing features. It happens, that in a series of oppositions making use of various bases of comparison, we have the same distinguishing features—e.g., *dom:domek*, *pies:piesek*, *ogród:ogródek*. In these oppositions, one of the members always contains the element *-ek*, which corresponds to the element *zero* in the other member. Oppositions which have the same distinguishing features are called proportional oppositions because their members remain in the same proportion to each other. It sometimes happens, however, that alongside a series of proportional oppositions, an incomplete opposition appears containing only one member, that with the distinguishing feature *zero*, but there is no word in the language built on that base and ending in the opposing positive feature, in this case the element *-ek*. This lacking member of the opposition constitutes an empty slot, or gap, in the system. This empty slot may be filled in by a form having features which are consistent with the structure of the system. This member must

have a base of comparison identical to that of the unpaired member and a positive distinguishing feature similar to those in the remaining proportional oppositions. The series of proportional oppositions of the type *dom:domek*, *pies:piesek*, *ogród*: *ogródek* is such that an empty space exists alongside the word *semafor*, which must be filled in by the word *semaforek* containing the distinguishing feature *-ek*, as in the words *domek* (little house), *piesek* (little dog), *ogródek* (little garden).

It is difficult to decide whether the word *semaforek* in Polish exists or not, but it is certain that this word is possible as a neologism, which through repeated use by many people may become a permanent part of the language. A language system includes not only elements clearly established in it, but also possible, potential elements which the speaker can create on the basis of already existing proportional elements. Because of this and other factors which will be discussed further on, language is an open, elastic system, adapting itself to the constantly arising needs of the individual and society. Consequently, a living language is never a stagnant, fixed system. In every period of a language's history, there exists alongside the stabilized part of the system, part in the process of change, either coming into being or dying out, and thus, composed only of possible, potential elements. The description of a language in a particular period must, therefore, present both its productivity and its gradual transformation.

Note: Terminology introduced in Chapter 4: the expressive function of speech gives information about the sender; the impressive function influences the emotional state of the receiver; the semantic function calls attention to phenomena existing independently of the sender and the receiver. Language is a system of conventional, semantic, two-class signs. These three properties of language correspond to three functions of language signs: the diacritic function of differentiating the form of signs on the basis of their conventional features, the semantic function of simple signs of the first class, i.e., words, which designate phenomena in the surrounding world, the syntactic function which involves combining simple signs into complex signs of the second class, i.e., sentences. These three functions are carried out by three hierarchically arranged component systems of language. The phonological system, fulfilling the diacritic function of differentiating signs, constitutes the base. The semantic system comprises the next level of the hierarchy, which is composed of designating phoneme sets, i.e. words, which refer to reality. At the highest level of the hierarchy, we have the syntactic system consisting of means for generating sentences out of words, sentences which constitute independent complex signs. The description of language must include four parts: phonology, semantics, syntax, and linguistic stylistics, which is concerned with the role of language in realizing the impressive and expressive functions of speech. The mutual independence of these component systems of language is indicated by the various forms of the speech disorder called aphasia. Aphonia involves the loss of the phonological system, alexicalism, of the semantic system, and agrammaticism, of the syntactic system. Particular descriptive linguistics investigates each system of language individually; general descriptive linguistics determines the common features of all languages and on this basis constructs a model of languages, a model of languages' simplified, general form. Living languages function by filling in empty slots, or gaps, in the system with new forms; dead languages have no new forms.

CHAPTER 5. THE PHONOLOGICAL SYSTEM

Before proceeding to the analysis of the phonological system of language, we must present a brief description of the functioning of the human speech apparatus, which is, in fact, the respiratory apparatus which only secondarily came to be utilized for communication. This apparatus may be compared to a fife or pipe. Its lower part contains two large bellows—the lungs—which by means of muscle contractions produce a stream of air leading to the trachea (or windpipe) and larynx, which protrudes upward. The larynx, at the upper end of the trachea, may be opened or closed by means of the vocal chords, i.e., two movable membraneous lobes located in the larynx. When the vocal chords are drawn together, the larynx is closed; when drawn apart, the larynx is open. From the larynx, air passes on to the pharynx, and, thence, to the oral and nasal cavities, which together constitute the resonating chamber of the speech apparatus. The nasal cavity may be closed off. The soft palate ends in an elongated lobe hanging loosely downward, which we call the uvula. Thus, when the uvula is drawn back so that it adheres to the back wall of the pharynx, the passage to the nasal cavity is cut off and air flows through the oral cavity only, the shape of which changes with the position of the lower jaw, tongue and lips.

An apparatus so constructed can produce two basic types of sound, depending on whether it is closed or open. The former are consonants and the latter, vowels. Consonants, the articulation of which involves various ways of closing the speech organs, are extremely various. The tightest closure occurs in consonants such as *p*, *t*, *k*, which are called stops. Let's observe, for example, the movement of the speech organs during the articulation of the consonant *p*. By looking in a mirror, we see that in producing this consonant, the lips are drawn together tightly and the air, forced from the lungs under pressure, undergoes compression behind the closed lips. Suddenly, the lips open energetically and the air explodes with a kind of popping sound similar to that produced when a bottle of effervescent liquid is uncorked. The articulation of the consonant *t* is similar, the single exception being that, in this case, the barrier closing the speech apparatus is produced by a closure between the tip of the tongue and the back wall of the upper alveolar ridge. The consonant *k* is articulated by means of a closure which takes place further back in the mouth and which involves drawing together the back part of the tongue and the velum. Thus, not only the degree of opening but also the point of articulation is a basic feature of consonants.

A less tight closure of the speech apparatus can be observed in the articulation of spirants such as *f*, *s*, *χ*. The articulation of spirants is characterized by a narrow constriction formed by the speech organs. When the consonant *f* is produced, this constriction is formed between the lower lip and the upper alveolar ridge, while the articulation of the consonant *s* involves a constriction formed between the tip of the tongue and the base of the alveolar ridge; when *χ* is produced, the constriction is formed between the back of the tongue and the velum. A stream of air from the

lungs is released through this constriction and the friction of the air stream against the walls produces the particular murmur which characterizes this consonant. In the case of spirants, the speech apparatus functions like a whistle, as is particularly striking in the articulation of the consonant *s*.

Consonants such as *r*, *l*, *m*, *n*, called sonants, have a still broader articulation. In the pronunciation of *r*, the wide channel through which air is released from the lungs is opened and closed at short intervals by the vibration of the tip of the tongue. In the articulation of *l*, the tip of the tongue is pressed against the upper alveolar ridge while its lateral sides are lowered completely. Finally, in the articulation of the nasal consonants *m* and *n*, the oral cavity is completely closed off and air is released through the nose. Thus, all of the consonants just discussed are sonants, but each in a different way. Sonority is their common feature. In the pronunciation of sonants, the vocal chords in the larynx are normally drawn together and air, released by pressure from the lungs, sets the chords into vibration, setting up waves in particles of air and thus producing an effect of sonority. Stops and spirants may be either voiceless (*p*, *t*, *k*, *f*, *s*, *χ*), i.e., pronounced with the vocal chords drawn apart, or voiced (*b*, *d*, *g*, *v*, *z*, *γ*), i.e., produced with the vocal chords drawn tightly together. The sonants are normally voiced.

Normally, vowels articulated with the mouth more or less open are also voiced. The articulation of *a* involves the broadest opening of the mouth, *o* and *e*, medium, and *u* and *i*, the least. Thus, the oral cavity forms a resonant chamber the shape of which, different in the production of each vowel, gives the vowels their particular acoustical character. Vowels differ from one another in pitch and volume. The further back the tongue is drawn, the fewer vibrations per second and the lower the pitch of the vowel; the further foreward the tongue is advanced, the greater the number of vibrations and the higher the pitch of the vowel. In pronouncing a series of vowels in turn—*u*, *o*, *a*, *e*, *i*—one notices a gradual shift of the tongue from the back to the front of the mouth and one hears the pitch gradually rise. Similarly, in proceeding from the relatively broad (*a*) to the successively narrower (*o*, *e* and finally *u*, *i*), we proceed from vowels of greater volume (stronger) to those of lesser volume (weaker).

The human speech apparatus—owing to the fact that the oral cavity can be opened and closed to various degrees, that the lips, tongue and uvula can be placed in various positions, that the vocal chords can be drawn together or apart, that various other changes are possible—can produce an enormous variety of sounds the various features of which, passing directly into one another through an infinite number of stages, constitute a continuous scale. Only certain, selected features of this scale play a role in any particular instance of communication. These are distinguished as semantic, diacritic features, as opposed to all other acoustical features of speech, which are non-diacritic and have, therefore, no bearing on language communication.

A basic difference exists between language signs and acoustical images, which have been discussed previously. Acoustical images, whether in an infant's babbling

or in music, imitate reality, and their features, flowing along the continuous sound scale, refer to that reality. The diacritic features of language signs neither imitate reality nor directly refer to it, but, instead, enable the receiver to identify the sign which he hears with one which he has previously memorized, by distinguishing that sign from all other signs occurring in the same language system. Upon hearing the word *kość* (bone), he must identify it with one of the words he has already heard before, referred to some reality, and memorized. In order to do this, however, he must, first of all, differentiate the word *kość* from other Polish words, especially those which are acoustically similar, such as, for example, *gość* (guest).

The fluid features of the continuous scale are unsuitable for this purpose. It is necessary to select from this scale a certain number of distinct features, standing in opposition to one another and which—being imposed by social tradition—will serve to keep language signs distinct, thus making it possible to identify them. These features, called distinctive features, are not only conventional, but are also in opposition to each other. Since their only function is the differentiation of signs, they constitute elements of signs only in so far as they differentiate signs from one another. Thus, the smallest basic element of a language sign is a distinctive feature which is in opposition to a contrasting feature in another sign of the same system, and thus, serves to distinguish both signs. For example, the opposition between the absence of voicing (*k*) and voicing (*g*) at the beginning of the words *kość* and *gość* distinguishes them from each other. In Polish, these features are distinctive and constitute an opposition.

In order for a language sign to be recognized by a receiver, it must not only differ from all other signs of the system, but also it must be separated from both the sign that proceeds it and that which follows it in the stream of speech. Certain sound features must, therefore, fulfill the function of separating or delimiting signs. In Polish, the accent on the penultimate syllable, which informs the receiver that the following syllable is the final one, constitutes such a delimiting feature in that it keeps words apart.

Thus, we have two kinds of diacritic features in language signs—distinctive features and delimiting features. They function on two different planes, which fulfill an important role in the structure of language—the associational plane and the textual plane. A given language sign must be differentiated from all other signs of the system which are not included in the text, signs which are associated with the given sign only in the minds of the persons in communication. This is precisely the function of distinctive features, which function within the system of these associations, i.e., on the associational plane. A language sign, however, must also be separated from the signs by which it is preceded and followed, which together with it constitute the text. It is the delimiting features, functioning on the textual plane, which separate signs from one another.

As we see, diacritic features are determined by mutual oppositions. Voiced sounds constitute distinctive features only in those languages in which they are in opposition

to voiceless sounds, and accented syllables constitute delimiting features only when they are in opposition to unaccented syllables. Each language has a limited number of diacritic features and, it follows, of possible oppositions. The total of these features and oppositions constitutes the phonological system, which in each language possesses its own particular structure.

Certain diacritic features occur simultaneously, being embodied in a single sound wave section, but differing on the plane of the various acoustical properties of this wave. Such a set of simultaneously occurring diacritic features is called a phoneme. A phoneme is a complex product of language, in contrast to diacritic features, which constitute its simple, indivisible elements. A diacritic feature is a member of one opposition only, in which it is opposed to a contrasting feature, while a phoneme is a component of as many oppositions as there are diacritic features comprising it; for on the basis of each of these acoustical properties, the phoneme is in opposition to some other phoneme. For example, the Polish phoneme *p* is a set of five simultaneously occurring—i.e., synchronic—distinctive features and, therefore, it is a member of five interphonemic oppositions: (1) a stop as opposed to a spirant *f* (e.g., compare *para*:*fara*), (2) oral as opposed to nasal *m* (*mama*:*papa*), (3) labial as opposed to laminal *t* (*pył*:*tył*), (4) voiceless as opposed to voiced (*pył*:*był*), (5) hard as opposed to soft *p'* (*pasek*:*piasek*).

The phoneme is the least complex language product; it differs, not only from simple elements like diacritic features, but also from more complex products, which are composed of phonemes, like words and sentences. All of these elements, hierarchically arranged and representing increasingly inclusive language units, have two features in common.

Firstly, each of them is more than the sum of its component parts, for the whole includes not only an assortment of elements, but also a structure, i.e., the defined relationship of these elements to one another. A phoneme is a structure of diacritic features, not simply their sum; it is a whole in which particular diacritic features are embedded, standing in a certain constant relationship to one another. This structure of the phoneme results from the simultaneity of diacritic features and from the hierarchical dependence of certain such features, deriving from other, more basic, features. The fact that Polish *p* is a stop means that it may be voiceless in opposition to voiced *b*, because Polish sonants and open phonemes are always voiced. Consequently, it is not diacritic features, but phonemes which constitute the basic elements of all the subcodes of a language. Phonemes correspond to the letters of various types of writing as well as to the signals of these or other systems of visual communication. Letters also have diacritic features by means of which they are distinguished, but these features do not directly correspond to the diacritic features of phonemes. A letter as a whole corresponds to a phoneme as a whole. It follows, that phonemes are the basic elements of any code, while the diacritic features of phonemes are the elements only of the principal subcode of language, spoken language.

The second feature common to all complex products of language is, in certain cases at least, the vagueness of the boundaries separating them from one another in the text. It is sometimes difficult to define the boundaries of words and sentences and to establish the boundaries of a phoneme separating it from preceding and following phonemes. The diacritic features of which a phoneme is composed, which usually occur simultaneously, do not always begin and end at the same moment. For example, in the pronunciation of the Polish phoneme soft *p'* (e.g., *piasek*), the placing of the blade of the tongue close to the palate, which produces the acoustical effect of "softness", continues after the opening of the lips and the explosion of air. This "softness", which is later than the articulation of the lips, gives the impression of the phoneme *j*; the question therefore arises, whether the word *piasek* begins with the soft phoneme *p'*, or with the phoneme group *pj*. Difficulties of this type result primarily from the fact of language evolution. In Old Polish, soft *p'* infallibly appeared in the word in question, but it has been gradually transforming into the phoneme group *pj*. In the transitional period, the pronunciation fluctuates between the two possibilities, which accounts for the unclear situation.

Despite these sporadically occurring vaguenesses in the demarcation of boundaries, phonemes function as coherent wholes, differentiating words not only by means of their diacritic features, but also by their sequence, as can be plainly seen in the opposition of the three words *tak:kat:akt*, composed of the same phonemes.

A phoneme may be conceptualized in two ways: as a separate individual entity or as a set of features. Both concepts are pertinent, but each from its own perspective. If we consider phonemes as elements of an abstract code which may be realized, with the aid of various information channels, by various subcodes, then we will look upon phonemes as entities, differing from one another, but devoid of all concrete features. In studying phonemes from this point of view, we may define only their statistical properties—the number of phonemes in a given language system and their percentage distribution, i.e., the frequency of occurrence of each phoneme in relation to that of other phonemes in a given text. These static properties of a phoneme are not determined by the subcode in which the phoneme is realized. This concept of the phoneme, however, is not adequate for linguistics. A phoneme must be studied, not only as an abstract entity which enters into the make up of a code, but also as a set of concrete diacritic features standing in opposition to the features of other phonemes of the same system.

If we look at phonemes from this latter perspective, we must define their relationship to the sounds produced by the human speech apparatus. By sound, we mean the sum total of features of any sound produced by a human being; a phoneme, on the other hand, is the set of diacritic features of this sound, which serve to distinguish and separate words. The relation between a phoneme and a sound may be represented schematically, as in Fig. 3.

The shaded in area of Fig. 3 represents the set of diacritic features of the sound which together constitute the phoneme, while the sound itself is represented by an

outer circle which includes all of the sound's acoustical features. From the point of view of phonetics, which investigates the sounds of human speech by means of natural methods, there is no difference between the diacritic and the non-diacritic

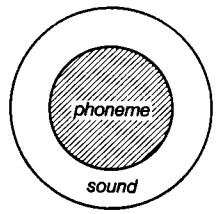


Fig. 3. Sound and phoneme

features of a sound. Both arise as a result of articulatory movements of the human speech apparatus, both constitute acoustical properties of the air waves emitted by the mouth of the speaker. The difference between these two categories of features, therefore, is not determined by the nature of the features themselves, but by their function in the process of communication. Diacritic features enable the receiver to distinguish a given language sign from all other signs of a given system and to separate this sign from those which precede and follow it in the text, at the same time reinforcing the association of the word heard with a word previously memorized; non-diacritic features, on the other hand, do not fulfill these functions. Only diacritic features appear in the speech of all members of a given speech community, and only these features are repeated in speech over long periods of time, i.e., they are both shared and permanent. Non-diacritic features, on the other hand, are different in each sound. Each speaker produces different non-diacritic features, and it is these features which enable us to recognize his voice; these features are different in each moment of the speaker's life, and it is on the basis of these changes that we recognize his emotional state.

And now, a concrete example. We telephone. The conversation proceeds normally. Both parties understand each other. Suddenly understanding is interrupted. Some name is mentioned which the receiver cannot hear correctly. The sender must break the name up into phonemes, giving each phoneme as the first letter of a name: *a* as in *Anne*, *d* as in *David*, etc. The question arises, as to why the receiver had no trouble in understanding the phonemes of the common words and names, while he did have trouble in identifying the phonemes in the surname, although the degree of precision in the sender's pronunciation remained more or less the same. The answer is simple. The common words and names were already familiar to the receiver and, therefore, even an incomplete number of diacritic features was sufficient for the identification of these words with those already memorized; the surname, however, was new and, therefore, all the diacritic features of the phonemes composing it had to be realized.

This example shows that in the colloquial and careless speech of the speaker, a certain number of diacritic features is left out, and the receiver fills them in when

identifying what he hears with previously memorized, carefully pronounced words. The same thing takes place in reading. We decipher an illegible letter by identifying whole words with those already memorized, on the basis of their general shape, and only later do we recognize individual letters of the alphabet, filling in the features lacking according to memorized letter models.

Colloquial pronunciation, just like careless writing, is simplified by leaving out certain diacritic features of phonemes which, as we see, are not all necessary for communication. It is thanks to this fact that we are able to understand children learning how to speak and foreigners who distort the language, although in neither case are all the diacritic features realized. These dispensable diacritic features are called redundant features, and the phenomenon of excessive features is called redundancy (from Latin *redundo* 'to overflow the banks', 'to be in excess'). This phenomenon can be represented schematically, as in Fig. 4.

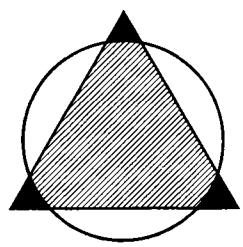


Fig. 4. Sound and phoneme in colloquial speech

Here, the circle represents the sound, i.e., the set of acoustical features actually produced by the speaker, while the triangle represents the phoneme, conceptualized as a set of diacritic features. The greater part of the triangle, shaded in in the figure, lies within the circle. These are the diacritic features actually produced by the speaker, and therefore entering into the composition of both the phoneme and the sound. The three apexes of the triangle, which are totally filled in in the figure and which extend outside the circle, represent the diacritic features of the phoneme not realized by the speaker, but filled in by the receiver, i.e., redundant features which turn out to be unnecessary for communication. The parts of the circle lying outside the triangle, not shaded in in the figure, represent the non-diacritic features of the sound, which do not constitute part of the phoneme.

In colloquial speech, we do not reproduce model phonemes, just as in freehand writing we do not write model letters. We speak in simplified abbreviations of phonemes, leaving out what is unnecessary in a given situation, and the same is true of writing. It is the receiver who imposes the complete phoneme model on the abbreviations contained in the sound waves of speech, filling in what was left out. This process may be compared to copying an illegible manuscript in script, type or print. This fact is extremely significant. Because, in colloquial speech there are no model phonemes, it is impossible to extract them by means of purely natural methods, just as it is impossible to establish the model shape of letters on the basis of careless

handwriting or to determine the complete names of institutions solely on the basis of abbreviations such as IPA or YMCA if we are not already familiar with the meaning of these abbreviations. Phonology therefore, i.e., the science of the phonological system of language, in order to establish phoneme models, must make use of its own particular method, based on the establishment of semantic oppositions between units of speech. Thus, for example, we say that the words *kość* (bone) and *gość* (guest) have different meanings and that their initial sections constitute a semantic opposition. Not until the next stage of investigations is the fact established, on the basis of observations of the process of speaking, that this opposition involves a certain difference in articulation connected with a definite acoustical difference. Thus *k* is pronounced with the vocal chords open, while *g* is pronounced with the vocal chords closed and vibrating. Consequently, *k* is acoustically voiceless, while *g* is voiced. These are diacritic features of both phonemes. Proceeding in this way, we gradually determine all of these features.

Having determined the diacritic features of one phoneme, we must determine those of other phonemes, for what we are trying to describe, are the oppositions existing among them. It is impossible, therefore, to study isolated phonemes; instead, we must simultaneously analyse all the phonemes of a language, all the oppositions existing among them and, thus, the entire phonological system. Only when we have described the entire system have we described in complete form all the oppositions, all the diacritic features and all the phonemes of a given language. To determine the phonemes of a language, we must analyse the structure of the entire language. It is this which constitutes the difficulty of phonological investigations.

The second basic difficulty in such investigations stems from the complexity of the relationship between the articulatory and the acoustical aspects of all sound features, and, consequently, of their semantic features, which, as diacritic features, enter into the make-up of phonemes. Only the acoustical aspect of diacritic features is of real significance in communication. Sounds, in the form of sound waves, reach the ear of the receiver, but only the diacritic features of the sound waves evoke a reaction in his acoustical apparatus and nervous system enabling him to distinguish a given language sign from all others and to identify it with one of the signs previously memorized. The difficulty lies in the fact that the same acoustical feature may be evoked by different articulatory movements, that one movement may substitute or compensate for another in producing the same diacritic acoustical feature. There is no strict correlation between the articulatory and acoustical features of sounds, but it is only the latter which have meaning in communication. It is as though the following principle held true in language communication: "Articulate as you wish and as you can, as long as the air waves you produce contain the diacritic acoustical features necessary for distinguishing and identifying language signs." At the same time, the description of acoustical features is much more difficult than the description of their articulation, which is why linguists have traditionally limited themselves to describing the movements of the speech organs. Only in the last

few years has precise research begun on the acoustical aspects of the diacritic features of phonemes; they have not yet, however, lead to generally accepted conclusions.

In order to properly evaluate the conclusions of phonology, it must be kept in mind that at its present stage of development, this science provides descriptions of diacritic features, of phonemes and of the phonological system of language that are based on two assumptions which are valid only in part. In the first place, phonology describes model phonemes such as those which are presented in Fig. 4. In the second place, phonology assumes a strict correlation between the articulatory and acoustical aspects of diacritic features, and having described the articulatory aspect, considers this description sufficient. The descriptions of contemporary phonologists, based on these assumptions, are valid in principle, but in the future they will probably be refined. Recognizing this fact, we shall now attempt to present the formulations concerning the structure of the phonological system of language established by the science of phonology.

First, we will consider the important question of the relationship between sounds and phonemes. We know that diacritic features constitute an unchangeable and permanent element of sounds, which are realized in various forms. Sounds having the same diacritic features and differing only in their non-diacritic features are called variants of a single phoneme, i.e., its allophones. Sounds, for example, containing the Polish phoneme *n* have a dual nature, depending primarily on the phonemes adjacent to them. Normally, they are lamino-nasal consonants (apico-alveolar closures)—e.g., *nasz*; in the position before the dorsal consonants *k*, *g*, *χ*, however, as a result of assimilation by these consonants, dorsal *ŋ* usually appears (dorso-velar closure)—e.g., *baŋk*—although the pronunciation *bank* is also possible. The fact that *baŋk* and *bank* have the same meaning, are the same word, proves that the difference between *ŋ* and *n* is not a phonological opposition in Polish and that the sounds *ŋ* and *n* are merely variants (allophones) of a single phoneme.

Each language has different features of sounds which serve to distinguish and separate words from one another, i.e., they have different diacritic features and phonemes. In Polish, for example, nasal vowels constitute distinct phonemes, for words in the nasal vowel—*ǫ* (written *ą*) differ in meaning from words distinguishable from them only in that they have the ending *-o* or *-om*, e.g., instrumental singular *żoną*, vocative singular *żono* and dative plural *żonom*. In Polish, it follows, nasality or the absence of nasality are diacritic features of vowels, while quantity is not a diacritic feature of vowels, for there is no case in which Polish words can be distinguished on the sole basis of the fact that in one word the vowel is short, while in another it is long. This is why we may, for example, pronounce the word *pył* with either a short *y* or a long *ȳ*, without changing the word. In Latin, however, the converse is true—while nasal vowels do not exist, since there is no phonological opposition of the type *-ą:-am*, at the end of words, quantity is a diacritic feature of vowels, for in Latin, pairs of words exist which differ from each other on the sole

basis of the quantity of a single vowel, e.g., nominative singular *Roma* 'Rome', and ablative singular *Romā*, 'from Rome'; Present tense *venit*, 'he comes', and Past Perfect *vēnit*, 'he came'. This difference of phonological features constitutes a great difficulty in learning foreign languages. If one knows only one's native language, one is aware of only those diacritic features functioning in that language to distinguish and separate words. All other features of sounds escape one's attention. In beginning to study a foreign language, one at first hears and repeats only those features of that language which constitute diacritic features in ones native language. The phonemes of the foreign language are identified with those of the native language. On the other hand, one fails to hear those diacritic features in the foreign language which are not found in one's own language and, consequently, one is unable to distinguish between certain words. For a Pole, the English words *man* and *men* sound the same. In order to master a foreign language, one must be aware of what sound features serve to distinguish and separate words in it, i.e., what diacritic features and what phonemes the language comprises.

Diacritic features are, in the first place, distinctive features of sounds, i.e., features which distinguish words from one another. Thus, for example, the voicing of the consonant *z* in the word *koza* (goat) is a phonological feature, since it distinguishes that word from the word *kosa* (scythe), in which *s* is voiceless. As we see, distinctive diacritic features are differences between sounds which are connected with differences in meaning. The existence of such a difference presupposes the existence of two sounds which embody, on the basis of a given feature, a contrast, or opposition. The voicing of the sound *z*, therefore, is a diacritic feature, since voiced *z* is contrasted with the voiceless sound *s*. No short vowels exist in Polish, however, because Polish does not make use of long vowels. Opposition between phonemes are of various types, which will now be described in turn.

On the basis of the number of distinctive features occurring in a given opposition we may distinguish three types of opposition: privative, gradual and equipollent.

Privative oppositions are based on the presence or absence of a single feature. Here, for example, is included the opposition between the Polish phonemes *p* and *b*. The voiced quality of the phoneme *b*, produced by drawing together and vibrating the vocal chords—which is lacking in the phoneme *p*—in this case constitutes the distinctive feature. The phoneme possessing the positive feature of the opposition is called the marked member, while the phoneme characterized solely by the absence of this feature is called the unmarked member. We consider as the unmarked member that which appears in the greater number of positions, while that appearing in fewer positions is the marked member. In the previously described opposition *p*:*b*, the phoneme *p* is the unmarked member because it appears not only before voiceless consonants and vowels (*bapka*), but also at the end of words (*słup*), while the phoneme *b* is a marked member because it appears only before voiced vowels (*był*) and voiced consonants (*dobry*); it does not appear at the end of words,

where *p* always appears (e.g., *domp*, written *dqb*). The positive feature in the opposition *p*:*b* is voicing, which appears in the marked phoneme *b*, while *p* is distinguished by the absence of this feature.

Gradual oppositions, to which we will now proceed, are those in which the members are in contrast with each other on the basis of varying degrees of intensity of a given feature. They constitute chains of several oppositions in the members of which the given feature appears in greater and greater intensity. The opposition of the vowels *u*:*o*:*a*, for example, constitutes such a chain, in respect to the degree of volume, from the minimum degree of intensity of this feature in the vowel *u*, through the medium member *o*, to the maximum member *a*.

Equipollent oppositions are based on the contrast between two different but functionally corresponding features, one of which appears in one phoneme, and the other in the second phoneme. Take, for example, the opposition of the consonants *p*:*t*, which contrast with each other on the basis of their points of articulation. Both of these consonants have the same acoustical feature, resulting from closure of the speech organs; each of them, however, is different, for the articulation of *p* involves labial closure, while the articulation of *t* involves apico-dental closure. The three types of oppositions described here are presented graphically in Fig. 5.

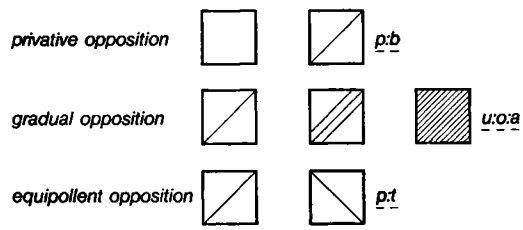


Fig. 5. Types of oppositions

In the first case, we have the absence of a feature in opposition to an already existing feature (represented by a square with a diagonal); in the second case, we have the greater and greater intensity of a single feature (represented by more extensive hatching in the square); in the third case, we have the opposition of two different features (represented by two squares with diagonals in opposite directions). In the three types of oppositions presented here, privative oppositions most clearly differentiate the distinctive feature. The comparative base of the opposition, i.e., the set of phonological features common to both phonemes, actually exists here as an unmarked member and, consequently, we have: marked member = unmarked member + distinctive feature. This fact infinitely facilitates the analysis of the marked member and the differentiation of the distinctive feature. For example, in the opposition of the Polish phonemes *p*:*b*, the comparative base is identical with phoneme *p*, and *b* = *p*+voicing.

Privative oppositions may be divided into proportional and isolated oppositions. An opposition is proportional when the difference between its members is identical with the difference between the members of some other opposition in the same system. An opposition is isolated when the difference between its members is the only one of its type in the system. Oppositions which are privative and proportional are called the phonological categories of a language. It sometimes happens, however, that alongside a series of proportional oppositions constituting a category, occurs a phoneme corresponding to the unmarked member of a category the marked member of which is missing. This missing member of the opposition constitutes an empty slot in the system which may be filled in by a phoneme possessing the features determined by the structure of the category. It must have a comparative base identical with that of the single member and a positive differentiating feature like that in the remaining members of the category. In Polish, for example, we have the phonological category of softness composed of a series of proportional privative oppositions in which the unmarked member, which is hard, is opposed by a marked member differing from it only in its softness[1]: *p*:*ṕ* (*pysk*:*pisk*), *b*:*b'* (*był*:*bił*), *v*:*v́* (*wyć*:*wić*), *m*:*ḿ* (*motać*:*miotać*), *n*:*ń* (*ran*:*rań*). Consequently, there exists in the Polish phonological system an empty slot next to the phoneme *r*, which could be filled in by the soft *ŕ*, which constitutes, with the phoneme *r*, the opposition *r*:*ŕ*, proportional to the previously mentioned oppositions. Such phonemes like Polish *ŕ*, which have not yet been established in the language but which are possible as forms filling in empty slots in the system, are called potential phonemes. As an example, we have the freshly borrowed Polish words of the type *bridż* (alongside the older form, *brydż*), *risotto*, *ring*, in which the potential phoneme *ŕ* already appears. This same potential phoneme *ŕ* distinguished the meaning of the borrowed word *trik* (trick) from the native word *tryk* (ram).

Thus, besides normal phonemes already established in the language, potential phonemes, which arise to fill in empty slots in the system, also enter into the composition of the phonological system. Owing to these potential phonemes, the phonological system functions, i.e., it flexibly adapts itself to increasingly new needs, which do not disrupt the system, but supplement it. This productivity takes place primarily as a result of the stabilization of borrowed words in the language. They introduce new phonological combinations which may be adapted to the existing system only through elaborating that system by filling in empty slots.

The structure of phonological systems will be further discussed in Part IV (Chapter 14), dedicated to the typology of languages.

---

[1] In Polish, we have one phoneme *i*, which appears in two phonetic variants, i.e., as a closed *i* appearing at the beginning of words and after soft consonants (*idę*, *pisk*), and as a more open and more back *y*, appearing after hard consonants (*pysk*).

Note: Describe the structure and functioning of the human speech apparatus. What types of sounds is it capable of producing? The diacritic features of sounds are either distinctive, distinguishing language signs, or delimiting, separating such signs from one another. A set of simultaneously occurring diacritic features is called a phoneme. Dispensable diacritic features are called redundant features. Allophones, or variants of a phoneme, are sounds having the same diacritic features and differing only in their non-diacritic features. Privative oppositions are based on the presence or absence of a single feature; gradual oppositions are based on varying degrees of intensity of a single feature; equipollent oppositions are based on the contrast between two different features. Proportional oppositions exist between the phonemes of two or more pairs of phonemes, while isolated oppositions exist only between the phonemes of a single pair. Oppositions which are privative and proportional constitute phonological categories. A phoneme which has not yet become established in a language, but which is possible as a form filling in empty slots within a phonological category, is called a potential phoneme.

Chapter 6. THE SEMANTIC SYSTEM

Phonemes, as a set of diacritic features, differentiate words on the associational plane and separate them from one another on the textual plane. The phonological system, which represents the most basic level in the hierarchy of the total language system, is limited to three functions. We will now proceed to the next level—to the semantic system—which must also be examined in two aspects—associational and textual.

The elements of the phonological system—diacritic features and phonemes—do not directly refer to contiguous reality; only more complete forms of a higher order, i.e., words composed of phonemes, refer to such reality. This relation to various objects is what we call the semantic function of words. Fundamental to the analysis of the semantic system of language is the question as to what causes words to refer to definite phenomena in our surroundings, what links words to these phenomena' how do they fulfill the semantic function? The semantic function is fulfilled in three ways, by means of three techniques of designating phenomena: denoting, referring and ordering. Thus, in all the languages of the world, there are three basic components of the semantic system: denoting words, the most numerous; referential words, or pronouns; ordering words, or numerals. Each of these three categories of words is capable of designating the same phenomena, each, however, in a different way.

Denoting words designate certain phenomena by virtue of the fact that they are part of the lexical system of language. On the basis of centuries of experience, society has created a classificatory system which it passes on by means of tradition and in which all of the phenomena accessible to the senses and thoughts of a society's members are divided, on the basis of certain characteristics, into classes. Each class is associated with a certain set of phonemes together with which it forms a word consisting of two parts: the thing designated, i.e., the meaning content of the word, encompassing a given class of phenomena, and the thing which designates, i.e., the

form of the word, composed of phonemes. This entire system of mutually limiting word usages is called the lexical system. Its general structure is similar in all languages. The classes of phenomena related to various words never exist all in the same semantic plane. They are always found in various planes, as a result of which a given word is in contrast not only with a word adjacent to it in the same plane, having a completely different usage, but also with a word in a higher plane, having the same usage and including other usages as well. The Polish words, for example, *krzesło* (chair), *stół* (table), *fotel* (armchair), *stołek* (small table), *taboret* (stool), designate related groups of phenomena, classes existing in the same plane with mutually limiting usages; all of these words are, however, at the same time, in contrast with a word existing in a higher plane—*sprzęt* (furnishings) designates a group of objects including in its usage all of the groups encompassed by the series of words listed above.

The lexical system of language embodies a world-view which is imposed on the individual by social tradition. Our sense perceptions are ordered on the basis of characteristics introduced by this system, in such a way that each set of impressions may be fitted into a class defined by a single word. The lexical system is established in childhood. Words impose a certain choice of features upon the child on the basis of which he classifies his experience. In this way, the child acquires a scheme for ordering the world. Language cooperates in this acquisition, segmentation and ordering of experience which, without language, would take another form. Many words have completely arbitrary usages which do not correspond to any class of phenomena existing in reality. The Polish word *jarzyna* (legume), e.g., encompasses a great variety of plants related solely by the fact that they are edible and planted in spring—*marchew* (carrots), *kapusta* (cabbage), *sałata* (lettuce), *groch* (sweet peas), etc.—but does not include *ziemniaki* (potatoes), which were introduced into Poland after the meaning of the word *jarzyna* had already become established. Such a word imposes upon our minds the existence of a class of phenomena which does not, in any objective sense, exist.

Once established, the lexical system embodies an image of the world used by all the members of a given society, and it is because of this commonly shared image that individuals are able to understand one another. This is accomplished by means of a complicated translation of one system of signs into those of another system, which we call recoding (code switching). The sense impressions of the person speaking are translated into a word composed of an acoustical image and the representation related to it. As a result of this operation, in a sense, a new word arises, which enters into the composition of the text created by the sender. The new word appearing in the text has certain features in common with the word existing in the system, which it reproduces. It is formed by the same set of phonemes, while the meaning content which it designates, its so-called textual meaning, is partly identical, and partly different from the value of the word reproduced, a value which it has by virtue of the system, on the associational plane. This is because of the fact that,

within the context of the system, the value (semantic invariance) of the word is determined solely by the structure of the system itself, i.e., by the value of both the adjacent words and those at a higher level of generality. In the textual plane, however, two new factors help in defining the meaning content: the meaning of the words which together with the given word constitute the context, and the consituation, i.e., the external circumstances in reference to which the sentence appears.

The textual meaning of the word rarely coincides with its value as determined by the structure of the system, as it does, for example, in reference to the words *human* and *mammal* in the abstract definition—"Every human is a mammal". Normally, the meaning of a word is not identical with its value. Differences between meaning and value involve three types of changes; narrowing, broadening or transference of usage. Most frequently, the context and consituation add certain new features to the value of a word as determined by the system, thus narrowing its usage. Such cases involve a narrowing of the content designated by the word as can be seen, for example, in the sentence—"I'm going to town" in the meaning "I'm going to Kraków"—where the invariant value of the word *town* has a smaller number of features and a greater range of usage than the invariant value of the word Kraków. This change takes place as a result of the influence of the consituation, where the closest city at the moment of speaking was Kraków. Had the nearest city been Warsaw, the word *town* would have meant *Warsaw*. In other statements, the opposite phenomenon occurs, i.e., the set of features is impoverished, thus broadening the range of usage of the word in relation to its value, e.g., *sails on the sea* meaning *sailboats*. As we see, in this case, a part (sail) stands for the whole (sailboat).

Metaphorical usages are also found in texts, in which case the meaning of a word lies in a different sphere from that of its value. There are several possibilities: metaphor sensu stricto, in which the meaning, although varying in its usage, has, nevertheless, certain features with the value (*Jan is a lamb.*—John is similar to a lamb); metonymy, in which the meaning is associated with the value on the basis of a causal connection, spatial or temporal (*I'm reading Mickiewicz.*—I'm reading the works of Mickiewicz).

The words constituting the text are recoded by the sender into articulatory movements, which in turn evoke sounds, which reach the ear of the receiver and finally, in the form of nervous stimuli, the speech center of his brain. Here, the word heard is identified with one of the memorized words belonging to the language system. This process is made possible by the diacritic features of the phonemes of which the word is composed, features which differentiate this word from all other words. On the other hand, this process greatly accelerates the formation of the structure of the lexical system, which involves ordering words in our subconscious in the order of their frequency in the text, beginning with the most frequent and proceeding to the least frequent. In searching for memorized words, we begin with the few classes of forms most frequently used. If the word heard belongs to one of these classes,

which happens more than fifty percent of the time, we identify it with the word already known within a tiny fraction of a second.

A word sought in this way changes its value as determined by the structure of the lexical system, which is shared by all members of society, under the influence of those same factors which shaped the meaning for the sender, i.e., the context and the consituation. Because the same semantic factors influenced both the sender and the receiver, the meaning of the word is the same for them both. It includes the same meaning contents and, through their mediation, refers to the same stimuli in the external world. When I say the word *table*, everyone directs his glance toward the center of the room in search of this object; when I say the word *snow*, everyone sees a patch of white, etc.

In all the languages of the world, denoting words are divided into two categories: common nouns and proper names. The value of a common noun is a class including a large number of like members, e.g., *human*, while the value of a proper name is a single individual, e.g., *Sophocles*. Only the system of common nouns is fundamental to language. It is the only classificatory system encompassing the entire world which man possesses and, thus, its significance reaches far beyond purely language phenomena. Essentially, all of the sign systems used by man are based on language; language is the basis of man's thinking and acting. Proper names, on the other hand, do not comprise a coherent system, and play a secondary role in language.

After denoting, the second technique of realizing the semantic function of language is referring. We may use a finger to point or indicate. The vision of the receiver follows the line of the extended finger, falling on some object to which the sender is calling attention. Referential words, i.e., pronouns, fulfill the same function as do a pointing finger or signposts at crossroads, for referential words comprise a conventional system of means for referring in typical situations of language communication. The systemic value of pronouns, which occur in all the languages of the world, involves invariant directions of reference in a situation consisting of three basic elements—sender, receiver and everything existing apart from them, which makes up the consituation. The first person pronoun, Polish *ja* (I), refers to the sender, the second person pronoun, *ty* (you), refers to the receiver, and the third person pronoun, *ten, tamten, ów, on* (this, that, this one, he), refers to that which, existing apart from the sender and the receiver, constitutes, in the broadest sense, the consituation of the act of speech. In a text, the pronoun referring to a certain person or thing, calls the sender's and the receiver's attention to that person or thing, thus becoming a semantic sign equivalent to a denoting word. The pronoun *ja* may have the same meaning as the name *Adam*, and the pronoun *ten*, as the denoting word *kapelusz* (hat). The textual meaning of both categories of words may be the same, but their value in the language system is different. The denoting words—*Adam, kapelusz*—have as their value a class of phenomena, while the pronouns—*ja, ten*—have as their value a certain direction of reference.

The third technique serving the semantic system of language is ordering, i.e., designating a certain object by indicating its place in a series. A person always orders on the basis of some objectively existing series of elements by setting the designated object in a definite place in this series. The parts of the body constitute such a series. A shorter series is represented by the paired members, like ears and hands, and a longer series by fingers and toes. Numerals constitute a series of elements of the same type. The difference lies in the fact that, in the former case, conventional words stand for material parts of the body. Numerals form a series by virtue of the fact each successive numeral calls to mind all the preceding ones. And so the numeral *five* reminds us of the series of numerals *one, two, three, four,* which preceds *five* in the series of numerals. The invariant value of the numeral *five,* therefore, consists solely in the fact that it reminds us of the series of four numerals following which it appears as the fifth. Neither the ordinal number *fifth,* nor the cardinal number *five* exists apart from the series which it evokes in the memory.

An infinite number of objects in the world constitute series. Attention may be called to such an object which is part of a series by giving its place in the series, which we do when we give a number which occupies the same place in the series of numbers as the given object occupies in its series of objects. Because the place in the series is identical, the given number becomes a semantic sign for the object, equivalent to a denoting or referential word. In the cloakroom, in asking for one's hat, a person may say to the attendant—*Hat*! (denoting), *That one*! (referring), *The third one*! (ordering), or, more accurately, *The brown hat!, That hat!, The third hat!* In being asked in a shop—*How many apples*? I answer—*Five,* and this number becomes the sign for a certain group of fruit.

Despite its richness, the semantic system of language is not sufficient for designating the ever new meaning contents formulated by the human mind. The semantic system must, therefore, be productive, i.e., it must constantly be creating new signs for designating new meaning contents as well as newly interpreted old meaning contents. This semantic creativity appears in several forms; in order to understand it, however, we must take into consideration certain properties of the structure of the lexical system of language, which includes denoting words and constitutes the most important part of the semantic system.

The lexical system does not strictly limit the number of words in the system and allows for the situation in which two or more words correspond to one and the same class of phenomena, so-called synonymy, as well as the situation in which one word corresponds to two or more classes of phenomena, so-called homonymy or polysemy. Such words in Polish as *chorągiew* (banner) and *sztandar* (flag) are synonyms, for they refer to the same object. A more difficult problem is involved in homonyms, in relation to which the question arises as to how we can determine whether we are dealing with one or more classes of phenomena. The majority of classes of phenomena designated by homonyms are located at a distance from one another in the lexical system, as manifested in the fact that the homonym is simulta-

neously opposed to various groups of words. The Polish word *zamek* (castle, lock, lock on a gun), for example, is such a homonym. This word refers to three classes of phenomena and, as the sign for each of them, is adjacent to different words and is opposed to them by virtue of different features. When *zamek* means 'castle', it is adjacent to words like *dom* (house), *pałac* (palace), *kamienica* (apartment-house), *dwór* (manor); when *zamek* means 'lock', it is adjacent to words like *zakrętka* (turnbuckle), *zatrzask* (latch), *zamykać* (to lock); when *zamek* means 'lock on a gun', it is adjacent to words like *lufa* (barrel), *kolba* (butt). In German we have the same homonymy as in Polish, but in English we have two words—*castle* and *lock*, just as in French, where *château*, 'castle', differs from *serrure*, 'lock in a door'.

The existence of synonyms and homonyms proves that the lexical system of a language has a rather loose structure, significantly looser than that of the phonological system. The number of elements in this system, i.e., of words, cannot be even approximately determined. As a result, new words may be continuously introduced to the lexical system, not only for new meaning contents but also for old meaning contents which already have equivalents in the newly adopted words. New words for these same meaning contents may be introduced into the system and become established as synonyms of older words, so long as they differ from these older words in connotation. The existence of the word *słońce* (sun) did not prevent the formation of another word for the same meaning content (although there is only one sun!), the word *słoneczko* (sun), which has a softer tone; despite the existence in Polish of the words *jeść* (eat) and *spożywać* (consume), the word *konsumować* was borrowed (from Latin *consumere*) and became established in the language because of its connotation of grandeur. We will discuss these phenomena more fully in Chapter 8, which is devoted to the stylistic system of language; here, however, we will discuss only the types of expansion of the lexical system, involving denoting words, at the same time pointing out that the two remaining semantic systems—those of pronouns and numerals—have a more coherent structure and cannot so easily be expanded.

The simplest form of enriching the lexical system is to introduce new elements originating from outside the system. Of primary importance are natural word sources, which involve the stabilization of speech forms which originally were not part of language—on the one hand, we have one-class signs of children of the type *mama, baba, tata, papa*, etc.; on the other hand, we have onomatopoeic images, in which the speaker imitates the sounds of the external world by means of speech sounds (e.g., *kukułka* (cuckoo), *tik-tak* (tic-tock), *bum* (boom), *jojczeć* (to say oh! oh!)), as well as movements and shapes (*dyndać* (dangle), *bimbać* (to not give a hang), *zygzak* (zigzag)). We will return to this question in our discussion of language stylistics.

The second type of forms which play a part in the enrichment of the lexical system of language includes all types of words borrowed from other languages as lexical calque or translation loan words, which involve the use of elements of the native

language composed on the model of a foreign form. These questions will be discussed in detail in Part III (Chapter 12), devoted to historical linguistics.

The second basic form of productivity of a language's lexical system, its adaptation to new needs, involves attaching new semantic values to old words. In favorable conditions, a new textual meaning of a word may become fixed in the language as a new value. The word *świeca* originally meant 'light', 'light source'. In Old Slavic, this word is used to mean 'campfire'. When rolls of wax with a wick inside for burning were introduced in Poland, these rolls were called *świece* (candles). When, in modern times, the idea of units of light measurement arose, these units were also called *świece*, whence the expression *żarówka o sile piętnastu świec* (a fifteen candle light). In each case, the textual meaning of the word was adapted to new needs, following which the language system accommodated it, adapting the defined textual meaning as a lexical value. Sometimes, a new meaning for a whole group of words used metaphorically in texts becomes stabilized in a similar way. Figures of speech arise in this way, designating meaning contents completely different from the value of the words of which they are composed, e.g., *biały kruk* (literally—white raven) means 'an extremely rare book'; *zamki na lodzie* (literally—castles on ice) means 'unrealistic plans'; *gałązka oliwna* (literally—olive branch) means 'intentions for peace'. These figures of speech, which, because they vary from language to language, cannot be translated from one into the other, are called idioms or idiomatic expressions (from the Greek *idíōma* 'special custom').

