Linguistics as an Exact Science

New Ways of Thinking, Hence of Talking, about Facts Vastly Alter the World of Science, Emphasizing the Need for Investigation of Language

By Benjamin Lee Whorf

THE revolutionary changes that have occurred since 1890 in the world of science — especially in physics but also in chemistry, biology, and the sciences of man — have been due not so much to new facts as to new ways of thinking about facts. The new facts themselves of course have been many and weighty; but more important still, the realms of research where they appear — relativity, quantum theory, electronics, catalysis, colloid chemistry, theory of the gene, Gestalt psychology, psychoanalysis, unbiased cultural anthropology, and so on — have been marked to an unprecedented degree by radically new concepts, by a failure to fit the world view that passed unchallenged in the great classical period of science, and by a groping for explanations, reconciliations, and restatements.

I say new ways of *thinking* about facts, but a more nearly accurate statement would say new ways of *talking* about facts. It is this *use of language upon data* that is central to scientific progress. Of course, we have to free ourselves from that vague innuendo of inferiority which clings about the word "talk," as in the phrase "just talk"; that false opposition which the English-speaking world likes to fancy between talk and action. There is no need to apologize for speech, the most human of all actions. The beasts may think, but they do not talk. "Talk" *ought to be* a more noble and dignified word than "think." Also we must face the fact that science begins and ends in talk; this is the reverse of anything ignoble. Such words as "analyze," "compare," "deduce," "reason," "infer," "postulate," "theorize," "test," and "demonstrate," mean that whenever a scientist does something, he talks about this thing that he does. As Leonard Bloomfield has shown, scientific research begins with a set of sentences which point the way to certain observations and experiments, the results of which do not become fully scientific until they have been turned back into language, yielding again a set of sentences which then become the basis of further exploration into the unknown. This scientific use of language is subject to the principles or the laws of the science that studies all speech — linguistics.

As I was concerned to point out in a previous article, "Science and Linguistics," in The Review for April, we all hold an illusion about talking, an illusion that talking is quite untrammeled and spontaneous and merely "expresses" whatever we wish to have it express. This illusory appearance results from the fact that the obligatory phenomena within the apparently free flow of talk are so completely autocratic that speaker and listener are bound unconsciously as though in the grip of a law of nature. The phenomena of language are background phenomena, of which the talkers are unaware or, at the most, very dimly aware — as they are of the motes of dust in the air of a room, though the linguistic phenomena govern the talkers more as gravitation than as dust would. These automatic, involuntary patterns of language are not the same for all men but are specific for each language and constitute the formalized side of the language, or its "grammar" — a term that includes much more than the grammar we learned in the textbooks of our school days.

From this fact proceeds what I have called the "linguistic relativity principle," which means, in informal terms, that users of markedly different grammars are pointed by their grammars toward different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers but must arrive at somewhat different views of the world. (A more formal statement of this point appears in my article of last April.) From each such unformulated and naïve world view, an explicit scientific world view may arise by a higher specialization of the same basic grammatical patterns that fathered the naïve and implicit view. Thus the world view of modern science arises by higher specialization of the basic grammar of the western Indo-European languages. Science of course was not *caused* by this grammar; it was simply colored by it. It appeared in this group of languages because of a train of historical events that stimulated commerce, measurement, manufacture, and technical invention in a quarter of the world where these languages were dominant.

The participants in a given world view are not aware of the idiomatic nature of the channels in which their talking and thinking run, and are perfectly satisfied with them, regarding them as logical inevitables. But take an outsider, a person accustomed to widely different language and culture, or even a scientist of a later era using somewhat different language of the same basic type, and not all that seems logical and inevitable to the participants in the given world view seems so to him. The reasons that officially pass current may strike him as consisting chiefly of highly idiomatic *façons de parler*. Consider the answers that were at one time given even by learned men to questions about nature: Why does water rise in a pump? Because nature abhors a vacuum. Why does water quench fire? Because water is wet or because the fiery principle and the watery principle are antithetical. Why do flames rise? Because of the lightness of the element fire. Why can one lift a stone with a

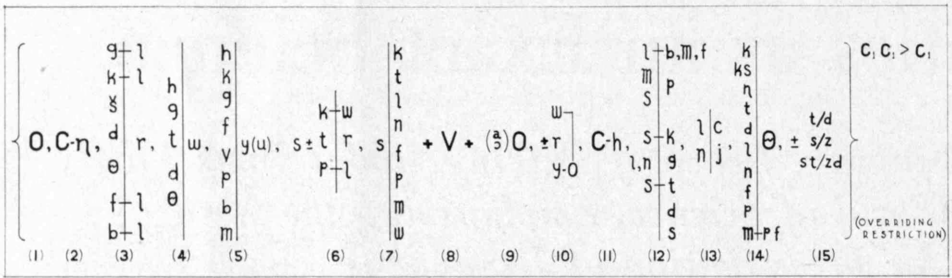


Fig. 1. Structural formula of the monosyllabic word in English (standard midwestern American). The formula can be simplified by special symbols for certain groups of letters, but this simplification would make it harder to explain. The simplest possible formula for a monosyllabic word is C+V, *and some languages actually conform to this. Polynesian has the next most simple formula,* O, C+V. *Contrast this with the intricacy of English word structure, as shown above.*

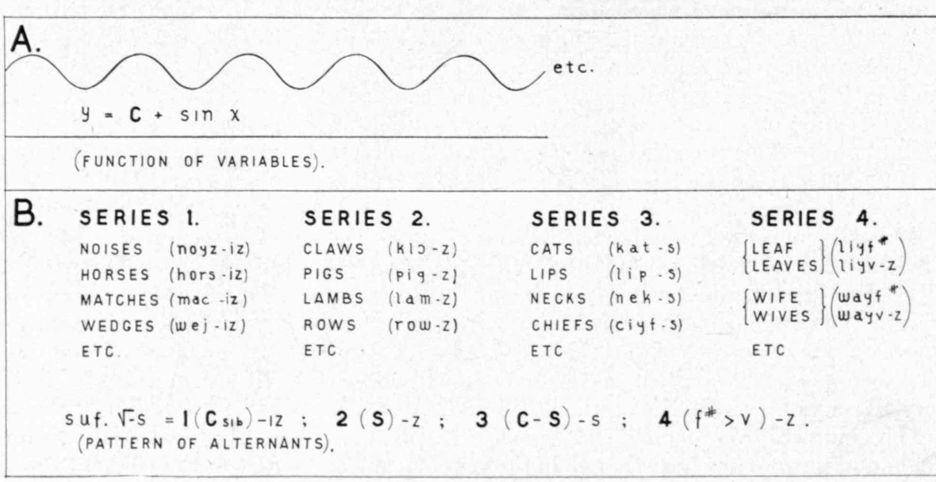
leather sucker? Because the suction draws the stone up. Why does a moth fly toward a light? Because the moth is curious or because light attracts it. If once these sentences seemed satisfying logic but today seem idiosyncrasies of a peculiar jargon, the change is not because science has discovered new facts. Science has adopted new linguistic formulations of the old facts, and now that we have become at home in the new dialect, certain traits of the old one are no longer binding upon us.

We moderns are not yet in a position to poke fun at the wiseacres of old who explained various properties of water by its wetness. The terminology which we apply to language and cultural phenomena is often of a piece with the wetness of water and nature's abhorrence of a vacuum. The researches of linguists into the ways of languages many and diverse are needed if we are to think straight and escape the errors which unconscious acceptance of our language background otherwise engenders. An increasing contribution from linguistics to the general philosophy of science is demanded by the new ways of thinking implied by those new realms of science cited at the beginning of this essay. It is needed for science's next great march into the unknown.