The third basic form of productivity of the lexical system involves the creation of new words on the basis of the structure of the system. The most important forms coined in this way are those which involve filling in the empty slots in the system, occurring in categories of proportional opposition, as previously described (cf. p. 42). In discussing this subject, we must begin by defining the morpheme as the smallest element of language possessing a distinct and constant function in language communication. This definition requires further clarification. In comparing the following Polish words with one another: *rękᵃ:za-ręcz-ony:ręc-e:rąk:rącz-ka*, we observe that the root morpheme occurs in five different forms, each of which is composed in part of other phonemes. The question arises as to why these five groups of phonemes are considered one and the same morpheme. They are considered as such primarily because of the fact that they have the same value, they designate the same group of objects. This criterion, however, is insufficient. The forms *człowiek* (person) and *ludzie* (people) are in the same relation to each other as *rękᵃ* and *ręc-e*, but the elements *rękᵃ-* and *ręc-* are considered the same morpheme, while the forms *człowiek* and *ludzi-e* are considered different morphemes. In order to resolve this problem, attention must be called to the fact that each of the various forms of the morpheme *rękᵃ-:ręcz-:ręc-: rącz-:rąk-* occurs in a different and strictly defined phoneme environment and, furthermore, that the differences among these forms as well as the differences in their phoneme environments are repeated in other series of Polish words. Thus, the series of forms *rękᵃ:za-ręcz-ony:ręc-e:rąk* is equivalent

to the series *mękˈa:zaˈmęczˈony:męcˈe:mąk*. The vowel in these two roots, as in many others, occurs either in the form *ę* or *ą*, while the final consonant either in the form *k*, *cz* or *c*. Thus we may say that semantically indivisible elements of language are considered to be one and the same morpheme (1) to the extent that they have the same function in language communication and (2) to the extent that the phonological differences existing among them are repeated in a manner proportional to other morphemes of the same language. Different language elements belonging to the same morpheme, e.g., *rąk:ręc-* are called allomorphs, or morpheme variants.

A group of two or more phonemes which, depending on the structure of the word, substitute for one another in the same position in a single morpheme is called an alternating series, or morphoneme; the phenomenon of such phoneme interchanges is called morphological alternation. In order to define an alternating series, it is necessary to determine the phonemes of which it is composed and the position in the morpheme in which the phonemes alternate with one another. Only such phoneme interchanges which are repeated in the same position in the morpheme in a series of morphemes are considered to represent alternation. The sum of defined morphological alternations constitutes the morphonological system of a language. In the Slavic languages, for example, the alternating series *k:cz:c* and *ch:sz* occur only at the end of morphemes, e.g., Polish *rękˈa:ręczˈny:ręcˈe* and *uchˈo:uszˈny*, for only in this position are interchanges among these consonants repeated in a proportional manner in a series of morphemes, thereby constituting part of the morphonological system. The interchanging of the phonemes *k:cz* and *ch:sz* in other positions, however, as, for example, at the beginning of the morpheme, are isolated and do not constitute part of the system—compare Polish *kosˈa* and *czesˈać* or *chodzˈić* and *szedˈł*. Thus, the elements *kos-* and *czes-* as well as *chodź-* and *szed-*, although historically related, are not actually considered as constituting part of the same morpheme.

As we see, the structure of the morpheme is, in a sense, similar to that of the phoneme (cf. Fig. 4). The phoneme consists of a set of constant features realized as a series of sounds, which constitute its variants. A morpheme, likewise, consists of a set of features realized in its different variants. The variants of both phonemes and morphemes are to a large degree conditioned by their phoneme environment. The previously cited elements, *rękˈa:ręczˈny:ręcˈe:rąk:rączˈka* are variants of one and the same phoneme which, aside from its designating value, is characterized by three constant formal features: (1) the morpheme begins with the phoneme *r*, (2) the *r* is followed by a vowel which alternates in two forms *ę:ą*, (3) the morpheme ends with an alternating consonant series *k:cz:c*. As we see, the morpheme is characterized, not by phonemes, for they may change, but by an alternating series, or morphonemes. The Polish morpheme *senˈ:snˈu:śńˈe* (written *śnie*), for example, is a set of three alternating series (morphonemes) *s:ś*, *e:zero* (absence of vowel), as a result of which, certain variants of this morpheme have no single common

phoneme. In further discussions we will define the morpheme as the smallest set of alternating series (morphonemes) possessing a distinct constant function in language communication; the alternating forms of the morpheme will be called its variants.

At the next level of complexity, we have an entity composed of morphemes—the syntactic member—i.e., a set of morphemes which, together, fulfill the function of one member of a sentence—predicate, subject, object, adjectival modifier, adverbial modifier. In other words, a syntactic member is a set of morphemes which, as a whole, functions as the sign for a certain phenomenon, at the same time defining its own position within the sentence. Signals which separate syntactic members, i.e., diacritic features enabling the listener to divide the sentence into parts, exist in all languages. The difference rests in the fact that, in some languages, the means for deliminating syntactic members are the same for all such members, while in other languages, each type of member has its own distinctive deliminating signals which, above all, are different in the predicate than in the remaining sentence members. In the first case, we have word languages, while in the second, wordless languages. A word is a syntactic member isolated in the flow of the sentence by means of the same delimiting features, or signals, which serve to isolate other syntactic members of the same sentence. Generally speaking, a word is a syntactic member constructed according to the same scheme as are all the other syntactic members of a given language. If, however, the means for delimiting a given syntactic member are different from those which serve to delimit adjacent syntactic members, then these means do not constitute specific word-delimiting features, but, instead, constitute means for delimiting various syntactic members, such as predicate, subject, etc. If there are no general means for delimiting words, then, there are no words. The concept of syntactic member is a broader concept, which includes the concept of a word. A syntactic member is a word when it is delimited by means of certain generally applied delimiting signals. Here, we will be exclusively concerned with word languages, which are more familiar to us; our discussion of wordless languages will be found in Part IV (Chapter 16).

Polish is an example of a word language, because all of its syntactic members, regardless of their function in the sentence, have a similar structure, both in respect to morphology and to phonology. All of the syntactic members consist of a stem, which is related to a certain objective phenomenon, and which is followed by an ending embodying a series of syntactic functions. Forms constructed in this way carry the accent on the penultimate syllable, which signals the end of one form and the beginning of the next one. If we take, for example, the sentence—*Matk-a widz-i córk-ę*—each of its members, subject, predicate, direct object, is constructed in accordance with this general scheme and, consequently, is a word. As we see, in Polish, a word is characterized, first, by its division into stem and ending, and, secondly, by the accent on the penultimate syllable. Similar relationships obtain in other Slavic languages as well, with the single difference that the accent falls on the

first syllable (Czech and Slovak) and signals the beginning of the word, or it may fall on any syllable (variable accent, as in Russian, Bulgarian, etc.) where it defines only the number of words in the sentence, e.g., Russian—*Já ne znál bólee dóbrogo čelovéka.*

In analysing sentences in word languages, we observe that, in addition to words, they are composed of free morphemes which are neither independent words nor parts of other words. In the Polish sentence, for example—*Ojciec i matka nie mieszkali już w dużym mieście*—we have a series of forms (*ojciec, matka, mieszkali, dużym, mieście*) all possessing the features of words (division into stem and ending, accent); alongside of them, however, appear free morphemes which are not words (*i, nie, już, w*). Free morphemes (the prepositions *w, przy, za, z*; the conjunctions *i, a, więc*; particles *nie, tak*) are primarily means for constructing sentences. Thus, they will be described together with the syntactic system of language; here, however, we are concerned solely with words.

In word languages similar in type to the Slavic languages, each word is composed of two principal parts, i.e., a stem and an ending. The ending is a morpheme which defines the role of the word in the sentence, while the stem designates a certain objective phenomenon by means of denoting, referring, or ordering. The stem is often composed of many morphemes, among which we differentiate the root and affixes. The root is that morpheme in the stem which may appear alone by itself; affixes, on the other hand, cannot appear alone, but only in connection with roots. Affixes attached to the beginning of a root are called prefixes; those attached to the end of a root are suffixes, and those which appear within the root are called infixes. In the Polish expression, for example, *Widzę pra-dziad-k-a.*, the morpheme *-a* is an ending characterizing the word as a direct object in the sentence, the morpheme *-dziad-* is a root which may appear as the independent word *dziad*, the morpheme *pra-* is a prefix, and *-k-* is a suffix. Infixes do not occur in Slavic languages; however in the Latin word *fra-n-g-ō* 'I break', we have the first person singular ending *-o*, the root *frag-* (compare the past perfect *frēg-i*, past participle *frac-tus*) as well as the nasal infix *-n-*.

In the Slavic languages, as well as in many other languages, different endings are added to the same stem, e.g., Polish *ran-a, rani-e, ran-y, ran-ą*. Thus, a group of forms which have a common stem are considered a single word, which simply appear in various forms of inflection. The same relation exists between the word and its inflectional forms as between a phoneme or a morpheme and its variants. A stem characterizes the word, while the endings constitute its variants. Not only positively existing phonemes, e.g., *ran-a, ran-y...*; *nos-a, nos-y, nos-ów...*, can be endings, but the lack of a phoneme also constitutes an ending, a so-called zero ending (e.g., genitive plural *ran*, nominative singular *nos*), for such an ending differentiates a given word from other inflectional forms of the same word, just as phonologically existing endings differentiate the various inflectional forms of other words (e.g., in the genitive plural *ran*, the zero ending fulfills the same function as the ending *-ów* in the genitive plural *nos-ów*). Many words possess similar inflectional forms constructed by means

of the same endings and morphological alternations, e.g., *wóz, woz-u...*; *mróz, mroz-u...* The set of inflectional forms which are repeated in a parallel manner in a series of words is called a paradigm. Paradigms which are based on the categories of case, number and gender are declensional paradigms, while those based on the categories of person, number, mode, time, voice are conjugational paradigms. A group of words inflected according to the same declensional paradigm is a declension, while a group of words inflected according to the same conjugational paradigm is a conjugation. In the Slavic languages, as in Latin and Greek, there are several declensions and conjugations. They constitute the inflectional system of language, in which empty slots may appear, especially in connection with borrowed words, which originally had no inflectional forms. The borrowed form of a noun becomes the nominative and, depending on how it ends, is included in one of the existing declensions. The lacking forms of the other cases, however, constitute the empty slots in the system. They are filled in by neologisms, the form of which is determinated by the structure of the system, i.e., of the paradigm. The nominative form undergoes the same changes in declining as do the nominative forms of the other nouns of the same declension. Thus, according to the native paradigm *słup, słup-a, słup-owi,* the inflectional forms *semafor, semafor-a, semafor-owi,* are created for the borrowed form, *semafor.* In this way, the inflection system demonstrates productivity, i.e., by increasing the number of inflectional forms of words.

The inflection system demonstrates productivity in yet another way, by transferring word-units with inflections possessing a single stem into suppletive inflections including two or more stems. In such cases, a group of stems is included within one of the paradigms, thus creating one inflection and, consequently, one word. Because the singular and plural forms of nouns usually have the same stem (e.g., *gość:goście, koń:konie*) and are, therefore, inflectional forms of a single word, the singular form *człowiek* and the plural form *ludzie,* included in the same paradigm, are also treated as the forms of a single word although they have different stems. It is clear, therefore, that suppletive inflections are always based on single stem inflections, on the basis of which they are interpreted as variants of a single word. Suppletion plays a large role in the comparison of adjectives. On the basis of the similarity of the stem in the comparison of the adjectives *długi:dłuższy,* etc., the forms *dobry:lepszy; zły:gorszy* are treated as forms of a single word. The suppletive inflections of personal pronouns, although based on several roots (e.g., *ja, mnie, my, nas; ty, tobie, wy, was*) are treated as the variants of a single word on the pattern of the uniform declension of nouns (e.g., *pan, pan-a, pan-owie, pan-ów*). The same holds true in the case of suppletive verbs, e.g., *jestem:byłem; idę:szedłem,* which are treated as one word by being included in the paradigm represented by inflections of the type *kocham:kochałem; niosę:niosłem,* etc.

While suppletion involves the inclusion of various stems in one word, in cases of inflectional syncretism, one form constitutes two variants of a word. Because in many Polish declensional paradigms, the genitive and accusative have different

inflectional forms, (e.g., *wozu:wóz*; *ręki:rękę*), the form *pana* (which is the same in both the accusative and the genitive) is interpreted, on the basis of this system, as two formally identical inflectional variants of the word *pan*, as a genitive (*dom pana*), and as an accusative (*widzę pana*). Thus, it can be said that each Polish noun has seven cases, but that in the inflections of the word *pan*, the genitive and accusative are syncretic, i.e., formally identical.

Having discussed in brief the variants of words which together constitute the inflection system of language, we will now proceed to the word derivation system of language, which is based on the proportional oppositions of words. Those oppositions, in which the formal and semantic differences are repeated in series of word pairs (e.g., *mądr-y:mędrzec* = *głup-i:głupi-ec*; *kocha-ć:kochani-e* = *czyta-ć:czyta-nie*) constitute a system of word derivation. In the Slavic languages, as well as in many other languages, proportional oppositions constituting word derivation categories are based primarily on the opposition of suffixes, e.g., *dom* (absence of suffix): *dom-ek* (suffix *-ek*). The word derivation system demonstrates productivity by filling in empty slots in morphological categories with neologisms, the form of which is determined by the structure of the system. In Polish, there exists a morphological category consisting of a series of proportional oppositions of the type *złoto:złot-nik*; *las:leś-nik*, etc. Alongside the word *lot*, an empty slot existed which, at the beginning of the 20th century, was filled in by the word *lot-nik*, containing a suffix determined by the morphological category.

Suffixes frequently used to form new words are productive suffixes of a distinct and constant semantic value. In Polish, for example, the suffix *-ek* forms diminutives, as in *dom-ek* (little house); the suffix *-nie* is used to form names of actions, nomina actionis, as in *kocha-nie* (loving); the suffix *-ciel* occurs in names for the agents of actions, nomina agentis, e.g., *nauczy-ciel* (teacher); the suffix *-dło* for naming instruments, nomina instrumenti, e.g., *liczy-dło* (abacus); the suffix *-stwo* in collective nouns, e.g., *koleżeń-stwo* (comradeship); the suffix *-ość* in nouns designating certain abstract features, nomina abstracta, e.g., *białość* (whiteness), etc. Each of these productive suffixes possesses a constant, abstract value which is transformed, becoming concrete, depending on the meaning content of the root to which the suffix is attached. The mutually limiting values and usages of these suffixes constitute the basis of the word derivation system. Certain suffixes have no semantic value, having been introduced in order to transfer a word from a declension which is dying out to one which is in the process of expanding. In Polish dialects, for example, the suffix *-ak* in forms of the type *dzieciak*, *cielak*, etc. is becoming prevalent in place of *dziecię*, *cielę*; thus, words belonging to the vowel declension (*dziecię:dziecięcia*; *cielę:cielęcia*), which is dying out, are transferred to the vital declension of masculine nouns ending with a hard consonant (*wiatrak:wiatraka*). Such suffixes which do not influence the meaning content of the word, but only its inflection, are called structural suffixes.

Alongside suffix derivation, the second basic form of productivity of the word

derivation system involves coining words composed of two or more roots combined on the model of other compound words already existing in the language.

We have three basic types of such compound words: coordinate, attributive and possessive. In coordinate compound words, both members are of equal rank, for neither defines the other, and the entire complex may be substituted by a group of words linked by the coordinating conjunction "and" (Polish *i*), for example, *bialo-czerwony* 'biały i czerwony' (red and white). In attributive compound words one of the members fulfills an identifying function and the other, a differentiating function. One member identifies a given phenomenon with all the other phenomena included by the concept, while the other member expresses a particular attribute which differentiates the phenomenon designated from other phenomena of the same type, thus defining them. In the Polish words, for example, *ojcobójca* (patricide), *matkobójca* (matricide), *bratobójca* (fratricide), *królobójca* (regicide), the second member *-bójca* (killer) identifies the word with a certain more extensive group of persons, while the member, *ojco-*, *matko-*, *brato-*, *królo-*, distinguishes from among this larger group of killers, a certain smaller and more particular group. It should be noted, that in words containing productive, expressive suffixes, the suffix is an identifying member, and the root is the differentiating and defining member. Thus, in the word *złot-nik* (goldsmith), the identifying member is the suffix *-nik*, the exponent of a group of agents, while the differentiating member is the root *złot-*. In the word *wilcz-yca* (she-wolf), the suffix *-yca* is the exponent of a group of female creatures, while the differentiating member, *wilcz-*, is one of the alternating forms of the root which appears in the word *wilk* (wolf). Here, therefore, the root fulfills a function similar to that of the first member, while the suffix fulfills a function similar to that of the second member of compounds of the type *ojco-bójca*. In possessive compounds, not only does one member define the other, but the compound as a whole designates something beyond the usage of either of the members, i.e., it designates the possessor of the second member as described by the first member. For example, *krzywonos* does not mean 'krzywy nos' (crooked nose), but refers to a person having a crooked nose; a similar situation obtains with the meaning of the compound words *krzyworączka* (he of the crooked hand), *głowonóg* (cephalopod), *żółtodziób* (yellow-beak), etc.

Attributive compounds of the type *ojcobójca* represent a type called endocentric constructions, i.e., compounds which as a whole have the same function or belong to the same class of forms as do one or more of the members of which they are composed. The compound, for example, *ojcobójca*, belongs to the same semantic class as (*za*)*-bójca*. Possessive compounds, on the other hand, of the type *krzywonos*, represent a type called exocentric constructions, i.e., those compounds which, as total complexes, have a different function or belong to a different class of forms than do any of the members of which they are composed. The compound *krzywonos*, in its entirety, designates a person who has a crooked nose, and thus, the usage of this compound is different from that of both the words *krzywy* and *nos*.

Thus, we have a general outline of three systems—the morphonological system, the inflection system and the word derivation system—which together constitute, in certain languages at least, the morphological system, the productivity of which involves the creation of new words and word forms by filling in empty slots in the system. In the 20th century, yet another way of creating new words has become prevalent. Such words are letter and syllable abbreviations which have become so established in the language that they undergo case changes. As an example of a letter abbreviation we have *ujot* (UJ = Uniwersytet Jagielloński—Jagellonian University), *ujotu, w ujocie,* etc. *Polgos* (Polskie Wydawnictwa Gospodarcze—Polish Economics Publications) is an example of a syllable abbreviation. In certain societies, abbreviations have become very widespread.

Note: The process of finding equivalents for the signs of one code among the signs of another code is called recoding (code switching). The language system facilitates recoding by ordering words according to their frequency in the text. The value of a word is the invariant part of its function in communication. The values of denoting words are classes of phenomena (including a large number of entities in common nouns and a single entity in proper names). The value of referential words (pronouns) are certain directions of reference. The value of ordering words (numerals) are defined in a series. The textual meaning of a word is defined by its function in a given text. The meaning of a word may be broader or narrower or completely different (metaphor) than that of the value; the meaning of a word is determined by its context, i.e., by the other words in the sentence of which the given word constitutes a part, and by the consituation, i.e., the external circumstances in reference to which the sentence appears. Synonyms are words which differ in form but have the same value; homonyms are words which are identical in form but have different values. The phrase "natural word source"—refers to the introduction of non-language forms into language, for example, onomatopoeic images. Lexical calques are words or groups of words composed of elements of the native language which are arranged on the model of a foreign form. Morpheme—a group of semantically indivisible elements (allomorphs) of a language form having the same function in language communication and among which phonological differences are repeated in a manner proportional to other morphemes of the same language. Morphoneme—a group of two or more phonemes which, depending on the structure of the word, substitute for one another in the same position in a single morpheme. The sum of a language's morphonemes constitutes its morphonological system. A syntactic member is a set of morphemes which designates some objective phenomenon, at the same time defining its own relationship to other syntactic members. A word is a syntactic member constructed according to the same scheme as are all the other syntactic elements of a given language. What are the functions of the stem and ending in a word? What is the difference between a root and an affix (prefix, suffix, infix)? What is a zero ending? What is a paradigm? What is the inflection system of language? What do inflectional suppletion and syncretism involve? What is the word derivation system of language and what role do productive and structural suffixes play in this system? What are endocentric and exocentric constructions?

Particular semantic elements of words, i.e., semantemes, which enter into the composition of texts, refer to concrete phenomena in the surrounding world by means of the three techniques distinguished above—denoting, referring and ordering. This defines the limits of the second function, the semantic plane, of language. But above the semantic plane, stands the third and highest, syntactic plane, in which true, complex language signs, or utterences, are found. Semantemes (in Polish, stems of words of the type *ręk-(a)*, *lat-(o)*) cannot appear as independent utterances, but only as the building blocks of which complex signs of language are constructed.

Each utterance is a set of experienced impressions structured by the individual in a certain manner. Besides semantemes, therefore, which directly refer to these impressions, utterances must be composed of structural elements which somehow order these impressions. Thus, each language sign which is comprehensible in communication—that which we call an utterance—must be a complex sign containing not only semantemes, which are simply building blocks, but also certain structural elements, which constitute the syntactic system of language.

There is an infinite number of complex signs. Every utterance is something new, not handed down by tradition, and therefore, in its entirety, is not part of the system of language. The elements, however, of which utterances are composed, are handed down by tradition and, therefore, do constitute part of the system of language. Language includes both semantemes and syntactic agents, which order and combine these semantemes into total units of a higher order, into utterances. It is this difference between semantic and syntactic agents of language which we refer to when we say that language is a two-class system. Syntactic agents, which we will now discuss, in all languages involve intonation, modifying and concrete demonstratives.

The syntactic intonation accompanying any word uttered involves changes in the level of the basic pitch produced by a variation in the number of air vibrations in a given unit of time. Intonation is basically conventional, as demonstrated by the fact that it varies in certain details in different languages. Although syntactic intonation schemes, like other elements of language forms, undergo various individual changes in articulation, they exist as components of the social tradition of language. The function of intonation is to create utterances by separating them from adjacent utterances (delimiting function) and by defining their relationship to reality (modal function).

In Polish, as in many other languages, there are three syntactic intonations: declarative, interrogative, exclamatory. Declarative intonation starts low, rises, describing a curve, then, at the end of the utterance, falls—e.g., *wróciłeś do domu* (you came home). Interrogative intonation begins low and, rising, ends at the highest point—e.g., *wróciłeś do domu*? (did you come home?) Interrogative intonation constitutes, in fact, the first half of declarative intonation; it is a fragment

interrupted in the middle, leaving us in expectation of the second half, the answer. Exclamatory intonation starts high and then falls—e.g., *wróciłeś do domu*! (you came home!). In writing, punctuation is the equivalent of intonation (.?!). The three conventional intonations may be represented diagrammatically as follows:

Fig. 6. Three conventional syntactic intonations

In order for an utterance to function and be understood, its beginning and ending must be delimited by separating them from adjacent utterances. These limits are marked when one scheme of syntactic intonation ends and the next begins. Normally, this is the lowest point of intonation, the point at which the falling intonation, or cadence, of the previous utterance meets the rising intonation, or anti-cadence, of the next utterance. If we place a comma or period, and, consequently, the intonation limit, in an improper place, the sentence becomes incomprehensible.

In order for an utterance to be understandable, in order for it to be capable of designating reality, it must not only be delimited, but also its relationship to the reality represented must be defined. The designating part of particular words and groups of words are representations contained in our minds; thus, an act of thought must occur to set these representations in the place of reality. This act, connected with a language form, is called modality. Syntactic intonation is one form of modality; frequently, it is the only form. In the semantic system, a word is only the equivalent of a certain class of impressions and representations. In order to transfer a word into a concrete language sign, an utterance, we must pronounce it with a certain intonation which delimits it and provides it with a modality, i.e., informs the receiver as to the type of act of thought by which the word refers to reality. Thus, a declarative intonation tells us that the speaker intends the representative content of the utterance to stand for reality, that he is subjectively persuaded that the utterance express reality—e.g., *wróciłeś do domu* (you came home). Interrogative intonation express uncertainty. The speaker realizes that different states of reality may exist, and would like to know which is actual—e.g., *wróciłeś do domu*? (did you come home?) Commanding intonation tells us that the speaker considers the meaning content of his utterance to be non-actual and wants it to become actual—e.g., *wracaj do domu*! (come home!). Commanding intonation is one form of exclamatory intonation and, like other intonations of this type, fulfills an expressive function, i.e., it indicates the feelings of the sender, and an impressive function, i.e., it acts upon the will and behavior of the receiver.

In speech, we have utterances consisting of a single word; only because they are spoken with one of the three syntactic intonations are they utterances. The word *ogień* (fire) becomes an utterance when it is shouted *ogień*! Many words function in the same way: the particles *tak*, *nie* (yes, no), so-called impersonal verbs—*grzmi*

(it is thundering), *błyska* (it is lightning), *mży* (it is drizzling), *świta* (it dawns); the vocative form of nouns *Janie*! *Mario*! *Profesorze*!; forms of commands—*baczność*! (attention!), *precz*! (begone!); often the imperative form of verbs—*idź*! (go!), *słuchaj*! (listen!), etc. All of these examples, within a given consituation, are complete utterances demanding no supplementation; on the other hand, they are not complex, for they are composed of a single word, they contain only one semanteme. For this category of forms, intonation is the single and sufficient agent for introducing them into the flow of speech. One-word utterances, however, are not typical. Language utterances proper are complex and, therefore, besides intonation, are composed of other special syntactic agents: modifiers and demonstratives, which combine semantemes into units of a higher order. This has the enormous advantage that these complex utterances are, to a great extent, and sometimes completely, independent of the consituation, the concomitance of which is indispensable to the functioning of one-word utterances.

Modifying involves the fact that certain categories of words create places, adjacent, for certain classes of other words, they announce and require them. Here, the syntactic connection resolves itself into the relation of the word modifying and requiring to the word modified and required. The phenomenon of modifying is a result of the semantic incompleteness of modifying words, the fact that they do not provide full information. Just as one-word utterances, they must be supplemented; but while one-word utterances require supplementation on the part of the consituation, modifying words require supplementation on the part of the context, and, because the meaning contents of these words are incomplete, their supplementation must be strictly defined. The entire set, consisting of a modifying word and the places adjacent to it in a text, which are open for defined categories of modified words, is called the syntactic scheme. These schemes, passed on by social tradition, constitute part of the system of language. The role of the speaker is limited to choosing one of the words in a class modified and placing it in the text, next to the modifying word.

Because modifying results from the meaning of words, the division of words into various groups on the basis of their modifying function to a certain degree corresponds with their division on the basis of semantic features. Those words fulfilling the same function in the system of modifying and manifesting the appropriate semantic features of this function, are included in the same class of words, which is called a part of speech. Systems of parts of speech vary widely in different language groups. Here, for purposes of illustration, the system of parts of speech of the Polish language will be presented, which does not greatly vary from the systems of other Indo-European languages.

We will begin with denoting words. They may be divided into several groups on the basis of their role in the modifying system. First of all, we must distinguish words which, in designating separate phenomena, independent entities, are semantically complete and, to a great extent, provide exhaustive information. These are primary

words, or nouns, like *dom* (house), *wilk* (wolf), *białość* (whiteness), *pisanie* (writing). Because their meaning content does not demand supplementation, they do not modify anything but are themselves modified by secondary words which, because they designate phenomena which are not autonomous, require that their meaning content be supplemented by primary words.

Secondary words can be divided into adjectives and verbs. Adjectives, like *biały* (white), *wysoki* (tall), *ojcowski* (paternal), *zgniły* (rotten), designate the dependent part of phenomena, i.e., their properties. Adjectives modify a noun of the same case, number and gender; they make place for this noun adjacent in the text, which constitutes the syntactic scheme of the nominal group. By filling in the place modified with an appropriate noun, we obtain in the text nominal groups of the type *biały dom* (white house), *wysoki człowiek* (tall person), *ojcowskie dziedzictwo* (paternal inheritance), *zgniły owoc* (rotten fruit), etc. A determining relation exists among the words comprising a nominal group. This relationship involves an adjective (e.g., *biały*), which is a modifying and determining member and which adds new features to the meaning of a noun (e.g., *dom*), the modified and determined member, thus enriching the noun's meaning content and narrowing its usage. The group *biały dom* has a richer meaning content and a narrower usage than the word *dom*; however, these language members are similar to each other to the extent that they are equally capable of becoming utterances. With the proper intonation and in the proper consituation, of course, they may be understandable language signs; but although the group *biały dom* is a complex form, it may function as a separate utterance in the same conditions as the word *dom* can. It is the verb which fundamentally changes the situation.

Verbs designate states, e.g., *siedzi* (sits), or actions, e.g., *bije* (beats) which are abstracted from a certain situation. In reality a state or action never occurs independently but is always the state or the action of some entity. Thus, the verb alone does not provide full information about the situation to be represented and requires semantic supplementation. Verbs which designate a state—intransitive verbs—*siedzi* (sits), *leży* (lies), *idzie* (goes), *bieleje* (whitens), *płonie* (flames)—require a subject, i.e., an entity the state of which they designate. Verbs which designate actions—transitive verbs *bije* (beat), *czyta* (reads), *niesie* (carry), *kocha* (loves)—require an agens, i.e., the person who is the source of an action, and a patiens, i.e., an entity which is the intended object of the action. In Polish, the subject of intransitive verbs and the agens of transitive verbs are both expressed in the same way, by the nominative form of the noun, and are both referred to by the term subject, while the patiens, expressed in Polish by the noun in the accusative, is called the direct object.

In the Polish language, the subject, so defined, is expressed by the personal exponents of the verb, which refers to the subject, as in pronouns. In other words, the personal verb forms contain a pronoun subject. For example, the first person form *idę* (I am going) refers to the sender as to the subject; the second person form *idziesz*

(you are going) refers to the receiver; the third person form *idzie* (he, she, it is going) refers to anything or anyone which is neither the sender nor the receiver and is, therefore, so general in meaning that it requires a supplement; therefore, it modifies an adjacent noun in the nominative which is called the noun subject, e.g., *ojciec idzie* (father is going). Here the noun *ojciec* is the subject and the verb *idzie*, the predicate. A predicative relation exists between these two members. The total utterance, which contains a personal predicate, is called a sentence, as opposed to all other utterances, which are called clause equivalents. The difference between these two types of utterances must be discussed.

Every utterance expresses a certain modality, its meaning content has a certain relationship to reality, which is realized by means of syntactic intonation. Each utterance has a location in time and space, we know where and when its meaning content takes place, but in clause equivalents, this is determined by the consituation only, by which it is defined and limited. A sentence, on the other hand, is to a large extent independent of intonation and consituation, because it contains a personal predicate which both expresses modality and gives locality by means of its inflectional forms. Thus, we have five basic categories of predicates: mood, which involves modality expressed, not by intonation, but by changes in verb forms, the categories of person, tense and aspect, which place the sentence in time and space, and the category of voice, which determines the structure of the sentence.

In the Slavic languages, as in many others, we have three moods: declarative, conditional and imperative. The declarative mood expresses the fact that the speaker considers the meaning content of the sentence as conforming to reality, that he intends the sentence to stand for reality, e.g., *czytam* (I am reading), *czytasz* (you are reading). The conditional mood expresses the fact that the speaker considers the meaning content of the sentence to be non-actual, but possible, in certain conditions, and desirable, e.g., *czytałbym* (I would read), *czytałbyś* (you would read). The imperative mood indicates that the meaning content of the sentence is non-actual, but demanded, e.g., *idź*! (go!), *czytaj*! (read!).

The category of person places the subject of the action within the situation of language communication, i.e., places the subject in space by indicating whether the subject is the sender, the receiver, or someone or something else.

Just as the category of person places the sentence in space, so the categories of tense and aspect place a sentence in time. Both of these categories refer to both past and future time, but from different starting points. The exponents of the category of tense refer to time from the moment of speaking, in other words, from the temporal standpoint of the person speaking; they indicate either the past, *czytałem* (I read), or the future, *będę czytał* (I will read). In the first case, we have the past tense, in the second case, the future tense. The present tense is the unmarked member of the opposition. It indicates no particular direction and includes that which is neither past nor future. In practice, this tense designates two different things, an act contemporaneous with the act of speaking, e.g., *siadam* (I am sitting), or a per-

petual activity, capable of encompassing both the past and the future—*ziemia krąży dookola słońca* (the earth revolves around the sun). Exponents of aspect indicate time from a moment independent of that of speaking, but related to the moment when the action described by the verb ends. And so, a form in the perfective aspect refers to a time in the past beginning after the completion of an action which we see as something past, e.g., *przeczytałem* (I have read), *przeczytam* (I will have read). In the imperfective aspect, on the other hand, in thought we place ourselves before the moment of completion of the action and we regard the action as developing, e.g., *czytałem* (I was reading), *czytam* (I am reading), *będę czytał* (I will be reading). In this way, the categories of tense and aspect describe the place in time of the action of the sentence. The functioning of these categories may be diagrammed as in Fig. 7.

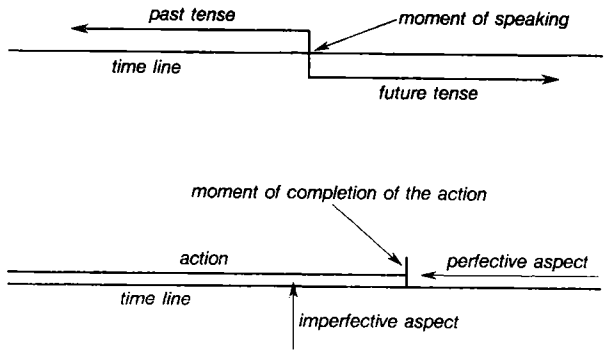


Fig. 7. Categories of tense (above) and aspect (below)

Just as the category of person indicates the subject, the category of voice establishes the relationship of a transitive verb to the subject and object. When the category of voice makes no changes in the syntactic scheme, so that the action flows from the subject to the object (direct object), we have the active voice, e.g., *mama myje córkę* (the mother is washing her daughter). When the syntactic scheme changes in such a way that the action proceeding from the subject flows back to the subject again, we have the reflexive voice, e.g., *mama myje się* (mother is washing 'herself'). When the direction of flow of action is reversed such that it reverts to the subject in the nominative, we have the passive voice, e.g., *mama jest myta przez córkę* (the mother is washed by her daughter).

In comparing the nominal group *biały dom* (white house) with the sentence *dom bieleje* (the house is whitening), it must be stated that the nominal group may be used as a clause equivalent, but this necessitates that both the intonation and consituation be functioning, e.g., *Co widzisz tam na skraju lasu?—Biały dom.* (What do you see there at the edge of the woods?—A white house.) A sentence, on the other hand, is an autonomous complex language sign independent of intonation and consituation owing solely to the fact that the inflectional form of the predicate *bieleje* expresses the indicative mood, the third person, the present tense and the im-

perfective aspect. Thus, the determining relation in the nominal group is different from the predicative relation existing between the members of the sentence. In both cases, we have a constitutive member which enters into a direct syntactic bond with higher order units, which represent the group externally and cannot be separated from it. In the nominal group, the constitutive member is the modified member *dom*, while in the sentence, it is the modifying member *bieleje*. In the sentence *dom bieleje*, the predicate *bieleje* is both the constitutive and modifying member and is, therefore, particularly important syntactically, while the nominal subject *dom* is a non-constitutive and modified member, syntactically less important. The relationships are different in the nominal group *biały dom*, where the defined member *dom* is constitutive and modified, while the defining member *biały* is non-constitutive and modifying. The structural elements of the sentence are concentrated in the predicate, while in the nominal group they are evenly distributed between the two members. The predicative relationship, therefore, is syntactically irregular with the predominance of the predicate over the subject, while the determinative relationship is one of syntactic equivalence between the determining member and the member determined.

Heretofore we have discussed intransitive verbs, which modify the subject only; now we will proceed to transitive verbs which open a larger number of places in a sentence. Transitive verbs of the type *kocha* (he, she, it loves) modify two nouns: one active, the point of departure of the actions, i.e., the agens, the second passive, the object of the action, i.e., patiens. In Polish, the agens, called the active subject, is in the nominative, just as the subject of intransitive verbs, while the patiens, called the direct object, is in the accusative. In Polish, the transitive verb *kocha* (he, she, it loves) modifies the subject in the nominative, e.g., *ojciec* (father), which answers the question who? and the direct object, e.g., *córkę* (daughter), which answers the question whom? Certain transitive verbs, for example, *daje* (he, she, it gives), also modify a noun in the dative, called the indirect object, which answers the question "to whom?" and designates the direction of the action.

Thus, we see that three types of Polish verbs—intransitive, transitive, and transitive modifying an indirect object—constitute the nuclei of three different syntactic schemes, which may be presented diagrammatically as follows:

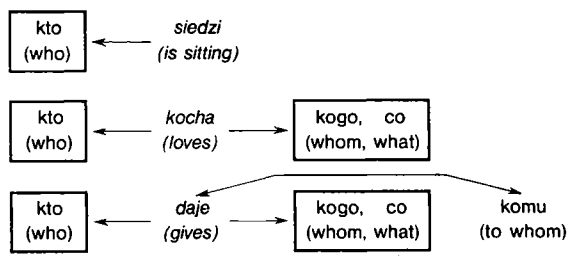


Fig. 8. Syntactic schemes

Each of these syntactic schemes may be realized by a practically infinite number

of concrete utterances, depending on what words belonging to modifying classes are used to fill in the places in the text. The first of the schemes presented may be realized in the sentences: *pies siedzi* (the dog is sitting), *człowiek siedzi* (the person is sitting); the second in the sentences: *matka kocha córkę* (the mother loves her daughter), *ojciec kocha syna* (the father loves his son); the third in the sentences: *ojciec daje książkę synowi* (the father gives his son a book), *brat dał jedzenie psu* (the brother gave the dog food). The same holds true for other schemes. The nucleus of any scheme is a sentence-generating word, the vehicle of the scheme, which modifies one or more places filled in by words of the modified class. Concrete utterances may be considered variants of the scheme which is realized in them. The relation between the syntactic scheme and the utterance is similar to that between the phoneme and sounds (cf. Fig. 4), the lexical value and meanings, the morpheme and its alternating forms, and the word and its inflectional forms.

The syntactic scheme determines either the form of the words of which it is composed, or their order in the utterance. In the latter case, it is the modifying word which determines the order of the words comprising the syntactic scheme. In French, for example, two places are created by the transitive verb—one place before the verb, for the subject, and the other after the verb, for the direct object—and this difference in position is the only difference between these two members of the sentence. In the two French sentences, for example, *Jean bat Paul* (John beats Paul) and *Paul bat Jean* (Paul beats John), the same form *Jean*, when situated before the verb *bat*, is the subject, while it is the direct object when situated after the verb. Similarly, the word *Paul*, depending on its position in the sentence, is in one case the object, and in the second, the subject.

In Polish, as in many other languages, the syntactic scheme determines, not by means of word order, but by means of the form of words alone, thus allowing for a wide choice as to word order. Two types of such schemes may be distinguished, called the syntagmatic relationships of agreement and government. In agreement, the words making up the syntactic scheme appear in the form of the same syntactic category. An adjective, for example, which modifies an adjacent noun agrees with it in gender, number and case, as a result of which, in the phrase *dobremu ojcu* (to the good father), both the noun and the adjective are in the dative masculine singular. In government, on the other hand, the words making up the syntactic scheme appear in completely different forms. The transitive verb, for example, modifies the direct object in the accusative, i.e., in a form in which it cannot itself stand, e.g., *kocham matkę* (I love mother). Thus, we say that the transitive verb governs the accusative. Sentence schemes involve government, while in nominal groups, which we will now discuss, agreement plays a huge role in syntactic schemes.

In Polish, as in many other languages, there are two basic types of nominal groups—e.g., those composed of two nouns, based on government, and those composed of a noun and an adjective, based on agreement. The first type contains, in the function of determining member, a noun in the genitive, which modifies another, adjacent,

noun. The genitive form *ojca* (father's), for example, modifies any adjacent noun which it determines (as an object modifier or possessive), e.g., *dom ojca* (father's house), *miłość ojca* (father's love), etc. In this case, the genitive is similar in function to an adjective *dom ojcowski* (paternal house) *miłość ojcowska* (paternal love). The difference, however, lies in the fact that the adjective agrees in gender, number and case with the noun it modifies, while the genitive modifies a noun in any case, e.g., *dom ojca, domowi ojca, w domach ojca.* Thus, we may say that a noun can modify an adjacent noun only in the genitive case, and, therefore, the noun modified governs the genitive, just as the transitive verb governs the accusative. In the structure of adjectival nominal groups, on the other hand, the most important function is fulfilled by the categories of gender and number based on agreement. These categories will presently be more thoroughly discussed.

The categories of number and gender are fundamentally word derivation categories which came to be used as syntactic agents in the construction of utterances. Various word derivation and lexical oppositions constitute the basis of the category of grammatical gender. In the Slavic languages, as in many others, physiological differences in the sexes play an enormous role in the gender of nouns, e.g., *pan* (gentleman) : *pani* (lady), *wilk* (wolf) : *wilczyca* (she-wolf), *koń* (horse) : *klacz* (mare). In languages of the Bantu family in Central and South Africa, differences in the shape of large and small objects or of round and elongated objects are the deciding factors in determining genders. In many languages, the classification into genders is based on the division of entities into higher and lower, i.e., into animate and inanimate, personal and impersonal, (who, what), male-personal and non-male-personal—e.g., Polish *dobrzy panowie* (good gentlemen) : *dobre panie, dzieci, domy* (good ladies, children, houses)—tribe-members and non-tribe-members. A word derivation or lexical gender becomes a grammatical, syntactic gender only when other words in a sentence change their form in order to agree with the gender of the noun, i.e., when agreement obtains. We have grammatical gender in Polish because adjectives, pronouns, numerals and verbs in certain tenses and moods change their forms to agree with the gender of the noun. Thus, in the singular we have three genders—masculine, feminine and neuter, e.g., *ten dobry pan szedł* (that good gentleman went), *ta dobra pani szła* (that good lady went), *to dobre dziecko szło* (that good child went); in the plural, we have two genders—masculine personal and non-masculine-personal, e.g., *ci dobrzy panowie szli* (those good gentlemen went), *te dobre panie* (*dzieci*) *szły* (those good ladies (children) went).

The function of grammatical gender involves the fact that in the syntactic scheme they connect words, one of which modifies the other. Thus, an adjective modifies a noun and agreement in gender indicates which adjective refers to which noun. If, in a certain language, an adjective has no gender, it must stand in a particular position in relation to the noun modified; in some languages of this type, it is always placed before the noun (as in English), while in others, always after the noun (as in Modern Persian and the Semitic languages).

The category of number is based on the word derivation category of collective nouns, each of which designates a set of entities which are related to one another in a certain way, e.g., *rodzeństwo* (sibilings). In certain cases, there is no semantic difference between a plural and a collective noun, e.g., *Stanisławowie Górscy* (plural), and *Stanisławowstwo Górscy* (collective noun). The plural of personal pronouns has a collective meaning. *My* (we) does not mean many times *ja* (I), but rather, *I* and *those* who together with *me* constitute a certain group. Likewise, *wy* (pl. you) means *you* (sing.) and *others* connected with *you.* The plural forms of verbs of the type *kochamy* (we love), *kochacie* (you love—pl.) possess the same characteristics. There is, however, a difference between plurals and collective nouns, a difference which rests in the fact that plurals are in syntactic agreement with other members of the sentence; e.g., *dobry pan* (good gentleman), *dobrzy panowie* (good gentlemen), *pan idzie* (the gentleman goes), *panie idą* (the ladies go), *ja jestem dobry* (I am good), *my jesteśmy dobrzy* (we are good), which is not true of collective nouns. Moreover, a singular noun modifies the adjacent numeral *jeden* (one), e.g., *jeden koń* (one horse), while in the plural it modifies numbers from *dwa* (two) upward, e.g., *dwa konie* (two horses). In languages which include dual number, the system differs to the extent that the singular modifies "one" (e.g., Old Slavic *jedinъ vlъkъ*), the dual number modifies "two" (Old Slavic *dъva vlъka*) and the plural modifies "three" or more (Old Slavic *trьje vlъci*). The category of number, therefore, is grammatical-syntactic, since it influences changes in the structure of the entire sentence.

The categories of gender, number and case in adjectives always function in accordance with the principle of agreement. The only category involving government in adjectives (and adverbs) is the category of degree. This is basically a word derivation category which involves designating differences in the degree of intensity of a given feature e.g., *mądry:mądrzejszy:najmądrzejszy* (wise:wiser:the wisest). The category of degree, however, is also a grammatical-syntactic category, since the form of the comparative degree—*wyższy, mądrzejszy* (taller, wiser)—is an element in the syntactic scheme appearing in a certain type of nominal group. This form modifies two places adjacent in the text—a place for nouns answering the question *kto* (who) and a place for nouns answering the question *od kogo* (than whom). The first of these places must be filled by a noun which agrees with the form of the comparative degree in gender, number and case (agreement), and the second place by a noun in the genitive, standing by itself in Old Polish, but today, with the preposition *od* (than), (government), e.g., Old Polish *syn wyższy oćca*, today, *syn wyższy od ojca* (the son taller than his father).

Heretofore, we have been discussing primary words—i.e., nouns which do not modify, but are themselves modified—and secondary words, adjectives and verbs which modify nouns. In addition, Polish, as do many other languages, includes tertiary words, i.e., adverbs like *dobrze* (well), *wysoko* (highly), *bardzo* (very), which modify secondary words—adjectives and verbs—and which are never themselves modified. Adverbs designate certain properties of features states and actions, and

are, therefore, semantically very dependent upon secondary words, which in turn are syntactically dependent upon primary words. Thus we have the syntactic series: adverb—adjective—noun, e.g., *bardzo dobry człowiek* (very good person), *wysoko kwalifikowany robotnik* (highly qualified worker) as well as the series: noun—verb—adverb, e.g., *ptak leci wysoko* (the bird flies high), *matka wychowała dobrze dzieci* (the mother brought up the children well), etc.

Only denoting words fulfill the function of verbs; a certain group of pronouns and numerals, however, fulfill the same role in the system of modifying as do nouns, adjectives and adverbs, as is schematically represented in the following table:

Words:	Denoting	Referential	Ordering
Primary	Nouns *dom, białość* (house) (whiteness)	Substantive pronouns *ja, ty, ten, on* (I) (you) (that) (he)	Cardinal numbers *dwa, trzy, cztery* (two) (three) (four)
Secondary	Adjectives *dobry, biały* (good) (white)	Possessive pronouns *mój, twój, nasz, wasz* (my) (your) (our) (your)	Ordinal numbers *pierwszy, drugi* *trzeci* (first) (second) (third)
Tertiary	Adverbs *dobrze, biało* (well), (whitely)	Adverbial pronouns *tu, tam* (here), (there)	Adverbial numbers *dwojako, trojako* (secondly), (thirdly)

The fact that verbs, basic to sentence generative syntax, are always denoting words, gives denoting words a primary importance in determining the general character of language communication. The role of pronouns and numerals, although important, is secondary.