The situation is not likely to be aided by the philosophical and mathematical analyst who may try to exploit the field of higher linguistic symbolism with little knowledge of linguistics itself. Unfortunately the essays of most modern writers in this field suffer from this lack of apprenticeship training. To strive at higher mathematical formulas for linguistic meaning while knowing nothing correctly of the shirt-sleeve rudiments of language is to court disaster. Physics does not begin with atomic structures and cosmic rays, but with motions of ordinary gross physical objects and symbolic (mathematical) expressions for these movements. Linguistics likewise does not begin with meaning nor with the structure of logical propositions, but with the obligatory patterns made by the gross audible sounds of a given language and with certain symbolic expressions of its own for these patterns. Out of these relatively simple terms dealing with gross sound patterning are evolved the higher analytical procedures of the science, just as out of the simple experiments and mathematics concerning falling and sliding blocks of wood is evolved all the higher mathematics of physics up into quantum theory. Even the facts of sound patterning are none too simple. But they illustrate the unconscious, obligatory, background phenomena of talking as nothing else can.

For instance, the structural formula for words of one syllable in the English language (Fig. 1) looks rather complicated; yet for a linguistic pattern it is rather simple. In the English-speaking world, every child between the ages of two and five is engaged in learning the pattern expressed by this formula, among many other formulas. By the time the child is six, the formula has become ingrained and automatic; even the little nonsense words the child makes up conform to it, exploring its possibilities but venturing not a jot beyond them. At an early age the formula becomes for the child what it is for the adult; no sequence of sounds that deviates from it can even be articulated without the greatest difficulty. New words like "blurb," nonsense words like Lewis Carroll's "mome raths," combinations intended to suggest languages of savages or animal cries, like "glub" and "squonk" — all come out of the mold of this formula. When the youth begins to learn a foreign language, he unconsciously tries to construct the syllables according to this formula. Of course it won't work; the foreign words are built to a formula of their own.

Variables and alternants: A shows by graph and by mathematical formula (equation) an interrelation of variables. B illustrates by extensible examples and by a pattern formula an interrelation of alternants. The formula means that the English suffix which is theoretically ("by root," √‾) a final s is actualized in any given case by one of four alternants: after a sibilant-ending consonant, by -iz; after any sonant (vowel or consonant), by -z; after any voiceless (nonsonant) consonant by -s; except that after the special alternant f#, it is actualized by -z, the f# alternating to v.



Usually the student has a terrible time. Not even knowing that a formula is back of all the trouble, he thinks his difficulty is his own fault. The frustrations and inhibitions thus set up at the start constantly block his attempts to use foreign tongues. Or else he even *hears* by the formula, so that the English combinations that he makes sound to him like real French, for instance. Then he suffers less inhibition and may become what is called a "fluent" speaker of French — bad French!

If, however, he is so fortunate as to have his elementary French taught by a theoretic linguist, he first has the patterns of the English formula explained in such a way that they become semiconscious, with the result that they lose the binding power over him which custom has given them, though they remain automatic as far as English is concerned. Then he acquires the French patterns without inner opposition, and the time for attaining command of the language is cut to a fraction (see Fig. 2). To be sure, probably no elementary French is ever taught in this way — at least not in public institutions. Years of time and millions of dollars' worth of wasted educational effort could be saved by the adoption of such methods, but men with the grounding in theoretic linguistics are as yet far too few and are chiefly in the higher institutions.