Because the information concerning factual states provided by a certain category of words is incomplete, thus necessitating supplementation, we have a modifying system and syntactic schemes based on it, all passed on by language tradition. It is in realizing and combining these schemes that we construct the basic framework of all types of utterances. Those connections among words which are based on modifying are abstract, for they result solely from the meaning of these words, which requires supplementation. In utterances, alongside these abstract connections, based on the relation among word usages, concrete connections appear, which give rise to special relationships.

As we have already seen, the feature which distinguishes language from many other codes is its ability to designate phenomena that are distant in time and space, outside the speaker's fields of perception. Utterances which refer to such non-perceivable phenomena must create a vision the space and time of which are distinct from the space and time within which the speakers are operating. The predicate of a sentence creates this vision by virtue of its categories of person, tense and aspect, which evoke a fictional space and time in our minds. In these conditions, concrete demon-

stratives become essential, in order to orient the speakers in this fictional sphere. Let's consider the sentence: *Juliusz Cezar przeszedl przez Rubikon, a potem wkroczyl do Rzymu.* (Julius Caesar crossed the Rubicon, and, after, entered Rome.) The meaning content of this sentence is immensely distant from us in time and space, such that the time and space of this sentence represent a completely distinct sphere from that which constitutes the consituation of the utterance. This fictional sphere creates two predicates in our minds—*przeszedl* and *wkroczyl*—which designate phenomena which took place in the distant past and at a distant place and, therefore, suggest the existence of a distinct time and space. In this distinct time, we are oriented by the preposition *po* (after), which establishes the time relationship between the phenomena presented by the two simple clauses making up the compound sentence. We are oriented as to the space referred to in these two simple clauses by the prepositions *przez* and *do*, which establish the spacial relationship between Julius Caesar and the Rubicon, on the one hand, and Julius Caesar and Rome on the other. The most important group of concrete demonstratives which orient us in the spacial, temporal and causal relationships of the fictional reality created by the sentence are represented in Polish, as in many other languages, by prepositions, to which we will now turn our attention.

Prepositions are syntactic exponents which determine what direction in relation to one another the meaning contents of the words contained in a sentence have. These directions are determined by prepositions in relation to the object designated by the noun adjacent to the prepositions, such that these directions refer to space *od domu* (from home), to time *od roku* (for a year) and to causality *drżal od nienawiści* (he shook with hatred).

There are three basic directions indicated by demonstrative prepositions in these various areas—approaching *do*, *ku* (to, toward), departing *z*, *od* (from), and standing in place *w*, *koło* (in, around)—and each of them may appear in two forms, depending on whether it indicates the very center of the object (complete contact), or whether it refers to the vicinity of the object (incomplete contact). Thus, the preposition *do* indicates the center of the object *do domu* (home), the preposition *ku* indicates the direction of the object *ku domowi* (toward home), the preposition *z* expresses a departing from the center of the object *z domu* (from home), the preposition *od* departing from the vicinity of the object *od domu* (from home), the preposition *w* calls our attention to the very center of the object *w domu* (at home), the preposition *koło* calls our attention to the vicinity of the object *koło domu* (near home).

The semantic relationship of the prepositions listed here, is presented in the following table:

	Complete contact	Incomplete contact
Approach	*do* (to)	*ku* (toward)
Departing	*z* (from)	*od* (from)
State of rest	*w* (in)	*koło* (around)

The directions indicated by the most important prepositions are presented in the following diagram:

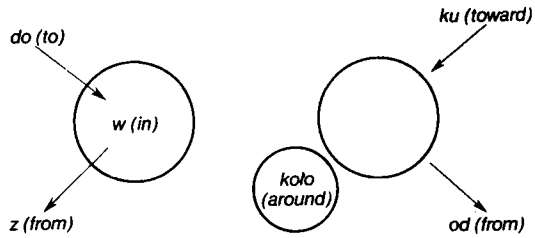


Fig. 9. Directions indicated by prepositions

Prepositions indicate certain directions in relation to the meaning content of the noun next to which they stand. In turning our attention in this direction, we come upon the meaning content of the neighboring word, thus recognizing the relationship of the two meaning contents in the temporal, spacial or causal sphere, e.g., *idę do miasta* (I'm going to town), *jabłko z drzewa* (apple from the three), *dom w mieście* (house in the city), *droga ku wsi* (road toward the village), *szedł do domu* (he went home), *drzewo koło drogi* (tree by the road). In this way, concrete exponents, together with the modifying system, combine simple language signs, i.e., words—strictly speaking, their semantemes—into complex signs, i.e., sentences.

In Polish, as in many other languages, prepositions interact with concrete case functions, which also involve indicating certain directions. In comparing the Polish expressions *w domu* and *w dom* ('home' and 'at home'), we notice that the former designates a state of rest and the latter designates approach, because, although the same preposition *w* appears in both, here the difference in meaning is connected with the difference in case. The locative (*w*) *domu* expresses a state of rest, while the accusative (*w*) *dom* expresses approach. In Latin, we have an example of concrete case functions without prepositions, e.g., accusative *Romam* (to Rome), ablative *Romā* (from Rome), locative *Romae* (in Rome). The concrete function of the Latin accusative is approach, that of the ablative, departing, and that of the locative, a state of rest.

In Polish, as in many languages, cases fulfill two functions: (1) that determined by the system of modifying, which we call the grammatical function and (2) the concrete function of indicating directions. These functions have different relationships to one another in different cases. The nominative has the grammatical function of the subject case—*ojciec idzie* (father is coming). The accusative fulfills primarily the function of the direct object case—*widzę matkę* (I see mother), but possesses a certain concrete function as well, e.g., *w dom* (as opposed to *w domu*), *pracował całą noc* (he worked all night). Today, the dative without a preposition is used only as the case of the indirect object—*dałem książkę synowi* (I gave my son a book); with the preposition *ku*, however, it has the concrete function of approach—*ku domowi* (toward home). The genitive is primarily the grammatical case of the possessive

attribute—*dom ojca* (father's house); it can be the case of one of the nouns in comparisons—Old Polish *syn wyższy oćca* (son taller than his father)—or the case of the direct object when the direct object involves a part of a whole—*daj mi chleba* (give me (a piece) of bread). The instrumental, aside from its newly acquired predicative function—*on jest królem* (he is a king)—is a concrete case expressing the presence of someone or something near or next to another person or thing, which is expressed either as the sociative instrumental—*ojciec z matką* (father with mother), *król wyjechał całym dworem* (the king left with the entire court)—the instrumental proper—*robię siekierą* (I am doing it with a hatchet)—or the instrumental of space, expressing the way by which an action takes place—*jechał drogą* (he went by road). The locative has a concrete function only, that of designating a state of rest at the center, and appears only with prepositions—*w domu* (at home). Four cases, the nominative, accusative, dative and genitive, have as their primary function that which is connected with the system of modifying and, therefore, they are called grammatical cases; the two remaining cases, on the other hand, instrumental and locative, primarily indicate directions and, therefore, are called concrete (semantic) cases. The vocative is not a case, but belongs to the class of impressive appeal signals. It will be discussed in the next chapter.

Having described the most important elements in the construction of simple sentences, we shall proceed to the syntax of the compound sentence, i.e., a set of several clauses the mutual semantic relationships of which are expressed by means of linguistic agents. The Slavic languages, as well as many others, have two basic agents for expressing these relationships and for constructing compound sentences—relative pronouns and conjunctions. These parts of speech will be described in turn.

In some languages relative pronouns developed from the anaphoric pronoun, while in other languages, from the interrogative pronoun—and they combine the functions of both of these types of pronouns. An anaphoric pronoun, in Polish, *on* (he), refers retrospectively in the text to a word which has already been stated, agreeing with it in gender, number and case—e.g., *Mój ojciec był prawdomówny. On nigdy nie kłamał.* (My father was truthful. He never lied.) In this case, *on* (he) refers to *ojciec* (father) and takes on the same meaning. Interrogative pronouns, in Polish, *kto, co,* (who, what), refer ahead in the text to the answer, which they announce, or modify. The interrogative refers to something which does not yet exist, which is going to be—i.e., to the expected answer. When this answer is missing, the interrogative pronoun becomes an indefinite pronoun—*ktoś, coś* (someone, something). Thus, while the interrogative pronoun refers ahead to the continuation of the text, which it modifies, and the anaphoric pronoun refers back in the text, the relative pronoun combines both these functions, simultaneously referring both ahead and back in the text. As a result, it may be replaced by a group composed of an anaphoric and an interrogative pronoun. And so, for example, the clause—*chłopak, którego nazywają Jankiem* (the boy whom they call Janek)—has the same meaning as the clause—*chłopak, co go nazywają Jankiem.* The various types of

indication performed by the pronouns described here, are presented diagrammatically in Fig. 10.

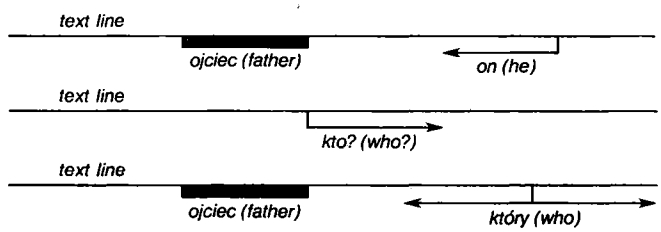


Fig. 10. Reference of pronouns in the text

The fact that the relative pronoun refers ahead as well as back in the text, enables it to connect clauses. Consider, for example, the clauses *weszli do miasta, które wyrosło przed nimi* (they entered the city, which rose before them). The relative pronoun *które* (which) refers back to a word belonging to the principal clause, (*do*) *miasta* (city), with which it agrees in gender and number (but not in case!). On the other hand, the pronoun refers to the clause following it, which it announces and of which it is a member, in this case the subject, as indicated by the nominative case. The relative pronoun is always a member of the subordinate, relative clause, modified by it, a member simultaneously referring to one of the members of the principal clause. In this way, it connects the subordinate clause with one of the members of the main clause.

Conjunctions differ from relative pronouns in that, rather than referring, they modify—in two places adjacent, that preceding and that following. The conjunction *i* (and) in no way determines what forms may fill these two places; various combinations are, therefore, possible. Other conjunctions make places only for clauses related to each other in certain ways—either coordinate clauses (parataxis) or subordinate clauses (hypotaxis). Coordinating conjunctions modify two clauses related to each other either as oppositions—*a, ale, lecz* (but)—alternatives—*albo...albo, bądź...bądź* (either ...or)—or results—*więc, przeto* (thus, therefore). Clauses which begin with subordinate conjunctions—*że, gdy, gdzie, choć, aby*, etc., (that, since, when, although, in order to, etc.,) modify subordinating clauses just as adjectives modify nouns, i.e., as a consequence of the incompleteness of the information given, which requires supplementation. Take, for example, the sentence—*Jan powiedział ojcu, że już dawno wrócił do domu* (Jan told his father that he had come home a long time ago.) The fragment of this sentence—*już dawno wrócił do domu* (he had come home a long time ago)—may act as an independent sentence; however, the clause—*że już dawno wrócił do domu* (that he had come home a long time ago)—owing to the presence of the conjunction *że* (that), modifies a subordinating clause. Conjunctions of the type *że* (that) connect subordinate clauses to the entire subordinating clauses which they modify, while relative pronouns connect them to only one of the words of the main clause, to which they refer.

The clause signalled by the conjunction is in essence subordinate because its sentence generating element, i.e., its verb predicate, provides incomplete information, in relation to the predicate of the subordinating clause, which supplements this information. This subordinateness, therefore, can be expressed not only by a conjunction, but also by a special category of subordinate predicate. This category exists in those languages in which the predicate of the subordinate clause has certain formal features differing from these of the principal clause. Thus, in German, the subordinate predicate is characterized by its position at the end of the sentence, as opposed to the subordinating predicate, which fills another position. In Latin, the formal feature of many subordinate clauses was the conjunctive mood, the tense and aspect of which depended on the tense of the subordinating clause (sequence of tenses). These facts facillitate our understanding of the function of the impersonal forms of the verb, i.e., the infinitive and participles. They often appear in the role of impersonal, dependent predicates, and thus, as equivalents of dependent clauses—e.g., *przyszedł zabijać* (he came to kill)—*przyszedł, aby zabijał* (he came in order to kill) (infinitive in the role of purpose clause); *ojciec kochający córkę* (a father loving his daughter)—*ojciec, który kocha córkę* (a father who loves his daughter) (participle n the role of relative clause).

Note: Semanteme—an element in a language form which refers to concrete phenomenon. Utterance—a language sign, comprehensible in communication, of necessity complex. What is sentence intonation and what are the different types of this intonation? What do the delimiting and modal functions of intonation involve? Modifying consists in the fact that modifying words (sentence-generating) make place adjacent for certain defined categories of other words. The set consisting of a modifying word and the places adjacent to it in the text, which are open for defined categories of modified words, is called the syntactic scheme. Words fulfilling the same function in the system of modifying and manifesting the appropriate semantic features of this function, are included in the same class, called a part of speech. How is the system of parts of speech constructed in the Polish language? What constitutes the difference between nominal groups and sentences? What constitutes the difference between the syntagmatic relationships of agreement and government? Concrete demonstratives (e.g., prepositions) indicate the spacial, temporal or causal relationships in respect to one another which the meaning contents of words appearing in a sentence have. List the categories of verbs and define their functions. What constitutes the difference between the grammatical and concrete (semantic) functions of noun cases? A compound sentence is a set of several clauses the mutual semantic relationships of which are expressed by means of linguistic agents. What does the functioning of anaphoric, interrogative and relative pronouns involve? What constitutes the difference between conjunctions and relative pronouns? What are the different types of conjunctions? Subordinate predication is characterized by the fact that it differs in certain formal features from subordinating predication.

Chapter 8. THE STYLISTIC SYSTEM

As we have previously mentioned, the speech of the adult is an orchestra of all the different types of signs. Heretofore, we have been discussing language signs only, i.e., arbitrary two-class semantic signals. The words which interact most closely with language signs are one-class arbitrary signals composed of language phonemes,

particles of the type *tak* (yes) (affirmative particle), *nie* (no) (negative particle), *czy* (interrogative particle), etc. Within a given consituation, particles always give the same information concerning the attitude of the sender in relation to reality. This becomes especially clear when they stand outside the sentence: *Czy wracasz do domu?—tak.* (Are you coming home?—yes.) *Czy byłeś w mieście?—nie.* (Were you in town?—no.) The question now arises as to the role played by the language system in the formation of the remaining signs of speech—symptoms, appeal signals and acoustical images. Three cases may be distinguished here.

The first case is that in which we make exclusive use of the continuous scale. In this case, the sounds produced by our voice have no conventional meaning. These segments of speech do not constitute part of language. Groans, humming a melody without words, animal calls—are all examples of this type of signs. Thus, the carriage driver urges his horse on by producing sounds which in linguistics are called clicks. In their articulation, the oral cavity is closed from two sides by labial and dorso-velar closures. Between these two points of closure, the air becomes rarified as a result of being drawn into the lungs. Suddenly, the lips are opened and air enters the oral cavity, producing the characteristic sound. In European languages, clicks do not constitute part of the diacritic scale. They are considered merely appeal signals for urging horses on, and other impressive functions. They are totally outside the realm of language.

The second case does not involve exclusively extra-linguistic segments isolated from the flow of speech (e.g., clicks). Within a given segment, the scale of diacritic sound features serves language signs, while the non-diacritic, continuous scale, simultaneously, serves symptoms, appeal signals and acoustical images. The same stream of sounds carries both types of signs. We hear diacritic features and, owing to them, we understand a sentence referring to a certain objective reality—which constitutes the function of semantic signals—while at the same time, the intonation of the voice in the continuous scale informs us about the emotional state of the sender—thus fulfilling the expressive function of symptoms—and influences our emotional state in fulfilling the impressive function of appeal signals.

The interaction of various types of signs proceeds further, however, leading to the creation of mixed, or compromise, forms, which constitutes the third possible symbiosis of the various types of signs being considered here. These mixed forms are composed of language phonemes operating in the diacritic scale, but which, because of their function, are extra-linguistic symptoms, appeal signals or acoustical images. It is this combination of language forms with extra-linguistic functions which gives the forms in question a mixed, compromise character. In respect to their form, they belong to the system of language, but in respect to their function, they do not. We consider these forms, therefore, as belonging to the periphery of language. The forms included in this category have two sources: elements of old, pre-language codes which have taken on the form of language; language signs which have come to fulfill on extra-linguistic function. These two categories of mixed forms, which

have a form composed of language phonemes, but which fulfill the extra-linguistic functions of symptoms, appeal signals and images, together constitute the stylistic system of language. We will discuss these forms in more detail, beginning with pre-language elements presented in language form.

Primitive sound symptoms constitute an entire group of mixed elements—exclamations fulfilling an expressive function, and which are represented in the form of conventional phonemes. These are so-called primary interjections like *ach*! *aj*! *ej*! *jej*! *oj*! *hu*!. Of course, a huge role is played by the expressive intonation of these interjections, which operates in the continuous scale; but in some of these interjections, the set of phonemes as well, fulfills the expressive function of symptoms. Thus, for example, the interjections *pfe*! *pfuj*! are expressions of contempt owing to the initial labial group *pf*, which is connected with a contemptuous inflation of the lips and a certain lip movement at the moment of expectoration—a gesture of contempt.

Appeal signals of the type *hop-hop*! or *halo*!, which we use to attract someone's attention in order to begin conversation, as well as commands of the type *hej-ho*! or *hu-hu*!, the repetition of which coordinates the movements of a team of workers, are both similar to primary interjections. These elements differ from commands addressed to a horse, involving some type of clicking, primarily in the fact that they are composed of normal language phonemes.

The next category involves onomatopoeia, i.e., acoustical images expressed in phonemes. The difference between the acoustical images produced by a child in the continuous non-diacritic scale and onomatopoeic words composed of phonemes is somewhat similar to the difference between realistic painting (e.g., a realistic painting of an eagle on top of a cliff) and heraldic painting (e.g., the coat-of-arms of Poland—a white eagle on a red field). The heraldic painting on coats-of-arms does not make use of the continuous scale of tones, as does realistic painting, but uses only certain, conventionally established, tones, i.e., pure tones of white, black, red, yellow, blue or green. In this form of painting, realistic tones are represented by the closest, though certainly not identical, conventional tones, with which parts of the heraldic shield are painted, in such a way that the borders between the colors are sharp, with no blending of colors. In such conditions, in order to identify the image on the coat-of-arms with the reality it symbolizes (e.g., a heraldic eagle with a real eagle), it is extremely useful to know the convention. The same may be said for onomatopoeic words. They are conventionalized images of real sounds represented in onomatopoeia by means of the closest, though certainly not identical, conventional phonemes, corresponding to the conventional tones of the coats-of-arms. The onomatopoeic word *kukulka* stands in the same relation to the real cry of the cuckoo as does the coat-of-arms of Poland to a real eagle. In both cases, identification is at least partly based on convention.

Onomatopoeic words, which make use of conventional phonemes standing for real sounds closely related to them, represent various acoustical phenomena, e.g.,

Polish *huczeć, hukać, buczeć, świstać, szumieć, wyć, kukać, klekotać*; German *heulen, klirren, knistern, pfeifen, paffen*; Latin *ululare, pipare*; English *hum, murmur, mumble, howl, miow*. It is clear that each conventional phoneme has its corresponding group of real sounds which the phoneme brings to mind and for which it may be substituted. The dental *s*, for example, may be substituted for various types of *świsty* (swishes), *gwizdy* (whistles), *syki* (whiz), and the phonological features of alveolar *sz* remind us of *szmery* (murmurs) and *szumy* (hums). It is not only the sound of phonemes which fulfills the onomatopoeic function, but also their arrangement and, especially, their repetition. In the word *kukułka*, for example, the group *ku* is repeated, which corresponds to the monotonously repeating cry of the cuckoo, while in the word *tik-tak*, the repetition is not complete, for a high *i* appears in the first syllable and a significantly lower *a* in the second, which corresponds to the impression made by the ticking of clocks, i.e., the rhythmic repetition of high and low sounds, The form *tik-tak* presents not only certain sounds, but also the rhythm or repeating, similar, though not identical, phenomena.

Phenomena, the rhythm of which is expressed in onomatopoeia may be not only audial, but also visual. In the latter case, we have to do with sound images of non-acoustical phenomena of the type *dyndać* (dangle), *bimbać* (to take it easy), in which the repetition of the consonant *d* or *b* produces the rhythm of a pendulum-like movement; or take, for example, the word *zygzak* (zigzag), in which the repetition of the dental consonant *z* and the dorsal *g* and *k* represents the course of a line which changes direction frequently, first to the right, then to the left. Here, the audial impression is associated with a visual one. A further step in this same process involves the transference of the onomatopoeia of an animal cry to the animal itself; thus, in Polish we have names for birds such as *kukułka* (cuckoo) and *czajka* (lapwing).

It should be noted that onomatopoeic words, in the broadest meaning, are significantly more frequent in many non-European languages than in European languages. They constitute a large percentage of the total vocabulary of the Austronesian languages of Indonesia and of the Austro-Asiatic languages of India and Indo-China, for example. A typical example is the word *dret-drot* (wavering, wobbling, vacillating) in the Khmer language of Cambodia. Here, the repetition of the same consonants represents the pendulum-like movement of something which vacillates. Similarly, the names of animals in all languages, but especially in the languages of primitive peoples, are to a significant degree onomatopoeic, like the Polish word *kukułka*.

Up to this point, we have been discussing mixed forms which originated outside of language and only secondarily took on certain features of language forms, thus becoming interjections (symptoms and appeal signals) or onomatopoeia (acoustical images). Next, we will proceed to mixed forms which are basically part of language, but which have secondarily taken on the extra-linguistic functions of symptoms and appeal signals. Besides their basic semantic function, these forms fulfill the secondary expressive function of representing the feelings of the sender (symptoms) as well

as the impressive function of influencing the feelings and behavior of the receiver (appeal signals). Usually, they are simultaneously symptoms and appeal signals, although typically, one of these functions predominates. The factors enabling language forms to acquire expressive and impressive functions, can be divided into categories.

A certain number of language forms owe their function as symptoms and appeal signals to their associations with extra-linguistic speech signs in the continuous scale, and especially, with the intonation which accompanies them. Intonation expresses the feelings of the sender and influences the affective state of the receiver, and the emotional content of intonation becomes associated with the form of words which always appear in this intonation. In this way, certain words originally possessing a semantic content, are transformed into expressive symptoms, and to a lesser extent, into impressive appeal signals. Such words are called secondary interjections of the type *Jezus Maria*! *Rety*! (Dear me!) *Gorze*! (Fire!) as well as closely related traditional swearing like *psiakrew*! (literally—dog's blood). These forms no longer possess a semantic content, but simply are associated in our minds with certain feelings expressed by the intonation which always accompanies such words.

Alongside secondary symptoms, secondary appeal signals appear in speech, which, besides a semantic function and secondary expressive function, have a basic impressive function. These secondary appeal signals—just as the previously discussed primary appeal signals—appear in two types of situations, namely, in opening conversations and in cooperation. If we wish to start a conversation with an acquaintance, we use the vocative form of the noun, e.g., *Janie*! *Tadziu*! *Panie Profesorze*! The vocative is, therefore, an impressive appeal signal the function of which involves calling the attention of the person named to the sender. Next, we have cooperation, in which one person directs the behavior of others by means of commands. The form of command traditionally included in the system of language is the imperative of the verb, e.g., *idź*! *rób*! *przynieś*! *połóż*! It is an appeal signal owing to its association with the impressive intonation which always accompanies it. The same is true in the case of other traditional forms of commands regardless of whether they are in the imperative, e.g., *padnij*! (stop by!) or not, e.g., *baczność*! (attention!).

Heretofore, we have been considering only such forms the impressive and expressive functions of which result from their being associated with a certain intonation which always accompanies them. Now we will proceed to forms which fulfill the same function owing to their being associated with a certain social consituation, in which they always appear. The system of language includes, alongside forms used by everybody, forms which are used only in certain social environments, which constitute their permanent consituation. This consituation, together with the emotional reaction which it evokes, is associated with a language form which appears always within the same social context, as a result of which, when such a form is used outside its environment, it evokes a vague memory of this environment as well as an emotional reaction associated with it. The word *belfer* for example, used by Polish pupils to refer to their teacher, evokes in us a memory of the school environment and brings

back feelings connected with it. In this way, a word connected with a certain environment becomes an expressive symptom of the emotional content with which the sender associates it and an impressive appeal signal which evokes a certain emotional content in the receiver.

When used within its proper environment, such a socially conditioned word does not evoke any affective reaction, for, in this case, it is a normal element of communication. This is what limits its use to a particular environment, for outside this environment, it may be used exceptionally, only for evoking affective reactions. A word of this type does not usually appear alone, but rather, in a group of many other words, which are used in the same environment and which evoke the same emotional reactions. Of course, every socially conditioned word corresponds in meaning to some word in common use, so that pairs of synonyms arise, differing only in affective tone. For example, the commonly used and emotionally colorless word *nauczyciel* (teacher) corresponds to the word *belfer*, used by pupils, which, when transposed from that environment to a different one, acquires specific affective overtones. Words of common usage and socially conditioned words constitute, in a sense, two subcodes of a single language, for the same semantic content may be expressed in one way or another within the same phonological and syntactic system. The commonly used words constitute the principal subcode and the socially conditioned words, the secondary subcode, which is called special language, or slang.

Social environments are hierarchically organized. In this situation, synonymous dublets—i.e., elements in a system of language possessing the same semantic or syntactic function and differing from one another only as regards their social usage—are also hierarchically organized and are evaluated higher or lower depending on how the environment in which they appear is evaluated. So long as the dublet is used in its proper environment, it does not fill an evaluating function; it takes on this function only when used in another environment. Such dublets remind the hearers, who know all the dublets of a language but simply do not use them in everyday speech, of the dublet's true environment. Because the evaluation of this environment and the feelings connected with it are associated with a given element of language, the use of this element evokes this evaluation or affective content, which becomes connected with either the person speaking or with the phenomena spoken about. If, therefore, a word from a less valued environment is used in one more highly valued, either the person speaking, the thing spoken about, or both are evaluated as negative and inferior; the situation is reversed, of course, when a word from a higher sphere is used in a lower one.

The words included in the slang of thieves evoke the most negative emotional reactions in us, e.g., *majcher*:*nóż* (knife), *kimać*:*spać* (sleep), *kumpel*:*towarzysz* (comrade), *melina*:*kryjówka* (hide-out). We listen to school slang with tolerance and pleasure, e.g., *buda*:*szkoła* (school), *belfer*:*nauczyciel* (teacher), *katabas*:*katecheta* (catechist), *pała*:*głowa* (head), *lufa*:*niedostatecznie* (failing grade). Hunting slang includes words which also have specific emotional overtones, e.g., *miś*:*niedźwiedź* (bear), *kot*:*zając*

(hare). Scientific language evokes a different kind of reaction in us, e.g., *erupcja*:*wybuch* (explosion), *intoksykacja*:*zakażenie* (infection), *insolacja*:*nasłonecznienie* (insolation). The same is true of administrative language, e.g., *etatowy*:*stały* (permanent), *apanaże*:*dochody* (income), *elaborat*:*wypracowanie* (elaboration—e.g., of a plan). Words from the latter two special languages evoke a feeling of importance and respect when transposed to another environment. Of all social subcodes, the most important is the language of literature and rhetoric, the forms of which evoke in us extremely positive affective reactions. It consists of a numerous set of words and phrases which are not used outside of poetry, theater and rhetoric, e.g., *rumak*:*koń* (horse), *gród*:*miasto* (town), *szata*:*suknia* (dress). Moreover, the syntax of literary or rhetorical language usually involves specific phonological and syntactic properties.

The third type of association to which words owe their function as symptoms and appeal signals, may be defined as semantic associations. In the minds of speakers, the borderline between the form and the meaning content of a semantic sign becomes obliterated. A word becomes for them the object itself and, consequently, the speakers are convinced that, in manipulating words, they are manipulating objects. This makes it possible to evaluate objects and to have a fictional control over them even, which we call verbal magic. In both cases, the word becomes the vehicle of action, i.e., an impressive appeal signal, although secondarily, it fulfills the expressive function of a symptom.

Because of the fact that the speaker subconsciously identifies the word with the object designated, he can, by manipulating the word, bestow certain positive or negative features on the object, thus raising it or degrading it in the hierarchy of things, which will then be connected with a certain emotional reaction, either positive (respect, good will) or negative (contempt, distaste). In this case, the word attributes certain characteristics to the object which it does not possess and thus, at least partially, the word loses its semantic function, becoming an instrument of evaluation—an appeal signal and a symptom. This phenomenon appears in the forms evaluating the first, second and third person.

The means of language evaluation in the first person is the so-called imperial plural, i.e., referring to oneself in the plural *we*, which in European languages is reserved for people of very high estate, above all, for monarchs. In this usage, the form *we* means the same as *I*, but connotes a notion of collectivity which gives the *I* an emotional overtone of importance and dignity.

The means of evaluation in the second person, the evaluation of the receiver in relation to the person speaking, define the norms of verbal courtesy in every society. In Polish, these norms decide what choice one makes among forms of the type *mówisz*:*mówicie*:*Pan mówi* (you say), depending on the social relationship of the persons speaking. Particularly elaborate forms of courtesy appear in the languages of the Far East—Japanese, Korean, Siamese, Javanese—as well as the Aztec language of Mexico. These forms evaluate the listener, defining his position in the social hierarchy and in relation to the person speaking, thus expressing as symptoms

various shades of respect and intimacy on the part of the speaker, and as appeal signals influence the emotional state of the receiver being evaluated by these forms.

The evaluation of phenomena in the third person, i.e., the evaluation of everything spoken about, is accomplished by metaphores which attribute certain fictional features to a given object, establishing their place in the hierarchy of things. If we say of someone *to orzeł* (he's an eagle), we evaluate him positively; if we say *to osioł* (he's a donkey), we evaluate him negatively. In both cases, we attribute certain features to the person which he does not possess, for he is neither an eagle nor a donkey, yet he is evaluated in terms of these fictional characteristics. We may evaluate by means of suffixes as well, which attribute certain fictional characteristics to an object. For example, in Polish, we have a category of words created by means of various suffixes indicating that the objects are smaller than the objects designated by the basic words. To the extent that this difference exists in fact, the phenomenon of evaluating does not occur, e.g., *stół*:*stołek* (table), *róg*:*rożek* (horn), When however, this difference does not exist in fact, the suffix becomes a means of evaluating. The words *słońce* and *słoneczko* (sun) are semantically identical, for there is only one sun. The suffix *-ko*, therefore, in the word *słonecz-ko*, does not designate a small sun, but simply appraises the sun as something small and nice. This is possible thanks to the fact that this suffix is associated in the minds of speakers with the feature of smallness, which is why it is capable of changing the sun into something having this property. It is in this way that we must understand the difference in function between the words—*wino*:*winko* (wine), *krowa*:*krowina* (cow), *gęba*:*gębusia* (face), *biały*:*białuśki* (white), *gruby*:*grubiuchny* (fat). In Polish, there also exists a category of words containing a suffix owing to which the words designate an object larger than the object designated by the basic word and, therefore, unpleasant. In cases where the difference in meaning content does not, in fact, exist, the suffix becomes merely an agent in negative evaluation, e.g., *dom*:*domisko* (house), *nos*:*nochal* (nose), *palec*:*paluch* (finger). Words which evaluate positively are called amelioratives; those which evaluate negatively are called pejoratives, e.g., *synuś*:*synal* (son).

It is the identification of the form and meaning content of words which lies at the basis of the type of language evaluation being discussed here, which leads to verbal magic. This is fictional action, the instrument of which is words. Because words are associated with phenomena, speakers can feel that words may be the instrument of direct action upon phenomena. A word or set of words which has this fictional power of action is called a spell or incantation. An example can be found in Part II of Mickiewicz's *Dziady* (*Forefathers*). There, the spell *a kysz, a kysz*! (begone!) is used, which is expected to force the spectres to disappear. The role of verbal magic in primitive societies, especially among the peoples of Oceania, is enormous. Words possessing magical power are taboo, i.e., their use is limited to those situations in which their action is desired. Sometimes absolute taboos obtain, which ban

the word from use. In place of the partially or totally taboo word, a word of the same meaning but devoid of magical power is used, i.e., a euphemism. Taboo embraces primarily the names of threatening creatures, which it is better not to summon. Thus, in the Melanesian languages the names of the dead and of anything once connected with them are taboo, so as not to call up ghosts. In the proto-Slavic language, the word designating the threatening bear (related to the Latin word *ursus*) became completely taboo and was substituted by a euphemism meaning 'honey-eater' (*miodoiad*), compare the Old Slavic *medv-ědъ* (*medъ*—'honey' and *jadętъ*—'eat') and Polish *niedźwiedź* (bear). This word, in turn, became taboo in hunting slang and was substituted by the word *miś* (teddybear).

In conclusion to these general considerations of the structure of language, the basic categories of elements comprising the semantic, syntactic and stylistic systems of every language will be listed. The categories of these elements closely correspond to what traditional linguistics defines as parts of speech. Because not all of the languages of the world are comprised of words, this general classification cannot take words into consideration, but only morphemes, which essentially function in all languages. Semantic morphemes, which refer to external reality, are called semantemes. Proceeding from these principles, we may distinguish six categories of elements in all the languages of the world:

(1) Denotative semantemes—(in Polish, the stems of nouns, adjectives, verbs and adverbs, e.g., *białość* (whiteness), *biały* (white), *bieleje* (whitens), *biało* (whitely).

(2) Referential semantemes (in Polish, the stems of pronouns, e.g., *ja* (I), *ty* (you), *on* (he)).

(3) Ordering semantemes (in Polish, stems of numerals, e.g., *dwa* (two), *trzy* three)).

(4) Syntactic morphemes (in Polish, prepositions, e.g., *w* (in), *przy* (at), *na* (on) and conjunctions, e.g., *i* (and), *a* (but), *że* (that), as well as inflectional endings.

(5) Particles (in Polish, one-class semantic signals, e.g., *tak* (yes), *nie* (no)).

(6) Interjections (in Polish, symptoms and appeal signals the forms of which consist of phonemes, e.g., *ach*! *oj*! *pfuj*! *halo*! *hop-hop*!).

Note: Particles are one-class signals composed of language phonemes. Acoustical symptoms, appeal signals and images composed of language phonemes constitute the stylistic system of language. What are the principles of functioning in the stylistic system of language of the following: primary and secondary interjections, onomatopoeia, the vocative and imperative, socially conditioned words, the imperial plural, forms of verbal courtesy, amelioratives and pejoratives, spells and incantations, and euphemisms.

PART III

HISTORICAL LINGUISTICS

CHAPTER 9. PRINCIPLES OF HISTORICAL LANGUAGE CLASSIFICATION

We have been discussing the methods of descriptive linguistics, which analyses the structure of each language independently, as though it were the only language in the world. Every science begins with description, but description alone is not enough, for the task of science is to demonstrate certain general propositions. In investigating processes, sciences establish general laws according to which these processes are carried out; thus, in analysing phenomena, sciences develop systems for classifying them. Linguistics is one of these latter types of sciences, for the languages which constitute the subject of linguistics exist permanently in time. The primary task of comparative linguistics, therefore, is to develop a system for classifying languages. Such a classification, however, may be of two kinds: historical—on the basis of common genesis, and typological—on the basis of presently existing common features. Consequently, comparative linguistics, which is concerned with the classification of languages, is divided into historical-comparative linguistics and typological-comparative linguistics. Historical linguistics, which constitutes our present concern, establishes an historical classification of languages according to the degree to which they have a common genesis, and then analyses the processes of language evolution which led to this division of languages, attempting to present, at least in part, the causes of particular evolutionary processes.

Everything in the world undergoes change; nature and people change, nor does language avoid the functioning of this general law. In the course of centuries, the system of language in a particular society undergoes changes which can be traced in texts which have been preserved. The totality of the enormous changes which the Polish language has undergone in the course of the last six centuries is illustrated by comparing the text *Kazania świętokrzyskie* (*Holy Cross Sermons*)—which reflects the state of the Polish language in the middle of the 14th century—with its translation into present-day Polish. Below is an excerpt from *Kazanie na dzień Bożego Narodzenia* (*Sermon for Christmas Day*), in which the archaic features of the language (transcribed into present-day writing) have been preserved; next to it, we have the same excerpt in present-day Polish:

Toć to i jeść prawda, iże idzie tobie krol zbawiciel, iżby nas ot wieczne śmirci zbawił. A trzecie idzie tobie ubogi, iżby ty w ubostwie nie styskował. Jakoż prorok Dawid, uznamionaw o jego silnem ubostwie, jeść świadeczstwo dał, rzeka: ...

To też i jest prawda, że idzie ku tobie król Zbawiciel, ażeby nas od wiecznej śmierci zbawił. A po trzecie idzie ku tobie ubogi, ażebyś ty w ubóstwie nie utyskiwał. Tak więc prorok Dawid, dowiedziawszy się o jego wielkim ubóstwie, dał świadectwo, mówiąc: ...

(And it is also true that the king Savior is coming to you to save us from eternal death. And thirdly, a poor man is coming to you, lest you complain of your poverty. Thus, the prophet David, recognizing his great poverty, gave witness, saying: ...)

Comparing the texts of these same sentences, in Old Polish and in present-day Polish, we note that all four of the component systems of the language, i.e., the phonological, semantic, syntactic and stylistic systems, have undergone changes in the course of these six centuries. The phonemes have undergone changes either in all the words in which they appear, or in all the words in which they appear in a particular phonemic environment. These are phonological changes based on the fact that certain diacritic features of phonemes are replaced by others. We have, for example, in *Kazania świętokrzyskie*, *krōl* (king), *ubōstwie* (poverty), both with a long *ō*, while today, these words are written *król*, *ubóstwie*, pronouncing them *krul*, *ubustwie*. This took place when the Old Polish long *ō*, written *o* or *oo*, changed, in the period when Old Polish duration disappeared in the second half of the 15th century, into closed *ȯ*, later written *ó*. In the 18th century, closed *ȯ* changed into *u*, traditionally written *ó*. In this case, the quantitative diacritic feature of duration was substituted by the qualitative feature of closed articulation. This change transformed *ō* into *ó* in all words, independent of position; in comparing the texts in question, however, other changes may be found which appear only in certain positions. Thus, the author of *Kazania świętokrzyskie* said *śmirci* (death), while today we say *śmierci*, as a result of the substitution of *e* for *i* in the position before *r*, when it belongs to the same syllable. He also said *świadeczstwo* (witness), which later, as a result of the assimilation of *cz* by the following *s*, produced *świadecstwo* and, finally, *świadectwo*; similarly, *ot wieczne* (from eternal) became *od wiecznej* as a result of the transformation of *t* into voiced *d* through assimilation by the voiced onset of the following word.

Changes in the semantic system constitute the second group, among which lexical changes should be mentioned first. Certain words have completely disappeared from use—*uznamionać* (recognize), *styskować* (complain)—others have changed their lexical value—e.g., *sklep* once meant 'vaulting', but from the end of the 17th century, it has meant 'a place for selling goods'. Words are composed of morphemes. Changes in the form of these morphemes which are independent of phonological changes, are defined as morphological changes. In *Kazania świętokrzyskie* we have forms of the

genitive of feminine adjectives ending in *-e—ot wieczne* (from eternal), the past participle ends in *-w—uznamionaw* (recognizing)—the past tense is expressed in a compound form—*jeść dal* (gave)—and the form of the present participle, active voice, ends in *-a—rzeka* (saying)—while today, the genitive form of feminine adjectives ends in *-ej* (*od wiecznej*), the past participle ends in *-wszy* (*dowiedziawszy*), the past tense is a non-compound form (*dal*) or a contraction (*dałem* from *dal jeśm*), and the present participle ends in *-ąc, -ący* (*mówiąc, mówiący*).

The third group of differences between present-day Polish and the language of *Kazania świętokrzyskie* involves syntactic changes. Thus, in *Kazania świętokrzyskie* we have *idzie tobie* (is coming to you), while today we say *idzie ku tobie, idzie dla ciebie.* As we see, the traditional function of the dative, that of indicating the goal toward which or for which something happens, has weakened, so that today, the form of this case must be reinforced by a preposition.

The fourth group of language changes consists of changes in stylistic habits. In Old Polish, for example, the custom prevailed of referring to the partner in conversation in the second person—*ty* (you), while today the third person form of adress—*pan* (mister)—prevails.

The phonological, lexical, morphological, syntactic and stylistic changes which accumulate in the course of centuries finally change a language to such an extent that, structurally, it becomes completely different from its historical predecessor. Who, for example, would recognize the Latin nouns (in the accusative) *patrem* (family head) and *matrem* (mother) in the present-day French words *mère* and *père*? Not until an historical study is carried out is it possible to confirm the fact that both of these words, as does the entire language system of French, derive from the transformation of original Latin forms.

This metamorphosis of particular elements of a language is an extremely gradual process. Only a very small number of language elements (phonemes, morphemes or entire words) undergo change at a given time, and these elements take the place of older elements imperceptibly, without hindering communication among people. Because, therefore, language changes occur in such tiny doses that they play no role in people's lives, people are seldom aware of them and, therefore, assume that language does not change.

We consider as "ancestor" and "descendants" those language systems which gradually replace one another through a process of small changes, without upsetting the conviction of society that it continues to speak the same language. We consider such language systems different evolutionary stages of the same language. Thus, the language system of Old Polish in the period of *Kazania świętokrzyskie* and that of Modern Polish in the 20th century constitute different evolutionary stages of Polish. Within such an historically defined individual language, the process of evolution gradually unfolds.

The fact that languages are ceaselessly changing leads to the question as to whether or not evolution, in transforming individual languages, simultaneously changes their

relationship to one another. Certain languages are more closely related and others more distantly related to one another. For a Pole, the most closely related languages are those which, like Czech or Russian, he can partially understand, even without studying them. Languages such as Lithuanian and Latvian, which cannot be understood by a Pole without special study, but the structures of which are similar to that of Polish, are more distantly related to Polish. Languages like Hungarian or Turkish are completely different from Polish and, therefore, afford significantly greater difficulty to a Pole in learning them.

The question arises as to whether the relationship presently existing among languages has always existed, or whether, considering the course of language evolution, these relationships were not in the past different from those which presently obtain. In fact, historical investigations have proven that the differences between such closely related languages as Polish and Czech were once much smaller than they are today. When we compare the Polish nouns and proper names written in Polish oldest existing text, *Bulla gnieźnieńska* (*The Gniezno Bull*) 1136, with contemporary Czech words, the differences appear insignificant. The further back we go in time, the greater the similarity between Polish and Czech. This phenomenon may be traced, on the basis of existing texts, all the way back to the 12th and 11th centuries A.D. Because of the lack of older texts, we cannot carry investigation further into the past. If, however, we imagine the courses of evolution of Polish and Czech as two oblique lines which approach each other and which, in the period represented by the oldest existing text, are already very close, we are led to suppose tthat in prehistorical times these lines approached each other even more closely unti finally, they merged. In other words, the fact that—the further we go into the past, the greater the similarity between such closely related languages as Polish and Czech—leads to the assumption that these languages arose as a result of different courses of evolution occurring in different regions occupied by the 'ancestor', or proto-language, common to both of these languages.

The same factors as those obtaining in the relation between Polish and Czech force us to postulate the existence of a proto-language common to all Slavic languages—i.e., Polish, Lower Lusatian, Upper Lusatian, Czech, Slovak, Slovene, Serbo-Croatian, Macedonian, Bulgarian, Ukrainian, Russian and White Russian. The proto-language common to all of these languages and from which they all developed is called proto-Slavic.

Proto-Slavic will always be a postulation, an extremely probable scientific hypothesis, for no texts written in this language have been found. The oldest preserved Slavic texts, i.e., copies of the translation of the Holy Scripture, the work of Constantin-Cyril and Methodius, c. 863, although linguistically very close to proto-Slavic, have already taken on a definite Macedonian-Bulgarian tone and cannot be considered a survival of proto-Slavic.

Linguistics is not always at such a disadvantage, however, for sometimes the proto-language from which various languages derived can be directly studied in

the form of preserved texts. Such a fortunate situation exists in the case of the Romance languages, the proto-language of which is Latin, strictly speaking, vulgar Latin, known from inscriptions and texts in the vulgate. The Romance languages—which include Rumanian, the now extinct Dalmatian dialects on the Adriatic, Italian, Sardinian, the Raeto-Romanic dialects in the Alps and on the Isonzo, French, Provençal, Catalonian, from the vicinity of Barcelona, Spanish and Portuguese—clearly differ from one another today. The further back we go into the past, however, the fewer these differences, so that, as we approach the first centuries A.D. these differences disappear almost completely and all of the Romance languages merge into vulgar Latin.

It is true that the vulgar Latin of particular provinces of the Roman Empire differed somewhat from one another, but these were minor, dialectical differences which did not disrupt the over-all unity of the language. Not until the period between the 6th and 10th centuries A.D. did these differences undergo a significant increase, which lead to the division of Latin into the series of Romance languages. Because this entire process can be traced in written texts, it is of particular significance to linguistics.

Thus we see that one language system may have a greater number of "descendants" as a result of the fact that it evolved differently in the different areas in which it was used. Such a group of languages arising as a result of the uninterrupted evolution, in different regions, of a single, common proto-language is called a language family. The languages constituting one family, those deriving from a common proto-language, are called related languages. The history of a language family may be presented in the form of a genealogical tree, analogous to that of a human family. Here the proto-language fills the place of the common ancestor, from which the particular languages differentiated. Below is the diagram of the genealogical tree of the Romance family:

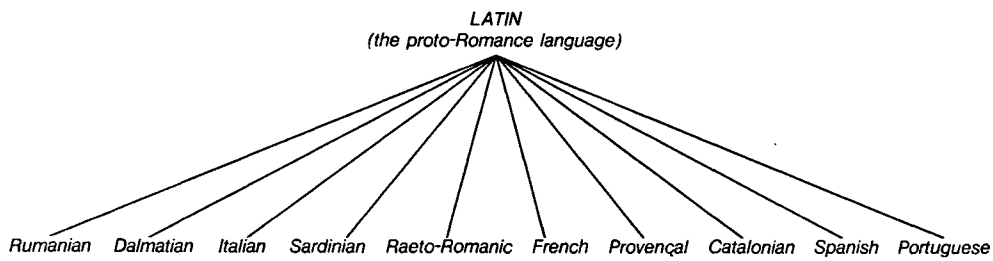


Fig. 11. Genealogical tree of the Romance family

Different degrees of relationship may arise among the languages of a single family, depending on the length of the period of common development. The longer this period, i.e., the latter the differentiation from one another, the closer their relationship. In this way, various groups of languages have become differentiated within a single family, groups which continued to constitute a relative unity at a time when

significant differences had already arisen among the groups themselves. Thus, single-degree and multi-degree sets can be distinguished among language families. The segmentation of a single-degree family, i.e., the Romance family, is represented by Fig. 11, which distinguishes only two periods—earlier and later. The proto-language existing in the earlier period, in this case Latin, has split in the course of centuries into a series of dialects having a parallel evolution, dialects which we may observe in the later period, that in which we live. In analysing multi-degree families, however, at least three periods must be distinguished. Languages of the latest period, usually that in which we live, derive from first-degree proto-languages, which existed lin the intermediary period, and which in turn derived from a single, common proto anguage of the earliest period. The Slavic language family exemplifies a two-degree-family; its genealogical tree can be represented as follows:

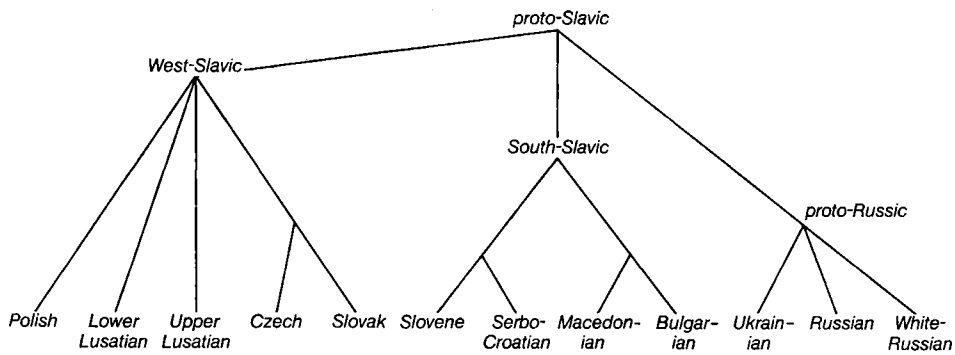


Fig. 12. Genealogical tree of the Slavic language family

The proto-Slavic language first split into three language groups—the proto-Russic or East-Slavic group, the South-Slavic group and the West-Slavic group. Only later did three Russic languages branch off from proto-Russic to form White Russian, Russian and Ukrainian; the South-Slavic group split up into Bulgarian-Macedonian, Serbo-Croatian and Slovene; the West-Slavic group became differentiated into the Czecho-Slovakian branch of languages, the Lusatian languages (including two distinct languages—Lower-Lusatian and Upper-Lusatian) as well as Polish (including the Kashubian dialects). It follows, that the Slavic languages are not all related to one another to the same degree. Ukrainian, for example, is more closely related to Russian than to Polish because, together with Russian, it derived from the proto-Russic group, which constituted a relative unity up to the 12th century A.D., while the ancestors of Polish and Ukrainian lost linguistic contact with each other and constituted part of separate language groups somewhere around the 6th century A.D.