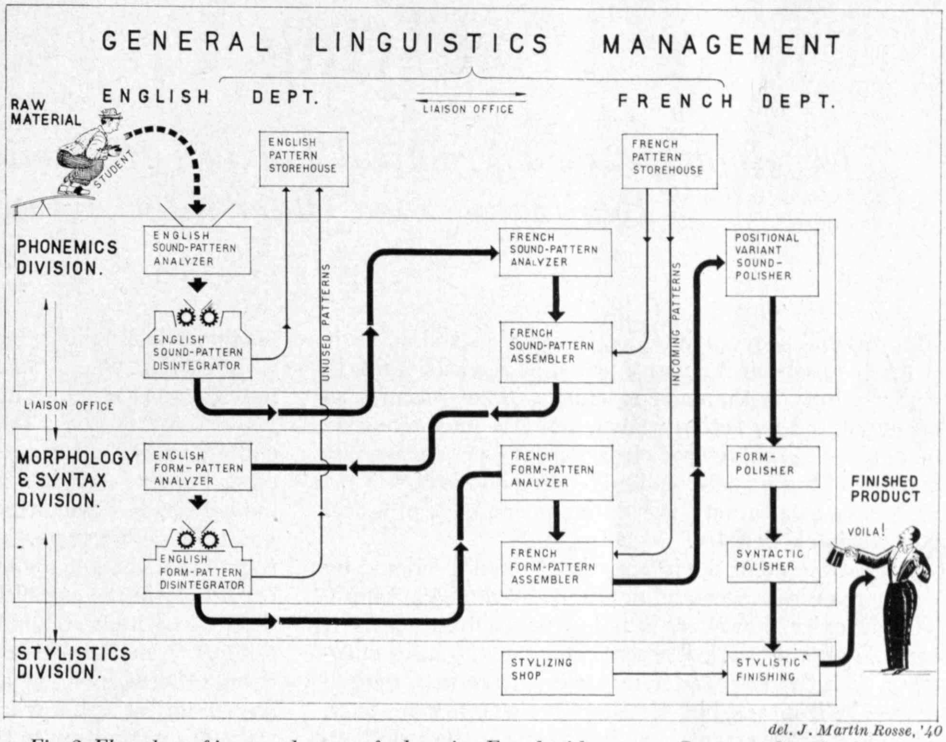


Fig. 2. Flow sheet of improved process for learning French without tears. Guaranteed: no bottlenecks in production.

Let us examine the formula for the English monosyllabic word. It looks mathematical, but it isn't. It is an expression of pattern symbolics, an analytical method that grows out of linguistics and bears to linguistics a relation not unlike that of higher mathematics to physics. With such pattern formulas various operations can be performed, just as mathematical expressions can be added, multiplied, and otherwise operated with; only the operations here are not addition, multiplication, and so on, but are meanings that apply to linguistic contexts. From these operations conclusions can be drawn and experimental attacks directed intelligently at the really crucial points in the welter of data presented by the language under investigation. Usually the linguist does not need to manipulate the formulas on paper but simply performs the symbolic operations in his mind and then says: "The paradigm of Class A verbs can't have been reported right by the previous investigator"; or "Well, well, this language must have alternating stresses, though I couldn't hear them at first"; or "Funny, but *d* and *l* must be variants of the same sound in this language," and so on. Then he investigates by experimenting on a native informant and finds that the conclusion is justified. Pattern-symbolic expressions are exact, as mathematics is, but are not quantitative. They do not refer ultimately to number and dimension, as mathematics does, but to pattern and structure. Nor are they to be confused with theory of groups or with symbolic logic, though they may be in some ways akin.

Returning to the formula, the simplest part of it is the eighth term (the terms are numbered underneath), consisting of a *V* between plus signs. This means that every English word contains a vowel (not true of all languages). As the *V* is unqualified by other symbols, any one of the English vowels can occur in the monosyllabic word (not true of all syllables of the polysyllabic English word). Next we turn to the first term, which is a zero and which means that the vowel may be preceded by nothing; the word may begin with a vowel — a structure impossible in many languages. The commas between the terms mean "or." The second term is *C* minus a long-tailed *n*. This means that a word can begin with any single English consonant except one — the one linguists designate by a long-tailed *n*, which is the sound we commonly write *ng*, as in "hang." This *ng* sound is common at the ends of English words but never occurs at the beginnings. In many languages, such as Hopi, Eskimo, or Samoan, it is a common beginning for a word. Our patterns set up a terrific resistance to articulation of these foreign words beginning with *ng*, but as soon as the mechanism of producing *ng* has been explained and we learn that our inability has been due to a habitual pattern, we can place the *ng* wherever we will and can pronounce these words with the greatest of ease. The letters in the formula thus are not always equivalent to the letters by which we express our words in ordinary spelling but are unequivocal symbols such as a linguist would assign to the sounds in a regular and scientific system of spelling. (*Continued on page 80*)

LINGUISTICS AS AN EXACT SCIENCE

(*Continued from page 63*)

According to the third term, which consists of two columns, the word can begin with any consonant of the first column followed by *r*, or with *g*, *k*, *f*, or *b* followed by *l*. The *s* with a wedge over it means *sh*. Thus we have "shred," but not "shled." The formula represents the fact that "shled" is un-English, that it will suggest a Chinese's pronunciation of "shred" or a German's of "sled" (*sl* is permitted by term 7). The Greek theta means *th;* so we have "thread" but not "thled," which latter suggests either a Chinese saying "thread" or a child lisping "sled." But why aren't *tr*, *pr*, and *pl* in this third term? Because they can be preceded by *s* and so belong in term 6. The fourth term similarly means that the word can begin with a consonant of the first column followed by *w*. *Hw* does not occur in all dialects of English; in ordinary spelling it is written backwards, *wh*. If the dialect does not have *hw*, it pronounces the spelled *wh* simply as *w*. *Thw* occurs in a few words, like "thwack" and "thwart," and *gw*, oddly enough, only in proper names, like Gwen or Gwynn. *Kw*, ordinarily spelled *qu*, can have *s* before it and therefore belongs in term 6.

The fifth term indicates that the word may begin with one of the first-column consonants followed by *y*, but only when the vowel of the word is *u;* thus we have words like "hue" (*hyuw*), "cue," "few," "muse." Some dialects have also *tyu*, *dyu*, and *nyu* (e.g., in "tune," "due," and "new"), but I have set up the formula for the typical dialects of the northern United States, which have simple *tu*, *du*, *nu* in these words. The sixth term indicates pairs that can commence a word either alone or preceded by *s*, that is, *k*, *t*, or *p* followed by *r*, also *kw* and *pl* (think of "train," "strain"; "crew," "screw"; "quash," "squash"; "play," "splay"). The seventh term, which means the word can begin with *s* followed by any one of the consonants of the second column, completes the parts of the word that can precede its vowel.

The terms beyond the eighth show what comes after the vowel. This portion is rather more complex than the beginning of the word, and it would take too long to explain everything in detail. The general principles of the symbolism will be clear from the preceding explanations. The ninth term, with its zero, denotes that a vowel can end the word if the vowel is *a* — which means (1) the vowel of the article "a" and the exclamation "huh?" and (2) the vowel of "pa," "ma," and the exclamations "ah!" and "bah!" — or the vowel can end the word if it is the *aw* sound, as in "paw," "thaw." In some dialects (eastern New England, southern United States, South British) the vowel ending occurs in words which are *spelled* with *ar*, like "car," "star" (*ka*, *sta*, in these dialects), but in most of the United States dialects and in those of Ireland and Scotland these words end in an actual *r*. In eastern New England and South British dialects, but not in southern United States, these words cause a linking *r* to appear before a vowel beginning a following word. Thus for "far off" your Southerner says *fa of;* your Bostonian and your Britisher say *fa rof*, with a liquid initial *r;* but most of the United States says *far of*, with a rolled-back *r*. For some dialects, term 9 would be different, showing another possible final vowel, namely, the peculiar sound which the Middle Westerner may notice in the Bostonian's pronunciation of "fur," "cur" (*fə*, *kə*) and no doubt may find very queer. This funny sound is common in Welsh, Gaelic, Turkish, Ute, and Hopi, but I am sure Boston did not get it from any of these sources.