The relationship among languages comprising a single family results from the process of differentiation of an original linguistic unity into a series of distinct systems. The evolution of languages, however, does not always lead to their divergence, to the differentiation of a language zone. Sometimes, the very opposite occurs,

the result of language evolution is such that different languages gradually and unobtrusively become more similar to one another, thus integrating a language zone. While the divergence of languages is a consequence of the loss of geographical contact among them, convergence of languages results from the establishment of such contact, from the mutual influence of these languages. A group of neighboring languages which, owing to mutual influences, become more similar to one another in phonological, morphological or syntactic structure, is called a league of languages. Languages belonging to the same league are affiliated languages. Linguists have claimed, for example, the existence of a Balkan league of languages composed of Modern Greek, Albanian, Bulgarian and Rumanian. The first and second of these languages constitute distinct language categories, Bulgarian belongs to the Slavic family and Rumanian to the Romance family—all four of these languages, therefore, were very different from one another in structure between the 6th and 10th centuries A.D., when they came into geographical contact. In the course, however, of the last several centuries, owing to the mutual influences of their syntactic and inflectional systems, a series of innovations occurred making these languages more similar to one another. This process of convergence may be diagrammatically represented as follows:

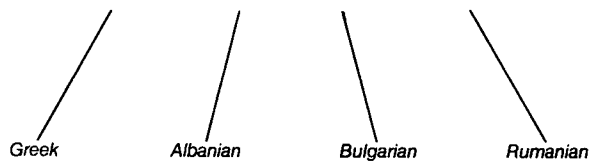


Fig. 13. The Balkan league of languages

The processes of language divergence and convergence may follow one another alternately, as a result of which languages become more or less similar to one another in various combinations.

The claim that a language family arises as the result of differentiation of a zone, while a league results from its integration, takes us from the field of language history to that of language geography. From the point of view of this latter discipline, the evolution of language on the territory of a given country is the concern of wave theory. Wave theory claims that every language change arises in a particular defined locality and then spreads in all directions over the territory inhabited by a given society. In dispersing, these changes extend into the territory of ever wider circles, similar to the circles of waves on the surface of a body of water radiating out from the place into which a stone has been thrown. It is this comparison which determined the name of the theory. Wave theory became the take-off point for the development of linguistic geography, or dialectology, which studies linguistic differentiation of territories as the result of language evolution. This discipline establishes the precise location of particular linguistic features over a given territory.

The boundaries of these features, so-called isoglosses—to the extent that we

are not dealing with a recent resettlement of people—rarely overlap with one another. Normally, they intersect a country in various directions. Dialectology distinguishes three types of such isoglosses—isophones, isomorphs and isolexes.

The first of these are boundaries dividing regions in which the same phoneme developed historically in different ways. For example, the old consonants *š, ž, č, ǯ* (written *sz, ż, cz, dż*) persisted without change in certain regions of Poland (e.g., Great Poland), for example *szyć, żaba, czas, jeżdżę*, while in other regions, they changed into *s, z, c, ʒ* (dz), for example *syć, zaba, cas, jezdzę*, which we call Masovian pronunciation. Isomorphs are boundaries delimiting regions in which different morphemes are used to fulfill a given function—i.e., different endings, suffixes or prefixes. For example, on one side of an isomorph, Polish country people say *chodźmy* (let's go), and on the other side, *chodźwa*, on one side, *robiłbym* (I would do), and on the other, *robiłbych* in the same meaning. Isolexes delimit regions in which completely different words are used to refer to the same thing—e.g., *klepisko* and *bojowisko* (both meaning 'threshing floor') or *kokot* and *kurak* (both meaning 'rooster'). After establishing linguistic boundaries, dialectology proceeds to establish their mutual relationships. In comparing the courses of these boundaries with one another, we observe that in certain parts of the country, these boundaries are more dense than in other parts, sometimes forming wide, intricate bundles, and at other times narrow, sharply defined bundles. These bundles of isoglosses constitute delimiting belts separating from one another dialects which are prevalent in relatively uniform linguistic zones.

The time dimension of language evolution is the concern of the theory of the genealogical tree, while the space dimension of evolution within a given country is the concern of wave theory, on the basis of which dialectology establishes the extent of various language changes, which are presented as isophones, isomorphs and isolexes. Both of these concepts, although valuable, are too one-sided and, therefore, the most valuable form of historical linguistics is embodied in their synthesis—historical dialectology. Starting with an analysis of the course of isoglosses representing the present-day state of a language, and then going back in time, historical dialectology describes the gradual differentiation and integration of language zones. Proceeding thus, it determines in the most effective way the degrees to which particular languages and dialects are related and affiliated by comprehensively classifying them.

Note: What is the task of historical linguistics? What is involved in phonological, lexical, morphological, syntactic and stylistic changes in language? What is a language family and what is considered its proto-language? What is the difference between a language family and a language league? What is the difference in the perspective on linguistic history provided by the theory of the genealogical tree and wave theory? What problems does linguistic geography deal with? What is the difference between isophones, isomorphs and isolexes? What is historical dialectology?

CHAPTER 10. RESEARCH METHODS OF HISTORICAL LINGUISTICS

Historical linguistics makes use of three methods for reconstructing the evolution of languages—the philological method, the method of internal reconstruction and the comparative method.

The philological method involves comparing texts written in the same language but originating in different periods. Thus, in comparing the elements of a language which fulfill the same function in progressively later texts, we can trace their gradual transformation. A comparison, for example, of the declension of nouns in Polish texts from the 14th century (e.g., *Kazania świętokrzyskie—Holy Cross Sermons*), and from the 15th century (e.g., *Biblia Królowej Zofii—Queen Sophia's Bible*), etc., has enabled us to establish the evolution of Polish declensional forms in the course of the last six centuries.

The method of internal reconstruction (part of which involves determining relative chronology) is based on the existence of variant elements in the language of a particular period. By analysing the language system, this method determines which of these elements is older and which younger. This method appears in several forms. In the first form, conclusions are drawn on the basis of phonological variants. Thus, in Polish, at the end of the stem we have the alternation of voiced and unvoiced consonants: *p*:*b*, *t*:*d*, *k*:*g*, *s*:*z*, etc.—e.g., *śnek* (written *śnieg*—'snow') alongside *śneg-u*, *gus* (written *guz*—'lump') alongside *guz-a*. The fact that the voiced form *b*, *d*, *g*, *z*... occurs in front of all endings beginning with a vowel, and the unvoiced form *p*, *t*, *k*, *s*... occurs at the end of words only, leads us to conclude that the voiced form appeared earlier in all forms of the stem, later becoming unvoiced in the coda. Thus, a change took place at the end of the word: $b \geqslant p$, $d \geqslant t$, $g \geqslant k$, $z \geqslant s$, etc... i.e., *śneg* $\geqslant$ *śnek*, *guz* $\geqslant$ *gus*... (In linguistics the symbol $\geqslant$ means "changed into" and the symbol $\leqslant$ means "originated from".) Traditional spelling substantiates this conclusion.

For historical morphology, which investigates the evolution of the morphological system of language, the method of exceptional forms is important. This method functions on the principle that, if we have two forms with the same meaning, one regular, conforming to the normal general morphological type of the language, while the other form is abnormal, exceptional, this latter form must be considered a relic and, therefore, the older form. Thus, in present-day Polish we have the alternate forms *w niebiosach*:*w niebiesiech* (in the heavens), *w Prusach*:*w Prusiech* (in Prussia). The first of these forms, with the ending *-ach* (borrowed from the declension of feminine nouns), are normal and later; the latter, with the ending *-ech*, rare today, are abnormal and earlier.

A related method is that which deals with forms that are dying out. This method maintains that, if there are two synonyms, one of which is in the process of extinction while the other is developing, then the form which is dying out, is older. In English, for example, two forms appear in the function of the second person singular—

thou and *you*. The first form is older and is becoming extinct; the second form, in the singular at least, is new and vital.

We will now consider the historical-comparative method. It can be considered an extension of the method of internal reconstruction. While the latter method determines which of two variant elements of the same language is older, the historical-comparative method aims at determining which of the elements of different, related languages is oldest. Before this method can be applied, it must be determined which languages are related to one another and which elements of these languages correspond to one another historically, i.e., derive from the same proto-language. These questions will be discussed in turn.

In order for a certain group of languages to be considered a family, it must be demonstrated that a constant correspondence exists among the phonemes of the words used in these languages, a correspondence which is consistent with the so-called phonetic law. In comparing Polish and Czech words, we see that Polish *ć* (written *ć* or *ci*) always corresponds to Czech *t* or *t'*; compare Polish *cichy, kochać, ściele*, with Czech *tichý, kochati, stelivo*. Polish *g* corresponds to Czech *h* in cognates; compare Polish *Praga, gawiedź, ginąć*, with Czech *Praha, havěd, hynouti*. The Polish group *ro* between consonants corresponds to Czech *ra* in those words in which the Russian form uses *oro*—compare Polish *krowa, droga, chronić*; Czech *kráva, dráha, chraniti*; Russian *koróva, doróga, xoronit'*, etc. Such constant phoneme correspondence cannot be merely a matter of coincidence. They can be understood only in the light of the assumption that the Polish and Czech words arose as a consequence of the transformation of proto-Slavic "word-ancestors" common to both languages—a transformation so constant that in certain defined positions a given phoneme developed in the same way in all words. Studies of old texts in Polish, Czech and other Slavic languages, totally confirm this assumption. Thus in *Bulla gnieźnieńska* (*The Gniezno Bull*) 1136, the later Polish *ć, dź*, are represented by the letters *t, d*,—e.g., *Techuta—Ciechuta, Radenta—Radzięta*. This proves that, in this case, the Czech language has preserved an earlier state and that in proto-Slavic, the consonants *d* and *t* occurred, pronounced softly before frontal vowels, in the words in question. This state was preserved without any basic changes in Czech, whereas in Polish, the soft *t'* and *d'* before the frontal vowels *i, e, ę*, changed into *ć, dź* during the second half of the 12th century. Historical studies have shown, however, that the Czech *h* derived from the older *g*, that in this case, it is the Polish language which has preserved the older form. As far as the Polish *ro* and the Czech *ra* between consonants are concerned, neither Czech nor Polish have faithfully preserved the original form. The Kashubian form *gard* and the Russian *górod* correspond to the Polish *gród, grodu*, and the Czech *hrad*. The Kashubian and Russian forms prove that, originally, in the combination in question, the vowel stood before the *r*, not after it, which is confirmed by the corresponding Lithuanian forms *gar̃das* and *gardìs*. Both in Polish and in Czech, therefore, the original form was transformed, in a different way in each language, but in both cases the change was consistently uniform in all words.

Just as between Polish and Czech, the strict correspondence of phonemes among all the Slavic languages can be determined; they can also be determined among all the Romance languages. This latter fact is of particular significance to us. Historical comparative studies supported by the philological method prove beyond a doubt that certain Latin phonemes recorded in written texts developed differently in different Romance languages, but in a specific, consistent way in each of these languages. In a given position, a particular Latin phoneme underwent the same development in all of the words of a given Romance dialect. It is because of this consistency in the development of particular Romance dialects and languages that the phonemes of these languages, originating from those of vulgar Latin, although different in each language, closely correspond to one another. For this reason, phonetic laws are, on the one hand, statements of the correspondence of phonemes in a given position, in two languages (e.g., "Polish *ć* corresponds to Czech *t* or *ť* in the position before original frontal vowels"), and, on the other hand, statements of the consistent evolution in a given language of phonemes in a certain position (e.g., "in the second half of the 12th century in Poland, soft *ť* changed into *ć* before original frontal vowels").

Phonetic laws are fundamentally different from natural laws, for the latter function always and everywhere, while phonetic laws function for only a certain period of time and within a human society inhabiting a certain territory. Within the limits of this time and space, a phonetic law functions without exception, i.e., it determines a given change in a given position in all words and for all members of a given society; after the law has expired, however, new forms may become incorporated within a language, forms constituting apparent exceptions to a given phonetic law because of the fact that they arose after the law had expired. These apparent exceptions may be either borrowed from foreign languages or may result from the productivity of the language system.

Because phonetic laws function only in particular territories, after such a law has expired, a language may borrow from neighboring languages, uninfluenced by this law, words the phonetic forms of which are inconsistent with the given law. It is on the basis of these inconsistencies with a law's functioning in a given language that we recognize borrowing. Thus, the Polish words *hańba* and *obywatel* are considered borrowings, for when compared with the corresponding Czech words *hanba* and *obyvatel*, we note that, in this case, the normal correspondence between Czech *h*, *t* and Polish *g*, *ć* before *e* is lacking and what appears instead is the completely atypical equivalence of Czech *h*, *t* and Polish *h*, *t*. In fact, studies of the Polish of the Middle Ages have shown that the Polish words in question have indeed been borrowed from Czech, whereas the indigenous Polish forms were pronounced *gańba* and *obywaciel* and demonstrated the normal correspondence between Polish *g*, *ć* and Czech *h*, *t*. The word *hańba*, however, was borrowed from Czech after the Czech *g* had already changed into *h*, while the word *obywatel* was borrowed after the Polish *ť* before original frontal vowels had already changed into *ć*. These words, therefore,

are only apparent exceptions to the Polish phonetic law according to which *g* remains unchanged and *t́* changes into *ć*.

The second source of apparent exceptions is the productivity of a language's morphological system. This phenomenon is normal and constant in language (cf. Chapters 4 and 6). Sometimes, however, associations on the basis of new proportions arise, resulting in new morphological forms—e.g., originally, the genitive singular of the noun *syn* was formed according to the proportion *miód*:*miodu* = *syn*:*synu*. This proportion, however, changed, and a new form of this case arose, *syn-a*, according to the proportion *wilk*:*wilka* = *syn*:*syna*. The latter is, of course, a new form.

Innovations of this type may lead to forms which are inconsistent with expired phonetic laws which once functioned in a given language. For example, in the prehistoric period of the proto-Slavic language "jať" *ě* (a vowel of the *e* type, but long and broad) in Poland, changed before hard, apical consonants *t*, *d*, *s*, *z*, *n*, *l*, *r*, into the vowel *'a* when the preceding consonant was soft, while in other positions, *'e* was introduced, e.g., *las*:*w lesie*. In conformity with this law, the accepted forms in the 15th and 16th centuries were *krzasło*:*w krześle*; the vowel change *'a*:*'e*, however, was no longer vital in that period and was replaced by new inflectional forms. Thus, in accordance with the proportion *w niebie*:*niebo* = *w krześle*:*krzesło*, the form *krzesło* arose in the period in which the law in question had expired. This form, after a certain period of coexistence with the old form *krzasło*, superseded it, creating an apparent exception to the relevant phonetic law.

The concept of phonetic law constitutes the foundation of the comparative-historical method, for on the basis of these laws, we determine the morphemes which in various languages correspond to one another. Only these forms, or their individual elements, i.e., morphemes, appearing in the languages compared, are considered historically corresponding, i.e., originating from common proto-forms which—possessing a similar semantic function—have a phonological form consistent with the phonetic laws of a given language. For example, the Polish form *krowa*, the Russian *koróva* and the Czech *kráva* are considered as corresponding to one another, for they are identical in meaning and their phonological form is consistent with phonetic laws, however, the Polish words cited above—*hańba* and *obywatel*—do not correspond historically to the Czech *hanba*, *obyvatel*, although they are borrowings for they are not consistent with Polish phonetic laws.

In applying the methods presented here, a distinction is made among elements of different origin in one language, and we come to the conclusion that all languages are constantly intermingling with one another. For this reason, we cannot say of any language of a later period that it is an absolute continuation of some proto-language of an earlier period, for each language is a combination of elements deriving from various proto-languages. In the Polish language, for example, alongside elements deriving from proto-Slavic, we have words representing continuations of proto-German, Latin, Persian and Turkish forms. Of course, the percentages of

elements of various origin are different in different languages. In Polish, for example, the great majority of elements derive from proto-Slavic, but in English, the percentages of elements of Germanic and Romance origin are more or less the same. In such conditions, it is necessary to set up certain tests for distinguishing among particular families of languages. Because there exist no languages of completely uniform origin, the problem rests on the fact that we must find in a series of linguistic mixtures—beginning with languages with an overwhelming predominance of forms derived from proto-language *A*, through an intermediary stage, to languages predominating in elements derived from proto-language *B*—some delimiting point enabling us to distinguish the *A* language family from the *B* language family.

The correspondence of morphemes comprising the morphological systems of the languages compared constitutes such a criterion for distinguishing language families. Here, as previously defined, we take correspondence to mean the similarity of semantic function and conformity to phonetic laws, a conformity substantiating the origin of a group of languages in common proto-forms. The comparison of the paradigms of Polish, Russian and Old Slavic illustrates what is meant by the correspondence of morphemes and their common origin:

	Polish	Russian	Old Slavic	proto-Slavic
Nominative	*ryb-a* (fish)	*rýb-a*	*ryb-a*	**ryb-a*
Genitive	*ryb-y*	*rýb-y*	*ryb-y*	**ryb-y*
Dative	*rybi-e*	*rýb-e*	*ryb-ě*	**ryb-ě*
Accusative	*ryb-ę*	*rýb-u*	*ryb-ǫ*	**ryb-ǫ*
Instrumental	*ryb-ą*	*rýb-oju*	*ryb-ojǫ*	**ryb-ojǫ*
Locative	(*w*) *rybi-e*	*rýb-e*	*ryb-ě*	**ryb-ě*

All of the endings of these inflections, very similar in semantic function, have phonological forms which conform to the phonetic laws of these particular languages, and, therefore, they must be considered as corresponding descendants of proto-Slavic endings. This fact is of great significance, for while vocabulary constitutes a completely open system, absorbing new elements from all sides, the morphological system and syntactic morphemes, to a certain degree at least, constitute a closed entity, which accepts new forms with difficulty. While there are no languages the entire vocabulary of which derived from a single proto-language, languages in which all of the elements of the morphological system, including syntactic morphemes, derived from a single language are common. This fundamental uniformity and permanence of the morphological systems and syntactic morphemes of languages are such that they constitute a reliable basis for classifying languages. By a language family we mean a group of languages in which the elements of the morphological system (morphological alternations, affixes, pronouns, numerals) and syntactic morphemes (prepositions, conjunctions and particles) all, or in a significant majority of cases, correspond to one another, having derived from a common proto-language. The origin of vocabulary does

not have such a determining significance. Only the composition of the basic stock of root words—e.g., *wilk* (wolf), *dom* (house), *jest* (is), *wie* (knows)—is taken into consideration. Linguistics, having applied these principles for centuries, includes English in the Germanic family of languages, for, although the vocabulary of English is composed of approximately fifty percent Germanic and fifty percent Romance elements, the entire English morphological system, including syntactic morphemes, is of exclusively Germanic origin, and all of the most basic words as well derive from this source. We proceed in this same way in other doubtful cases.

Having determined the set of languages belonging to a single family, having defined the family's extent, we proceed to study the components of such a family by working out a classification of the languages and dialects belonging to it. Thus, we enter the field of linguistic geography, which utilizes the method of linguistic atlases. These atlases present the linguistic differentiation of the territory studied on hundreds of maps. Such atlases already exist for a series of Romance countries—France (the atlas of J. Gilliéron), Catalonia, Italy and Rumania. The first attempt in this direction on Polish territory was the *Linguistic Atlas of the Polish Carpathian Lowlands* (*Atlas językowy polskiego Podkarpacia*), Kraków, 1934, compiled by M. Małecki and K. Nitsch. Work is presently being carried on to compile atlases for all Polish and other Slavic territories.

In the first stage of this type of study, we have monographical descriptions of the local dialects of particular country regions. In comparing these monographs, the basic features differentiating the various dialects studied can be determined. On the basis of this comparison, a questionnaire of one to three thousand questions concerning all forseeable linguistic differences is made up. Next, six localities of approximately equal distance from one another, around thirty to forty kilometres, depending on location, are established on the map. Linguists are then sent out to make inquiries, using the questions found in the questionnaire, among the local population inhabiting the localities marked on the map. Having collected a corpus of material on one question, the data is then transferred onto a single map which, in this way, provides a picture of linguistic differentiation in respect to that single detail. All of the questions in the questionnaire are dealt with in turn. Thus, a linguistic atlas is compiled containing as many maps as there were questions; each map presents the extent of differentiation of one aspect of language structure or another, thereby giving the course of isophones, isomorphs and, most important, isolexes in a given region. Having established the linguistic boundaries, we proceed to determine their bundles, thereby determining the delimiting belts between dialects and languages.

Linguistic geography maintains that absolute linguistic homogeneity exists nowhere, that even in a single village, dialectical differences sometimes occur. Differentiation, therefore, is not the same everywhere. On areas surrounded by delimiting belts it is less marked and, consequently, those regions characterized by relatively little linguistic differentiation are defined as homogeneous dialects. Dialects which are mutually comprehensible without special study are considered as belonging to a single language.

While dialectology is concerned with the distribution of languages and dialects at the present time, historical dialectology studies their distribution in the past, determining changes in the courses of language boundaries over the centuries. In achieving this goal, it makes use of methods already familiar to us—philology and relative chronology—as well as its own special geographical method. In this latter method, conclusions are drawn concerning the courses of language boundaries in the distant past on the basis of the present courses of these boundaries. The correspondence between the courses of present-day isoglosses and their bundles and tribal, political and church boundaries in past centuries is also established. Certain conclusions as to the chronology and even the causes of linguistic changes which are reflected in the isoglosses studied can be drawn on the basis of this correspondence.

The comparative method, supplemented by the results of linguistic geography and historical dialectology, enables us to compare related languages and the corresponding forms within them deriving from common forms in the proto-language. It is clear, however, that only in a few cases can these proto-forms be found in written form; normally, the linguist must reconstruct them. He accomplishes this by choosing the oldest elements from among related, corresponding forms and then arranging them to obtain reconstructed forms, identified by an asterisk. In comparing Old Slavic *jamъ* and Polish *jem*, alongside *jadł* we reconstruct the proto-Slavic **jěmъ*. The question now arises as to the criteria on the basis of which we choose those elements which are oldest, closest to a common proto-form, from among related forms. This is accomplished on the basis of three principles, which will be presented in brief form.

The first principle of reconstruction can be defined as the exclusion of unconditioned splitting. This method states that, if one phoneme of language *A* corresponds to two or more phonemes of language *B* and it is impossible to state the conditions and causes of this splitting, then language *B* is older, for the assumption that the several phonemes found in language *B* could have merged into the single phoneme of language *A* is easier to accept than the assumption of the unconditioned and unmotivated splitting up of the single phoneme of language *A* into the several phonemes of language *B*. Thus, for example, in Serbo-Croatian (language *A*) we have the same vowel *a* as a form corresponding to two Old Slavic phonemes *ъ* and *ь*, and the two Russian vowels *o*, *e* (languages *B*)—e.g., Serbo-Croatian *san*, *dan*, Old Slavic *sъnъ*, *dьnь*, Russian *son*, *deń*, Polish *sen*, *dzień*. Thus, the Old Slavic form is in this case oldest, and therefore it must be assumed that in proto-Slavic two separate vowels existed, which later merged, e.g., back *ъ* (jor) and frontal *ь* (jer). In this way, we reconstruct proto-Slavic **sъnъ*, *dьnь*.

The second principle in the reconstruction of proto-forms is based on linguistic geography and involves inferring the age of elements from their distribution over a territory. Two basic principles operate: (1) A given element is older, regardless of its distribution, if it is found in peripheral areas of the territory of a given language

family. These are peripheral archaisms, which can be explained only in terms of a common heredity from a proto-language. Among the Slavic languages, for example, only Old Slavic and Polish have nasal vowels—e.g., Old Slavic *pętъ*, *dǫbъ*, Polish *pięć*, *dąb*—while in all the other Slavic languages an oral vowel occurs here—e.g., Russian *pjat'*, *dub*; nevertheless, we assume that nasal vowels existed in the proto-Slavic period (proto-Slavic **pętъ*, **dǫbъ*), because Macedonia (the cradle of Old Slavic) and Poland occupy the periphery of the Slavic zone. This conclusion is substantiated by other data as well; (2) Where peripheral archaisms are not involved, the older form is that which covers the greater area. Thus, in Polish, the root syllable of the word *biorę* contains the vowel *o*, while in all the other Slavic languages the vowel *e* occurs here, e.g., Russian *berú*, Old Slavic *berǫ*, Serbo-Croatian *berem*, etc. We conclude from its wider distribution that the vowel *e* occurred in this word in proto-Slavic (proto-Slavic **berǫ*)—which in Polish was transformed into *'o*, the preceding consonant being soft before hard apical consonants *t*, *d*, *s*, *z*, *n*, *r*, *l*.

The third method enabling us to choose the oldest elements from among the forms being compared is based on the analysis of general linguistic developmental tendencies. It maintains that, in a series of languages, changes occurred in a particular direction and that, by analogy, a similar change took place in the language studied. In the Slavic languages, the alternation *k*:*cz*:*c* appears—e.g., *ręka*:*rączka*:*ręce*. The softening of *k* into *cz* or *c* can be noted in a whole series of languages, while the reverse process has never been observed. We conclude from this fact that among the three Slavic phonemes in question, *k* is the original form, while *cz* and *c* evolved from *k* before frontal vowels.

Every language change should be investigated by means of the greatest possible number of methods so that the results of one method may be checked and corrected by those of other methods. From the standpoint of methodology, the history of languages falls into two periods—the historical period from which we have written texts and to which we can, therefore, apply the philological method, and the prehistorical period from which we have no written texts. In dealing with the prehistorical period, only the method of internal reconstruction and the comparative method, including linguistic geography, are applicable, as a consequence of which, our picture of this prehistorical period must be less clear than that of the historical period. Despite lacks resulting from inconsistencies between writing and pronunciation, the philological method, which cannot be applied to the prehistoric period, is the most reliable method in historical linguistics. Only this method, based on the analysis and chronology of texts, enables us to present changes in the total system. A text, although it may include only a few pages of print, graphically imperfect, enables us to establish the general outlines at least of the phonological, morphological and syntactic systems which a language in a given period makes simultaneous use of. By comparing the conceptualization of this system with those taken from later and earlier texts, we obtain a profile of the evolution of a language system.

This excellent method, however, has one great disadvantage. Its range of applicability is limited, since very few languages have a long tradition of writing.

The remaining methods of historical linguistics are much less reliable. The method of internal reconstruction (relative chronology) generally is of secondary significance. It enables us to determine a small number of unconnected facts, but does not enable us to reconstruct the entire evolution of a language. Furthermore, this method cannot be applied to many languages, for example Turkish, which are practically lacking in morphological exceptions. The geographical method does not give particularly reliable results and is of an auxiliary nature. Next to the philological method, the most important is the historical-comparative method, which is indispensable. It has certain disadvantages, however, which limit its application. These limitations can be covered in four points:

(1) The comparative method makes it possible to reconstruct fairly reliably the elements of the proto-language, i.e., particular phonemes, words and endings; it does not, however, enable us to reconstruct the entire system. There are two reasons for this: firstly, it is not certain whether the reconstructed elements once functioned simultaneously or whether they also represent various periods; thus, it is uncertain whether or not they constituted a system. Secondly, it is not certain whether our reconstruction includes all of the elements of the proto-language or whether, besides the reconstructed elements, the proto-language did not possess certain phonemes, words and endings which have disappeared altogether in the surviving languages and, therefore, cannot be reconstructed. Thus, we cannot be sure that our reconstruction is complete. The comparative method presents the evolution of individual elements of a language; it does not provide, however, a complete and reliable picture of the evolution of the entire system of the language.

(2) In many cases, the historical-comparative method provides no criteria for distinguishing between elements that are borrowed and those that are indigenous, for prehistorical borrowings can be distinguished from indigenous elements only when they are in contradiction to phonetic laws, as previously discussed in reference to borrowings of the type *obywatel*. In other cases, it is impossible to distinguish early borrowings from indigenous words. Moreover, cases exist in which regular phonetic correspondence occurs in borrowings. Polish *król* (king), for example, corresponds in the regular way to Russian *koról'* and Czech *král*, although the word is a later borrowing from German *Karl*, i.e., *Karol Wielki* (the Great). Here, the strict transposition from one language to another substituted for strict phonetic development. These difficulties in uncovering word borrowings significantly decrease the value of vocabulary for comparative studies.

(3) The historical-comparative method does not always provide sufficient criteria for distinguishing processes arising from a common source in the proto-language from processes which merely had a parallel development in particular distinct languages of a family. Forms identical in structure are not necessarily continuants of a single common proto-form; if they are of the productive morphological type,

these forms, in particular languages may have arisen independently from one another, although from identical elements. For example, the forms of the genitive singular in Polish *słowa* (words) and Russian *slóva*, although they correspond phonetically, do not derive from a common proto-Slavic form, but are both morphological neologisms formed according to the proportion *lato:lata* = *słowo:słowa*. The corresponding form in proto-Slavic, **slovese*, died out in both Polish and Russian.

(4) The historical-comparative method yields reliable results in studies of those language families which include numerous languages characterized by elaborate morphological and, especially, inflectional systems; in other cases, however, it is difficult to use. In studies of the Indonesian languages of Borneo, Java and the other Sunda Isles, the comparative method meets with two difficulties. First, the absence of an elaborate inflectional system forces the linguist to base his conclusions on the comparison of words, which may include mutual borrowings occurring at a relatively late period. Second, a huge percent of Indonesian words were influenced by factors operating outside the sphere of phonetic laws, factors such as expressive duplication, onomatopoeic distortion, childish language and the language of courtesy. Despite these difficulties, it appears highly probable that the Indonesian languages constitute a coherent family, the phonological system of which does not greatly differ from that of the common proto-language.

In summary, it may be maintained that the comparative method, although it does not allow for the accurate reconstruction of the system of the proto-language, does give a certain over-all idea of this system and is almost completely reliable as concerns certain elements and details. We cannot, therefore, study reconstructed languages in the same way that we study present-day languages or languages recorded in texts; this is not, however, necessary. The reconstruction of proto-languages is merely a means for achieving the primary aim of historical linguistics, i.e., the historical classification of languages and the presentation of their evolution. We are concerned with being able to state which languages derive from a common proto-language, i.e., constitute a family, next, which languages within a family had a longer common development, and which diverged from one another earlier, and, finally, what changes these languages underwent in the prehistorical period of their evolution.

The classification of the dialects of a single family proceeds on these same principles. Those dialects involving the greater number of significant common linguistic changes, are considered more closely related to one another, while those involving fewer such changes, are more distantly related. The matter is complicated by the fact that, frequently, dialects and even whole languages, in their development, alternately converged and diverged in various combinations, which, given the insufficiency of written texts, seldom allows for the elucidation of all particulars. We do not, therefore, yet have a precise and detailed historical classification of the languages of the world; we do have a general outline of this classification, however, which in time will most surely be perfected.

Note: Describe the philological method, the method of internal reconstruction (relative chronology) and the historical-comparative method. What are the advantages and disadvantages of these three methods for investigating the evolution of language? In what sense do phonetic laws function without exception? What is the difference between phonetic laws and natural laws? What are so-called apparent exceptions to phonetic laws and in what manner do they arise? What principles do we apply in reconstructing the forms of the proto-language of a particular family? Why do we establish the degree of relationship among languages primarily on the basis of their morphological systems? How is the geographical method applied in studies of language evolution?

CHAPTER 11. HISTORICAL CLASSIFICATION OF THE LANGUAGES OF THE WORLD

The first impression one has upon examining a map of the languages of the world is of a hopeless chaos and confusion of dots designating the territories in which various languages are used. An enormous effort on the part of linguists was necessary to make order of this chaos, to understand the causes for these particular language boundaries. In the course of centuries, language zones have shifted and superimposed themselves upon one another, crowding each other out and transforming one another. In order, therefore, to understand the present-day configuration of language zones, we must look back in time and study the main routes along which peoples moved in disseminating languages into new territories. This process led to the genesis of families of languages.

Three stages may be distinguished in the history of the formation of these families. In the beginning, the proto-language was used in a relatively small area, owing to which fact, it retained its homogeneity. The second stage involved a period of rapid territorial expansion of the proto-language, which superimposed itself on the zones of neighboring languages effecting certain initially insignificant language changes. In the third stage, centrifugal forces take over and a relatively homogeneous territory splits up into a series of distinct languages.

The Romance family of languages provides an excellent example of these successive stages of development, the whole of which is, in this case, fully illustrated by written documents. In the first stage, between the 7th and the 4th centuries B.C., the proto-language of the family, Old Latin, was used in the city of Rome, on the seven hills. On this small territory, the language was completely homogeneous and distinct from other dialects. In the second stage, between the 3rd century B.C. and the 4th century A.D., owing to the expansion of the Roman Empire, the vulgar Latin of the general populace inhabiting the capital, which had derived from Old Latin, extended into the northern part of the Balkans, into southern and western Europe as well as to the northern coast of Africa. Not until the fall of the West Roman Empire, in the 5th century A.D., did the homogeneity of Latin undergo fragmentation and, in its place, as a result of different development in various conditions, the ten Romance languages arose, which have been discussed previously (cf. pages 98–99). We treat the evolution of other families of languages, more poorly documented, in

a manner consistent with the pattern of development of the Romance family, which is substantiated by written texts.

Besides families, we distinguish other, looser language groups, which we call leagues. They are made up of languages of different origin which, owing to centuries of mutual contact and influence, have become closer to one another, not only as regards vocabulary, but also in the structure of their phonological, morphological and syntactic systems. The Balkan league has already been cited as an example of this type of grouping (cf. page 100).

Cycles, or chains, of languages constitute even looser groupings. In this case, one family borders another and, as a result of the shifting boundaries between them, a transitional region of vital mutual contact arises. The language system of such a zone demonstrates features originating from both families. In this way, both families are connected with each other like links in a chain by the transitional zone of the languages.

The languages shown on the map of the world constitute two great strata—the older, which underwent expansion before the 15th century A.D., and the younger, which became superimposed on the previous stratum after the 15th century. These two strata disseminated along different routes and in different conditions. The languages of the earlier stratum disseminated primarily along land routes from the central zone, i.e., Asia and neighboring territories. The newer stratum, on the other hand, is represented by the languages of Europe, which from the beginning of the 16th century, were carried to the furthest ends of the earth, primarily by sea routes. Therefore, it is actually necessary to keep in mind two linguistic maps of the world—a 15th century map and a present-day map. Each of these two maps is elucidative on the basis of a somewhat different type of phenomena. (Cf. maps at the end of the book.)

The migration of ethnic groups gave rise to the older stratum of languages having their origin in the central zone of the inhabited world and reaching out toward its extremes. Asia and the adjacent territories constituted the central zone. Thence came migrations moving in four directions—East across the Bering Strait into North America and further into Central and South America as far as Tierra del Fuego; South-East through Indo-China, the islands of Indonesia and Oceania, to New Guinea, Tasmania and Australia; South-West to the southern-most extremes of Africa; West into Europe. Families of languages disseminated along these long migratory routes, crowding and superimposing themselves upon one another, which led to the rise of language leagues and cycles connected to one another in very intricate ways. Those families which were crowded out retreated to the furthest corners of the earth. They were superimposed on by the language families which came after, and these in turn by new languages again. Normally, expanding families absorbed only part of the territory of the languages which they superseded, partially overlapping with these languages, which resulted in a pattern similar to the scales of a fish or the tiles on a roof, one over the other. The shape of the earth's land

masses caused languages to disseminate in the four directions described, which resulted in the formation of four cycles of language families. These cycles will be discussed in turn, beginning with the languages of America and the oldest stratum of the languages of Asia.

In the ice age, the territory of Siberia and North America was covered by a huge glacier. As the glacier retreated toward the north, various hunting tribes driven out of Central Asia by neighboring tribes, migrated north along the track of the receding glacier. Siberia retained its character as a territory gradually occupied by tribes retreating from aggressive neighbors up until the 17th century, as we will see later. The second territory to which peoples retreated when driven out of Asia was at the opposite extreme of Asia, the Caucasus Mountains among which the populations driven from the plains hid. In this way, the oldest language stratum of Asia available to us was preserved, on the one hand, in the Caucasus, where it is called the Japhetic group, and, on the other hand, in the eastern extremes of Siberia, where it is called the Paleo-Asiatic group. Except for these territories, this old stratum of language disappeared under the pressure of expanding languages, primarily Altaic, by which it was superseded. Before this happened, however, this archaic language stratum, driven out of Asia, was carried to the neighboring continent. The Japhetic group of languages reached all the way to the western-most extremes of Europe, while the language group of eastern Siberia began to penetrate into America.

Before the coming of the white man at the beginning of the 16th century, America was inhabited by tribes which in the course of the last several millennia had crossed from Asia into Alaska over the narrow and often frozen Bering Strait. The imigrants arrived in small groups, practically each of which spoke a different language. The first tribes had crossed into Alaska from Siberia so long ago that for ages there had been no languages left in Asia having anything in common with the dialects which they spoke. This older stratum, characterized primarily by rich vowel systems and scanty consonant systems, persisted up to the coming of the white man on areas, far removed from the Bering Strait, to which they had been driven by later arrivals. The principal territory of this older language layer was South America, the lowland part of which, east of the Andes, was totally occupied by these languages, while in the eastern part of North America languages of similarly structured phonological systems were introduced. Because the languages of this archaic layer were distributed over territories lying on the Atlantic, they are referred to as the Atlantic group.

Before the coming of the white man, in the eastern lowland territory of South America, several hundred languages were used which, in the opinion of scholars, constituted about one hundred distinct families. Although we include them all in the Atlantic group because of the characteristic features of the structure of their phonological, semantic and syntactic systems, it is nevertheless possible to distinguish several strata among them.

The oldest stratum, the Paleo-American, included a great number of small families of languages used by peoples who, inhabiting isolated recesses of the Uruguay,

Paraguay, Amazon and Orinoco river basins, remained at the lowest level of cultural development. The largest of these families, Ges-Tapuya, until the coming of the white man, commanded practically the entire area of the Brazilian Uplands. The languages of three great families introduced by tribes of a somewhat higher culture, which included the cultivation of certain plants, were superimposed on this archaic linguistic foundation. The first of these families, the Tupi-Guarani, proceeding along water routes, occupied a large region in the southern basins of the Amazon and Paraguay rivers, as well as the Atlantic coast today belonging to Uruguay and Brazil. The upper Amazon basin between the Madeira and Rio Negro rivers was to a great extent occupied by languages of the Arawak family, which also extended into part of the Orinoco basin as well as to the Lesser and Greater Antilles. Languages of the third great family, Carib, were used primarily in the region between the lower Amazon and the Caribbean Sea. On the Lesser Antilles and Haiti, they superimposed themselves on the dialects of the Arawak family, previously used in these regions.

In contrast to the older, Atlantic stratum, which has no equivalents in Asia, the languages of the newer stratum, occupying the territory along the Pacific coast—thence their name, the Pacific group—demonstrate a distinct structural similarity to the Paleo-Asiatic languages of Siberia and the Japhetic languages of the Caucasus. It seems probable that in the distant past a league of languages existed on the territory of Asia which was characterized by a series of structural features, the most important of which was the unusual wealth of consonants and the scantiness of vowels. The migrations of various ethnic groups drove the languages of this league to opposite ends of the continent—to the Caucasus and to Siberia. From Siberia they then crossed the Bering Strait into Alaska and then spread out towards the southern range of Cordilleras and the Andes as far as Tierra del Fuego.

In South America, languages of the Pacific group, having lost in this region several of their characteristic features as a result of foreign influences, occupied the foothills of the Andes and the southern-most parts of the continent including Tierra del Fuego. In the entire area of the foothills of the Andes between 30° southern latitude and the Equator, the Kichua language is used, and in the vicinity of Lake Titicaca the Aymara language is used. These two great languages of the Andes civilization belong primarily to the Pacific type, as do the languages of the Chibcha family in Columbia and Panama.

The languages of the high civilizations of Central America and Mexico comprise two distinct strata. Here, languages of the family Mixtec-Zapotec-Otomi belong to the older Atlantic stratum, while the languages of the Maya-Quiche family in Central America and of Uto-Aztecan in Mexico—the main representative of which is Nahuatl or Aztec—belong to the younger, Pacific stratum.

In North America, the Pacific types of language had a strong influence on the older Atlantic language stratum. The older type of phonology was retained by languages which, until the coming of the white man, occupied the eastern part of

the continent—the dialects of the Algonquian family between the Atlantic, the upper Mississipi and the Hudson bay, the Iroquoian family in the basin of the St. Lawrence, and of Lakes Ontario and Erie as well as in the Allegheny Mountains, and several other less wide-spread families. Certain archaic features of phonology and distinctive syntactic features characterize the Penutian family in California. The Pacific type, newer and expanding, is represented by languages of the Hokan, Jakon-Takelma and Oregon families occupying the Pacific coast and the foothills of the Cordilleras, by the Kwakiutl-Salish-Kwileut, as well as Tsimshian, in the state of Washington and British Columbia, and by the dialects of the great Na-Dene family in the greater part of Canada and Alaska.

The northern coast of America, the islands of the Arctic Ocean and Greenland are occupied by Eskimo dialects, and the Aleutian islands south of the Bering Strait are occupied by the related Aleutian dialects. The languages of the Eskimo-Aleutian family were the last to be introduced into America from Asia and demonstrate in their structure features common with the American languages of the Pacific type as well as with the Paleo-Asiatic and Ural languages of Asia.

Directly to the west of these languages, languages of the Paleo-Asiatic group are found, which comprise two families—Chukchi-Kamchadal in the north eastern extremes of Asia and on Kamchatka and the Ainu-Gilyak family on the islands of Yezo and Sakhalin. Further to the west, the previously used Paleo-Asiatic dialects were wiped out by the expansion of the Ural-Altaic languages, so that, by the 19th century, only the Yenisei-Ostiac language remained on the Yenisei and the Kotian language on the Agul river.

More can be said concerning the history of this ancient language stratum of South West Asia, characterized by many structural features. The languages belonging to this stratum form here an extensive league which is called the Asianic or Japhetic. In the course of the last three thousand years B.C., the Japhetic languages, familiar to us from cuniform texts, spread over the enormous territory of South Western Asia. Included in this group were the Sumerian language on the lower Tigris and Euphrates, the Elamite language in the vicinity of Susa, and the Cassite language to the north of Susa, the Mitannian and Churic languages in northern Mesopotamia, the Vannic language on the Lake Van, the proto-Hittite language in Central Asia-Minor and many others. In the east, the Japhetic languages reached India. The Burushaski language exists today at the place where the Japhetic languages bordered with the Pamir languages. In the west, three thousand years B.C., the expansion of the Japhetic family reached across the Mediterranean Sea to the Iberian Peninsula where the Basque language in the Pyrenees today represents the survival of that group of the world. In the 9th century B.C., the Etruscans carried their Japhetic language by sea from Asia Minor to northern Italy, where it was used to the 1st century A.D. At the present, the principal center of the Japhetic languages is the Caucasus. Two groups may be distinguished here: the South-Caucasian group (Georgian, Laz, Mingrelian, Svan), and the North-Caucasian group, which includes two sub-

groups: West-Caucasian (the Abkhazian, Ubyk and Adyghe-Qabardi dialects) and East-Caucasian (Chechen, Daghestan and Samurian).

The basic outline of the history of South-East Asia, Oceania and Australia, to which we will now proceed, involves the gradual shifting of peoples from north to south, from Central Asia which was the take off point for these migrations, all the way to Tasmania, which represented their furthest extent.

The extreme culturally backward population of Tasmania, which died out in the 19th century, in various parts of the island, spoke five various related dialects constituting a distinct family, which had originally reached the southern extremes of Australia. Thence, they were driven out by languages of the next layer.

As late as the 19th century in the vicinity of present day Melbourne, languages of the Paleo-Australian family are spoken (Kulin, Kolijon, Buandic) and on the river Murray dialects of the Narrinyeri family are spoken. These areas constituted the southern periphery of the continent. They did not undergo the process of integration which affected the rest of the continent and led to the formation of the great Central-Australian family. For a certain period, this family occupied nearly all of Australia with the exception of its southern-most parts. The proto-language of this family had been carried from New Guinea across the islands of the Torrens Strait and the York peninsula. Through expansion, this proto-language imposed on the whole territory of originally highly differentiated language foundation, a basically uniform grammatical system. The unification of vocabulary did not proceed so far, which led to the formation of several groups of dialects distinct in this respect.

Between the two great families—Central-Australian in the south and the Austronesian in the north—stretches a zone of about 150 small language families which, on the basis of their geography, are divided into two groups—North-Australian and Papuan. The Aranda family, which forced its way from the north into the very center of Australia, and about thirty small families carried from New Guinea to the country of Arnhem on the coast of the Carpentaria Gulf and the York Peninsula, belonged to the North-Australian group. In all of these territories these languages were superimposed on the dialects of the Central-Australian family.

New Guinea—except for its northern coast—is occupied by one hundred small Papuan language families, all very distinct from one another. These languages constitute the survivals of an older stratum which had once occupied the whole of Melanesia and Indonesia before the dissemination of Austronesian languages in these territories. Traces of the old relationships are represented by numerous islands of Papuan languages in these regions. East from New Guinea in New Britain, in the Louisiade Archipelago, on the Solomon Islands (on Bougainville and Savo), New Caledonia and on the Loyalty Islands occur several distinct families of Papuan. Moving west from New Guinea we find distinct Papuan language families on the islands of Halmahera, Timor, Sumbawa and, further, on the island of Engano of the western coast of Sumatra. North of New Guinea on the Andaman Archi-

pelago, dialects of the Andamanese family belonging to the same archaic stratum, are used.

A significant part of the older stratum of Papuan dialects was driven out of use by the expansion of the great Austronesian family. This family came from the south-eastern part of Indochina where dialects of the Cham group belonging to this family are used to the present day. It is from this area that, in the last centuries B.C. tribes swarmed which, travelling by ocean routes, occupied in the course of one thousand years the enormous territory of islands between Africa and America, Asia and Australia, imposing their own language on these regions. Today the Austronesian (or Malayo-Polynesian) family is divided—except for the Cham group—into four sub-families: Indonesian, Melanesian, Micronesian and Polynesian. The Indonesian sub-family encompasses Madagascar, where the Malagassi language is used, the Sunda Isles, the Philippines, Formosa and the Mariana Islands. On the Sunda Isles, the most important languages are: Javanese on Java (the Old-Javanese language, Kavi, having a literary tradition dating from the 9th century A.D.) as well as Malayan from the coast of the Malay Peninsula and neighboring islands, which today, under the title "Indonesian language", is the official language of Indonesia. In contrast to the Indonesian languages which, in principle, preserved the structure of the family's proto-language, the remaining dialects underwent deep changes under the influence of the Papuan language foundation on which the expansive languages coming out of Indonesia in the first centuries A.D., were superimposed. Clear traces of Papuan influences appear in the dialects used on the Moluccas, and still traces in the languages of the Melanesian sub-family on the coast of New Guinea and on the archipelagoes located further to the east, as far as Fiji, including the Micronesian sub-family north of Melanesia. In the first millennium A.D., from the Solomon Islands and New Hebrides, originated the expansion of one group of Melanesian dialects which occupied the previously uninhabited territory of Polynesia. Before colonialization by the whites in the 19th century, the Polynesian sub-family, thus created, occupied an enormous triangle on these islands the apexes of which were New Zealand (Maori language), Hawaii (Hawaiian language) and the Easter Islands.

The Cham group was strongly influenced by neighboring languages of the Austro-Asiatic family, which occupied the greater part of India and Indo-China before it was limited, by the expansion of the Dravidian and Sino-Tibetan languages, to a series of language islands within this territory. Today, the Austro-Asiatic family is divided into two sub-families occurring to the west and to the east of the Ganges delta. The sub-family of the west in India includes two groups: Nahali, south of the Narbada River, and the Munda languages between the Narbada, the lower Ganges and Bengal Bay. The sub-family of the east is divided into five groups: Khasi in Assam, the Palaung group in the northern parts of Indo-China, the Mon-Khmer group, the principal representative of which are two languages having a long literary tradition—Mon in southern Burma and Khmer in Cambodia—the Malaka

group on the Malay Peninsula, and the Nicobar on the Nicobar Island north of Sumatra.

In India, the Dravidian languages, the source of which was Baluchistan, superimposed themselves on the Austro-Asiatic languages. Driven south-east to Deccan by the Indo-European expansion, the Dravidian family broke up into four groups: Brahui, Tamil-Kurukh, Kui-Gondi and Telugu. The Brahui language in Baluchistan represents the remnant of the original territory of the family. Languages of the Tamil-Kurukh group comprise two sub-groups. The southern sub-group includes Tamil (the south-east part of the peninsula and the northern part of Ceylon), Malayalam (the south-west coast of India) and Kanarese (southern Deccan). The northern sub-group, transformed under the influence of the Munda dialects, includes the Kurukh language on the Mahanadi, and Molto on the lower Ganges. In northern Deccan, the Munda languages were superimposed upon by Dravidian dialects of the Kui-Gondi group, strongly influenced by the Munda languages, and in turn superseded by the Indo-European languages. Eastern Deccan and the coast of Bengal Bay are occupied by the Telugu language, which arose from the mixture of Dravidian elements with the substratum of Munda languages.

The eastern part of the Austro-Asiatic zone was broken up into many tiny linguistic islands by the expansion, from the north, of the huge Sino-Tibetan family, which includes two sub-families: Tibeto-Burmese and Sino-Thai.

The first of these occupied Tibet in antiquity, and on the slopes of the Himalayas, in Assam and in Burma superimposed itself on the Austro-Asiatic languages. Influences of this Austro-Asiatic stratum lead to the breaking up of the Tibeto-Burmese subfamily into four parts: (1) the very archaic Tibeto-Himalayan group, (the Old Tibetan language, known from the 7th century A.D., and Modern Tibetan as well as the Himalayan dialects in Nepal and Buton), (2) the group of northern Assam on the Brahmaputra, (3) the group of southern Assam, (4) Arakan-Burmese in Burma, the main representative of which is Burmese, which has a literary tradition reaching back to the 11th century A.D.

The Sino-Thai sub-family also expanded from the north into the south. The isolated Lati languages spoken by 500 people on the Song-ka River on the border of China and Tonkin, appears to be a remnant of the original language stratum in this region. Lati was superimposed upon by the Thai language, which in the first centuries A.D., occupied southern China. Between the 10th and the 13th centuries, under the influence of the expansion from the north of the Altaic tribes, the Chinese language shifted toward the south, driving related languages into Indo-China, where they intermingled with the languages of the Austro-Asiatic stratum and were influenced by them. The Sino-Thai sub-family broke up into four groups: (1) Vietnamese (the Vietnamese, or Annamese, languages in Vietnam and Muong language in Tonkin, strongly transformed by the influence of the Austro-Asiatic stratum), (2) the Thai group (the Siamese languages in Siam, the Lao language in Laos and the now extinct Ahom language in Assam, as well as Miao and Dioi in southern China),

(3) the Karen group in Burma on the lower Salween and Irrawaddy, (4) the Chinese language, which has a literary tradition reaching back to the 14th century B.C. and which today includes a number of northern and southern dialects, quite distinct from one another, spoken by around 800 million people.

The languages of Africa and South-West Asia, to which we will now proceed, constitute three overlapping strata. In the very south, languages of the Khoin family were used; they were impinged upon by languages of the Negro league from the north, which retreated from their northern territories as a result of the southern expansions of the languages of the Semito-Hamitic family. In the course of the last several millennia these three strata have shifted toward the south.

In the last millennium B.C., the dialects of the Khoin (meaning—'man') family were used by the dwarf-like peoples inhabiting regions near the equator. This fact is evidenced by language relics which have survived to the present. In the tropical forests on both sides of the equator, between the Albert and Victoria Lakes and the Gulf of Guinea, live scattered groups of Pigmies. They use languages borrowed from neighboring tribes, but they have introduced into these languages certain phonological features and words deriving from their previous dialects, which belonged to the Khoin family. Between the southern-most extreme of Lake Victoria and Kilimanjaro, there is a small territory occupied by the Kindiga (Hatsa) language, and to the south of this language, the Sandawe language. Both of these languages belong to the Khoin family, although they are saturated with influence of the surrounding languages, which arrived later, and which drove the main part of the Khoin dialects to the southern-most extremes of Africa. In the 17th century, the two principal languages of the Khoin family—San, the language of the Bushmen, and Nama, the language of the Hottentots—occupied the part of the continent south of the Kunene, Okovango, and Waal rivers. Today San, representing the pure family type, is used by the Bushmen of the Kalahari Desert, while the Nama language, which has clearly been influenced by foreign elements, is used by the Hottentots, inhabiting the territory between the Kalahari and the Atlantic. The characteristic feature of the Khoin languages is the phonological use of clicks, inspiratory consonants, which have been discussed previously.

The hundreds of languages used by the Negros of Africa belong to several separate families, which, however, as a result of mutual influences, have become so similar to one another in respect to vocabulary and the structure of the system, that they constitute a secondary entity, which is called the Negro league. In the last millennia B.C., the Negro languages occupied the part of Africa north of the equator, i.e., primarily in the Sudan and the present-day Sahara, which then was covered by abundant plant life. In the center of the Sahara, south of present-day Tunis and directly north of the Tropic of Cancer extends the Tasili Mountain Range. Paintings on the cliffs of these mountains show that this region was once fertile and well irrigated and that the population was originally of the Negro type. Not until the later stages of development of this culture was it strongly infiltrated by white peoples,

who introduced the Semito-Hamitic languages. Around the beginning of the first century A.D., the Sahara began to change into desert, which forced the Negroid peoples to the south. This migration lasted to the 15th century. Only a very scarce population remained on the Sahara, a people the great majority of which spoke Semito-Hamitic languages. The peoples of the Negro league crowded into the Sudan, next occupying southern Africa, partially liquidating the territory of the Khoin family.

Today, the territory of the Negro league, generally speaking, can be divided into two parts: On the one hand, we have the Sudan, i.e., the southern part of the former territory of the league, on which several families were crowded, including several hundred languages, part, indiginous Sudanese languages, part languages which have retreated from the Sahara. The second part constitutes a territory extending south of the equator, highly homogeneous linguistically. It arose as a consequence of relatively recent expansion. One of the group of Negro languages, Bantu (meaning 'the people'), occupying, in the first centuries A.D., territories at the southern extremes of the league in the regions of the equator, probably between the Benue River and the Cameroons, was driven by other Negro peoples to the south and east. It settled in an area extremely scarcely populated by a short-statured people of the Khoin family. In these conditions, the tribes of the Bantu, moving along two paths to the west and to the east of Tanganyika Lake, by the 17th century, had occupied all of South Africa except for the Khoin region described above. Their move toward the south continues. Bantu dialects south of the Limpopo River demonstrate influences of the Khoin language stratum. The most wide-spread of the Bantu languages is Swahili, which had its source on the island of Zanzibar and which became the language of government and commerce on large territories of East Africa.

Part of the Sudanese languages demonstrate a relationship to the Bantu group and together with it comprises the Bantu family, the languages of which are characterized by the use of prefixes of various origin to fulfill various semantic and syntactic functions. A very distinct position within the separate Bantu family is occupied by the Kwa languages (Yoruba, Ewe and others) which extend in a long belt along the Gulf of Guinea from the lower Niger, all the way to the St. Paul River. The remaining Bantu languages of the Sudan, more similar to the Bantu languages in structure, constitute the so-called West Bantu group, which can be divided into four subgroups: (1) the sub-group used in the basin of the Benue River, (2) the subgroup of the Togo province on the lower Volta, (3) The Voltaic sub-group in the upper basin of the White Volta and the Black Volta, (4) the Atlantic sub-group between the Atlantic and the Senegal and Gambia rivers. The last group also includes the Fula language, the territory of which, because of recent expansions, comprises a series of linguistic islands stretching from Senegal to Lake Chad.

Among the remaining language families of Sudan, the number and boundaries of which have not been accurately established, four can be clearly distinguished: (1) to the west, in the region of the upper tributaries of the Niger, Senegal, and Gambia,

languages of the Mande family (Ngo-Nke), the most extensive of which are Mandinge and Soninke, are found. Typologically, they markedly deviate from the remaining languages of the Negro league, (2) the Songoi family in the valley of the central Niger, above and below Timbuktu, (3) in the center of Sudan, the Kanuri-Tubu languages are found, including Kanuri on Lake Chad and Tubu further to the north on the Sahara around the cities of Bilma and Bardai, (4) in the Nile basin, languages of the Nile family are used, which can be divided into three groups. In the Nile Valley, from Aswan in the south to Khartoum, the Nubian language prevails, the main representative of the North-Nile or Nuba-Kunama group. The tributary region of the White Nile and the Blue Nile is occupied by languages of the Central-Nile group, the most important of which are Dinka and Nuer. Languages of the South-Nile group (Masai and others), occur among the Rudolph, Albert and Victoria Lakes and in the mountains of Kenya and Kilimanjaro. This group arose as the result of Hamitic influences superimposed on a Nile layer, which is why it is sometimes called the Nile-Hamitic group.

The source of the Semito-Hamitic family was in Arabia and the Syrian Desert. From this area, in the course of the last several thousand years, it spread out in three directions: north to Mesopotamia and Syria, east to Egypt and the Sahara, and south across the Bab-el-Mandeb Strait to Central Africa. As a result of the Semitic expansion in the two latter directions, a third linguistic stratum, the Semito-Hamitic, arose in Africa. On the peripheries of this language zone occur languages belonging to the older strata, which, however, have been strongly influenced by the Semito-Hamitic stratum, e.g., dialects of the Nile-Hamitic group. Even more distinct influence of the newer layer can be found in the languages of the mixed Hausa-Musgu group, between the Niger and Lake Chad and the Logon River, the most important of which is Hausa, the language of commerce of northern Africa. The languages of this group arose as a synthesis of Bantu forms and archaic Semitic elements, perhaps from the period of the Tasili culture.

The Hamitic languages represent a further degree of Semitization. Their systems are of Semitic origin, but they have been immensely transformed by the influence of the Negro stratum. The eastern part of Africa, between the Red Sea and the Gulf of Aden and the Nile, Lake Rudolph, Kenya and Kilimanjaro, was occupied by the Kushitic languages, the most important of which today are Bedawje, Somali and Galla. The Libyan and Numidian dialects, the present-day continuants of which are the Berber dialects such as Tamashek, the language of the Tuaregs, Rifian in Morocco, and others, were used on the Sahara in the last millennia B.C. Next to the Kushitic and Berber groups, we have the third Hamitic group, represented by Egyptian, which has a literary tradition reaching back to the third millennium B.C. The last developmental phase of this language is represented by the Coptic language which disappeared from use in the 16th century A.D., but which is still the liturgical language of the Christian-Coptic Church in Egypt.

The true Semitic languages can be divided into two groups—eastern and western

As a representative of the eastern group, we have the Akkadian language, which arose in the third millennium B.C., as a result of the superimposition of Semitic dialects in the region of Mesopotamia on the Sumerian and other Japhetic (Asianic) languages. The Akkadian language covered lower Mesopotamia (i.e., Babylonia) and part of the upper basin of the Tigris (Assyria). Akkadian texts written in cuniform appeared between 2500 and 500 B.C., the language died out around 500 B.C.

The West-Semitic group can be divided into two sub-groups—northern and southern. The Northern-West sub-group included the Canaanite language used in later Palestine in the second and first millennia B.C. We have: (1) the Ugaritic language, known from inscriptions from the 14th century B.C. discovered in Ras-Shamrah on the coast of the Mediterranean Sea; (2) the Phoenician language in Phoenicia and her colonies on the North African coast (Carthage and others); (3) the Moabite language south-east of the Dead Sea; (4) Hebrew, the classical form of which is preserved in the Old Testament, between the 11th and 2nd centuries B.C., and the present-day form of which (Modern Hebrew) is used in Israel. After the Canaanite languages, the second member of the Northern-West sub-group of the Semitic languages is Aramaic. It includes those western dialects which occupied Syria in the 10th and 9th centuries B.C., and Palestine and Phoenicia in the 3rd and 2nd centuries B.C., superseding the Canaanite languages. The East-Aramaic dialects occupied Mesopotamia from the 6th century B.C. to the 10th century A.D., and, in the region between the Wan and Urmia lakes, has survived to the present day.

The Southern-West sub-group of Semitic languages can be divided into three zones—South-Arabic, Ethiopian, and North-Arabic. The South-Arabic dialects, which today are found in the eastern extremes of the province of Hadhramaut and on the island of Sokotra, are known from inscriptions discovered in Yemen, and originating from the period between the 9th century B.C., and the 6th century A.D. The expansion of the South-Arabic dialects across the Red Sea to Africa in the last centuries B.C., lead to the formation, on a Kushitic stratum, of the Ethiopian group, which can be divided into two sub-groups; northern (the Ge'ez language in the first millennium A.D., and its present-day continuant, the Tigrinia language, together with the closely related Tigre language); southern (the Amharic language and others). The North-Arabic, i.e., Arabic languages, owing to the expansion which developed along with that of Islam from the 7th century A.D., superseded practically all the other Semito-Hamitic languages and occupied Iraq (ancient Mesopotamia), Syria, practically all of Arabia and a significant part of North Africa, and temporarily reached even Europe, on the Iberian Peninsula and Sicily. Classical Arabic, of which we have records dating from the first centuries A.D., today is a literary language only, while Modern Arabic dialects, different in each country, are used in everyday communication.

Again we return to the question of the languages of North Asia and Europe. The old, highly differentiated linguistic stratum of these territories, respresenting Paleo-Asiatic and Japhetic languages, disappeared in the course of the last 4000 years as

a result of the expansion of three great families: Altaic, Uralic, and Indo-European.

Proto-Altaic, the source of which was some part of the territory of the present-day Gobi Desert, initially broke up into two groups—Turkic and East-Altaic. The first language to split off from the western periphery of the Turkic group was Chuvash, which moved far to the west and today is used along the Kama and Volga rivers in the vicinity of Kazan and Ulianov. The territory of Old Turkish, from which all of the remaining Turkic languages derive, in the 8th century A.D., reached the region to the south of Baikal Lake on the Orkhon River. The Yakut language split off from its northern periphery. The territory of this language, on Baikal Lake, was part of the nation of Jenghiz Khan, as a result of which Yakut has a large number of borrowings from Old Mongolian. In the 14th or 15th century, the Yakut people deserted their territory on Baikal Lake, which then was occupied by Buriat dialects of the Mongol group, and wandered along the Lena in the vicinity of present-day Yakutsk. Here the Yakut language zone arose, which expanded at the cost of the Tungus and Yukaghir dialects.

The remaining, Central-Turkic dialects shifted toward the west. Around the 9th century A.D., they occupied East and West Turkistan, and, in the 14th and 15th centuries, Asia Minor, part of Thrace, the Crimea and several smaller territories in eastern Europe and Iran.

The present-day Central-Turkic dialects can be divided into four groups. The southern group includes Osmanli, i.e., the language of the present-day Turkish nation in Asia Minor and Thrace, as well as the Turkoman dialects; the western group includes Kirghiz, Bashkir and Tartar; the central group includes Uzbek and Sart; the eastern group includes Uighur and the Sayan dialects.

The East-Altaic group initially broke up into two sub-groups—Mongol and Manchu-Tungus. In the 12th and 13th centuries A.D., all of the Mongolians, who then inhabited the eastern part of their present-day territory on the Chinese border, spoke a single language, within which appeared only minor dialectical differences. This language, known to us from 13th century texts, is called Old Mongol. It constituted the proto-language of the family. In the 13th century, as a result of the conquests of Jenghiz Khan, the Old Mongol language occupied a huge territory in Central and West Asia, which, after the fall of the Empire in the 14th century, began to differentiate linguistically. In the period between the 14th and 17th centuries, Old Mongol broke up into six Mongol languages. Three of them, at a greater distance from the center of the family, constitute peripheral languages. They differ distinctly from one another and from the central languages, as a consequence of the fact that they have preserved many of the features of Old Mongol (so-called peripheral archaisms), which died out in the central territory, as well as influences of a foreign language stratum. Among these languages, in the west is the Mogol language of Afghanistan, transformed by influences of the Iranian dialects. South of Lake Kuku-nor, in the vicinity of the Sino-Tibetan dialects, is

the Mongor language. East of the Nonni River in Manchuria, the Dahur language is predominant. These are all remains of the 13th century Mongolian expansion. The Kalmuk language, occurring in the west in the vicinity of the city of Kobdo and the source of the Irtysh River, is one of the three central languages. In the 17th century, part of the Kalmuk people migrated to the lower Volga and Don rivers, creating the Kalmuk language zone. North of Baikal Lake, the highly differentiated and highly developed Buriat dialects are used. The entire territory of Inner and Outer Mongolia is occupied by the Chalcha language, which constitutes the center of the family. The Chalcha zone is distinctly homogeneous; only at its most south-western extremes are individual dialects containing archaic features found, such as Ordos, Tsachar and others.

The source of the Manchu-Tungus group, which extended to the east and north of the Mongolian territory, was the basin of the Sungari River in Manchuria. In the first century A.D., the linguistic ancestors of the present-day Tungus people, deserted their original fatherland and moved in a north-western direction to the basin of the upper Amur and Lena rivers. Here, under the influence of the Paleo-Asiatic linguistic substratum, the Tungus language arose, differing clearly from the Manchu language; Tungus is still used on the Sangari River in Manchuria, and in the 17th and 18th centuries, it was the language of the ruling class in China. In the course of several centuries, in several migratory waves, the Tungus dialects occupied the huge territory between the Sea of Japan and Okhotsk and Baikal Lake and the Yenisei, becoming superimposed on Paleo-Asiatic dialects. In the last centuries, the Yakut language has occupied the territory on the Lena, forcing itself like a wedge between the Tungus dialects.

The relation of the Korean language—which occupies Korea, on the west bordering the territory of Manchuria—to the Altaic family is unclear. Probably proto-Korean was related to proto-Altaic and, together with it, derived from a common proto-language, which certain scholars have called proto-North-Asiatic. In Korea, proto-Korean was superimposed on some Paleo-Asiatic language, by which it was deeply transformed. Certain scholars suppose that Korean is related to Japanese, which is a mixed language and arose as a result of greater or lesser influences on the part of a language close to Korean. The grammatical structure of both Korean and Japanese is closest to the Altaic languages, but, as a result of the fact that Korea, in the 4th century A.D., and Japan, in the 7th century, adopted the culture and script of China, both languages abound in vocabulary borrowed from Chinese.

The most important member of the Uralic family is the Finno-Ugric group. Its source, in antiquity, was the basin of the Volga and Oka rivers. Among the languages of this region, the first to become differentiated were the Ugric dialects which moved east to the foothills of the Urals and the Ob River basin, where two languages of this group, Ostyak and Vogul, are used to the present. The first to separate from the Ugrian group were the Magyars, and after wandering far, to the Caspian Sea and the Sea of Azov, underwent Scythian and Turkish influence and

settled, in the 9th century A.D., on the Danube and Tisza rivers, where they became superimposed on Slavic peoples. The Hungarian language, which has a literary tradition reaching back to the 13th century, constitutes the synthesis of these various elements.

The Finno-Mordvin dialects, which remained after the departure of the Ugrian peoples, underwent further differentiation. The first to break off was the Permian group, today including the Zyrien language on the western slopes of the Urals, in the basin of the Pechora and Vichegda, as well as the Votyak language in the Kama basin. After the Permian group, the Finnish dialects broke off, so that only the Mordvin and Cheremiss dialects remained in the old Volga nation. The Finnish languages occupied a large territory on Lakes Onega and Lagoda (the Karelian and other languages), next in Estonia (Estonian), Kurland (Livonian) and Finland (Suomi). The Lapp language on the Kola Peninsula and in northern Scandinavia also belongs to the Finnish group. The Lapps adopted this language, abandoning their original tongue, perhaps related to the Samoyedic languages, to which we will now proceed.

As has been indicated previously, the whole of Siberia is a kind of backwash area, the northern periphery of the inhabited world to which peoples driven out by their neighbors retreat. In the upper basin of the Yenisei, on the northern slopes of the Sayan Mountains, the Kamassin language is used, the territory of which represents a remnant of the original territory of the Samoyedic group which—alongside the Finno-Ugric group—constitutes the second member of the Ural family. Of the Samoyedic group, only the Kamassin language remains on its original territory, while the remaining languages shifted to the north under the pressure of Turkish tribes, and occupied the basin of the central and lower Yenisei (Ostyak-Samoyed and Yenisei-Samoyed) as well as the coast of the Arctic Ocean from the White Sea to the mouth of the Khatanga to the east (Yurak and Tavgi).

Samoyedic influences, reaching far to the east, were superimposed upon a substratum of Paleo-Asiatic languages which, in combination with other influences, led to the formation of the Yukaghir language, today spoken on the coast of the Arctic Ocean east of the lower Jena as well as in the basin of the Kolyma and Omoloy. Further to the east, the penetration of Ural influences is demonstrated by the Eskimo language.

Turkish, Mongolian, and Finno-Ugrian tribes, during their wanderings in the west, came into contact with languages of the great Indo-European family, the expansion of which, a continuous process for the last 5000 years, has completely transformed the relationship among the languages of the world. The enormous number of languages belonging to this family constitute thirteen groups, each of which is derived from an intermediary proto-language, which are merely continuants of the dialects of the proto-Indo-European language, the proto-language of the entire family. The groups of the Indo-European family are as follows:

(1) the Indic group, including, in antiquity, the Vedic language and Sanskrit,

which derived from it, today including the various languages of Hindustan and Pakistan;

(2) the Iranian group, including present-day Persian, Afghan, Ossetic;

(3) the Tokharian languages, which became extinct in the 8th century A.D., used in Central Asia and in Chinese Turkistan;

(4) the presently extinct Anatolia group, the main representatives of which were Hittite and the Luvian language used in Asia Minor in the second millennium B.C., and known from texts from the 14th and 13th centuries B.C.

(5) the Thraco-Armenian group which, in antiquity included the Thracian dialects in the Balkans and Phrygian in Asia Minor, and the only representative of which today is Armenian;

(6) Greek, the oldest preserved form of which is the language of the texts written in so-called Linear B—1450 to 1200 B.C.—and next, the language of Homer; the Macedonian language was closely related to Greek;

(7) Albanian, the last remnant of the Daco-Mysian dialects used in antiquity in the land of the Daco people and in the south-east part of the Balkans;

(8) the Illyrian and Messapian dialects in the Adriatic region;

(9) the Italic languages, including three sub-groups in antiquity (a) Latinic languages, represented by Latin (today continued by the Romance languages), (b) Osco-Umbrian dialects, (c) the Venetic language;

(10) the Celtic languages—today Irish, Scots, Welsh and Breton;

(11) the Germanic languages—German, English, the Scandinavian languages and Gothic, now extinct;

(12) the Slavic languages (cf. genealogical tree on page 99);

(13) the Baltic languages—Lithuanian and Latvian and Old Prussian, which died out in the 17th century.

Our hypothesis concerning the relation of all these languages is based on the fact that the elements of their semantic and syntactic systems, word endings, vowel alternations, grammatical morphemes, numerals, pronouns and basic vocabulary, can to a great extent be traced to a single proto-source. To illustrate, we will present the conjugation of the verb *jest* (is), which was preserved in many language groups:

Proto-Indo-European forms	Vedic	Hittite	Armenian	Greek	Latin	Gothic	Lithuanian	Old Slavic
Sing.								
1st per. **és-mi*	*ásmi*	*esmi*	*em*	*eimì*	*sum*	*im*	*esmì*	*jesmъ*
2nd per. **és-si*	*ási*	—	*es*	*essì*	*ess*	*is*	*esì*	*jesi*
3rd per. **és-ti*	*ásti*	*eszi*	*ē*	*estì*	*est*	*ist*	*ẽsti*	*jestъ*
Plur.								
1st per. **s-més*	*smás*	—	*emkh*	*eimès*	*sumus*	*sijum*	*ẽsme*	*jesmъ*
2nd per. **s-té*	*sthá*	—	*ēkh*	*estè*	*estis*	*sijuth*	*ẽste*	*jeste*
3rd per. **s-énti* **s-ónti*	*sánti*	*asanzi*	*en*	*eisì*	*sunt*	*sind*	*ẽsti*	*sątъ*

The correspondence, in conformity to phonetic laws, of forms of the oldest European languages, has made it possible to reconstruct proto-Indo-European inflection, which was based on changes of word endings—1st per. sing. *-mi*, 2nd per. sing. *-si*, 3rd per. sing. *-ti*, 1st per. pl. *-mes*, 2nd per. pl. *-te*, 3rd per. pl. *-enti*:*-onti*. The forms of proto-Indo-European inflection were also characterized by root vowel inflection and the place of accent. In the singular, the root vowel appeared in the form of *e* (the root form being *es-*), and the accent fell on the root; in the plural, however, the root vowel appeared in the form of *zero* (the root form being *s-*), while the accent fell on the ending. These properties of proto-Indo-European inflection survive in vestigial form in modern Polish, where the survival of the old alternation of the root vowel is represented by the opposition of the forms *jes-t* and *s-q* ('is' and 'are').

It seems probable that the proto-Indo-European dialects were spoken in the third millennium B.C. in Central and Eastern Europe between the Rhine and the Dnieper The conquering Indo-European tribes, having a pastoral culture and a patriarchal social system, rapidly increased their territory, as a result of which language contact became less close. The first manifestation of this process was the gradual loss of contact between the central dialects, which elaborated the Indo-European system, and the peripheral dialects to which, consequently, these innovations did not reach.

The first to lose contact were the dialects of the eastern and southern peripheries spoken by the predecessors of the Tokhars and Hittites. They also preserved the most ancient form of the phonetic and inflectional systems of the proto-language. The Hittite language (and probably the same was true of the Tokharian languages) possesses in the plural and dual declensions the continuant of the proto-Indo-European form of the nominative only; it lacks the old forms of the remaining cases. This fact seems to prove that, not until the differentiation of the proto-Hittite dialects, (and probably proto-Tokharian as well), can we speak of the completion of the proto-Indo-European declensional structure through the formation of dependent cases, i.e., those other than nominative, plural and dual, in which the endings of these cases have a different form in different Indo-European languages—which also argues for the late date of the process.

Some time after the separation of the Tokhars and Hittites, a break in contact between Italo-Celtic dialects of the western periphery of the remaining Indo-European dialects took place. The central dialects accomplished an important elaboration of the system of conjugation, an elaboration in which not only the proto-Tokharian and proto-Hittite dialects took part, but also the dialects from which the Italic and Celtic languages developed. The old conjugation system involved the use of two series of endings in the conjugation of the active voice (activum)—one for the present tense (sing. *-mi*, *-si*, *-ti*, 3rd pers. pl. *-nti*), the other for the remaining formations (sing. *-m*, *-s*, *-t*; 3rd pers. pl. *-nt*), while in the conjugation of the passive-reflexive voice (medio-passivum), only one series of endings was used for all formations. This system was preserved in the dialects from which the Hittite and Tokharian languages as well as the Italic and Celtic languages developed, while in the remaining Indo-European

dialects, which occupied a more central position, in the inflection of the passive-reflexive voice, two series of endings, analogous to the conjugation of the active voice, were created—the newer series (sing. *-mai, -sai, -tai*, 3rd pers. pl. *-ntai*) being used for conjugating the present tense, and the old series for the conjugation of the remaining formations. The second difference of a similar extent and, therefore, most likely arising from this same period, concerns the form of the genitive singular of nouns the stem of which ends in *-o*. In the central dialects, the ending *-o-so* or *-o-syo* became established in the forms of this case—e.g., sanskrit *vṛkasya* (wolf), Old Prussian *deiwas* from **deiwo-so* (of god)—however, there is no trace of this form ending either in the Hittite or Tokharian group nor in the Italo-Celtic group.

Some time after the splitting off of the Italo-Celtic dialects, a break in contact between the proto-Germanic dialects of the northern periphery of the language zone, the proto-Greek dialects of the southern zone and the remaining, more central, dialects occurred, which left its mark on the extent of two important sound changes. The first of these concerns the evolution of dorsal consonants, the system of which presents itself originally in the proto-Indo-European language as follows:

		Unvoiced	Voiced	Voiced aspirated
Palatal		*k'*	*g'*	*g'h*
Velar	unlabialized	*k*	*g*	*gh*
	labialized	*k*^{u}	*g*^{u}	*g*^{u}*h*

Palatal and velar consonants formed a parallel opposition differing from each other in the point of articulation, the palatal being pronounced more toward the front, and the velar more toward the back of the oral cavity. Among the velar consonants, however, appeared a privative opposition, the unmarked member of which involved unlabialized consonants. These latter included within a single phoneme, besides the main articulation, which involved closure of the back part of the oral cavity, an additional articulation involving the rounding of the lips, or labialization (expressed in writing by the letter ^{u}), which is what constitutes the positive distinguishing feature in the opposition.

The complicated system was simplified everywhere, but in two different ways. Thus, there arose two groups of Indo-European languages, the centum group (from Latin *centum* 'hundred') and the satem group (from Old Iranian *satəm* 'hundred'). In centum languages (Tokharian, Hittite, Italic, Celtic, Greek, Germanic), the opposition of point of articulation disappeared, and only the labial:non-labial opposition remained. Thus, the series *k'* and *k* merged into a single *k* (in Latin written *c*), while the labial *k*^{u} (in Latin written *q*) retained its distinctiveness. In the satem languages, on the other hand, (Indic, Iranian, Armenian, Albanian, Slavic and Baltic), the center of which is represented by our family, the labial:non-labial opposition disappeared and opposition based on point of articulation was extended,

which led to the transformation of dorsal palatal consonants into alveolar affricate consonants. Thus, the series *k* and *k*$^{u}_{\sim}$ merged into a single *k*, while *k'* was transformed into an affricate of the type $\check{c}$, which later changed into the spirants—Indic *ś*, Lithuanian *š*, and in other satem languages, into *s*. This whole evolution can be presented schematically as follows:

Proto-Indo-European	Latin (centum language)	Lithuanian (satem language)
k', e.g., **k'ṃtom* 'hundred'	*c*, e.g., *centum* 'hundred'	*š*, e.g., *šimtas* 'hundred'
k, e.g., **kerpō* 'I cut'	*c*, e.g., *carpo* 'I rip'	*k*, e.g., *kerpù* 'I cut'
k$^{u}_{\sim}$, e.g., **k*$^{u}_{\sim}$*ot* 'what'	*q*, e.g., *quod* 'what'	*k*, e.g., *kàs* 'who', 'what'

The second change, comparable in extent, involves the more or less consistent transformation of the original *s* into *š* following *i*, *u*, *r*, *k*. Traces of this process have been noted in central languages—Indic, Iranian, Armenian, Albanian, Slavic and Baltic.

Only after the introduction of the last two innovations did the central dialects break up into four separate groups: the Indo-Iranian group in the east (later divided into two groups—Indic and Iranian), the Thraco-Armenian in the south (represented today by the Armenian language), the Daco-Mysian in the west (represented today by Albanian) as well as the Balto-Slavic group in the north (later divided into Baltic and Slavic). Thus, around the year 2000 B.C., the grouping of the Indo-European dialects was accomplished. This grouping is presented in the following diagram:

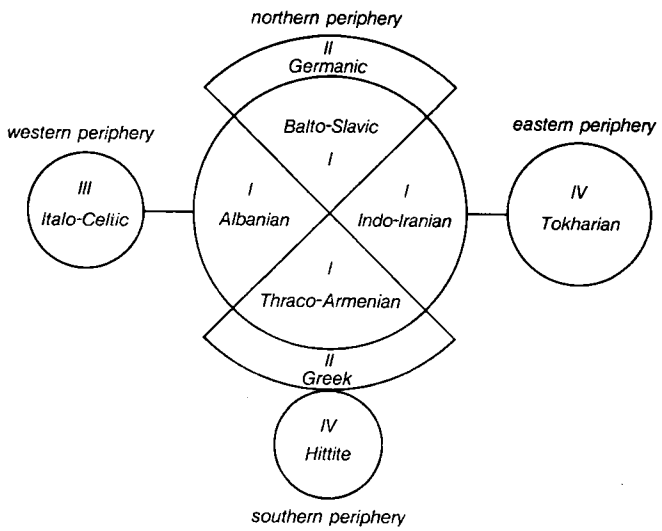


Fig. 14. Grouping of the Indo-European dialects *c.* 2000 B.C.

The Tokharian and Hittite groups, indicated on the diagram by the numeral IV, separated first and most completely from the center of the family; next come the

Italo-Celtic dialects (III); even closer to the center are Greek and the Germanic languages (II); the very nucleus is represented by the groups which were the last to become differentiated—Indo-Iranian, Thraco-Armenian, Albanian and Balto-Slavic.

In this way, the Indo-European family of languages broke up into distinct dialectical groups. Their further differentiation was a consequence of their enormous vitality, as a result of which they mixed with the dialects of conquered peoples, which brought with them deep and various changes in the structure of the language of the conquerors. The Indo-European expansion proceeded in four directions; east, south, west and north. Each of these directions of conquest will be discussed in turn.

The Tokhars, who occupied east Turkistan, represented the first eastward moving Indo-European wave. In the 7th century A.D., two Tokharian languages existed, known to us from translations from Buddist texts—Tokharian language A on the border of the Gobi Desert in Turfan and Tokharian B, more to the west, in the Kucha Oasis. Both of these languages died out in the 8th century, driven out of use by the Turkish dialect of the Uighurs.

Earlier, the Indo-Iranian dialects divided into two groups—Indic and Iranian. Around the 15th century B.C., the Indic group occupied the Indus basin, where, in the following centuries, the sacred poetry written in the Vedic language arose. Around the 5th century B.C., the Vedic language was transformed into classical Sanskrit, which constituted the literary language of India for two thousand years. Simultaneously, in the last centuries B.C., the Middle Indic languages—Pali and Prakrits—arose out of the Old Indic dialects, closely related to the Vedic languages. Around the 10th century A.D., they developed into the Modern Indic languages, the most important of which are Hindi, Urdu and Bengali. The influence of the Dravidian language substratum is clearly evident in the development of the Indic group.

In antiquity, Old Persian, spoken in South-Western Iran and known to us from inscriptions from the 6th and 5th centuries B.C., was part of the Iranian group. To the north of Old Persian, the Median language was spoken and, further to the north and east, dialects from which the Avestan language derived in the 7th century B.C. The Avestan texts, the sacred books of the Zarathustrans, were written in this language. Still further to the north, from Prut to East Turkistan, the Scythian-Sarmatian dialects were spoken. In the first millennium A.D., of greatest significance in Iran was the continuant of the Old Persian language—Middle Persian or Pehlevi—which, around the 9th century A.D., developed into Modern Persian. The Kurdish dialects in North-Western Iran are continuants of the old Median language, while the North-East part is occupied by the Afghan language (Pushtu). In the Scythian-Sarmatian zone, the most important languages were Sogdian in West Turkistan and Sakian in East Turkistan which, around the 9th century, were supplanted by the Turkic dialects. From this great zone, the Ossetic language in the Caucasus has persisted to the present, continuing the Sarmatian

dialects as well as Yagnobi, at the foot of the Pamir Mountains, deriving from dialects closely related to Sogdian.

The first Indo-European wave moving towards the south carried with it dialects of the Anatolic group, i.e., Luvian-Hittite, which, occupying Asia Minor around 2000 B.C., were strongly influenced by the Asianic language substratum. Hittite, the most important language of this group, was used in the Hittite state, which included the basin of the present Kazil Irmak, especially in the capital, Hatusas. Hittite is known from cuniform texts originating from the 14th and 13th centuries B.C. as well as from hieroglyphic texts primarily from the 14th to the 8th centuries B.C. We distinguish cuniform Hittite from hieroglyphic Hittite on the basis of certain details different in these two languages. In the southern part of Asia Minor, from Lycia to Cilitia, the Luvian language was spoken, while in the northern part, in Paphlagonia and Pontus, the Palaic language, related to Luvian and Hittite, was spoken. These languages died out in the first millennium B.C. Two closely related languages, Lycian and Milian, which derived from Luvian, persisted in the south-west extremes of Asia Minor up to the first centuries A.D. The Lidian language, spoken more to the north, was also closely related to these two languages and, like them, is known from texts written in the Greek alphabet from the 5th and 4th centuries B.C.

It is difficult to establish the place in the Indo-European family occupied by the language known as pre-Greek, or Pelasgian, which was spoken in Hellas in the first centuries of the second millennium B.C. Certain words from this language, borrowed by the Greeks, were preserved in the Greek language.

Greek tribes, coming from the north in two waves, superimposed themselves on the Pelasgian substratum in the first half of the second millennium B.C. The older wave, including the Achaean-Aeolic and Ionic-Attic dialects, dominated unchallenged in Greece in the middle of the second millennium. We are acquainted with the Old Achaean dialects from Mycenaean texts from the period between 1450 and 1200 B.C., which were preserved primarily on Crete and on the Peloponesus and were written in Linear B. The second Greek wave, in the 12th century B.C., introduced the Doric dialects, which occupied the Peloponesus, Crete and Rhodes and superseded the northern-west dialects as well. In the 9th century B.C., the Greek alphabet arose, based on the model of Phoenician. The dialectical division of Greece lasted up to the 1st century B.C. and left its mark on the forms of literary language preserved in masterpieces, beginning with the *Illiad* and *Odyssey* from the 8th century B.C. Not until the 3rd century B.C. did a universal Greek language take over—Koiné—a synthesis of Attic and Ionic elements. It drove the remaining dialects out of use and temporarily occupied Asia Minor. It is from this language that present-day Modern Greek was derived. To the north, the closely related Macedonian language bordered with the Greek dialects; it died out in the first centuries A.D.

The third wave arriving from the north, after Luvian-Hittite and Greek-Macedonian, was made up of predecessors of the Thraco-Armenian group. They origi-

nated on the northern coast of the Black Sea, which, until the 8th century B.C., was inhabited by the Cimmarians, as well as on the western coast. In the second millennium B.C., the Thracian dialects took over the territory between the Danube and the Aegean Sea and the Sea of Marmara. In the 13th and 12th centuries B.C., the related Phrygian dialects occupied a great part of Asia Minor. The Indo-European expansion proceeding from this area in the 6th century B.C. in the region of Lake Van, overlapped with Vannic dialects of the Japhetic group, thus creating the beginning of the Armenian language. The Thracian dialects died out around the 1st century A.D., as did the Phrygian dialects in the 5th century A.D. The Old Armenian language (Grabar) is known to us from Christian literature which developed from the beginning of the 5th century. The strongly differentiated Modern Armenian dialects became the basis for two literary languages. The West-Armenian language is used in Europe among emigrants, while East-Armenian is the official language of the Armenian Republic.

North of the Greek-Macedonian and Thracian zone, the dialects of two separate language groups existed in antiquity. Daco-Mysian and Illyrio-Messapian. The Dacian dialects were used in Dacia (present-day Rumania), and the Mysian dialects in the central part of the Balkan peninsula, while the Illyrian dialects commanded the territory of present-day Yugoslavia. The Messapians, linguistically related to Illyrians, crossed the Strait of Ostranto in the 10th century A.D., occupying Apulia and its neighboring territory. The Messapian language is known primarily from short inscriptions, around 270 of which have been preserved from the period between the 6th century B.C. and the 1st century A.D. They are written in a local alphabet of Greek origin. All of these dialects died out in the first centuries A.D. Only one of the Mysian dialects escaped romanization, having shifted from the center of the Balkan Peninsula toward its south-western parts; this dialect was transformed into present-day Albanian, used in Albania and neighboring countries. The huge majority of Albanian vocabulary is made up of words borrowed from Latin, Slavic, Turkish and Modern Greek. The literary tradition of this language reaches back to the 17th century.

We shall now proceed to Western Europe. Certain scholars suggest that, at the beginning of the second millennium B.C., the first Early Italic wave of Indo-European tribes came into Italy from the north. Their language, influenced by the pre-Indo-European substratum, underwent deep changes, just as did the dialects of the Luvian-Hittite group. In the historical period, the language of the Siculians in the eastern part of Sicily constituted a remnant of the Early Italic stratum, as did certain elements preserved in the vocabularies of the Volscians, Sabines, Etruscans, and Ligurians, and perhaps in Latin as well.

In the middle of the second millennium B.C., Italy was overrun by dialects of the Italic group. From the general group of Italo-Celtic tribes, the first to break off were the Latins, who crossed the Alps and the Apennines and settled on the lower Tiber. Of the Latin dialects used in the last centuries B.C., three are known: that of Faleria (Falerian language), that of Praeneste, and the clearly distinguishable

Latin of Rome. The second wave consisted of Osco-Umbrian tribes. From the 14th century B.C., the Oscian language dominated southern Italy, while central Italy—except for Latium—was overrun by the Sabellian dialects. The north-eastern part of the peninsula was occupied in the 12th century B.C. by the Umbrian language. The third Italic group was represented by the Venetic language which, from the 11th century B.C., was used north of the mouth of the Po River along the coast of the Adriatic to Tagliamento. In the first millennium B.C., three groups of Italic languages—Latin, Osco-Umbrian and Venetic, initially differing from one another—influenced by one another and by a common language substratum, became distinctly similar to one another, forming something like an Italic league. Between the 2nd century B.C., and the 1st century A.D., Latin, the original Latin dialect of Rome, took over the entire peninsula, supplanting the other Italic dialects. Between the 1st and the 5th centuries A.D., Latin occupied practically all of Western Europe and part of the Balkans. After the fall of the first Roman Empire, in the period between the 5th and the 10th centuries the Romance dialects, deriving from vulgar Latin, broke up into ten separate languages—Rumanian, Dalmatian, which died out in the 19th century, Italian, Sardinian Raeto-Romanic, in the Alps, French, Provençal, Catalonian, Spanish and Portuguese.

At the same time that the Italic tribes were pushing toward the south, the Celts were heading toward the west. The first tribes to break off from the larger group spoke Goidelic, and, around the 10th century B.C., occupied Great Britain, to be driven further on into Ireland around 400 B.C. by related tribes. In the 5th century A.D., the Goidelic peoples of Ireland occupied the Isle of Man and Scotland (where previously the non-Indo-European language of the Picts had been used), as a result of which three closely related Goidelic languages emerged—Irish, Manx and Scots. The further expansion of the English language, beginning in the 16th century, limited the Scots language to the Hebrides, while Manx disappeared from use in the middle of the 20th century. Irish, limited at the beginning of the 20th century to the western extremes of Ireland, is today gaining in significance owing to the support of the Irish nation.

The Celtic tribes which remained after the retreat of the Goidelic peoples in the last millennium B.C. occupied Great Britain, Gaul, the Po Valley, the land of the Czechs and part of the Danube Valley and reached across the Balkans to Asia Minor. The progress of Romanization and Hellenization, however, drove the Celts out of the European continent entirely in the 1st century A.D. Only in Great Britain was the Brythonic group of Celtic dialects preserved; the expansion of the Germanic peoples, however, from the 5th century A.D., broke these dialects up into three parts. Thus emerged the three Brythonic languages—Welsh, in Wales, Cornish, in Cornwall, which died out in the 18th century, and Breton, in Brittany.

At the northern extremes of the Indo-European territory we have the Germanic group, the source of which was Jutland, Southern Scandinavia, and the countries of the lower Elbe. In the first centuries A.D., the Germanic dialects broke up into

three groups: eastern, northern and western. From the eastern group we know only the language of the Goths, who, from Scandinavia, through the territory of present-day Poland and Ukraina, reached the northern coast of the Black Sea in the beginning of the 3rd century A.D. The Gothic translation of the Bible by Bishop Wulfila dates from the 4th century. In the 16th century, the Gothic language was still being used in the Crimea.

Between the 1st and 7th centuries A.D., the North-Germanic tribes spoke a proto-Nordic language, fairly homogeneous throughout Scandinavia. As a result of territorial expansion between the 8th and 10th centuries, the North-Germanic dialects began to differentiate so that, at present, this group includes five separate languages, three of which: Icelandic in Iceland, the language of the Shetland Islands and the language of the Norwegians (Landsmaal), very closely related to one another—constitute an archaic set. The Swedish language in Sweden and Finland holds an intermediary position. The Danish language in Jutland and the Danish Islands as well as its variant, used in Norway as a literary language, Riksmaal, are very highly developed.

In the 1st century B.C., the West-Germanic tribes, a linguistic unit, occupied Jutland and the territory between the Elbe and the Rhine rivers, from the southern part of which the southern tribes had only recently been driven out. On these newly occupied territories, important phonetic changes took place breaking the West-Germanic territory up into two parts: southern and northern. The further stages of development of the southern sub-group will be called High German (the Bavarian, Swabian, Alemannic and South-Frankonian dialects), while the old northern sub-group broke up, as a result of expansions in the 5th and 7th centuries A.D., into three languages: (1) Low German in northern Germany (Old Saxon, today—Platt-deutsch), in Holland (Dutch) and in northern Belgium (Flemish); (2) the Frisian language in northern Holland and, as a result of expansion to the east, on small, remotely situated areas as far as Schleswig; (3) the English language, derived from the Anglo-Saxon dialects in the 5th to the 11th centuries and which, in the 12th and following centuries, was deeply influenced by the French language. From the 5th century on, English continuously extended its domain on the British Isles.