Can one-syllable words end in *e*, *i*, *o*, or *u*? No, not in English. The words so spelled end in a consonant sound, *y* or *w*. Thus, "I," when expressed in formula pattern, is *ay*, "we" is *wiy*, "you" is *yuw*, "how" is *haw*, and so on. A comparison of the Spanish *no* with the English "No!" shows that whereas the Spanish word actually ends with its *o* sound trailing in the air, the English equivalent closes upon a *w* sound. The patterns to which we are habituated compel us to close upon a consonant after most vowels. Hence when we learn Spanish, instead of saying *como no*, we are apt to say *kowmow now;* instead of *si*, we say our own word "see" (*siy*). In French, instead of *si beau*, we are apt to say "see bow."

Term 10 means that *r*, *w*, or *y* may be interpolated at this point except when the interpolation would result in joining *w* and *y* with each other. Term 11 means that the word may end in any single English consonant except *h;* this exception is most unlike some languages, e.g., Sanskrit, Arabic, Navaho, and Maya, in which many words end in *h*. The reader can figure out terms 12, 13, and 14 if he has stuck so far. A small *c* means *ch* as in "child"; *j* is as in "joy." Term 13, which contains these letters, expresses the possibility of words like "gulch," "bulge," "lunch," and "lounge." Term 14 represents the pattern of words like "health," "width," "eighth" (*eytθ*), "sixth," "xth" (*eksθ*). Although we can say "nth" power or "fth" power, it takes effort to say the unpermitted "sth" power or "hth" power. "Hth" would be symbolized **eycθ*, the star meaning that the form does not occur. Term 14, however, allows both *mθ* and *mpf*, the latter in words like "humph" or the recent "oomph" (*umpf*). The elements of term 15 may be added after anything — the *t* and *s* forms after voiceless sounds, the *d* and *z* after voiced sounds. Thus, "towns" is *tawnz*, with *wnz* attained by term 10 plus 11 plus 15; whereas "bounce" is *bawns*, with *wns* by 10 plus 12. Some of the combinations resulting in this way are common; others are very rare but still are possible English forms. If Charlie McCarthy should pipe up in his coy way, "Thou oomphst, dost thou not?"; or a Shakespearean actor should thunder out, "Thou triumphst!" the reason would be that the formula yields that weird sputter *mpfst* by term 14 plus term 15. Neither Mr. Bergen nor Mr. Shakespeare has any power to vary the formula.

The overriding factor applicable to the whole expression is a prohibition of doubling. Notwithstanding whatever the formula says, the same two consonants cannot be juxtaposed. While by term 15 we can add *t* to "flip" and get "flipt" ("flipped"), we can't add *t* to "hit" and get "hitt." Instead, at the point in the patterns where "hitt" might be expected, we find simply "hit" (I hit it yesterday, I flipt it yesterday). Some languages, such as Arabic, have words like "hitt," "fadd," and so on, with both paired consonants distinct. The Creek Indian language permits three, e.g., *nnn*.

The way the patterns summarized in this formula control the forms of English words is really extraordinary. A new monosyllable turned out, say, by Walter Winchell or by a plugging adman concocting a name for a new breakfast mush, is struck from this mold as surely as if I pulled the lever and the stamp came down on his brain. Thus linguistics, like the physical sciences, confers the power of prediction. I can predict, within limits, what Winchell will or won't do. He may coin a word "thrub," but he will not coin a word "srub," for the formula cannot produce a *sr*. A different formula indicates that if Winchell invents any word beginning with *th*, like "thell" or "therg," the *th* will have the sound it has in "thin," not the sound it has in "this" or "there." Winchell will not invent a word beginning with this latter sound.