East of the Germanic peoples, the territory of the Balto-Slavic group extended in the basin of the Odra, Vistula and Neman rivers. In the first millennium B.C., this group was divided into two related groups, Baltic and Slavic. Among all of the presently existing groups of this language family, the Baltic group is in many respects today the least changed in relation to the proto-Indo-European language. It includes the western set, composed of two extinct languages—Prussian, between the lower Vistula, the Pregel and the Baltic, and Yatving, on what is now northern Masovia and in Suwałki. The Yatving language died out in the 14th century, and Prussian in the 17th century. The eastern set of the Baltic group includes the Lithuanian and Latvian languages used in the two republics of the USSR. Their literary tradition reaches back to the 16th century. In the Middle Ages, on the territory of

what is now Kurland, the Kur language was used, intermediary between Lithuanian and Latvian.

The Slavic group, as a result of expansion occurring from the first centuries A.D., near the end of the first millennium, broke up into several groups. From the most homogeneous eastern group—i.e., Russic—three languages were derived in the period between the 13th and the 14th centuries: Russian (two groups of southern and northern dialects), Ukrainian and White Russian. The southern group, in the Balkans, broke up into two sets at an early period. In the east, the Bulgarian-Macedonian dialects were used, from which, in the 9th century—owing to the efforts of Constantin-Cyril and his brother, Methodius—the liturgical and, later, the literary language of Old Slavic arose as well as the later literary languages—Bulgarian (in the 19th century) and Macedonian (1946–1960). The second, south-west set includes the Serbo-Croatian language (three sets of dialects: Štokavian, Chakavian and Kaykavian) and the Slovene language at the foot of the Alps. At the end of the first millennium, the West-Slavic group broke up into two sets: Czechoslovakian and Lech, and the Lusatian dialects representing a position intermediary between them. In time, two languages emerged from the first of these sets—Czech and Slovak. The Upper Lusatian language is more closely related to Czech and Lower Lusatian, to Polish.

Three languages arose from the Lech set—Polabian, on the lower Elbe, which died out in the 18th century, Pomeranian, the last survivals of which today are the Kashubian dialects, and Polish.

Heretofore, our discussion has been in reference to the map representing the distribution of the languages of the world in the 15th century A.D. We shall now proceed to the map representing the present-day situation. In the course of the last 470 years, enormous changes of a completely new type have taken place. The old migration moving along ancient routes, primarily land routes, gave rise to cycles of languages in which one overlapped the other like the scales of a fish. New expansions, setting out from Europe and moving primarily along water routes, have given rise to large isolated language zones on distant continents.

The Romance languages were the first, chronologically, to spread. The expansion of Spanish and Portuguese to America began around 1500 and led to the formation of Latin America. Linguistically, Brazil is Portuguese except for the inaccessible parts of the Amazon Basin, in which American Indian languages are still spoken. The rest of South and Central America, together with Mexico, is the domain of Spanish, which is interrupted by mere islands of native languages—Guarani in Paraguay, Kichua in the Andes, Maya in Central America, Aztec in Mexico, and many others. New French language zones arose in Canada, on Haiti and in certain former French colonies in Africa.

Later, but even more impressive, has been the expansion of the Germanic languages. The Dutch language provided the foundation of Afrikaans, presently used in the Union of South Africa. The expansion of English, which began in the first years

of the 17th century, has extended over practically all of North America, Australia, New Zealand and the southern extremes of Africa, plus numerous smaller territories scattered all over the world.

Of the Slavic languages, only Russian has achieved a comparable expansion. It has extended along eastern land routes and today is used in the entire northern part of Asia which is included in the Soviet Union.

Note: The three stages of development of the Romance family. What constitutes the difference between a language cycle and a language league or family? Along what routes did the expansion of the language families of the older stratum (to 1500 A.D.) move; along what routes did those of the younger stratum move? Language strata in America before the coming of the white man. American-Asiatic language contacts. Language strata of Australia, Oceania and South-East Asia. Language strata of Africa and West Asia. History of the expansion of the Altaic, Uralic and Indo-European families.

CHAPTER 12. THE CAUSES OF LINGUISTIC CHANGE

Having presented the classification of languages—at least its general outlines—historical linguistics then proceeds to analyse the changes which led to the presently existing differentiation of languages and to determine the causes of these changes. As already stated in Chapter 9, all parts of the language system undergo change; keeping this fact in mind, we can distinguish five basic categories of linguistic change: phonological, morphological, lexical, syntactic and stylistic change. These will be discussed in turn.

The evolution of the phonological systems of languages is determined, on the one hand, by the structure and function of the systems themselves—which we call the phonological sphere of evolution—and, on the other hand, by factors functioning outside the system, i.e., the structure of the human speech organs, general psychological predispositions on the part of people and the influence of other phonological systems made possible by contacts with other social groups—all of these factors are included in what we call the phonetic sphere of language evolution. Because the relation between these two spheres—phonological and phonetic—is different in different epochs of the development of languages, depending on conditions not always clear to us, the general evolutionary trends of phonological systems may be only to a certain extent determined. The strict and concrete phonological laws with which we have hitherto been concerned—e.g., the fact that proto-Indo-European *e* changed in Indo-Iranian into *a*—compare Latin *que* (and), Avestan *ča*, Sanskrit *ca*—which occurred only in a particular period and on a particular linguistic zone. The general trends of evolution which constitute our present concern do not necessarily operate at all times and in all places; they do, in favorable conditions, occur with a certain degree of probability.

The phonetic sphere is the basic take-off point for linguistic change. The general psychological predisposition to execute all possible movements with the least possible

effort—the so-called tendency of economy of effort—produces numerous simplifications in phonological systems. Certain combinations of these phonological features are difficult for the human speech organs to articulate and are, therefore, less stable than other combinations and, consequently, easily undergo change. The influence of foreign languages often transforms phonological systems. The significance of the phonological sphere in the evolution of language should not, however, be underestimated. Language is a system of signs in the service of social communication, i.e., it fulfills a definite function. The phonological system fulfills the function of differentiating signs in a given language. The purposive and intelligent functioning of the whole society, although not carried out completely consciously, aims at guaranteeing the clarity and efficiency of this function of the phonological system. Each speaker aims at being intelligible, at producing words which are phonologically distinct and, at the same time, as far as possible, not allowing words to become identified, i.e. for new cases of homonymy to arise. This tendency is a consequence of the existence and functioning of the phonological system and is conditioned by its structure. Society guards the phonological system, constantly threatened by the phonetic sphere, and, even if it succumbs to influence from this sphere, it frequently modifies the effects of this influence. The character of this modification depends, of course, on the structure of the phonological system, particularly, on the empty slots occurring in the system.

Owing to this phonological tendency, phonetic processes undergo three basic changes: (1) The phonetic process may be delayed in order to avoid the identification of two phonemes or the excessive density of phonemes within a given narrow articulatory zone by the introduction of new phonemes to this zone. (2) A change of direction in the phonetic process may take place owing to which the phoneme undergoing transformation does not become identified with another phoneme, as would have threatened if the phonetic tendency had been realized, but instead, changes direction and fills in empty slots within the system—thus preserving its individuality and completing the system. (3) When, as a result of phonetic processes, new phonemes arise which threaten identification with old phonemes or excessive density in a given articulatory zone, a shift in the articulation of the old phonemes may take place, which guarantees against identification and excessive density.

Phonological changes, conditioned by the factors presented here, can be divided into two large categories—changes which are conditioned by, i.e., dependent on and those which are unconditioned by, i.e., independent of the phonetic environment in the word. Conditioned changes occur in certain positions and they are caused by the influence of the adjacent phoneme upon the phoneme which undergoes change. Conditioned changes may be divided into three sub-groups—the phenomena of assimilation, dissimilation and metathesis.

Assimilation involves decreasing the difference between two (directly or indirectly) adjacent phonemes by extending the proper articulation of one to include the other. We have regressive assimilation in which the preceding phoneme

is assimilated by the succeeding phoneme—e.g., *babka* ⩾ *bapka* (voiced *b* becomes voiceless *p* by being assimilated to the succeeding voiceless *k*); we also have progressive assimilation in which one phoneme is assimilated by the phoneme preceding it—e.g., *twój* ⩾ *tfuj* (voiced *w* becomes voiceless *f* by becoming assimilated by the preceding voiceless *t*).

Dissimilation is the opposite of assimilation and involves an increase in the difference between two (directly or indirectly) adjacent phonemes by changing the articulation of one of two phonemes sharing a common articulation. Thus, for example, the group *dl* changed into *gl* in the Lithuanian and Latvian languages, e.g., original **edlā*—compare Polish *jodla* (fir)—became Lithuanian *ẽglė* and Latvian *egle*. In this case, in place of the two apical consonants *d* and *l*, we obtain, by dissimilation, a group composed of dorsal *g* and apical *l*.

Metathesis, involves arranging two phonemes in the opposite order. Thus, for example, from the earlier **błcha* (compare Slavic *blcha*, Old Slavic *blъχa*, Lithuanian *blusà*), we have in present-day Polish—as a result of metathesis of the group *lch* into *chl*—the word *pchla* (flea).

Of the three types of conditioned changes presented here, the most frequent is assimilation, which also constitutes the starting point for the two remaining types of changes. Assimilation is the phenomenon most closely connected with the phonetic sphere. It is basically a result of the tendency to minimize effort which leads to abandoning a particular articulation in a phoneme which for some reason is weaker and substituting it by the articulation of an adjacent, stronger phoneme. Thus, for example, a consonant appearing between two vowels is naturally weaker, because it is alone, while the vowels surrounding it are stronger, because there are two of them. This results in the assimilation of the consonant by the vowels. If the consonant is voiceless, the influence of the vowels leads to voicing being imposed on the consonant, as a consequence of which the vibrating of the vocal chords is not interrupted between one vowel and the other, which decreases the effort of articulation. If, on the other hand, the consonant is voiced, the vowels, which are always more open than such a consonant, force it to be more open—which is also consistent with the law of minimum effort. These two basic phenomena occur in various form in different languages. In their typical form they occur, for example, in the western Romance languages. Thus, for example, in Spanish, the consonants *p*, *t*, *k*, when occupying a position between vowels, changed into *b*, *d*, *g*, which, in their further evolution, became voiced spirants—compare Spanish *lobo* with vulgar Latin *lupu* (wolf), *vida* with *vita* (life), *amigo* with *amicu* (friend), and French *louve* with Latin *lupa* (she-wolf).

The purely phonetic functioning of positional assimilation undergoes great modification owing to the functioning of the phonological sphere. The fact is evident in the process of monophthongization of diphthongs, which involves the assimilation of one member of a diphthong by the other. In the diphthongs *ei*, *ou*, the terminal elements *i* and *u*, precisely because as terminal elements they attract the attention of the speaker, were psychologically and phonetically stronger, which led to the

legressive assimilation of the first member by the second, *ei*, *ou* into $\bar{\imath}$, $\bar{u}$. The development of the majority of Indo-European languages proceeded in the same direction; when, however, the vowels $\bar{\imath}$ and $\bar{u}$ appeared in the phonological system of a given ranguage and the vowels $\bar{e}$ and $\bar{o}$ were lacking, the phonological sphere, in order to preserve the functional distinctiveness of the diphthong, evoked a change in the direction of assimilation from regressive to progressive such that the terminal members *i* and *u* were assimilated by the initial members *e* and *o*, as a result of which the diphthongs *ei* and *ou* changed into $\bar{e}$ and $\bar{o}$, filling in empty slots in the system.

A classical example of this type of phonological evolution is provided by the Celtic dialects. In the proto-Celtic language five short vowels occurred, *i*, *a*, *e*, *o*, *u*, and only three long vowels—$\bar{\imath}$ (from proto-Indo-European $\bar{\imath}$, $\bar{e}$), $\bar{a}$ (proto-Indo-European $\bar{a}$ and $\bar{o}$ in the interlude and onset) and $\bar{u}$ (from proto-Indo-European $\bar{u}$ and $\bar{o}$ in the coda). The proto-Celtic vowel system can be represented in the following manner:

$$\begin{array}{cccc} i & u & \bar{\imath} & \bar{u} \\ e & o & + & \\ & a & \bar{a} & \end{array}$$

This system contained, therefore, two empty slots corresponding to the long middle vowels $\bar{e}$ and $\bar{o}$. Because the regressive assimilation of the diphthongs *ei*, *ou* would have led, in Celtic, to the loss of their phonological distinctiveness, progressive assimilation took place under pressure from the phonological sphere, which led to the monophthongization of *ei* into $\bar{e}$ toward the end of the pre-Christian era, and of *ou* into long $\bar{o}$ at the beginning of the modern era. Thus, parenthesizing the new long vowels which, having arisen as the result of the monophthongization o diphthongs, filled in empty slots in the system. We may present the Celtic vowel system of the first centuries A.D., as follows:

$$\begin{array}{cccc} i & u & \bar{\imath} & \bar{u} \\ e & o & + (\bar{e}) & (\bar{o}) \\ & a & \bar{a} & \end{array}$$

In comparing the two Celtic systems, before and after monophthongization, it should be noted that the latter system has a more regular structure in that each short vowel has a corresponding long vowel.

A similar course of monophthongization of diphthongs can be noted in the Umbrian dialects and in the Armenian language. In all of these languages, *ei* changed into $\bar{e}$, and in Umbrian *ou* changed into $\bar{o}$, and the new vowels $\bar{e}$, $\bar{o}$ in all cases filled in empty slots in the system, preserving their distinctiveness. In all of these languages, just as in Celtic, the course of monophthongization was modified by the intervention of the phonological sphere.

The phonological sphere, however, does not only modify the process of assimilation. The intervention of the phonological sphere sometimes completely inhibits assimilation by introducing preventative dissimilation or metathesis. Preventative

dissimilation is, in fact, a consequence of intervention on the part of the phonological sphere, which in guarding the language against a threatening case of assimilation, leads to a break in the continuity of the articulatory movement, thus inhibiting assimilation. For example, the group *dl* was easily assimilated into *ll*, from which *l* was derived, as happened in the East-Slavic and South-Slavic languages, e.g., Old Slavic *ralo* (Polish *radło*). But such assimilation leads to the identification of certain forms which have hitherto been different, homonymy of entire words or particular morphemes. Thus, a subconscious disinclination to homonymy and the resulting possibility of misunderstanding produces in speakers a tendency to increase the differences between sounds, i.e., to the dissimilation of sounds—as occurred, for example, in the change from *dl* to *gl* in Lithuanian and Latvian.

Another type of preventative dissimilation involves the development of a phonetic element between two closely related phonemes, an element which separates them, thus inhibiting assimilation. A certain proto-Indo-European change may be cited as an example of this type of dissimilation: The original groups *tt*, *dd*, *ddh* in the proto-language had already changed on the boundary of the two phonemes into *tst*, *dzd*, *dzdh*, e.g., Hittite *atsteni* (you eat) from the original **at-te-* from the stem *ed-*, Greek *édomai*, Sanskrit *ádmi* (I eat). Here, *s* has been inserted between the two consonants *t* to avoid their fusing into a single *t*, which would obliterate the boundary between the morphemes.

Metathesis also results from the intervention of the phonological sphere aimed at guarding the system against the functioning of phonetic factors, above all, against assimilation. This explains the change from **błcha* into Polish *pchła*, cited above. In Polish, the voiced consonant *l*, in undergoing assimilation by the succeeding voiceless consonant, became voiceless and disappeared, e.g., *japko* from the earlier *jabłko*. Similarly, in the form **błcha*, the consonant *l* before voiceless *ch* should have become voiceless through assimilation and disappeared, in which case **błcha* should have produced **pcha*. Intervention on the part of the phonological sphere, however, prevented this change, and led to the placing of *l* at the end of the group and before the vowel, where there was no danger that it would become voiceless as a result of which **błcha* became *pchła*.

Changes which are not conditioned by the phonetic environment, to which we will now proceed, are in part the result of social transformation and the concomitant influence of one language upon another. With the expansion of a language into a neighboring territory hitherto occupied by another language, there arise in this newly conquered territory two linguistic strata: that of the older, conquered language, i.e., the linguistic substratum, and that of the in-coming, conquering language, i.e., the superstratum. In these conditions, the strata undergo mutual influence, the intensity of which depends upon various social factors. Substratum influence explains, among other things, the disappearance of closed *a* in literary Polish. In the past, in the eastern territories of the former Polish State, White Russian and Ukrainian constituted the linguistic substratum, while the superstratum was Polish.

The Russic languages had no closed *å* and, consequently, in becoming polonized, the Russic nobility introduced open *a* in place of closed *å* into their Polish. Because of the political predominance of these eastern territories, this pronunciation was extended in the first half of the 18th century to the literary language of the whole of Poland.

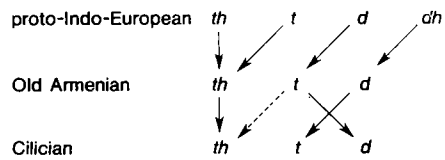
Just as in the phonetic functioning of assimilation, the phonetic functioning of the substratum evokes a reaction on the part of the phonological sphere of a language which leads to extensive modification of phonetic processes. A classical example is provided by the history of consonant shifts in various languages. The source of these changes was probably the influence of the substratum. When a language in which the opposition of voiced and voiceless consonants occurs, e.g., *p*:*b*, *t*:*d*, *k*:*g*, is superimposed upon a language lacking this opposition, the conquered populace, in adopting the language of the conquerors, do not adopt the voiced:voiceless opposition, but instead substitute it by an opposition familiar to them. As we see, the source of this phenomenon is the phonetic sphere—in this case, the difficulty which arises in languages, in which the voiced:voiceless opposition occurs, of coordinating vocal chord articulation with the articulation of stops. The phonological sphere immediately reacts, however, by substituting the voiced:voiceless opposition by some other phonological opposition.

In several Indo-European languages, the opposition of voiceless and voiced consonants *p*: *b*, *t*:*d*, *k*:*g*, was replaced by the strong: weak opposition *P*:*p*, *T*:*t*, *K*:*k* (capital letters indicate strong consonants). This is a completely natural phenomenon since voiceless consonants are normally strong and voiced consonants weak. In languages, however, lacking sound shifts, only voicing and its absence are diacritic features; strength and weakness, i.e., the difference between the greater or esser strength of expiration, are phonetic features only, of no significance in differentiating words. Here, the sound shift involves the fact that the hitherto diacritic opposition, voiced:voiceless ceases to be diacritic and what was hitherto a phonetic difference only, strong:weak, becomes a diacritic opposition; this happens because the former of these oppositions was unfamiliar to those adopting the new language, while the latter was familiar.

The change of the oppositions *p*:*b*, *t*:*d*, *k*:*g* into *P*:*p*, *T*:*t*, *K*:*k*—i.e., of the previously voiceless into strong and of voiced into weak, voiceless—is the first stage of sound shifts in Indo-European languages, which was preserved, e.g., in Hittite. In the further stage of development, strong consonants become aspirants, since the excess of expiration accompanying their articulation, escaping from the mouth after the initial explosion has taken place, creates an aspirant similar to the consonant *h*. Consequently, the opposition strong:weak is transformed into the opposition aspirated:unaspirated, and thus *P*:*p*, *T*:*t*, *K*:*k* changes into *ph*:*p*, *th*:*t*, *kh*:*k*. This stage of development was basically preserved in the Armenian language. In a still further stage of development, sounds aspirated by opening the aspiration constriction are transformed into spirants, as a consequence of which the aspirated:

unaspirated opposition becomes a spirant: stop opposition; in this case *ph*:*p*, *th*:*t*, *kh*:*k* becomes *f*:*p*, *þ*:*t*, *χ*:*k*. This stage was preserved in its basic form in the majority of the Germanic languages, e.g., Gothic.

It is characteristic that sound shifts are repeated in many languages. Sometime after the first shift, another shift follows. Such a course of development may be observed in the Germanic and Armenian languages. In Armenian, the first sound shift took place in the prehistoric era and was reflected in Old Armenian. The second shift, called Cilician after the name of the province Cilicia, in the dialects of which the shift occurred, was introduced towards the end of the first millennium A.D. The evolution of the system of stops from proto-Indo-European through Old Armenian to the Cilician dialects may be presented as follows:



As the result of a prehistoric shift, proto-Indo-European *t* changed into *th*, thus becoming identified with proto-Indo-European *th*; proto-Indo-European *d* changed into *t*, and proto-Indo-European *dh* into *d*. As a result of the Cilician shift, *d* changed regularly into *t*, while *t*, contrary to expectation, (dotted line), did not change into *th*, since this would have led to the identification of the Old Armenian series represented here by *th* and *t*. In this situation, intervention on the part of the phonological sphere occurred. Old Armenian *th* remained unchanged and *t* retained its separateness, becoming *d*.

Because of the coherence of the phonological system, one set of changes leads to another, i.e., they create conditions in which certain constant psychological tendencies evoke phoneme changes. Thus, the proto-forms of the present-day Polish words *babka*, *główka*, in proto-Slavic were pronounced **babъka*, **golvъka*. After the disappearance of yer (*ъ*), the groups *bk*, *vk* arose, which as a result of the psychological tendency to greater and greater assimilation of forms which are adjacent and similar, changed into *pk*, *fk*. The creation of the variant *f* in groups of the type *fk* made possible, in turn, the assimilation of the group *tw* by *tf*.

Here is another example. In proto-Slavic, before frontal vowels, consonants were pronounced softly, while they had a hard pronunciation before back vowels. Because, however, these features were conditioned by the succeeding vowel, they were non-diacritic, for they had no independent role in the differentiation and delimiting of words. In Polish, however, the soft terminal yers disappeared, leaving soft consonants in a pure coda—for example, proto-Slavic **gostь* ⩾ Polish **gość* ⩾ *gość*. In addition, the vowels *e*, *ě*, before the hard apical consonants (*t*, *d*, *s*, *z*, *n*, *r*, *l*), changed into *o*, *a*, while the preceding consonants remained soft, e.g., *biorę*, *biały* (from proto-Slavic **berą*, **bělъ*). Owing to these processes, softness in consonants

ceased to be determined by the sound of the succeeding vowel and, consequently, it became a diacritic feature, which led to the formation of a new series of soft consonant phonemes. This change was a simple phonological consequence of the preceding processes. Processes of this type, proceeding from one another, constitute part of the internal evolution of language.

Changes of the morphological and syntactic systems of language are, in part, a simple consequence of phonological changes. This fact is clearly evident in the phenomenon of morpheme boundary shifts, i.e., morphological absorption. Let's take, for example, the inflection of masculine stems ending in *-o* and compare their division into stem and ending in the proto-Indo-European era (just after the proto-language had broken up into dialects) with that of the Old Slavic language:

	proto-Indo-European	Old Slavic
Nominative sing.	**wl̥kᵘ̯o-s*	*vlъk-ъ*
Accusative sing.	**wl̥kᵘ̯o-m*	*vlъk-ъ*
Instrumental sing.	**wl̥kᵘ̯o-mi*	*vlъk-omь*
Dative pl.	**wl̥kᵘ̯o-mos*	*vlъk-omъ*

As long as the stem of the word ended, in the majority of case forms, in *-o*, the part of the word succeeding this vowel was clearly distinguishable as an ending; when, however, the *-o* ending of the stem underwent certain phonetic changes, as for example in the transformation of the previous endings *-o-s* and *-o-m* into *-ъ*, the phoneme *-o* ceased to be a distinct stem ending and in those forms in which the *o* persisted, it became a part of the ending, which in turn was extended to other stems, e.g., Polish *rękom* from *ręka* (stem in *-a*). In this way, the boundary between the stem and the ending shifted; in other words, the ending absorbed part of the stem. Certain scholars use the term perintegration in referring to morphological absorption.

Sometimes phonetic processes of the type described above lead to a total change in the syntactic system of a language. Thus, in the Romance and Celtic languages, as well as many others, all endings disappeared as a result of phonetic reduction and, consequently, prepositions and word order must be used in place of endings to express syntactic relationships.

Phonetic processes also create new proportions in a language which enable new forms to arise. Thus, in proto-Indo-European, nouns having a stem ending in *-o* (e.g., **wl̥kᵘ̯o-s*, **wl̥kᵘ̯o-m*) were clearly distinct from nouns with a stem ending in *-u* (e.g., **sūnu-s*, **sūnu-m*), such that it was impossible for one type to influence the other. As a consequence, however, of the phonetic evolution of the nominative and accusative, the singular forms of both paradigms became identified since both the endings *-os*, *-om* and *-us*, *-um* changed into *-ъ*, which later disappeared. This identity of the nominative and accusative singular forms of both paradigms created new proportions in the language the application of which produced new forms, e.g., *wilk*:*wilka* = *syn*:*syna* (in place of the previous *synu*) or *syn*:*synów* = *wilk*:*wilków* (in place of the previous *wilk*).

The question now arises as to why the new proportions and neologisms created on their basis prevail over the older forms of a language. This happens either because of the numerical predominance of the newly used proportions or because of their greater clarity. Thus, the form *syna* won out for the first of these reasons, for proportions of the type *wilk*:*wilka* were much more frequent in Polish than were proportions of the type *syn*:*synu*. Forms of the type *wilków*, however, won out through the function of the second of these causes. The genitive plural of the noun *wilk* was originally also the form *wilk*; the feature of this form, however, differentiating it from other forms of the declension—i.e., the ending *zero*—was not distinctly characteristic, since it also appears in the nominative singular—which is why it was substituted by the ending *-ów*, used only in the genitive plural and transferred from forms of the type *synów*, consistent with proportions like *syn*:*synów* = *wilk*:*wilków*.

We should also keep in mind the phenomenon of agglutination (from the Latin *agglūtino* 'I stick on to'), which involves the fusion of two words to form a single word. In the Old Polish of the 14th and 15th centuries, for example, a compound form of the past tense existed consisting of participles ending in *ł* plus an auxiliary form of the verb *jeśm* (present-day *jestem* 'I am'), e.g., *dał jeśm*, *dał jeś*. Because these two forms were always used together, the sense of their individuality was lost so that the auxiliary word lost its accent and became attached to the preceding form of the verb, as its ending, at the same time undergoing reduction—thence the forms *dałem*, *dałeś*.

Lexical, i.e., semantic, changes involve substituting the old value of a word by one of its textual meanings, which then becomes its new value (cf. Chapter 6). Consequently, these changes always represent either an extension, a narrowing or a transference of the earlier value. Such changes are possible primarily because of the lack of continuity in the transmission of language. The child who is learning how to speak is not presented with the lexical system in prepared form, but must infer its entirety from what he hears around him. In these conditions, it often happens that children attach a different value to a word than that which it has in the lexical system of the older generation. For if a certain word is often used by adults in one of its particular textual meanings, then it is this meaning which attracts the child's attention and becomes established in his memory as the lexical value of the word, while the previous value existing in the minds of the adults, disappears in the younger generation. Thus, for example, the French word *soûl* originally meant 'replete'. In time, it became applied to intoxicated persons in the sense 'replete with wine', and children, hearing this expression, directly associated the word *soûl* with an intoxicated person, attributing to the word the meaning 'intoxicated' or 'drunk'. In the same way, the Polish word *babka* originally meant 'mother's mother' and later, 'old woman' in general. It could be used ironically and metaphorically for designating a young girl (conventional Polish word for 'girl'—*dziewczyna*). In the meantime, this metaphorical meaning was impressed upon youth and the word *babka* came to mean girl in their language, e.g., *fajna babka*, 'a beautiful girl'.

If the transmission of language were not discontinuous from generation to generation, semantic shifts would not occur at all, just as they would not be possible if groups of words did not undergo dislocation. For as long as a certain word is associated with a certain group of forms, its lexical value, maintained and defined by the values of the other words of the group, maintains a certain constancy. If, however, as a result of phonetic changes or for some other reasons, the group of words is broken up, the various elements originally constituting this group, no longer supporting one another, easily undergo semantic changes. Thus, for example, in Latin, the word *captivus* was clearly associated with the verb *capere* 'to capture' and with the participle *captus* 'captured' and, consequently, its meaning, 'captive', had a guaranteed permanence. For various reasons, however, the group of words in question was broken up in the Romance languages and in these new conditions the continuant of the Latin *captivus*—French *chétif*—completely isolated, took on the new meaning, 'weak', 'cheap', 'poor'. Another example: the basic element of the Polish word *łabędź* (swan) is *łab-* from **alb-* (white), compare Latin *albus* (white). The original meaning of the word *łabędź* was 'white bird'; as a result of the fact, however, that the group of words containing the element **alb-* was broken up in the Slavic language zone, the word *łabędź* became isolated and today designates black swans as well.

The lack of continuity in linguistic tradition and the breaking up of word groups create the necessary conditions for semantic shifts, but do not cause them. Several basic causes may be given for this phenomenon. Certain changes, rather few in number, have a purely linguistic character. They are caused by the structure of commonly used sentences in which a given word plays a special role. For example, the Latin word *magis* (more), placed at the beginning of the sentence, as was the case in Latin, created the impression of a conjunction, which is why its continuant in French *mais* took on the meaning 'but'. The Polish word *chocia* was primarily a participle form of a verb in Old Slavic *χošto* '*chcę*' (I want) and, therefore, meant *chcąc* (wanting). In time, the words *chocia* and *chocia-ż* became concessive conjunctions. This type of shift involves impoverishment of the meaning content as a result of which the denoting word becomes a syntactic morpheme, a preposition or a conjunction.

Another type of semantic change is a consequence of changes in culture. A word is linked to a certain social function which may be fulfilled by various objects, and the word passes from one object to another fulfilling the same function. Thus, for example, in the 17th century *pióro* meant 'quill', which was at that time used for writing, while today it means 'pen', a completely different object, but serving the same function. Similarly, the Latin word *papyrus* from Greek *pápȳros*, referred to a certain water plant 'papyrus', as well as to the writing material made from this plant, while its Polish continuant *papier* today refers to a material made primarily from wood, but also used for writing. In these and many other cases the same word referred successively to various objects which in the course of centuries replaced one another in fulfilling the same function within society. Changes of this type are

rather frequent; the most numerous semantic shifts, however, are caused by the passing of words from the language of one social sphere to that of another.

National languages, like Polish or French, include elements of two different types, i.e., each of them includes a general language used by everyone, plus special languages—jargons of particular professions and social groups. The general language and social languages share a common phonological and syntactic system; they differ from one another primarily in vocabulary. Thus, one of the most important causes of meaning changes in words is the fact that they are constantly being borrowed from special languages by the general language and vice versa.

This leads to important consequences. When words pass from general language into jargon, their meaning becomes narrower and more precise, while it becomes more general when words pass from jargon into general language, with the accompanying impoverishment of the meaning content. The cause of this phenomenon is not difficult to determine. Jargons functioning within fairly rare and precisely defined situations limit the meaning of words to these narrow possibilities; general language, on the other hand, used in much more variagated situations, broadens the meaning of words. Here are a few examples. The meaning of the Latin words *ponere* (to put), *cubare* (to lay), *trahere* (to pull) and *mutare* (to change) underwent great contraction in the language of French peasants in the early Middle Ages as a result of their being limited to situations connected with agriculture and livestock raising. Thus, in the French language, their continuants took on new meanings: *pondre* (to lay eggs), *couver* (to brood, to hatch) *traire* (to milk) and *muer* (to moult). An opposite case is the Latin word *adripare* (to land), used by sailors, which, when borrowed by the general language, had its meaning broadened—thence the French *arriver* (to arrive). Similarly, in Polish, the word *osnowa* (warp) has been used among weavers to refer to parallel threads into which the weaving apparatus weaves the transverse woof, *wątek*. These words, transferred from the jargon of weaving into the general language took on a more general, abstract meaning, 'thread', e.g., thread of thought.

Cases exist of the successive contraction and broadening of a word's value, which results in a kind of metaphore. Thus, for example, the Old High German word *marahskalk* (stable boy, groom), passing in the early Middle Ages from the general language into the special language of the French royal court, took on the narrower meaning 'royal functionary in charge of horses—equerry'. This same word, next passing back into the general language broadened its meaning in another direction, becoming a high military title, French *maréchal*. This word, borrowed into Polish, was even further broadened to include civilian dignitaries, e.g., *marszałek sejmu* (Marshal of the Sejm (Parliament)).

Changes in the stylistic system of language are usually connected with lexical changes, with the social wanderings of words which to a great extent are evoked by affective factors. For example, words used in the administrative spheres of society are frequently adopted by the general language, primarily because of the prestige they possess as "elegant" words. Words from lower social spheres, on the other hand, are drawn

into general language owing to the freshness of their imaginative and affective associations. Such a word which has passed into a new environment is matched by a word previously established in that environment having the same lexical value and, therefore, constituting a synonym. One of these synonyms will either become unnecessary and disappear, or will change its meaning. In the latter case, the synonym originating in the language of "higher" social prestige takes on a more dignified value, while the word from the socially "lower" dialect, cast aside and degraded, takes on a vulgar or trivial connotation. In certain of the French dialects of Switzerland 'room' is referred to as *païlé*; since the introduction of the literary *chambre*, however, the word *païlé* took on the meaning of 'tiny attic'. Similarly, a Sabaudian uses the literary words *père* and *mère* in referring to his parents, reserving for animals the older dialectical words *pâre* and *mâre*.

The evolution of the function of the Polish word *pan* (mister) illustrates another type of change involving both the lexical and the stylistic systems of language. The word was derived by shortening the word *żupan*, borrowed from the Turkic languages. In the first millennium A.D., *żupan* $\geqslant$ *pan* referred to an official of the Avar khagan who collected tithes. After the fall of the Avar nation in the 8th century A.D., the word in question lost its original specific meaning and in time became the name of a municipale functionary, i.e., *kasztelan* (castellan), e.g., *pan krakowski* (Castellan of Kraków). Passing, in the 15th century, from the narrow language of administration into the general language, the word *pan* replaced the older *gospodzin*, taking on that word's principal meaning 'master of the house', 'owner', 'ruler'. Initially *pan* was used as a form of address only in reference to *Pan Bóg* (God) and *pan król* (king); while *ty* (you) was the form of address for all other persons. Only later did *pan* become the subject of courtesy forms containing the verb in the third person, e.g., *Czy pan spał*? (literally—Did the gentleman sleep?) instead of *Czy spałeś*? (Did you sleep?). As a form of courtesy belonging to the stylistic system of language, the word *pan* (sir, gentleman) has spread extensively so that today, everyone is a gentleman (*pan*).

Note: What is the difference between conditioned and unconditioned changes in language forms? What roles do the phonetic and phonological spheres play in the general evolution of language and, particularly, in the processes of assimilation, dissimilation and metathesis? The significance of linguistic substrata and superstrata for the course of unconditioned changes. What is involved in the internal evolution of language? What causes morphological absorption (perintegration)? The creation of new forms on the basis of new proportions. What is involved in agglutination? What are the causes of shifts in the semantic values of words?

PART IV

TYPOLOGICAL LINGUISTICS

CHAPTER 13. FOUNDATIONS OF LANGUAGE TYPOLOGY

As has been previously mentioned, (Chapter 2), the languages of the world constitute a group of codes having numerous common features owing primarily to the fact that all are phonetic, two-class, arbitrary semantic codes. These features distinguish language codes from all other codes. The proof that all language systems are essentially parallel and possess similar basic features fulfilling the same function is the fact that certain texts, for example, the Bible, have been translated into hundreds of languages. Thus, the same content is expressed by means of hundreds of different language systems—which proves that all of these systems are capable of fulfilling the same functions, primarily semantic, involving the presentation of certain objective phenomena. While, however, all languages fulfill the same functions, they do so by means of agents differring from one another to a great extent.

In the language systems of the world, we may distinguish primary features, i.e., those common to all languages, and secondary features, those characterizing certain languages only. Primary features have been described in Part II; secondary features constitute our present concern. Primary features constitute a kind of general framework model of language, a model which is not productive but which must be supplemented and elaborated by secondary features, which differ from language to language. Those agents which are partially different in different languages but which nevertheless, to a certain extent, fulfill the same function, are functionally identical—e.g., Polish *ja*, Latin *ego*, French *moi*, English *I*. Languages which make use of a great number of functionally identical secondary agents are categorized as the same type of languages in that they elaborate the general language model in similar ways.

The notion of the function of a particular element of language leads to permanent relationships between the position of this element and that of other elements of the text and system of the same language, as well as to the relationship between this element's frequency and that of other elements in the text and system. The task of descriptive linguistics ends at the moment when all the elements of the four component systems of a language have been determined, since this constitutes the

definition of all the relationships existing among the elements of a language—which constitutes the description of the language system, the determination of the language's structure. This marks the take-off point for typological investigations of languages. Once a series of languages described in a manner more or less compatible with the principles presented here is available, typological linguistics proceeds to compare these languages. As in every comparison, the first step involves finding elements of two languages which are entirely or at least partially identical typologically and distinguishing them from the elements which are basically different in the languages being investigated. The identity of the elements compared must of course be an dentity of essential features. What is essential in language is the functioning of its components, presented by descriptive linguistics. Typological linguistics, which constitutes the continuation of descriptive linguistics, must also be based on the concept of linguistic function. Typological comparison involves comparing the functions of language elements, their relation to other elements of a given system. The concept of function must form the basis for the concept of identity. The elements of two languages are typologically identical when they have—each within its own system—the same linguistic function, i.e., they occupy the same place in the text and in the system, and in both places demonstrate the same frequency. This must be further explained.

In the first place, typological identity must be distinguished from historical identity. Thus, for example, Armenian *t* is historically identical to Greek *d*, because both of these phonemes constitute the continuation of the same phoneme, proto-Indo-European *d*, as is evidenced by the constant correspondence of Armenian *t* and Greek *d* in words which are etymologically identical. For example, Armenian *tur* (gift) and Greek *dōron* from proto-Indo-European **dōrom*; Armenian aorist *etu* (I gave) and Greek *édōka* from proto-Indo-European **e-dō-*; Armenian *otn* (leg) and Greek accusative singular *póda* from proto-Indo-European **podm̥*, etc. Typologically however, Armenian *t* is identical to Greek *t*, because both of these phonemes have the same basic function, they occupy the same place in the system of dental stops in both languages and stand in basically the same relationship to other dental stops in their respective systems. This is a consequence of the fact that the system of dental stops of the Old Armenian language of the 5th century A.D. and Old Greek from the 5th century B.C., were very similar in structure—which can be schematically presented as follows:

t

d *th*

The voiceless and unaspirated phoneme *t*, in both systems remains, on the one hand, in opposition to voiced *d* and, on the other hand, to aspirated *th*. The fact that Armenian *t* was phonetically weak, while Greek *t* was phonetically strong is in this case of no significance, for weakness and strength were not phonological features of these languages and had no influence on the structure of their systems.

In order, however, for the principle of functional identity formulated above to be of significance in concrete investigations, it must be more precisely defined. First of all, in comparing the functions of two different languages, i.e., the internal relationships among the elements of these respective systems, we cannot take the entire language systems into consideration. Language systems which are identical constitute the same system and, therefore, are the concern of descriptive linguistics. Typological linguistics compares only different systems, the elements of which, contained in systems of different structures, naturally have different positions within these structures and, therefore, different functions. Thus, if we were to take the function of language elements in relation to entire systems, always different, we would never find functionally identical elements belonging to different languages—and typological linguistics would not exist. In comparing the function of elements belonging to different languages, we must consider this function as the relationship of the elements compared to only a certain part of their respective systems. To this end, we must compare only parts of language systems and the functions of certain elements within these parts.

The question now arises as to what part of the system has to be considered. Only elements possessing some common feature can be compared. Only those parts of systems the elements of which are characterized by certain distinctly common features can be compared. Thus, for example, we can compare the system of stops of two languages. All stops have a distinct common feature in the articulatory closure of the speech apparatus and in their acoustical momentality. Consequently, all the stops in a language, because they possess a distinct common feature, i.e., a basis for opposition, and create oppositions among themselves, standing in strictly defined relationships to one another, have an easily defined function. In these conditions we can compare the systems of stops of two languages, we can demonstrate the structural similarities and differences between them, thereby establishing the degree of functional identity in the stops of both languages.

Typological linguistics requires that we distinguish among the four components of the language system: the phonological, semantic, syntactic and stylistic systems. In each of these four component systems, still smaller sets of elements can be distinguished which constitute coherent sub-systems possessing distinct common features, such as stops. Only these sub-systems can be rigorously compared with one another in determining the degrees of functional identity of their elements.

As we have seen, historical linguistics classifies languages according to the degree to which they have a common genesis, and thus to the degree of their relationship (language families) or affinity (leagues and cycles). Typological linguistics, however, does not take the origin of languages and their elements into account, but considers only the existence at a given moment of time of similarities and differences among the agents by which languages realize the same functions. Thus, while historical linguistics divides languages into families, leagues and cycles, typological linguistics distinguishes types of languages according to the degree of identity of their secondary

features, according to the actually existing similarity among the linguistic agents of various languages. Historical linguistics establishes the extension of families and leagues, while typological linguistics establishes the extension of types distinguished on the basis of structural similarities. The zones circumscribed by these two branches of linguistics sometimes overlap; sometimes, however, they differ from one another completely. Thus, for example, the Indo-European family does not today constitute a typological unity—it includes languages possessing very different secondary features and making use of very different agents, despite the fact that all originated from the same proto-language. The same is true of many other language families. There is more congruence between the zones of language types and language leagues, for leagues arose from the phenomenon of many languages becoming closer to one another and the consequences of this phenomenon are still evident today. Even when the zones of language groups distinguished by historical and typological linguistics do coincide, each of these disciplines studies these groups from a completely different perspective. Historical linguistics is concerned with origins, while typological linguistics is concerned with similarities and differences. Typological linguistics maintains that every language function can be realized by an enormous number of various agents; it determines these agents, therefore, as well as their distribution in time and space.

Note: What is the difference between the primary and secondary features of languages? General model of language as a set of primary features. Languages which have a greater number of secondary features in common are considered to be of the same type. What does the typological identity of the elements of different languages involve?

Chapter 14. PHONOLOGICAL TYPOLOGY

As we know, the diacritic features of sounds, which in creating oppositions among one another differentiate words, constitute the basis of the phonological system. These oppositions, and the features which make them possible, are in every language in part primary, i.e., universal, and in part secondary, i.e., characteristic of a certain group of languages only. Primary oppositions are based on contrasts between those acoustical features which are clearest and easiest to perceive. Owing to this fact, they are the first to become established in the function of differentiating words in the initial period of individual language acquisition, i.e., in the speech of the child. On the other hand, they are the last to disappear in cases of gradual language deterioration in aphasia. Furthermore, they appear in all, or nearly all, of the languages of the world as phonological features serving to differentiate words.

The clearest and most easily perceivable difference between sounds is the difference between vowels—pronounced with the speech apparatus open—and consonants—the essence of which involves different types of closure. This difference is used in

all languages to differentiate words. Vowel systems are always based on the opposition of at least three phonemes. Normally, they are: frontal high *i*, back high *u* and low *a*. The striking differences among these vowels, which are the most distinct from one another, constitute distinctive features where ever they are found. Among consonants, we have oppositions between the nasal phonemes *m*, *n*, in which the air is expirated through the nose only, and oral consonants, in which the air is expirated through the mouth. Among oral consonants, differences in the degree of opening of the mouth constitute diacritic features. The opposition is established between the broadest of these consonants, fluids, spirants, which are narrower, and stops, which are the narrowest. The fluid consonant—articulated as a trilled *r* or a lateral *l*—does not constitute a phonological feature in many languages, in which it is neither trilled nor lateral, but constitutes the only fluid phoneme of the system and the sole diacritic features of which are oral and semi-open articulation. The same is true of the single spirant phoneme of many languages which most frequently, but not always, is *s*. The system of stops, however, almost always includes at least three phonemes—labial *p*, apical *t* and dorsal *k*. In these phonemes, differences in the point of articulation constitute the distinctive oppositions differentiating words.

The distinctive oppositions described here appear in nearly all the languages of the world, which is why they are called primary oppositions. The phonological systems of some languages which are composed almost exclusively of primary, universal oppositions and as a result include no more than twenty phonemes, are called impoverished systems. Such an impoverished system functions, for example, in the Aranda language used by the tribes inhabiting the very center of Australia. This system can be represented as follows:

		CONSONANTS			
VOWELS	SEMI-VOWELS	nasal		oral	
			fluids	affricates	stops
i u	*w*	*m*	*l r*	*č*	*p*
a	*j*	*n*			*t*
					k

Leaving aside the semi-vowels *w*, *j*, which can be defined as variants of the asyllabic vowels *u*, *i*, and leaving aside the fact that in place of a spirant we have the affricate *č*, only one secondary opposition appears in the Aranda system, i.e., the distinction between the fluid phonemes lateral *l* and trilled *r*.

Impoverished systems consisting of 13 to 20 phonemes function in the languages of Australia and Polynesia (e.g., the Hawaiian language has 13 phonemes) and are in contrast with rich systems which are not based on primary oppositions only but which also include a significant number of secondary oppositions, as a result of which the number of phonemes occurring in them is sometimes significantly higher. Certain limits do, however, exist.

As we know, particular phonemes are not directly connected with a designated meaning content, but serve to distinguish and delimit certain larger sets of phonemes

which are directly connected with meaning content. Thus, the smallest semantically indivisible set of phonemes permanently linked to a given function in language communication, is called a morpheme. Defined in this way, a morpheme is a universal linguistic phenomenon. Morphemes appear in all languages, but the average length of morphemes, i.e., the average number of phonemes of which the morphemes of a given language are composed, varies from language to language. This is related to the degree of richness of the phonological system. It can be demonstrated that the average morpheme length is inversely proportional to the number of phonemes in the phonological system of a given language; the greater the number of phonemes in the system, the smaller the average number of phonemes in the morphemes of that system, and vice versa. This is a consequence of the fact that, the more phonemes in the system, the more distinctive features each of them possesses. In Aranda, the phoneme *p* has only two distinctive features—oral as opposed to nasal closure (*m*), and labial as opposed to apical closure (*t*); in Old Greek, Armenian and many other languages, on the other hand, besides these two features, the phoneme *p* has two more distinctive features—voiceless as opposed to voiced (*b*) and unaspirated as opposed to aspirated (*ph*). It can be assumed, therefore, that the total number of distinctive features of the phonemes of which a morpheme of average length is composed, is similar in all languages. It approaches the minimum number of features necessary for distinguishing a given morpheme from all other morphemes of the system. If the number of morphemes is similar in the languages compared, the minimum number of necessary distinctive features, which constitute units in the system of language information, must also be similar.

Thus, to a certain extent, a given number of features may be realized by fewer phonemes, the more distinctive features each of these phonemes possesses. If there is a large number of phonemes in the system, each of these phonemes possesses a great number of distinctive features and, consequently, a small number of phonemes is sufficient for distinguishing a given morpheme from all others—and vice versa—if there are fewer phonemes in the system, each phoneme possesses fewer distinctive features and the average morpheme contains a larger number of phonemes. Two extreme types appear in the languages of the world. The first is characterized by systems containing a small number of phonemes and a large number of phonemes in the average morpheme. Characteristic examples are Aranda and Hawaiian, each having a system of 13 phonemes, while the average length of their morphemes is 4 phonemes.

Languages of the opposite type, characterized by systems containing a large number of phonemes and using a small average of phonemes in their morphemes appear in the western extremes of North America in the coastal areas of the Pacific Ocean as well as in the Caucasus, where the North-Caucasian languages—such as Lakh, Archi, the Adyghe-Qabardi dialects and the now extinct Ubyk language—exemplify this type. The systems of these languages contain 45 to 75 phonemes and the average length of morphemes fluctuates between 1.25 and 1.40 phonemes. The phonological

system of the Chinook language on the North American Pacific coast contains 45 phonemes, while the average length of its morphemes barely equals 1.25 phonemes. For purposes of illustration we will present a Chinook form fulfilling the function of predication: *a-m-L-a-x-cg-ā'm-x* (you were used to taking it from her): *a-* the exponent of the past perfective tense, *-m-* (you), *-L-* (a lateral spirant pronounced *tl* 'it'), *-a-* (her), *-x-* (a dorsal spirant similar to Polish *ch*) indicates that "it" refers to "her", *-cg-* (taking) is the root of the form, *-ā'm-* the exponent of mood, *-x-* indicates that the activity was habitual. In this form, 8 morphemes are formed by means of 10 phonemes!

Now, when we multiply the number of phonemes in the system of a given language (*S*) with the number of phonemes in the average morpheme of that language (*M*), the product is similar for different languages. Thus, for Aranda $13 \times 4 = 52$ and for Chinook $45 \times 1.25 = 56.25$. Because the number of distinctive features of each phoneme is proportional to the number of phonemes in the system, the magnitude *S*, designating this number, is also proportional to the number of distinctive features in each phoneme, for the more phonemes, the more distinctive features each of them possesses. Thus, the fact that *S* multiplied by *M* yields a similar product for different languages, proves that the number of distinctive features in morphemes is approximately the same in all languages, and that it is only their distribution among the particular phonemes of which morphemes are composed that differs.

Hitherto, we have been discussing two extreme types: the impoverished type containing 13 to 20 phonemes in the system and characterized by long morphemes (about 4 phonemes), and the overloaded type containing 45 to 75 phonemes in the system and characterized by very short morphemes (1 to 2 phonemes). Both of these extreme types are recessive. They occur in strictly circumscribed zones only and are slowly dying out, being replaced by expansive types averaging 20 to 45 phonemes in the system and characterized by medium morpheme length (averaging 2 to 3 phonemes). This intermediary type is supplanting other types because it saves effort on the part of the speaker and, therefore, is the most economic. The impoverished type produces excessively long morphemes and, consequently, long texts. The overloaded type makes use, in distinctive oppositions, of acoustical differences which are extremely minute and difficult to perceive. Both demand great effort, which is why they are giving way to the intermediary type, which does not share these disadvantages.

Primary universal oppositions exist between the three vowel phonemes *i*, *a*, *u*, and the 7 consonant phonemes *m*, *n*, *l* or *r*, *s*, *p*, *t*, *k*. In this primary system, composed of 10 phonemes, vowels constitute 30 percent of the system, and consonants, 70 percent. In all functioning language systems, besides primary oppositions, secondary oppositions occur, so that the number of phonemes is always greater than 10. Two types can however be distinguished. In certain languages, the majority of secondary oppositions are vowel oppositions, and in this case, the number of vowels significantly exceeds 30 percent of the total number of phonemes in the system. These systems which, relative to the primary system, are unbalanced on the side of vowels, i.e.,

which have more than 30 percent vowels and less than 70 percent consonants, are called vowel systems. All of these systems are elaborated by means of similar secondary oppositions, i.e., oppositions based on various degrees of opening of the oral cavity (*a*:*e*:*ẹ*:*i*), flat vs. round labial articulation, irrespective of the articulation of the tongue (*i*:*u*) as well as the opposition between purely oral and oral-nasal expiration (*e*:*ę*).

In consonant systems we have an analogous shift in the opposite direction. In this case, consonants represent more than 70 percent of the system and vowels less than 30 percent. The system is different than that of vowel systems in that among secondary consonant oppositions, we must distinguish between oppositions of expiratory phonemes and those of non-expiratory phonemes. The first are created by the expiration of air forced out of the lungs through the trachea, larynx and oral or nasal cavity. These consonants make use of this air stream in the production of sound. Phonemes of the second type, on the other hand, are not dependent upon expiration and, as a result, the air movement necessary for their production is achieved by other means. Thus, in the articulation of clicks, the air is drawn into the lungs from the oral cavity which is closed by means of two types of closure, one in the back part of the oral cavity and the second involving the lips. The lips are suddenly parted and the air is drawn into the vacuum which has been produced in the mouth—which produces the acoustical effect of a click. In other cases, the larynx is closed, and works like the piston of a pump which, rising, forces air out (ejective and glottalized phonemes) or, lowering, draws air in (injective phonemes).

Clicks are normal phonemes in the consonant groups of the Khoin family in southern Africa, while glottalized phonemes function in certain American Indian languages on the Pacific coast as well as in parts of Canada and Alaska (Kichua, Aymara, Maya, the Kwakiutl-Salish-Kwileut group, Tsimshian, the Na-Dene family, etc.) and also in the Japhetic languages in the Caucasus. Non-expiratory consonants are a recessive type, however, receding before the expanding type of expiratory consonants. In Africa, the transformation of clicks into expiratory stops can be observed. This evolution is a consequence of the tendency toward economy of effort in that non-expiratory phonemes demand additional energy in producing air movements, which is avoided in the articulation of expiratory phonemes.

French represents the vowel type of language; its phonological system can be represented as follows:

VOWELS			CONSONANTS			
orals	nasals	reduced	nasals	fluids	spirants	stops
i ü u			*m*		*f v*	*p b*
ẹ ọ̈ ọ			*n*	*l r*	*s z*	*t d*
e ö o	*ę ọ̈ ǫ*	*ə*	*ń*		*š ž*	
a	*ą*				*j*	*k g*

The phonological system of French contains 15 vowels, i.e., 45.5 percent of the total 33 phonemes, and 18 consonants, i.e., 54.5 percent. The consonant system is

relatively impoverished and, except for the use of oppositions involving voicing (*p*:*b*, etc.), does not greatly deviate from the primary system. The vowel system, on the other hand, is rich and highly elaborated. In French, the system of oral vowels includes three series: (1) frontal vowels—pronounced flat with the tongue advanced and the lips flattened (*a*, *e*, *ę*, *i*), (2) back vowels—pronounced round with the tongue shifted towards the back and the lips rounded (*o*, *ǫ*, *u*), (3) medium frontal vowels—pronounced round with the tongue advanced and the lips rounded (*ö*, *ǫ̈*, *ü*). Four degrees of opening of the mouth are of diacritic significance and, therefore, four degrees of full vocalism, i.e., volume (*a*:*o*:*ǫ*:*u*). The system of nasal vowels duplicates the system of oral vowels in two degrees of opening (broad *ą* and medium *ę*, *ǫ̈*,*ǫ*). Moreover, we have the independent reduced vowel *ə* (so-called silent *e*).

Polish is a typical consonant system. It can be represented schematically as follows:

VOWELS		CONSONANTS				
orals	nasals	nasals	fluids	spirants	affricates	stops
i u		*m*		*f v*		*p b*
e o	*ǫ*	*ḿ*		*f' v'*		*p' b'*
a		*n*	*l r*	*s z*	*c ʒ*	*t d*
		ń		*š ž*	*č ǯ*	
				ś ź	*ć ʒ'*	
				x		*k g*
						k' g'

In this table—as in the table representing the French system—potential phonemes (cf. Chapter 5) have been omitted, as well as *i̯*, *u̯* (written *j*, *ł*), which may be considered asyllabic variants of the phonemes *i*, *u*. We have, therefore, 39 phonemes, 6 of which are vowels, i.e., 15.4 percent and 33 of which are consonants, i.e., 84.6 percent of the total. This wealth of phonemes was achieved owing to the fact that a series of secondary sound features, based on oppositions that are less clear than primary conditions, were used phonologically to differentiate words. Among vowels, besides the extreme *i*—*a*—*u*, we have the medium *e*, *o*, and alongside oral *o*, nasal *ǫ*. But it is in the consonant system that the greatest wealth of phonologically utilized secondary features are found. Among the spirants, five points of articulation have a distinctive function and, consequently, there are phonemes produced by the following types of constriction: labial-dental (*f*), apico-dental (*s*), alveolar (*š*), domal (*ś*) and dorsal (*x*). The spirants *s*, *š*, *ś* correspond to the affricates *c*, *č*, *ć*. Moreover, the number of consonants was increased by the elaboration of two phonological categories—voicing (*p*:*b*, *p*':*b*', *t*:*d*, *k*:*g*, *k*' :*g*', *c*:*ʒ*, *č*:*ǯ*, *ć*:*ʒ́*, *f*:*v*, *f*':*v*', *s*:*z*, *š*:*ž*, *ś*:*ź*) and softness (*p*:*p*', *b*:*b*', *k*:*k*', *g*:*g*', *f*:*f*', *v*:*v*', *m*:*ḿ*, *n*:*ń*). It is mainly these two categories which determine the fact that Polish is a consonant language.

There is no particular value in directly comparing the phonological systems of Polish and French. In order to achieve results that are objective and valuable we must compare these languages by referring each to the primary system, which in this case constitutes the basis for comparison. Because the primary system contains

70 percent consonants and 30 percent vowels, Polish, which contains 84.6 percent consonants and 15.4 percent vowels deviates from the primary system in the direction of consonants by 14.6 percent (84.6−70 = 14.6), while French, containing 54.5 percent consonants and 54.4 percent vowels, deviates from the primary system in the direction of vowels by 15.5 percent (70−54.5 = 15.5). Polish, therefore, is an extreme consonant system to the same degree as French is an extreme vowel system, for both of these languages deviate from the primary system by about 15 percent, though each in the opposite direction.

The vowel type includes many sub-types distinguished by their utilization of various vowel oppositions, and the same is true, to an even greater extent, in consonant languages. French is one of the most extreme vowel types. In some Adyghe dialects, however, belonging to the North-Western Caucasian group, a more extreme consonant system than that of Polish functions. In this case, 67 of the 70 phonemes are consonants, i.e., 95.7 percent, and only three are vowels, i.e., 4.3 percent. The deviation in the direction of consonants equals 25.7 percent (95.7−70 = 25.7). This is probably the most extremely consonant system in existence.

The distribution of the various phonological types in all the regions of the world has not yet been determined. For the present, we can only say that the languages of Australian and Polynesian natives represent impoverished systems, close to primary systems. The languages of the Indians of South America east of the Andes represent medium rich and decidedly vowel systems (20 to 35 phonemes). Fairly rich and decidedly consonant systems (27 to 48 phonemes) are found in languages of American Indians of the Pacific Coaste. Extremely rich and extremely consonant systems (47 to 70 phonemes) are characteristic of the North Caucasian languages. The present-day Indo-European languages of Europe have medium rich systems (25 to 40 phonemes). The languages of Western Europe (French, English, German) are vowel systems, while the Slavic languages are consonant systems.

The vowel and consonant features which we have been describing constitute the distinctive features of phonemes; prosodic features, however, which we will now discuss, involve the features of syllables and sets of syllables which are of significance in differentiating and delimiting morphemes and words. The question of syllables must be analysed in three aspects—phonetic, psychological and phonological, i.e., functional. A syllable is a phonetic reality. Speech is produced through expiration which increases and decreases rhythmically; in order to utilize this expiration economically in the flow of speech, sounds of greater acoustical volume alternate with those of lesser acoustical volume, as a result of which the acoustical volume of the flow of speech constitutes a wavy line corresponding to a certain extent to the line representing intensity of expiration. One section in the flow of speech, composed of one syllabic sound of greater volume, bounded by two sections which are acoustically weaker—constitutes a syllable. A syllable can be represented diagrammatically in the following way:

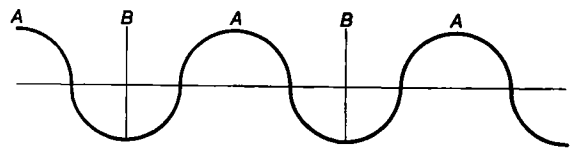


Fig. 15. Course of syllables (A = syllabic sound, B = syllable boundary)

The sounds of speech can be divided, on the basis of their function in the syllable, into three groups: (1) broad vowel (*a*) and medium vowels (*e*, *o*) are always syllabic—e.g., *ma-tko* or *mat-ko*, *mat-ce*, (2) stops, affricates and spirants are always asyllabic, (3) sounds of intermediary volume, i.e., closed vowels and broad, semi-open consonants, may be either syllabic or asyllabic. In Polish, semi-open *r*, *l*, *m*, *n*, are always asyllabic; narrow vowels, on the other hand, fulfill either a syllabic function, as *i*, *u*, or an asyllabic function, as *i̯*, *u̯* (written *j*, *ł*), depending on their phonetic environment, e.g., *wi-sus*, *u-szy* (*i*, *u* in the phonetic environment of consonants or between the consonant and the boundary of a word) versus *ja-ma*, *ład-ny* (*i̯*, *u̯* at the onset of a word before a vowel). Old Slavic included the syllabic *i*, *u*, *r̥*, *l̥* and the asyllabic *i̯*, *r*, *l*. Some languages include even syllabic (sonantic) *m̥*, *n̥*.

The basic problem concerning the structure of the syllable is that of its syllabic elements. In some languages, only vowels constitute syllabic elements, while in others the syllabic sonants *r̥*, *l̥* or even *m̥*, *n̥* can be syllabic elements. An analysis of the distribution of asyllabic elements leads to the problem of the structure of the syllable. If V (vocalis) signifies a syllabic section and C (consonans) an asyllabic section, we obtain the following types of syllables: V, CV—open syllables—and VC, CVC—closed syllables. In certain languages, rather few in number, like Old Slavic and Japanese, only open syllables occur; in other languages both open and closed syllables are found. In both open and closed syllables, the asyllabic section C may be composed either of a single phoneme or of many phonemes—which influences the number of phonemes in the syllable. The average number of phonemes in the syllables of a given language is an index of the development of the asyllabic sections of the syllables.

The regular wave-like nature of the flow of speech resulting from the succession of acoustically more intense (syllabic) and less intense (asyllabic) elements, creates the impression of rhythm. One beat of this rhythm constitutes a psychological syllable. It is a universal human phenomenon and is perceived by the speakers of all languages. A constant syllabic rhythm, because it is constant, is of no semantic significance; acoustical features which interfere with this rhythm, however, may have a semantic significance. This interruption of rhythm—exploited phonologically in differentiating and delimiting words—constitutes the prosodic aspect of syllables and their sets, primarily quantity and accent.

The category of quantity involves the phonological exploitation of differences in the duration of syllables, while the category of accent is based on the exploitation of differences in intensity. Where intensity involves the strength of expiration, the accent is dynamic; where intensity involves pitch (i.e., the number of vibrations

in a given unit of time), the accent is tonic. If not only the pitch within the syllable but also the course of this pitch is of functional significance in tonic accent—then the language in question includes the category of intonation. Many languages, (e.g., Serbo-Croatian), utilize the opposition between falling intonation—in which the pitch falls within the syllable—and rising intonation. Because prosodic features are connected with syllabic rhythm, the monotony of which they interrupt or differentiate, they refer to the entire group of syllables, for their character is determined by differences in quantity or intensity (in accent) between one syllable and another, adjacent to it, or by differences among the parts of a syllable in the course of intonation. Prosodic features cannot be defined in absolute magnitudes but only relatively to other syllables.

Languages can be divided into two groups—prosodic and non-prosodic—on the basis of their phonological exploitation of prosodic features. The former of these categories includes those languages in which the prosodic features function to differentiate words. Latin, for example, is a prosodic language containing forms which are differentiated from one another on the sole basis of the quantity of a single syllable, e.g., *venit* (he is coming) : *vēnit* (he came). Russian is also a prosodic language in which we find words differentiated solely by the point of accent, e.g., *múka* (torture) : *mukά* (flour). Non-prosodic languages, on the other hand, are those in which prosodic features are not distinctive but serve only to delimit words (moderately non-prosodic languages), or serve no function at all (extremely non-prosodic languages). Polish is a moderately non-prosodic language in which the accent on the penultimate syllable is the principal signal separating words from one another. The Turkic languages, on the other hand, are extremely non-prosodic languages in which the phonological signal delimiting words is not accent, but vocal harmony, which will be discussed in Chapter 16.

Prosodic languages can be divided into two categories—mora and non-mora—on the basis of the degree to which syllables are phonologically exploited. In non-mora languages, only the syllabic phoneme constitutes a prosodic element of the syllable, all other parts of the syllable being non-prosodic. In mora languages, on the other hand, not only the syllabic elements are of prosodic significance but also—in certain cases at least—the boundaries of the syllable. Prosodic but non-mora languages, such as Czech, are characterized by long syllables containing long vowels and short syllables containing short vowels. In mora languages, on the other hand, such as Latin, short syllables contain short vowels which are directly followed by the syllable boundary, e.g., *da-re* (to give) (both syllables short), while long syllables are of two types—i.e., they are long either intrinsically or because of their position. An intrinsically long syllable contains a long vowel, e.g., *rē-gī* (to the king) (both syllables intrinsically long), while syllables which are long because of their position are those in which the vowel is followed by some phoneme belonging to the same syllable, e.g., *al-tē* (high) (the first syllable contains a short vowel but is long because of its position; the second syllable is intrinsically long). In this way, mora languages embody a prosodic equality between the duration of one long vowel and two short

vowels. This fact is of important consequence, for while the prosodic unit in non-mora languages is the syllable, in mora languages it is the mora, i.e., the duration of a short syllable measured in psychological time; a long syllable equals two moras, i.e., two short syllables.

The following types of languages can be distinguished in respect to their phonological exploitation of syllables: (1) prosodic mora languages (Latin, Greek) exploit both the syllabic elements of syllables and syllable boundaries in distinguishing and deliminating words, (2) prosodic non-mora languages (Czech, Russian) are structured as the above languages except that they do not exploit syllable boundaries phonologically, (3) moderately non-prosodic languages (Polish) which exploit the features of syllables only for delimiting words, (4) extremely non-prosodic languages (Turkic) which do not exploit syllables phonologically at all.

Note: What distinctive oppositions are primary, i.e., occur in all the languages of the world? Why is there an inverse proportion between the number of phonemes in a system and the average number of phonemes in the morphemes of that system? What are the differences between impoverished, intermediary and overloaded phonological systems, and between vowel systems and consonant systems? Describe the syllable in its phonetic, psychological and phonological aspects. What are prosodic features, and what different types exist? What categories of languages can be distinguished in respect to the degree of phonological exploitation of syllables?

CHAPTER 15. SEMANTIC TYPOLOGY

The semantic systems of all the languages of the world—as we have already die cussed—include three distinct component systems: a system of denoting words, i.e., the lexical system, a system of pronouns and a system of numerals (cf. Chapter 6). These component systems will be discussed in turn.

Lexical systems are composed of many levels of words in such a way that, the higher the level, the smaller the number of words and the broader their usage. Words of the lowest level are directly associated with representations of various classes of phenomena, e.g., *table, chair, bench.* Each class of representations constitutes the value of a given denoting word. These are the words having the narrowest usage, the most concrete words. The next level is composed of words having a broader usage, words which are directly associated, not with the representations of phenomena, but with the most concrete words, and which are associated with phenomena only through the mediation of these concrete words. These are abstract words. The value of each abstract word is a certain group of concrete words together with the representations associated with them. A group of words of the first degree of abstraction therefore, has an established structure. Such a group includes one abstract word, superordinate, as well as a certain number of concrete words the usages of which, taken together, correspond to the usage of the superordinate word. The abstract word *furniture,* for example, constitutes such a group together with the subordinate, concrete words *table, stool, chair, armchair, bench, bed, bureau,* etc.

Abstract words of the first degree are subordinate to abstract words of the second degree, and so forth, all the way up to abstract words of the highest degree, such as *being*.

The difference between abstract and concrete words is reflected in their frequency in texts. Studies carried out recently have led to the establishment of a statistic law governing the structure of texts—called, after its discoverers, the Estoup-Zipf Law. This law maintains that, if we calculate the frequency of each word in a text, thereby obtaining a list of words in the order of decreasing frequency, the place number of each successive word on the list (word usage = u) multiplied by its frequency in the text (word frequency = f) is a constant, the magnitude of which depends on the length of the text. In a list established by Eldrige and including 43,000 of the word forms contained in the texts of American newspapers, the word occupying place number 500 in the list is repeated 10 times—according to the formula—$500(u) \times \times 10(f) = 5000$—while the word located at place number 1000 on the list is repeated 5 times, which also yields—$1000 \times 5 = 5000$.

The amount of selective information is inversely proportional to the frequency of the sign. The more frequently the word is repeated, the easier it is to predict its occurrence and, therefore, the less informative it is. Statistical analysis of a text enables us to establish the hierarchy of words from the least frequent, and therefore the richest in information, to the most frequent, the least informative. Data concerning contemporary French provides a good example. It had been determined that in French, the first, i.e., the most frequent 100 words represent 60 percent of the text, the first 1000 words 85 percent, the first 4000 words 97.5 percent and the remaining (i.e., from 4000 to 50,000 words) barely 2.5 percent. These facts prove that French words are divided into classes hierarchically arranged in respect to frequency.

This hierarchy is parallel to the previously established hierarchy of words determined according to the range of their usage from the most concrete, having the narrowest usage, to the most abstract, having the broadest usage. More abstract words are repeated more frequently precisely because of the fact that they can refer to a greater number of situations, which means that they are less informative. More concrete words, on the other hand, are less frequent; they appear in more specific contexts and, therefore, are richer in information. All of the facts seem to indicate that words are ordered in our minds according to frequency, and thus according to their degree of abstractness, from the most frequent and the most general to the least frequent and the most particular. The frequency and usage of a word are proportionally related, for the greater the frequency the broader the usage and vice versa, such that frequency serves as a measure of the word's degree of abstractness. Frequency determines the selective meaning of a word, the identification of the acoustical stimulus heard with the previously memorized form of the word, for these memorized forms are ordered in our minds according to frequency. We search for them in our memory in this order, beginning with the most frequent. The usage of a word often determines its semantic value, its relation to reality.

As we see, the lexical systems of the languages of the world are composed of series of word groups. Concrete words constitute the base of these groups, a base superordinate to which exist hierarchically arranged abstract words, increasingly broad in usage. The structure of these word groups is different in different languages, consider, for example, the word group equivalent in usage to the Polish word *rodzeństwo*, German *Geschwister*, English *siblings*:

	Hungarian	German	Polish	Malay
older brother	*bátya*	*Bruder*	*brat*	*sudarā*
younger brother	*öccs*			
older sister	*néne*	*Schwester*	*siostra*	
younger sister	*húg*			

As we see, in Hungarian, this group is broken up into four classes connected with four different words; in German and Polish the group is broken up into two identical classes; in Malay the group is not broken up at all. We can say, therefore, that in this respect, Polish and German are of the same type, which is in contrast to both the Hungarian and the Malay types.

By comparing the languages of the world in this way, the general similarities and differences among the lexical systems can be determined. At the same time it appears that these systems differ from one another in the degree of elaboration of their multi-level structure such that, in recessive systems, i.e., those which are dying out, there is less elaboration in comparison to the expansive types, which are gradually gaining predominance and in which this elaboration is continuously greater. New words constantly arise which occupy ever higher levels in the semantic hierarchy, they encompass greater and greater numbers of subordinate words. Thus, e.g., in the speech of Polish country people the word *zwierzę* (animal) means more or less what in literary language is referred to as *ssak* (mammal); this country speech, however, contains no word which could include the entire animal kingdom. This speech contains only groups of words of the first degree of abstraction: *zwierzę* (including the words *wilk* (wolf), *lis* (fox), *zając* (hare), *sarna* (deer), etc.), *ptak* (bird) (including *wrona* (crow), *wróbel* (sparrow), *jastrząb* (hawk), etc.), *ryba* (fish) and so on; but there is no group superordinate to these words. In Polish literary language such groups were created. The concrete words remained, but in place of the word *zwierzę*, the word *ssak* (mammal) was substituted. Next, a second degree group was created in which the words *ssak* (mammal), *ptak* (bird), *ryba* (fish), *gad* (reptile) and *płaz* (amphibian) were encompassed by an abstract word of the second degree—*kręgowiec* (vertebrate). The word *zwierzę* was placed above *kręgowiec* (vertebrate) as an abstract word of the third degree (i.e., *zwierzę* = animal). Above *zwierzę*, an even more abstract word was introduced—*organizm* (organism)—and still higher—*byt* (being).

The majority of languages in the world have undergone and are undergoing analogous transformations in their lexical systems. The semantic hierarchy of words in these languages becomes extended, the number of levels of increasingly abstract words increases to the most abstract, while the percent of concrete words decreases with the increase of the percent of general words, which occupy the increasingly greater part of texts composed in these languages. The furthest development in this direction has occurred in the lexical systems of the literary languages of Europe, both ancient (Greek, Latin) and modern (Italian, French, English, German, Polish, Russian, etc.). They represent the expansive, disseminating type. At the opposite extreme, we have the recessive lexical systems of peoples of low and medium levels of culture, in which a significantly larger percent of words are concrete.

Thus, for example, in the Hausa language in Central Sudan 311 equivalents for our word "big" have been recorded. They can be divided into several groups depending on which of the following meanings is used: (1) general, (2) for people and animals, (3) for young women, (4) for domestic animals, (5) for round objects, (6) for thick, long objects, (7) for thin, long objects (8) for blocks and fragments (9) for extensive spaces. The Bushmen of the southern extremes of Africa, a people of extremely primitive culture, use very few—about 150—words referring to general life, the rest of their vocabulary being composed of numerous names for the concrete phenomena occurring on their native savannah. In classical Arabic, the basis of which was the dialect of nomadic herdsmen, around 1000 words exist corresponding to our word "camel", used to refer to certain differences in size, coloring, race, etc. in these animals. Between these extreme types numerous intermediary degrees, of course, exist. The expansive type, rich in abstraction, is a superior instrument for formulating sentences having a general meaning content, which is why this type predominates.

Recessive and expansive features are also evident in pronoun systems, especially in the third person form, which refers to everything existing apart from the sender and receiver. In the recessive type, such as that which functioned in Latin, there are three third person pronouns, referring to that which is in the vicinity of the sender (Latin *hic*), in the vicinity of the receiver (Latin *iste*), and that which is at a distance from both sender and receiver (Latin *ille*). In proto-Slavic, *sъ* corresponded to the first of these pronouns, *tъ* to the second, and *ovъ* to the third. In the expansive type—which won out in the Romance languages which derived from Latin, and in the modern Slavic languages—only two pronouns function, which indicate objects which are nearer to and further from the sender, e.g., French *celui-ci* and *celui-là*, which correspond to Polish *ten* and *tamten* (this, that). The expansive type, consisting of two members, is simpler and easier to adapt, which gives it precedence over the more complicated and, therefore, more difficult recessive type, consisting of three members.

Typological differentiation also appears in the first person plural of pronouns. In the recessive type—which still functions in many American Indian languages, in the languages of the Manchu-Tungus and Austronesian groups, in many Austro-Asiatic

and Papuan dialects, in the dialects of the Paleo-Australian family, and in Nama, the language of the Hottentots—two nouns occur which correspond functionally to English *we*: (1) inclusivus, designating "I and you (or you plural)"—e.g., in the Ojibwa language of the North American Algonquian family we have *kīnuwi*—and (2) exclusivus, designating "I and this person (or these persons)", excluding the receiver (e.g., in Ojibwa—*nīnuwi*). In the expansive type, functioning in all other languages and gradually gaining predominance, we have only one pronoun, e.g., Polish *my*, Latin *nos* ("we" includes the function of both the inclusivus and the exclusivus, designating "I and you (or you plural) and this person (or these persons)"). The expansive type, having a single member which is simpler and makes use of a more general instrument for referring, is superseding the recessive type, which has two members and is more complicated and detailed.

The third part of the pronoun system in which the contrast between recessive and expansive features is evident concerns reference within a text, reference independent of the concrete consituation of the act of speech. In recessive systems, this referring is not highly developed, and pronouns function primarily within the consituation. The expansive type is characterized by highly developed systems of contextual reference. A pronoun which refers to something too far away and, therefore, imperceivable, at the same time fulfills the function of referring back along the line of text to a word that has already been stated. This is the so-called anaphoric pronoun—Polish *on*, Latin *is*, English *he*, etc. Thus, for example, in a text composed of two sentences—*Mój ojciec był prawdomówny. On nigdy nie kłamał.* (My father was truthful. He never lied)—the anaphoric pronoun *on* (he) refers to the noun *ojciec* (father). While the anaphoric pronoun refers back, the interrogative pronoun, Polish *kto*? *co*?, Latin *quis*? *quid*?, English *who*? *what*?, refers ahead in the text to something which is yet to be stated, i.e., to the reply. This expected meaning content is undefined, which explains why originally only a single common category of interrogative-indefinite pronouns existed. In time, only the expansive type of languages created separate interrogative pronouns (Polish *kto*? *co*?, English *who*? *what*?) and indefinite pronouns (Polish *ktoś*, *coś*, English *someone*, *something*), always, however, of common origin.

The further evolution of the expansive type led to the creation of relative pronouns—e.g., Polish *który*, English *which*, *who*. The relative pronoun exists in only some of the languages of the world. The recessive type, which lacks the relative pronoun, includes the American languages, the Uralic and Altaic languages in an earlier stage of their development, the proto-Indo-European language, etc. The relative pronoun refers simultaneously back in the text to one of the words in the subordinating clause and forward in the text to the subordinate clause of which it is a member—e.g., *Mickiewicz, który był największym poetą polskim, urodził się na Litwie.* (Mickiewicz, who was the greatest Polish poet, was born in Lithuania.) Here, the pronoun *który* (who) refers to *Mickiewicz* and at the same time constitutes the subject of the subordinate clause which it introduces. Thus, the relative pronoun

combines the functions of the anaphoric and the interrogative pronoun, and whenever it occurs, it arose either from the anaphoric pronoun, e.g., Old Slavic *iże*, Old Polish *jenże*, or from the interrogative pronoun, e.g., Latin *qui*, Polish *który*. This entire development, which was accomplished in the expansive languages, was a consequence of the growing needs of the syntactic system of language. Pronouns in part took over the function of connecting sentences which, in languages of the recessive type, is fulfilled exclusively by intonation.

Numerals, to which we will now proceed, constitute in all the languages of the world series of forms associated with one another in the memories of speakers; but the structure of these series varies from language to language. This structure depends upon the number of simple numerals constituting the first section of the series. The basis of this series is the magnitude corresponding to the final numeral of the simple series. Further simple numerals of this system constitute its multiples (ten... hundred... thousand... million..., etc.). By combining simple numerals, we get compound numerals, which occupy the remaining places in the series. We distinguish types of numeral systems on the basis of the magnitude of the final numeral in a simple series.

Among numeral systems, the most recessive is the binary system in which only two simple numerals occur—1 and 2; $3 = 2+1$; $4 = 2+2$. Thus, for example, in Australian dialects in the vicinity of Port Darwin the numeral '1' is *kalaguk* and the numeral '2' is *kalajilik*, the numeral '3' is represented by the compound *kalajilik kalaguk* (i.e., $2+1$) and the numeral '4' by the compound *kalajilik kalajilik* ($2+2$). The numeral four is the last one in the series; further, we have only 'many' and 'very many'. Binary or dual systems are connected with counting on the paired organs of the body such as ears, eyes, arms and legs, and are characteristic of the languages of Australia, the American languages of Tierra del Fuego and the Amazon Basin and San, the language of the Bushmen. The most recessive type, characterizing the languages of peoples having the lowest culture, is superseded by a relatively expansive system based on five, which originated with counting on the fingers of one hand. This system, developed through a transitional stage into a system based on twenty in some languages and a system based on ten in other languages. The system based on twenty originated with counting on both fingers and toes. Thus, for example, in one of the languages of the eastern part of Australia, in Pikumbul, the number five is designated by their word for 'one hand'; ten by the word for 'two hands' and next, as counting proceeds to the toes, twenty is designated by the word for 'two feet'. Gradually, the system based on twenty became autonomous of counting on fingers and toes and was further elaborated according to the formula $30 = 20+10$, $40 = 2\times20$, $50 = (2\times20)+10$, etc... Systems of this type function in the languages of South America (Tupi-Guarani, Arawak, Carib, Chibcha), Central America (Maya-Quiche), Mexico (Otomi-Mangue, Mixtec-Zapotec, Uto-Aztec) and in the languages of the Pacific coast of North America, in the Caucasian languages, and in Basque in the Pyrenees as well as in the language of West Sudan.

The influence of the linguistic substratum, a remnant of which is represented by the Basque language, explains certain forms of French numerals formed according to the system based on twenty: *soixante-dix* ($70 = 60+10$), *quatre-vingt* ($80 = 4\times20$), *quatre-vingt-dix* ($90 = (4\times20)+10$).

In the remaining languages, which occupy by far the greatest part of the world, the decimal system is used. In these systems, which originated with counting limited to the fingers of both hands, each number from one to ten has a corresponding simple numeral, while the remaining numerals are formed by combining these ten basic numerals. The decimal system, based on a series of numbers which is of average length, is expansive relative to both binary systems and systems based on five, constituting short series, and systems based on twenty, constituting long series. This is a two stage expansion. First, short series are superseded by decimal systems and systems based on twenty and, next, of these two competing systems, the decimal system gains predominance. It is the most expansive, for it is the most practical, intermediary between the extremes of the excessively short series of binary systems and the extremely long series of systems based on twenty.

Note: What is the difference between concrete and abstract words? What is the Estoup-Zipf Law? What does the hierarchy of words involve? In what direction is the evolution of lexical, pronoun, and numeral systems proceeding?

Chapter 16. SYNTACTIC TYPOLOGY

The smallest meaningful units of language forms are morphemes, which occur in all the languages of the world. Semantic morphemes, i.e., denoting, referring and ordering semantemes, constitute the semantic system of language. They are connected in various ways with extra-linguistic reality, with psychological representations and, through their mediation, with perceived stimuli from the external world. These representations appear in similar form in the minds of all the members of a given linguistic community, and owing to this fact, they take on the nature of social representations. Semantic morphemes are not, however, the essential signs of speech; they are merely the building-blocks from which these signs are constructed. In every language, therefore, the possibility exists of combining morphemes into higher order units. The agents fulfilling this function constitute the syntactic system of language (cf. Chapter 7). Owing to these agents, semantic morphemes enter into the composition of four types of sets, of increasingly higher order: (1) syntactic members, (2) nominal groups, (3) verbal groups, i.e., clauses and simple sentences, (4) complex sentences. We will now attempt to present, in general outline at least, the structure of these four types of forms in the languages of the world.

By syntactic member we mean the smallest set of morphemes which simultaneously fulfill two functions: the semantic function of designating certain extra-linguistic reality, and the syntactic function which involves the fact that a syntactic

member may, within a text, enter into the composition of higher order syntactic units by indicating its relation to the other elements of these units. Syntactic members appear in all the languages of the world; their structure, however, differs from language to language—which fact accounts for the differences in the over-all structure of syntactic systems. These differences in the structure of syntactic members, which have an enormous significance for the general typology of languages, are of two types. One involves the degree of integration of these members, while the other involves the order of morphemes within these members.

Four types of languages can be distinguished in respect to the degree of integration of their syntactic groups: isolating, agglutinative, inflected and alternating. Chinese, especially in its classical form, is a typical isolating language, as are the Thai languages of Indochina as well as many of the languages of Sudan and America. The majority are prosodic, intonational languages characterized by syllables rich in distinctive features. In these languages, this distinctly characterized syllable is either a full semantic morpheme referring to extra-linguistic reality, or an empty, syntactic morpheme expressing the relationships among full morphemes. Groups of semantic and syntactic morphemes constitute loose syntactic members various types of which are constructed along different principles.

In describing such languages the term "word" is superfluous. Some scholars use this term in the sense "morpheme", and others in the sense "syntactic member". Both notions are incorrect, since isolating languages are wordless. By "word" we mean a syntactic member constructed on the basis of the same scheme according to which all the other types of syntactic members of a given language are constructed. This constant and general scheme is the particular feature of words by virtue of which they exist as entities of a higher order than phonemes (cf. Chapter 6). In isolating languages, there are no features common to all types of syntactic members, regardless of their function in the sentence. In such languages, different agents distinguish various types of such members; the scheme for constructing the predicate particularly differs from that of all other members of the sentence. In French, for example, which like English is closely related to the isolating type, there is no uniformity in the morphological structure of syntactic members. In French, the predicate group—still to a certain extent possessing endings of an older type—is constructed completely differently from the remaining members of the sentence, which no longer have endings. The accent on the last syllable delimits groups of morphemes of extremely various and loose structure. Consequently French should be considered as a basically isolating language, containing only remnants of a previous (Latin) inflectional system. In the sentence—*Je vois la maison de mon père*—the elements separated from one another in writing by means of spaces represent individual morphemes constituting three syntactic groups completely differring in structure: the predicate group *je vois* (I see), the direct object group *la maison* (the house), and the noun attribute group *de mon père* (of my father).

In isolating languages, there are no words in the above sense; words do occur

in other types of languages, but in various forms, depending on the degree of integration of the semantic and syntactic functions within the particular morphemes of a word in a given language. In agglutinative languages, such as Turkic, there is no integration whatsoever, for each semantic function of a word is linked to a separate morpheme, which fulfills a single function. The word begins with a root possessing a lexical meaning only, which is followed by a series of suffixes each of which fulfills a single semantic or syntactic function. For example, the Turkish word *el-ler-im-den* (from my hands) is composed of the root *el-* (hand), the suffix expressing plural number *-ler*, the suffix of possession *-im-* (my), and the suffix of the ablative expressing departure *-den* (from). Two characteristic features of all Turkish words are: the initial position of the root, and vocal harmony, i.e., the accommodation of suffix vowel pronunciation to that of the root vowel. If the root contains a back vowel, only back vowels appear in the suffixes; if the root contains a frontal vowel, only frontal vowels appear in the suffixes.

Agglutinative languages are very wide-spread. They include the Altaic and Uralic languages, the Paleo-Asiatic languages, Japanese and Korean. Modern Chinese is slowly becoming an agglutinative language, without, of course, vocal harmony. The Dravidian, Austro-Asiatic and Austronesian families as well as a great number of the languages of Australia, America and Africa are included in this group. In these languages, the problem of words is limited to the problem of signals delimiting words and to the problem of morpheme order. In some languages, accent constitutes this signal, in others, vocal harmony, and in still others, particular morphemes occur which signalize the beginning of words. This latter agent is characteristic of the polysynthetic languages of North America and eastern Siberia.

Inflected languages, such as Polish, demonstrate a certain integration of functions, involving syntax only. The word is composed of two parts: the stem, which fulfills a semantic function only, and an ending which combines a series of syntactic functions. This means that the entire ending, at the same time, constitutes a member in a series of oppositions designating the word's position in the sentence. Thus, for example, in the Polish word *pan-ów* (gentlemen's), the ending *-ów* is both the exponent of the genitive in opposition to the nominative *pan-owie* as well as the exponent of the plural in opposition to the singular *pan-a*. All of the Indo-European languages which preserved the former structure—i.e., Sanskrit, Old Persian, Greek, Latin, etc. in antiquity, and the Baltic and Slavic languages with the exception of Bulgarian, today—belong to the inflected type of language. This type is extremely rare outside the Indo-European family.

In alternating languages, such as Arabic, all functions, both semantic and syntactic, are integrated within the word which, as a result, constitutes an individual morphological entity, usually composed of a root only. Here, consonants are the exponents of the semantic content while vowels, alternating between the consonants, fulfill the syntactic function. Thus, for example, an Arabic root characterized by the consonants *q-t-l* refers to killing; from this root, words containing various forms

of alternating vowels are created—*qatala* (he killed), *qatila* (he was killed), *ya-qtulu* (he was killing), *yu-qtalu* (he was being killed), etc. As we see, the consonants appear constantly in the same position, while the vowels change in various forms of the word, creating the alternating series *a-u-i-zero*. The most frequent form of the alternating type occurs in the Semitic languages, particularly in Arabic and Hebrew.

In classifying languages typologically according to the structure of their syntactic members, a second criterion, besides the degree of integration of functions, is the order of members. In each set of two morphological elements, the principal member (P) can be distinguished from the auxiliary member (A). Thus, the part of the word fulfilling the syntactic function is an auxiliary member (A_1) in relation to the lexical part, which in this case is the principal member. Within the lexical part, affixes constitute auxiliary members (A_2) in relation to the root, constituting the principal part. Finally, within the root, the identifying root (cf. Chapter 6) is an auxiliary member (A_3) like affixes, while the distinguishing root is the principal member (P). Thus, a uniform principle concerning the order of members in relation to their principal members generally operates in language, involving the placement of all auxiliary members either after the principal members or before them. In the first case, the structure of the word is represented by the scheme ($P+A_3+A_2+A_1$), while in the second case, by the scheme ($A_1+A_2+A_3+P$). Languages in which words are generated according to the first scheme are called postpositional languages, while languages of the second type are called prepositional languages.

The prepositional type is predominant in the languages of Africa and in those of Oceania and Indochina (Austronesian and Austro-Asiatic families), while in the remaining parts of the world, the postpositional type predominates. The Indo-European languages also belong generally to the postpositional type. Thus, in Polish, the word is constructed according to the scheme ($P+A_3+A_2+A_1$). In Polish, the ending is placed after the stem according to the scheme ($P+A_1$), e.g., *dom-em*, *dom-u*. The affix stands after the root as its suffix, e.g., *dom-k-a*, *dom-k-i* ($P+A_2+A_1$). The identifying member of stems composed of two roots stands after the distinguishing member, e.g., *ojco-bój-c-y*, *Nowo-gród-k-a* ($P+A_3+A_2+A_1$). On the other hand, the Swahili language of the Bantu family in Central Africa is of the prepositional type. In Swahili, morphemes having a syntactic function, primarily exponents of gender, stand before the root according to the scheme (A_1+P), e.g., *m-tu* (human being), *m-* is the exponent of the class of gender of people in the singular and *-tu* is the root. Similarly affixes occupy a position before roots as prefixes (A_2+P), e.g., the word *dżi-tu* (big man) which is composed of the prefix *dżi-*, the exponent of the group of large objects, and the root *-tu* (man). By analogy, the word *ki-meza* (small table) is composed of the prefix *ki-* which is the exponent of small objects and corresponds in meaning to the Polish suffix *-ik*, and the root *-meza* (table). The word *ki-meza* may be represented by the scheme (A_2+P), while the word *stol-ik* (small table) by the scheme ($P+A_2$). The Malay language of the Austronesian family is also prepositional; here we find the compound word *ōrang-ūtan* (man of

the forest) in which the first member *ōrang* (man) is the identifying member and the second member *(h)ūtan* (forest) is the differentiating member. Thus, we have the scheme (A_3+P), the opposite of the Polish *wielkolud* (*wielko-* = 'giant', 'big'; *-lud* = 'people') ($P+A_3$).

Syntactic members having a fairly integrated structure enter into the composition of higher order units, i.e., nominal groups and sentences. Nominal groups are composed of two members connected by a determining relation. One of these members (the determinant) defines the second member, i.e., increases the number of features it possesses, thereby decreasing its semantic range, e.g. *ojca* (father's) (determinant) *dom* (house), *bialy dom* (white house), *żywopłot* (hedge). A sentence, however, is a completely different kind of entity. Its nucleus is the predicate, which concretizes the sentence by locating it in space (the categories of person and number) and in time (the categories of tense and aspect), defines its relationship to reality (mood) and determines its structure (the category of voice) (cf. Chapter 7). In intransitive sentences, we have only one member outside the predicate, i.e., the grammatical subject modified by the predicate. It constitutes the basis upon which the meaning content expressed by the predicate develops, e.g., *ojciec śpi* (father is sleeping), *ojciec jest chory* (father is ill). In transitive sentences we have—besides the predicate—two members: the agens designating the take-off point of the action, and the patiens designating the object of the action, e.g., *ojciec* (father) (agens) *rąbie* (is chopping) (predicate) *drzewo* (a tree) (patiens—direct object).

In the languages of the world we have four basic syntactic relationships based on modifying—(1) the subject of the intransitive predicate, (2) the agens of the transitive predicate (3) the patiens (direct object) of the transitive predicate, (4) the determinant of the defining member. There is no language in which these four functions are expressed by four separate formal agents. In all languages, the same formal agent expresses two or even three of these functions; the integration of functions is, however, different in different languages. The typology of the structure of nominal groups and sentences may be of two types: one is based on the form of syntactic exponents only, the other on the extent of their functioning.

From the first standpoint, we may distinguish three main types of languages: positional, inflected and concentric.

In positional languages, syntactic relationships are expressed by the order of the syntactic members. Thus, for example, in the transitive English sentence—*Mark killed the lion*—the only agent defining the syntactic relation of the members is their order—agens, predicate, direct object. The same order appears in French, e.g., *Jean bat Paul.* This order is the most typical in positional languages, but it is certainly not the only one.

In inflected languages, the function of words—which in these languages basically constitute syntactic members—is determined by their form, i.e., case ending, prefixes, phoneme alternation, point of accent, etc. Thus, in Latin and Polish, the determining noun in a nominal group is distinguished by the ending in the genitive, e.g.,

Latin *domus patr-is*, Polish *dom ojc-a* (father's house); the subject of an intransitive sentence ends in the nominative, Latin *Marc-us dormit*, Polish *Marek śpi* (Mark is sleeping); the agens in a transitive sentence is also characterized by the nominative ending, and the patiens by the accusative, e.g., Latin *Marc-us necat leon-em*, Polish *Marek zabija lw-a* (Mark is killing the lion.)

Verbal and nominal groups are constructed differently in concentric, incorporating languages. The function of the non-constitutive members of these syntactic groups (subject, agens and patiens, in verbal groups and the determining member in nominal groups) is not expressed by these members alone in such languages, but by their pronoun substitutes, incorporated by the constitutive member of the group, i.e., by the predicate in a verbal group and by the determining member in a nominal group (cf. Chapter 7). Either the order or the morphological form characterizes the function of the pronoun within the constitutive member; the non-constitutive members of the group, however, are characterized neither by form nor by order, but solely by agreement in number and, in certain languages, in gender, with the pronouns of the constitutive member which determines each of them individually. This enables us to define the syntactic role of each non-constitutive member of the group, for each of them is the same type of member as is the pronoun incorporated in the constitutive member—with which it agrees in number and, sometimes, in gender. Thus, for example, in the concentric Ojibwa language of the Algonquian family, the nominal group has the following structure: *māpa enimi u-wītekēmākenan* 'that man his-wife'—i.e., that man's wife; compare *māpa* 'that', *enimi* 'man', *u-* 'his', *wītekēmākan* 'wife'. The determining noun *enịmi* 'man' is indicated here by the pronoun *u-* 'his', which is embodied in the determining constitutive member. Here is an example of a transitive sentence in the Ojibwa language: *nen-tawēmā u-kī-ness-ā-n wāwāškēššu,-wan* 'my brother he to kill it a deer'—i.e., my brother killed a deer; compare *nen-* 'my', *-tawēmā* 'brother', *u-* 'he' (pronoun prefix of the agens), *-kī-* (prefix of the perfective aspect), *-ness-* (root) 'to kill', *-a-* 'it' (pronoun suffix of the patiens), *wāwāškēššu* 'deer'. The pronoun prefix *u-* 'he' refers to the agens (brother), while the pronoun suffix *-ā-* 'it' refers to the patiens. This is characteristic of their syntactic function. Transitive sentences in the language of the Yuchi Indians living in Oklahoma are constructed in a similar way: *gɔnt'ɛ'nᵩ b'aχτɛ'wᵩnᵩ wɛ'hɛτnɛ* 'the man saw some horses', nominal agens *gɔnt'ɛ'-nᵩ* 'that man', nominal patiens *b'aχτɛ'-wənᵩ* 'that horse', predicate *wɛ'-hᵩ-τnɛ* 'his-he-to see'. Two pronoun prefixes appear in the predicate *wɛ'-hᵩ-τnɛ*—one referring to the patiens *wɛ-* and one referring to the agens *-hᵩ-*. Such a syntactic function of prefixes is determined by their position. The nominal members of the sentence, agens and patiens, are devoid of morphological exponents which would characterize their syntactic function; their function is determined by agreement in number and gender between the suffixes of the nominal members and the prefixes of the predicate: the suffix *-nᵩ* agrees with the prefix *hᵩ*, and the suffix *-wənᵩ* agrees with the prefix *wɛ'-*. The function of nominal members is the same as that of the pronoun prefixes with which they agree.

The inflected type of language includes archaic Indo-European languages (Sanskrit, Old Persian, Greek, Latin, etc.) and the modern Baltic and Slavic languages with the exception of Bulgarian, the Semitic and Japhetic languages, and the Uralic and Altaic languages. The concentric type includes, in the first place, the languages of the North American Indians (with the exception of the Penutian family in California) and in parts of Mexico (the Uto-Aztecan family), Central America (Maya-Quiche) and in the foothills of the Andes (Kichua). In the Old World, the languages closest to this type are those of the Northern-West Caucasian group (Abkhazian, Ubyk, the Adyghe-Qabardi dialects). The construction—*that man his-wife* (that man's wife)—is wide-spread in the Turkic languages. Even in America, many languages occupy a position intermediary between the concentric and positional types and between the concentric and inflected types, e.g., the Eskimo dialects. These are moderately concentric languages. The most expansive is the positional type, which is gradually supplanting all others. The positional type became established in the course of the modern era in the majority of Indo-European and Semito-Hamitic languages, and it has operated for centuries in the Sino-Tibetan, Negro and South American languages, and in many others as well.

Equal in importance to the formal typology of syntactic agents, described above, is the typology based on differences in the extent of functioning of these agents. We know that the languages of the world include four basic syntactic relationships between the non-constitutive members of verbal and nominal groups: (1) between the subject and intransitive predicate, e.g., *father is sleeping*; (2) between the agens and the transitive predicate, e.g., *father killed...*; (3) between the patiens and the transitive predicate, e.g., ... *(he) killed a deer*; (4) between the determining member and the determined member of nominal groups, e.g., *father's house.*

There is no language which expresses these four syntactic functions by means of four separate formal exponents. In all languages, one of these exponents (order, syntactic morpheme, alternating position or a form of incorporation) serves to express two and sometimes even three of the four syntactic relationships cited. Although integration occurs in all languages, six various types of integration may be distinguished in the languages of the world. If the letters *a, b, c* are used to indicate the formal exponents of syntactic relationships, the extent of their functioning in each of the six types of languages can be represented as follows:

Syntactic relationships of the non-constitutive member to the constitutive member:	Syntactic types					
	1	2	3	4	5	6
subject to predicate	*a*	*a*	*a*	*a*	*a*	*a*
agens to predicate	*a*	*b*	*a*	*b*	*a*	*b*
patiens to predicate	*b*	*a*	*b*	*a*	*b*	*a*
determining member to determined member	*c*	*c*	*b*	*b*	*a*	*a*

Type #1 includes Polish as well as many other languages. In Polish, the exponent *a* is the nominative ending, which is used to characterize both the subject of intransitive sentences—e.g., *matk-a śpi* (mother is sleeping)—and the subject (agens) of transitive sentences—e.g., *matk-a kocha córkę* (mother loves her daughter). In Polish, the exponent *b* characterizing the patiens, i.e., the direct object of transitive sentences, is the accusative ending—*syn kocha matk-ę* (the son loves his mother), and the exponent *c* which characterizes the determining member of nominal groups is the genitive ending—e.g., *dom matk-i* (mother's house). The form, of course, of these three exponents may vary (order, syntactic morpheme, alternation, forms of incorporation) but the distribution of their functions is the same in all languages of type #1, which is the most common, expansive type. This type includes the languages of the great Indo-European family, the Semito-Hamitic, Uralic and Altaic languages, the majority of the Sino-Tibetan languages, as well as a great number of languages on other continents, especially in Africa and South America.

In languages of type #2, the syntactic exponent *a* characterizes the subject of intransitive sentences and the patiens in the equivalent to our transitive sentences, while the agens is characterized by the exponent *b*, and the determining member of nominal groups by exponent *c*. In many Caucasian languages the exponent *a*, common to the subject and the patiens, is the *zero* ending of the so-called absolutive case, the exponent *b* is the ending of the so-called ergative case, related in meaning to Latin instrumental, while exponent *c* is the genitive ending. In languages of this type there is no formal difference between transitive and intransitive sentences. From our point of view, all sentences are intransitive. The formal subject, which agrees with the predicate, is the patiens, while the agens is treated as a kind of adverb of instrumentation. Alongside the sentence—*the deer is sleeping*—we have the sentence—*the deer was killed by means of father*. Both of these constructions are intransitive although the latter corresponds in meaning to our transitive sentence—*father killed the deer*. This type includes the Japhetic languages (Basque, the majority of the Caucasian languages, the ancient Sumerian language, etc), Old Tibetan, and the Paleo-Asiatic languages—all in the Old World. It also includes languages of the Central Australian family, several languages of South America (Moluche, Kunibo) and many of the languages of the North American Indians of the Pacific coast (Sioux, Kus, Sahaptin, Chinook).

In contrast with the languages of types #1 and #2—in which syntactic relationships are expressed by means of three exponents *a*, *b*, *c*—in languages of types #3 to #6 only two exponents *a*, *b*, function. This is because in languages of types #1 and #2, exponents *a*, *b*, function only in verbal groups and exponent *c* only in nominal groups, while in the languages of types #3 to #6, the same exponents *a*, *b*, function in both verbal and nominal groups.

In languages of type #3, two exponents appear which, for example, in the Hopi language of Arizona (Uto-Aztecan family) have the ending forms of two grammatical cases. In Hopi, therefore, there are only two forms of syntactic members

besides the predicate—the subject (including both the intransitive subject and the agens) in the nominative with the ending *zero*, and the determining member with the genitive ending *-t* which corresponds functionally to both our genitive attribute—*dom matki* (mother's house)—and to our direct object—*widzę matkę* (I see mother). In this case therefore, just as in type #2, only two constructions function: sentences, the non-constitutive member of which is in the nominative, and nominal groups, the non-constitutive member of which is in the genitive. The equivalent of our transitive sentence has the form of an intransitive sentence in which the patiens functions as a member determining the predicate. Similar relationships also operate in Indonesian languages, the single difference being that, in these languages, the syntactic exponent is order.

In languages of type #4 as well, there are only two syntactic exponents. In the Greenland languages of the Eskimo family these exponents take the form of two case endings, as a consequence of which there are only two forms of syntactic members: the subject (corresponding in meaning to our intransitive subject and to the patiens of transitive sentences) with the *zero* ending of the absolutive case, and the determining member (the equivalent of our genitive attribute and the agens of transitive sentences) with the genitive ending *-p*, e.g., the nominal group: *tihiania-p ihlu-a* (fox his-house, i.e., the fox's house), compare *tihiania-* (fox), *-p*, the genitive ending, *ihlu* (house), *-a* (his); the sentence: *tihiania-p ihlu takuba-a* (the fox saw the house). As we see, two constructions function here: a sentence having a non-constitutive member (subject) in the absolutive case, and a nominal group with a non-constitutive (determining) member in the genitive. The equivalent of our transitive sentence has the form of an intransitive sentence the subject of which is the patiens—*ihlu* (house)—while the agens—*tihiania-p* (fox's)—has the form of a member determining the predicate which, therefore, is closer to our noun in meaning—*takuba-a* (seeing). The construction "the seeing of the fox" corresponds to our construction "the fox saw". Systems of this type appear in certain Japhetic languages (the Northern-West Caucasian group and, to some degree, the Lakh language) as well as in many languages of North America (the Eskimo family, the Salish group, Taos), Central America (Maya-Quiche, Zoque), and South America (primarily dialects of the Arawak family).

In languages of types #5 and #6, there is only one basic form of syntactic construction because the equivalents of our intransitive sentences and nominal groups are constructed in the same way. In the Nass language of British Columbia (Nass-Tsimshian family), which is one of type #5, the exponent *a* in the form of a suffix *-tl* characterizes the non-constitutive member of general syntactic groups at the same time corresponding to our (type #1) determining member of nominal groups and to intransitive and transitive subjects. In Nass, the exponent *b*, in the form of a suffix *zero*, characterizes only the patiens of transitive sentences, thus it has a range of function similar to that of the Polish accusative ending. In the Tsimshian language of British Columbia, which represents type #6, the exponent *a*, in the form of the suffix *-ge*, also characterizes the non-constitutive member of general syntactic

groups. The only difference involves the fact that, here, the non-constitutive member corresponds to the non-constitutive member of languages of type #2 and, therefore, to our determining member of nominal groups, the subject of intransitive sentences and the direct object (patiens) of transitive sentences. In Tsimshian, the exponent *b*, on the other hand, in the form of the suffix *-sge*, is a feature of the agens of transitive sentences, thus it has a range of function similar to that of the ergative ending of the Caucasian languages. Type #6 includes several languages of North America (e.g., Tunika, at the mouth of the Mississipi) and South America (e.g., Guarani).

There is no difficulty in distinguishing among the recessive and expansive forms of the six syntactic types described here. It is clear that types #5 and #6 are the most recessive and that type #1 is the most expansive, Differences in their distribution proves this fact. Types #5 and #6 function in several American languages most of which are already dying out; the languages of types #4, #3 and #2 occupy clearly limited zones, while type #1 occupies most of the continents of the world. Historical considerations lead to the same conclusions. In many languages, two, three and even four of the types distinguished here function simultaneously in various parts of their systems. This is the consequence of evolution which usually, though not always, nvolves development in the direction from type #6 to type #1. The history of the North American Algonquian family is characteristic in this respect. The oldest grammatical constructions in the languages of this family are of type #6, younger constructions of type #5, even younger constructions of type #2, and the youngest, of type #1. The same is true elsewhere and the rare contrary cases are probably the result of substrata influences.

As is evidenced by these facts, recession in the syntactic structures in question is a two-stop process. Types #6 and #5 are the most recessive. They contain only one form of undifferentiated general scheme for the syntactic group, the non-constitutive member of which possesses the same constant exponent *a*. In type #5, the exponent *b* characterizes the patiens, while in type #6 it characterizes the agens—these, however, are exponents of very special and perhaps concrete usage. These types, surviving only in America, were the last vital types and were superseded by all other types within the boundaries of which another two groups arose. Types #4, #3 and #2 underwent the second stage of recession. They contain two types of schemes: the scheme of sentence in which the non-constitutive member (subject) is characterized by the exponent *a* and the scheme of the nominal group in which the non-constitutive member (determining) is characterized by the exponent *b* (types #4 and #3), or *c* (type #2). The determining member of the predicate is the agens in type #4 and the patiens (exponent *b*) in type #3. In type #2 the agens is the adverbial form of the instrumental. All of these three types are supplanted by type #1 which distinguishes three syntactic schemes: transitive and intransitive sentences and nominal groups. In general, therefore, it can be maintained that a given syntactic type is more expansive the richer it is in agens, the more syntactic schemes it distinguishes

and consequently the greater possibility it gives for various formulations of the relationships among syntactic groups.

Hitherto we have been discussing the structure of nominal and verbal groups only, which contain only one predicate. Now we will proceed to the question of syntactic structures of a higher order, i.e., polypredicative groups (which contain several predicates).

Combining large number of morphemes in a hierarchical unit of a higher order in order to express highly elaborated and complicated meaning contents is accomplished in different ways in various languages. In certain languages, this function is fulfilled by elaborate multi-stage nominal groups in which each member itself constitutes a complex group. They frequently take the form of long complex words. This is especially characteristic of polysynthetic languages (e.g., Eskimo languages) and of classical Sanskrit. Polypredicative groups, which appear in two various forms, are a much more common means for expressing highly elaborated contents. In both cases, the polypredicative group has a hierarchical structure; the structures of both types differ only by virtue of the fact that, in the first type, we have a main predicate which constitutes the basis of predicates which are in a dependent relation to it within the same sentence, while in the second case, we have a sentence composed of several clauses one of which is the main clause and the others, subordinate clauses.

In the majority of American Indian languages and in Ural-Altaic languages only the former type of polypredicative groups occurs, composed of a main predicate and several dependent predicates. Indo-European languages make use of a somewhat similar construction—accusative with the infinitive appearing in Latin and Polish as well as other languages, e.g., *quin tu illam iube aps te abire quolubet = zmuś ją odejść od siebie* (force her to leave you). In this sentence the Latin infinitive *abire*, Polish *odejść* (to leave) is a predicate dependent upon the main predicate Latin *iube*, Polish *zmuś* (force). Latin forms of the participle, in expressions with the ablative absolute, fulfill a dependent predicative function, e.g., *eo praesente* (if he were present) as well as the various constructions using gerunds and gerundives, while in Polish we find the adverbial participle construction in the role of dependent predicate—e.g. *szedł paląc papierosa* (he walked along smoking a cigarette), *wrócił zwyciężywszy wrogów* (he returned having triumphed over his enemies). In all of these cases the main predicative is in the personal form, constituting the semantic basis for the dependent predicate.

The tendency to combine a great number of morphemes in hierarchical structures led, in certain languages only, e.g., Indo-European languages, to the creation of two types of complex sentences: coordinate and subordinate. Coordinate compound sentences constitute loose sets connected by coordinating conjunctions of the type *i, a, lecz, lub*, in Polish, *et, sed, aut*, etc. in Latin, *and, but, or* in English. Conjunctions are loose morphemes that do not belong to particular words, morphemes possessing no semantic content but merely modifying in two directions, i.e., creating places in the text for words and clauses preceding and following them. Coordinating

conjunctions do not belong to any of the clauses of a sentence but, instead, they stand between such clauses, modifying both. As a result, both coordinate clauses are semantically complete and reversible. The clause standing before the conjunction can be placed after it without any change in its meaning and vice versa, e.g., *Father chopped with a hatchet and mother returned home.* = *Mother returned home and father chopped with a hatchet.*

Coordinate sentences do not constitute an integrated structure; they are a series of equal-ranked elements which are associated with one another. Subordinate sentences, on the other hand, constitute hierarchical sets including subordinating clauses and subordinate clauses. Subordinate clauses are of two types: one type is connected to the entire subordinating clause the constitutive member of which is the predicate, while the other type is connected to only one of the words of the subordinating clause, not included in the predicate. In Indo-European languages, as well as in many others, subordinate clauses which are connected to only one of the words of the subordinating clause include a relative pronoun which, constituting one of the members of the clause, simultaneously indicates one of the nouns of the subordinating clause thereby combining the two clauses into a hierarchical whole, as we have already discussed.

Subordinate clauses connected to the entire subordinating clause, which are characterized by the presence of the subordinate conjunction or by a special mood or tense in the predicate, modify, i.e., announce, the subordinating clause, for only in connection with this clause are they semantically complete, e.g., *Mother said that father returned home.* The subordinate clause *that father returned home,* semantically incomplete, fulfills the function of one of the members (in this case the direct object) of the subordinating clause, and owing to this fact, together with it, creates an integrated and irreversible structure.

Polypredicative structures with dependent predicates and coordinate clauses (parataxis) occur in the great majority of languages in the world, while subordinate sentences (hypotaxis) are newer constructions, limited to certain languages only, but at present distinctly expansive. Clauses connected by subordination are almost completely lacking in American Indian languages, in the old Ural-Altaic languages and in many others. Subordinate sentences developed immensely, however, in the 1st millennium B.C., first in Greek, later in Latin and in the Indo-European languages of modern Europe—thence influencing the languages of other families.

Note: What are syntactic members? What constitutes the difference among isolating, agglutinative, inflected and alternating languages? What are words? How is the word constructed in postpositional languages and in prepositional languages? What is the difference between nominal groups and sentences? How are syntactic relationships expressed in positional, inflected and concentric (incorporating) languages? Describe the functional integration of the formal syntactic exponents of the Indo-European family (type #1), the North-Caucasian languages (type #2), the Hopi language (type #3), the Eskimo language (type #4), the Nass language (type #5) and the Tsimshian language (type #6). What is the difference between dependent predicates and subordinate sentences?

Chapter 17. STYLISTIC TYPOLOGY

Because it is not our intention to present a comprehensive view of the various impressive and expressive agents occurring in the languages of the world we shall limit our discussion to the most important of these agents, i.e., special languages (cf. Chapter 8). These languages, limited to certain groups and characterized primarily by special vocabulary, evolve, changing their function in connection with the cultural development of societies. The earliest to appear in the development of man are secret languages and women's languages, known in societies having primitive cultures. Secret languages are systems of communication intended to be of limited comprehension, intentionally kept secret by the initiated who know the language—a group composed, for example, of men only, of women only, of members of a given tribe only or of members of a given profession only. Secret languages make use of either words taken from general language—used in another meaning, or distorted—or words made up or borrowed from other, sometimes extinct, languages. Thus, for example, tribes inhabiting the northern coast of New Guinea colloquially speak dialects belonging to the Austronesian family; their secret language, however, used only by men, includes many words from the Papuan dialects which previously occupied these regions. Men use their secret language when, for example, they are communicating information which women and children should not know about, e.g., information about war, which could evoke panic.

The difference between the men's languages and women's languages found in various parts of the world is especially evident in several languages of America and the adjacent parts of Asia—in which both sexes know both languages equally well, and even use the forms appropriate to the opposite sex in quoting another person's statements. Women's languages are usually more archaic because women are more sedantary. Among the Indian tribes of the Antilles, for example, the women's language contains word survivals from the Arawak language family spoken by the original populace, which was wiped out in the 15th century, while the masculine language is a pure Carib dialect spoken by the conquerors. In certain societies, the women's language is characterized by certain suffixes added to masculine forms, while in other societies the opposite is true and masculine forms originate from feminine forms.

In higher cultures, new special languages arise which, after the introduction of writing, take on the character of written cultural dialects. They are not yet literary languages, for a literary language is a written and spoken instrument of the culture of an entire society, it is a complete system of national language. Cultural dialects do not represent a national tradition of written language for cultural traditions constitute first an oral tradition, and later a written tradition of particular, relatively closed, professional and social groups. In this connection, we may distinguish three main types of cultural dialects: religious languages, the origin of which is connected with secret languages, administrative languages and poetic languages.

Among religious languages, used by priests during devotional ceremonies and preserved by them, we may distinguish three types: (1) sacred languages such as the Vedic in India or Avestan in Iran, which gradually spread within the priestly class, (2) missionary languages like Old Slavic or Gothic—established by missionaries on the base of a dialect which previously had no cultural tradition—for the propagation of religion, and (3) auxiliary religious languages, such as Old High German and Old Polish. Between the 8th and the 12th centuries A.D. the Old High German cultural dialects, with minor exceptions, were used only within the Church where they were limited, however, to an auxiliary function relative to the omnipotent Latin. Translations or paraphrases of various parts of the Bible, church songs, prayers, catechisms, monastic rules—these constitute the most numerous texts in this "monastic" writing. Great dialectical differentiation, combined with the absence of any survivals of a general written language, is connected with the limited and unofficial social function of this language. A similar auxiliary social function in churches (songs, the Gospels, sermons, prayers after the Mass) was fulfilled by the Old Polish dialects in the 14th century.

The second basic type of cultural dialects are administrative languages the tradition of which was preserved in bureaucratic centers publishing official proclamations and state documents in these languages. Thus, between the 6th and the 3rd centuries B.C., almost each of the Greek city-states had its own administrative language which arose on the base of the local dialect and which we know from inscriptions in stone having a legislative or commemorative meaning. Written Latin, between the 4th and the 2nd centuries B.C. had the same character.

The third basic type of cultural dialects are poetic languages, based on oral tradition preserved by groups of singer-narrators. Such a language, after the classical repertories of the singers had been written down, functioned for a certain period of time as the raw material of poetry, representing the traditional forms of versification. A typical example of this poetic language is the language of Homer, reflected in poetry in dactylic meter. Originally this was oral poetry performed by singers and wandering musicians; in time, however (from the 6th century B.C., at least) it became written poetry as well. Provençal lyrical poetry, which developed between the 11th and 13th centuries, represents a similar phenomenon in the Middle Ages.

As civilization advanced in a given society, cultural dialects having a narrow range of use were no longer sufficient. The need arose for a national written language which could serve as the cultural instrument for the entire nation. Thus, one of the cultural dialects already existing in such a society became extended in its range of use, its vocabulary enriched, its semantic and syntactic systems modernized and it became a literary language, satisfying a general need. In this way, in the 2nd and 1st centuries B.C., in Rome, under the influence of the revolution in thinking brought about by the influence of Greek culture, the Latin literary language (classical) developed from the administrative language of the Senate. In 16th century Germany

Luther used the state administrative language in his translation of the *Bible* and owing to the upheaval brought about by the Reformation, this language became the New High German literary language. In Poland, under the influence of the Reformation and Renaissance, the Polish literary language developed in the 16th century from the cultural dialect of the Kraków region, which had had the character of an auxiliary church language. Similarly, in the 18th century, under the cultural influence of the Enlightenment, the language of the Russian Orthodox Church, used in Moscow, basically an Old Slavic (Macedonian-Bulgarian) language but overloaded with local, Russic elements, developed into the Russian literary language.

These examples, which could be multiplied, prove that every society possesses a certain system of special languages, i.e., so-called linguistic sub-codes, limited to certain social environments and associated with emotional, expressive and impressive meaning contents. Such oppositions as secret language : open language; women's language : men's language; religious, administrative, or poetic cultural dialect : non-cultural dialect; literary language : local dialect—are always associated with various emotional reactions, thus constituting the basis of the stylistic system of a given language.

Note: What are special language? What types of special languages occur in societies primitive in culture? What are the functions of cultural dialects? What are the types of cultural dialects? How do literary languages arise?

CHAPTER 18. RECESSIVE AND EXPANSIVE LANGUAGE FEATURES

The various languages of the world deal with various problems of communication in different ways; certain of these solutions die out and others spread—thus, we have recessive features of language structures, which are rare and becoming rarer, and expansive features of language structures, which are common and becoming increasingly wide-spread. Two basic causes for the spread of expansive features can be distinguished. The first involves economy of effort. In many cases a given solution to a certain problem in communication prevails because it involves less effort than does the competing solution, and given a possibility of choice, the speaker usually chooses the more economic of the two.

As we have seen, systems in which—alongside expiratory phonemes—non-expiratory phonemes occur (clicks, glottalized consonants, ejectives, injectives, etc.), demanding additional effort on the part of the speech apparatus, are dying out. Systems based on expiration only, which does not require this additional effort, are expansive. Impoverished systems—including only 13 to 20 phonemes—are dying out, for they necessitate excessively long morphemes and, consequently, long statements—which increases the work of the speech apparatus. Overloaded systems—45 to 75 phonemes—

are also dying out, for in the diacritic function, they make use of acoustical differences which are extremely minute and difficult to perceive. Intermediary systems, on the other hand—20 to 45 phonemes—which are free of both disadvantages of the extreme systems, are expansive.

Similarly, numeral systems characterized by short, dual series are dying out, for in higher numbers they lead to the formation of extremely long numeral forms. Systems based on twenty are also dying out, for the excessively long series are difficult to memorize. The intermediary, decimal systems, which do not share the disadvantages of the other solutions, are expansive.

Systems which burden the memory with an excessive complexity of details, become extinct, while simpler systems, having a smaller number of elements which must be memorized, are expansive. Among third person pronouns, three-part reference (Latin *hic—iste—ille*) is dying out in favor of the simpler two-part reference (Polish *ten—tamten*; English *this—that*). In the first person plural, the distinction between two forms—inclusivus and exclusivus—is dying out in favor of systems which contain only the general form "we". Among agents for expressing syntactic relationships, techniques which are too complicated tend to die out: inflectional, alternating and incorporating—and positional techniques, requiring no additional morphemes and, therefore, less effort, are expansive.

Another group of changes involves the introduction of new means of communication, means enabling new tasks required of language by cultural progress to be realized more efficiently. These changes appear primarily in the development of lexical systems. Increasing numbers of abstract words arise, of an increasingly greater degree of abstraction, thus making possible the synthetic conceptualization of extensive parts of the external world.

Parallel changes also occur in syntactic systems. In the construction of simple sentences, type #1 is becoming increasingly common, for it is the only type which precisely differentiates three syntactic schemes: nominal group, intransitive sentence and transitive sentence—at the same time making possible a greater variety of syntactic constructions than do the remaining types, from #2 to #6.

The syntactic structure of the highest order, i.e., the complex sentence, is also expansive. Such sentences involve the syntactic organization of large numbers of morphemes. This problem is solved by means of the two-stage structure of complex words (morpheme—word), by the three-stage structure of simple polypredicative sentences containing dependent predicates (morpheme—word—simple sentence), by the four-stage structure of complex sentences (morpheme—word—clause—complex sentence). The multi-stage organization of morphemes and, consequently, their far-reaching segmentation in complex sentences is much clearer than are structures having fewer stages—which assures the increasing expansion of complex sentences.

It must be added, that we notice the expansion of types of languages having a rich system of subcodes (special languages), arranged hierarchically with literary language and, especially, its most important subcode—the language of art—at the top. This type

of system gives the greatest possibility for stylistic variation in the composition of various types of texts.

In studying the language map of the world, we notice that the features of the recessive and expansive languages presented are not evenly distributed over the entire territory of the inhabited world—in certain territories recessive features predominate, while in others, expansive features. This is especially striking on the language map of the 15th century, which presents the situation before the great expansion of the white races. In this period, recessive features appear in high concentration in the languages of America, in the oldest stratum of the languages of Asia and Europe (Paleo-Asiatic and Caucasian languages, Basque), in the languages of Australia and Oceania (Austronesian, Papuan and Andamanese languages) and in the dialects of the Khoin family in the southern extremes of Africa.

This was a recessive zone, in contrast with the zone of languages having an enormous predominance of expansive features, encompassing practically all of Asia, Europe and North Africa (Indo-European, Semito-Hamitic, Uralic, Altaic and Sino-Tibetan languages).

The expansive zone occupied the center of the inhabited world, and the recessive zone, its periphery. The expansive sphere was occupied by the languages of a small number of families only and was clearly homogenous as regards the structure of the language systems. The recessive peripheral zone was broken up into hundreds of families and was extremely differentiated as regards the typological structure of its languages.

The expansive zone was the result of relatively fresh conquests and language transformations; it can be assumed, therefore, that in the 4th and 3rd millennia B.C., this zone did not exist at all. From the 16th century, however, it spread incredibly, limiting the recessive zone in America, Australia, Asia and Europe, and in the southern extremes of Africa to tiny relic islands. In this period, the linguistic history of the world clearly proceeded from multiplicity to unity.

Note: What features of languages are considered recessive, and what features expansive? What causes the wide-spread distribution of expansive features? What parts of the inhabited world are occupied by expansive languages and what parts by recessive languages?

BIBLIOGRAPHY

Axmanova N. S., Mel'čuk I. A., Padučeva E. V., Frumkina R. M. *O točnyx metodax issledovanija jazyka (O tak nazyvajemoj "matematičeskoj lingvistike")*, Moscow, 1961.

Apostel L., Mandelbrot B., Morf A. *Logique, langage et théorie de l'information*, Paris, 1957.

Biuletyn Polskiego Towarzystwa Językoznawczego (Bulletin of the Polish Linguistics Association), No. I–XXX, Kraków, 1931–1963 (Bibliography of Polish Linguistics from No. XIV, 1955).

Bloomfield L. *Language*, New York, 1933.

Bühler K. *Sprachtheorie*, Jena, 1934.

Cherry C. *On Human Communication*, New York, 1957.

Chomsky N. *Syntactic Structures*, The Hague, 1957.

Cohen M. *L'écriture*, Paris, 1953.

Coseriu E. *Sincronía, diacronía e historia*, Montevideo, 1958.

Grammont M. *Traité de phonétique*, Paris, 1933.

Guiraud P. *Introduction à la linguistique statistique*, Paris, 1959.

— *Problèmes et méthodes de la statistique linguistique*, Paris, 1960.

Handel J. *Dzieje językoznawstwa* (History of Linguistics), Lwów, 1935.

Harris Z. S. *Methods in Structural Linguistics*, Chicago, 1951.

Heinz A. *Funkcja egzocentryczna rzeczownika* (The Exocentric Function of the Substantive), Wrocław, 1957.

Herdan G. *The Calculus of Linguistic Observations*, The Hague, 1962.

Hjelmslev L. *Omkring sprogteoriens grundlaeggelse*, Copenhagen, 1943, transl. from Danish to English by F. Y. Whitfield: *Prolegomena to a Theory of Language*, Baltimore, 1953, 2 nd ed.: Madison, 1961.

Hockett Ch. F. *A Course in Modern Linguistics*, New York, 1958.

Ingarden R. *Das literarische Kunstwerk*, Halle, 1931.

Jakobson R. *Kindersprache, Aphasie und allgemeine Lautgesetze*, Uppsala, 1941.

— "Kazańska szkoła polskiej lingwistyki i jej miejsce w światowym rozwoju fonologii" (The Kazan School of Polish Linguistics and Its Place in the General Development of Phonology), *Biuletyn PTJ*, XIX, 1960, pp. 9–34.

— "Poetyka w świetle językoznawstwa" (Poetics in the Light of Linguistics), *Pamiętnik Literacki* (Literary Memoir) LI, Wrocław, 1960, 2, pp. 431–474.

— *Selected Writings*, I: *Phonological Studies*, The Hague, 1962.

— *Essais de linguistique générale*, transl. to French by N. Ruwet, Paris, 1963.

— Halle M. *Fundamentals of Language*, Leiden, 1956.

Jespersen O. *The Philosophy of Grammar*, London, 1925, 3rd ed.: London, 1958.

Kotarbińska J. "Pojęcie znaku" (The Concept of Sign), *Studia Logica* VI, Warszawa, 1958, pp. 57–143.

Kuryłowicz J. *Esquisses linguistiques*, Wrocław, 1960.

Lee Whorf B. *Language, Thought and Reality*, Cambridge, Mass., 1956.

Lingua Posnaniensis, T. I–XV, Poznań, 1949–1972.

Martinet A. *L'économie des changements phonétiques*, Berne, 1955.

— *Eléments de linguistique générale*, Paris, 1960.

Meillet A. *Introduction à l'étude comparative des langues indo-européennes*, 7th ed., Paris, 1934.

Meillet A., Cohen M. *Les langues du monde*, 2nd ed., Paris, 1952.

Mieščaninov I. I. *Glagol* (The Verb), Moscow, 1949.
Milewski T. *Zarys językoznawstwa ogólnego* (Outline of General Linguistics), Lublin–Kraków, 1947—1948.
— *Wstęp do językoznawstwa* (Introduction to Linguistics), 4th ed. Warszawa, 1962.
— "Założenia językoznawstwa typologicznego" (Principles of Typological Linguistics), *Biuletyn PTJ*, XXI, Kraków, 1962, pp. 3–39.
Paul H. *Prinzipien der Sprachgeschichte*, 1880, 5th ed., Tübingen, 1920.
Rozwadowski J. M. *Wybór pism* (Selected Writings), T. III: *Językoznawstwo ogólne* (General Linguistics), Warszawa, 1960.
Safarewicz J. "Sur la fonction predicative du verbe personnel", *Biuletyn PTJ*, IX, Kraków, 1949, pp. 76–79.
Sapir E. *Language*, New York, 1921.
de Saussure F. *Cours de linguistique générale*, Paris, 1913.
Schaff A. *Wstęp do semantyki* (Introduction to Semantics), Warszawa, 1960.
Schmidt W. *Die Sprachfamilien und Sprachenkreise der Erde*, Heidelberg, 1926.
Shannon C. E., Weaver W. *The Mathematical Theory of Communication*, Urbana, 1949.
Smoczyński P. *Przyswajanie przez dziecko podstaw systemu językowego* (The Child's Acquistion of Language System Fundamentals), Łódź, 1955.
Trubeckoj (Trubetzkoy) N. S. *Grundzüge der Phonologie*, Praha, 1939.
Vendryes J. *Le langage*, Paris, 1921.
Weinreich U. *Languages in Contact. Findings and Problems*, New York, 1953.
Wiener N. *Cybernetics and Society*, 1950, 1954, 1956 print. in US.
Woodward P. M. *Probability and Information Theory, with Applications to Radar*, London, 1955.
Wundt W. *Völkerpsychologie*, I Bd.: *Die Sprache*, Leipzig, 1900, 3rd ed., Leipzig, 1912.
Zabrocki L. *Usilnienie i lenicja w językach indoeuropejskich i w ugrofińskim* (Reinforcement and Lenition in Indo-European languages and in Finno-Ugric), Poznań, 1951.
— "Przesuwka spółgłosek w dialektach ormiańskich" (Consonant Shifts in Armenian dialects), *Spraw. Pozn. Tow. Przyj. Nauk*, No. 1 (63), pp. 39–49.
— *Wspólnoty komunikatywne w genezie i rozwoju języka niemieckiego* (Speech Communities in the Origin and Development of the German Language), Part I.: *Prehistoria języka niemieckiego* (The Prehistory of the German language), Wrocław, 1963.
Zawadowski L. *Constructions grammaticales et formes périphrastiques*, Wrocław, 1959.
Zipf G. K. *Human Behaviour and the Principle of Least Effort*, Cambridge, Mass., 1949.

INDEX OE TERMS DEFINED

INDEX OF THE NAMES OF LANGUAGES

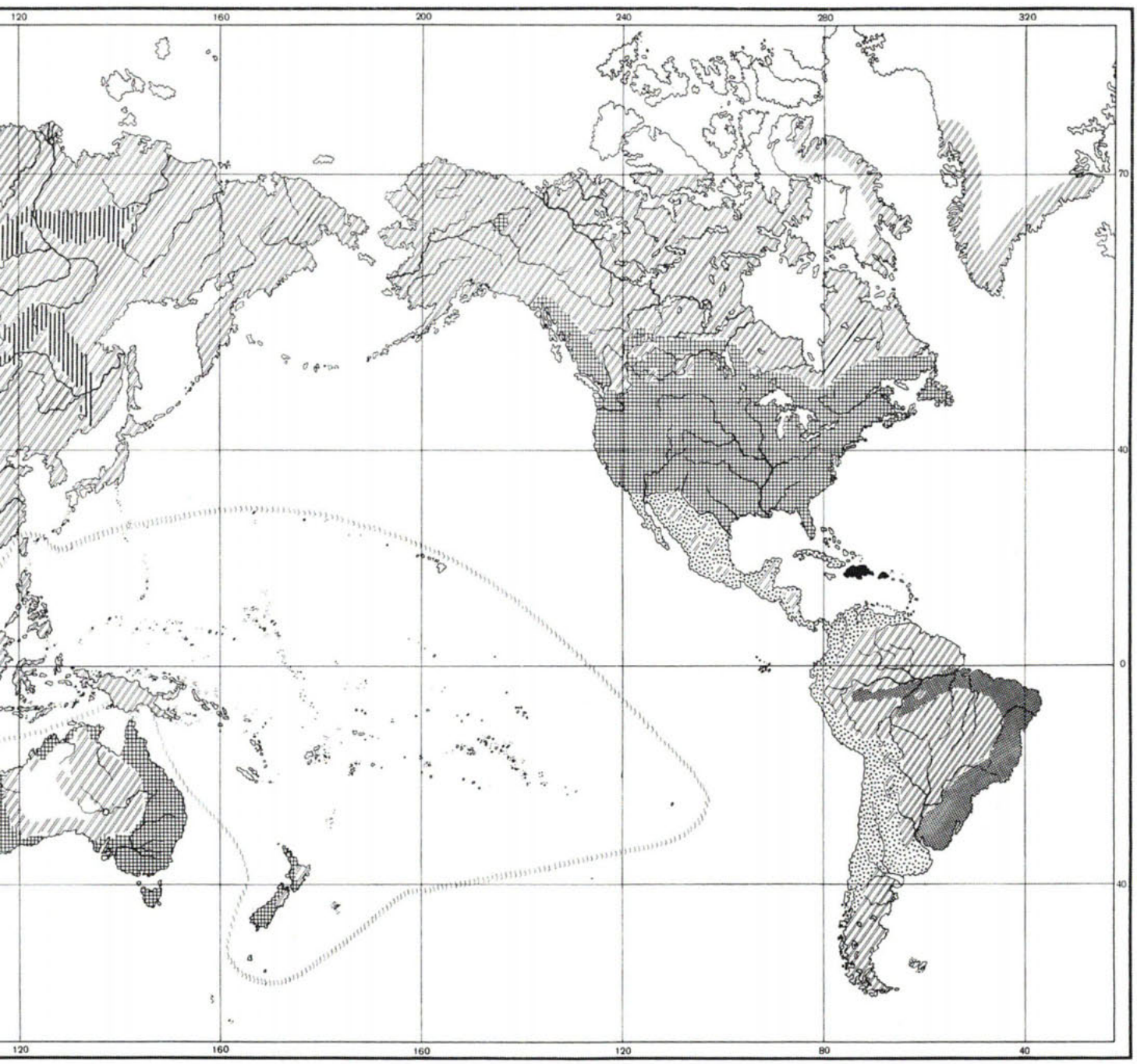
M I — Map of the languages of the world in 1500 A. D.
M II — Map of the languages of the world in 1900 A. D.

INDEX OF NAMES

MAP OF THE LANGUAGES OF THE WORLD IN 1900 A.D.

LEGEND

- Portuguese
- Spanish
- French
- English
- Russian
- Dutch (Low-Franconian dialect of the German group)
- Remaining Indo-European Languages
- Non-Indo-European Languages
- Extent of Austronesian Languages


120
160
200
240
280
320
70
40
0
40
120
160
160
120
80
40

MAP OF THE LANGUAGES OF THE WORLD IN 1500 A.D.

LEGEND

American and Palaeo-Asiatic Languages

1.Ges-Tapuya 2. Tupi-Guarani 3.Arawak 4. Carib 5. Kichua 6. Chibcha 7. Maya-Quiche 8. Mixtec-Zapotec-Otomi 9. Nahuatl (Aztec) of the Uto-Aztecan family 10. Algonquian family 11.Iroquoian family 12.Penutian family 13.Hokan 14. Yakon-Takelma and Oregon families 15.Kwakiutl-Salish-Kwileut 16. Tsimshian 17. Na-Dene 18. Eskimo languages 19. Aleutian languages 20. Chukchi-Kamchadal family 21. Ainu-Gilyak family 22. Yenisei-Ostiac group

Japhetic Languages

1. Caucasian languages 2. Basque

Australian, Papuan and Andamanese Languages

1. Tasmanian family 2. Paleo-Australian family 3. Narrinyeri 4. Central-Australian family 5. Aranda 6. Papuan languages 7. Andamanese family

Austronesian Family

1. Cham 2. Malagassi 3. Javanese 4. Malayan 5. Moluccas languages 6. Melanesian languages 7. Maori 8. Hawaiian

Austro-Asiatic Family

1. Nahali 2. Munda 3. Khasi 4. Palaung 5. Mon 6. Khmer 7. Malaka group 8. Nicobar group

Dravidian Family

1. Brahui 2. Tamil 3. Kanarese 4. Kurukh 5. Molto 6. Kui-Gondi 7. Telugu

Sino-Tibetan Family

1. Tibeto-Himalayan 2. North-Assamese group 3. South-Assamese group 4. Arakan-Burmese group 5. Vietnamese 6. Muong 7. Thai 8. Karen group 9. Chinese

Khoin Family

1. San 2. Nama

Negro League

1. Bantu group 2. Kwa group 3. Benue River group 4. Togo group 5. Atlantic group 6. Mande (Ngo-Nke) Family 7. Songoi family 8. Kanuri 9. Tubu 10. Nubian 11. Central-Nile group 12. South-Nile (Nile-Hamitic) group 13. Hausa-Musgu group (Bantu-Semitic mixture)

Semito-Hamitic Family

1. Bedawje 2. Somali 3. Galla 4. Berber group 5. East-Aramaic 6. South-Arabic group 7. Ethiopian group 8. Arabic

Altaic Family

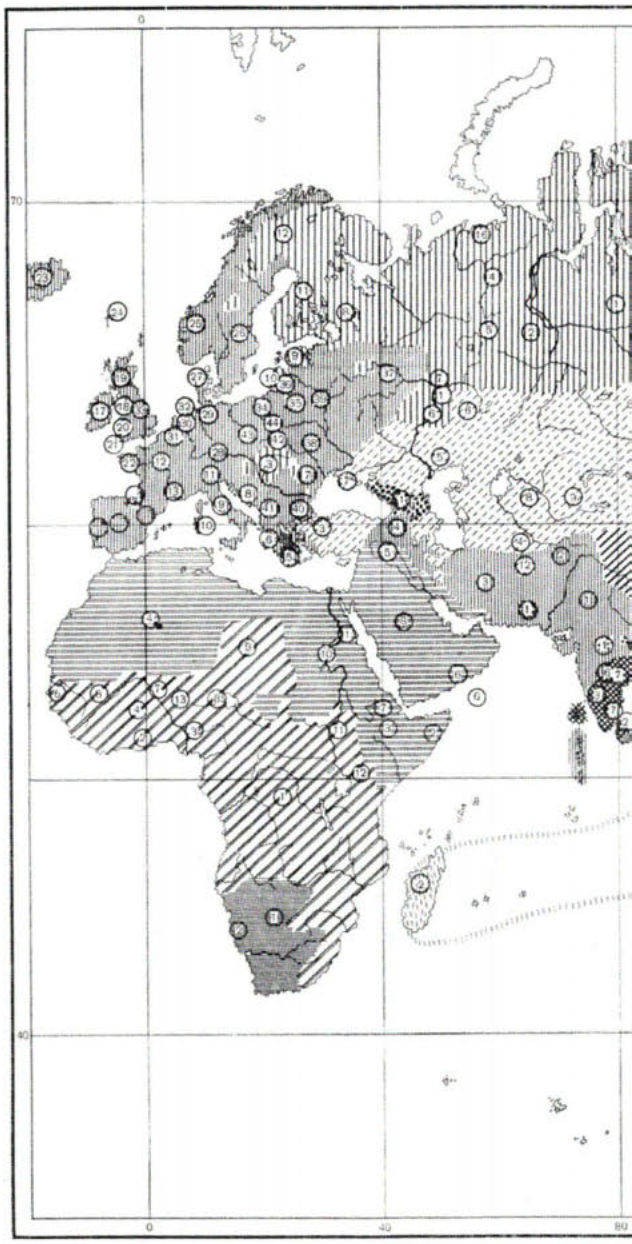
1. Chuvash 2. Yakut 3. Osmanli 4. Turkoman 5. Kirghiz 6. Bashkir 7. Tartar 8. Uzbek 9. Sart 10. Uighur 11. Sayan 12. Mogol 13. Dahur 14. Kalmuk 15. Buryat 16. Chalcha 17. Manchu 18. Tungus languages 19. Korean 20. Japanese

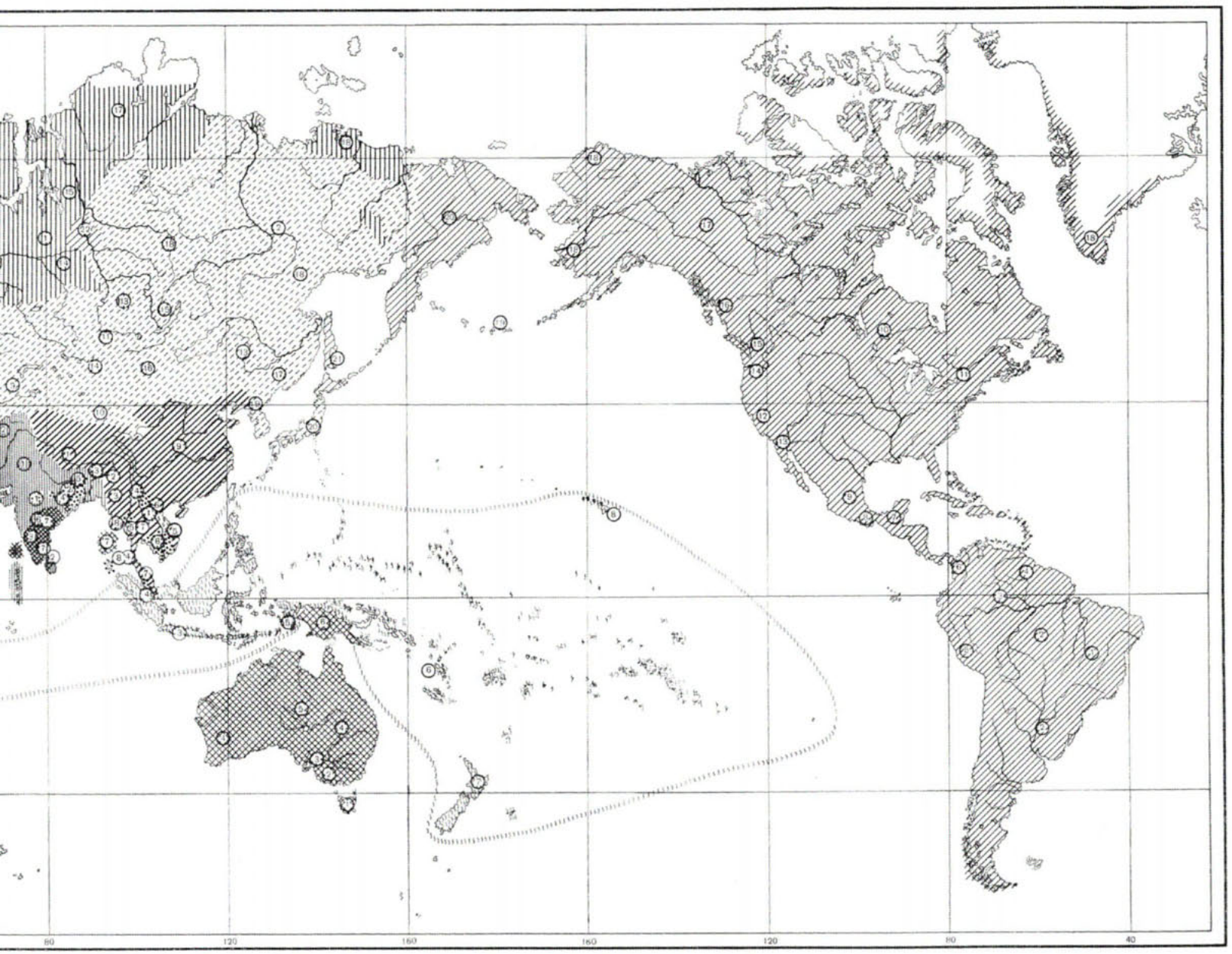
Uralic Family

1. Ostyak 2. Vogul 3. Hungarian 4. Zyrien 5. Votyak 6. Mordvin 7. Cheremiss 8. Karelian 9. Estonian 10. Livonian 11. Suomi 12. Lapp 13. Kamassin 14. Ostyak-Samoyed 15. Yenisei-Samoyed 16. Yurak 17. Tavgi 18. Yukaghir

Indo-European Family

1. Indic group 2. Afghan 3. Persian 4. Armenian 5. Greek 6. Albanian 7. Rumanian 8. Dalmatian dialects 9. Italian 10. Sardinian 11. Raeto-Romanic group 12. French 13. Provençal 14. Catalonian 15. Spanish 16. Portuguese 17. Irish 18. Manx 19. Scots 20. Welsh 21. Cornish 22. Breton 23. Icelandic 24. Shetland Islands language 25. Norwegian 26. Swedish 27. Danish 28. High German 29. Low German 30. Dutch 31. Flemish 32. Frisian 33. English 34. Prussian 35. Lithuanian 36. Latvian 37. Russian 38. Ukrainian 39. White Russian 40. Bulgarian- Macedonian group 41. Serbo-Croatian-Slovene group 42. Slovak 43. Czech 44. Polish




80
120
160
160
120
80
40