We can wheeze forth the harshest successions of consonants if they are only according to the patterns producing the formula. We easily say "thirds" and "sixths," though "sixths" has the very rough sequence of four consonants, *ksθs*. But the simpler "sisths" is against the patterns and so is harder to say. "Glimpst" (glimpsed) has *gl* by term 3, *i* by 8, *mpst* by 12 plus 15. But "dlinpfk" is eliminated on several counts: Term 3 allows for no *dl*, and by no possible combination of terms can one get *npfk*. Yet the linguist can say "dlinpfk" as easily as he can say "glimpsed." The formula allows for no final *mb*; so we do not say "lamb" as it is spelled, but as *lam*. "Land," quite parallel but allowed by the formula, trips off our tongues as spelled. It is not hard to see why the "explanation," still found in some serious textbooks, that a language does this or that "for the sake of euphony" is on a par with nature's reputed abhorrence of a vacuum.

The exactness of this formula, typical of hundreds of others, shows that while linguistic formulations are not those of mathematics, they are nevertheless precise. We might bear in mind that this formula, compared with the formulation of some of the English (or other) grammatical patterns that deal with meaning, would appear like a simple sum in addition compared with a page of calculus. It is usually more convenient to treat very complex patterns by successive paragraphs of precise sentences and simpler formulas so arranged that each additional paragraph presupposes the previous ones, than to try to embrace all in one very complex formula.

Linguistics is also an experimental science. Its data result from long series of observations under controlled conditions, which, as they are systematically altered, call out definite, different responses. The experiments are directed by the theoretic body of knowledge, just as with physics or chemistry. They usually do not require mechanical apparatus. In place of apparatus, linguistics uses and develops *techniques*. Experimental need not mean quantitative. Measuring, weighing, and pointer-reading devices are seldom needed in linguistics, for quantity and number play little part in the realm of pattern, where there are no variables but, instead, abrupt alternations from one configuration to another. The mathematical sciences require exact measurement, but what linguistics requires is, rather, exact "patternment" — an exactness (*Continued on page 82*)

LINGUISTICS AS AN EXACT SCIENCE

(*Continued from page 81*)

of relation irrespective of dimensions. Quantity, dimension, magnitude, are metaphors since they do not properly belong in this spaceless, relational world. I might use this simile: Exact measurement of lines and angles will be needed to draw exact squares or other regular polygons, but measurement, however precise, will not help us to draw an exact circle. Yet it is necessary only to discover the principle of the compass to reach by a leap the ability to draw perfect circles. Similarly, linguistics has developed techniques which, like compasses, enable it without any true measurement at all to specify *exactly* the patterns with which it is concerned. Or I might perhaps liken the case to the state of affairs within the atom, where also entities appear to alternate from configuration to configuration rather than to move in terms of measurable positions. As alternants, quantum phenomena must be treated by a method of analysis that substitutes a point in a pattern under a set of conditions for a point in a pattern under another set of conditions — a method similar to that used in analysis of linguistic phenomena.

Physics and chemistry, dealing with inanimate matter, require chiefly inanimate apparatus and substances for their experiments. As conducted today upon a large scale, they require highly wrought physical equipment at every step, immense investments in physical plant. Their experiments are costly to conduct, both absolutely and relatively to the number of scientists. Experimental biology uses much inanimate apparatus, too, but its fundamental apparatus is its experimental animals and plants and their food, housing, and growth facilities. These also are expensive in the quantities needed. No one grudges the expense, either here or in the physical sciences, so long as an increase in human knowledge and welfare is promised.

The apparatus of linguistics is much less expensive than that of these sciences, but it, too, costs money. The experimental linguist, like the biologist, uses and must have experimental animals. Only, his "animals" are human. They are his informants and must be paid for working with him. Sometimes he must make trips to Indian reservations or African villages where his informants live; at other times it is more economical to transport them to him. They provide the field for experimental investigation. They are apparatus, not teachers. It is as important to study in this way languages of Indians, Africans, and other aborigines as it is to study the English dialects of Brooklyn, Boston, Richmond, or London.

While informants are the basic apparatus, the linguist can improve and speed up his work with the aid of mechanical tools, just as the biologist studies his animals and plants with the aid of microscopes, x-ray machines, and other costly instruments. The linguist is aided by judicious use of good phonographic reproducing devices. Much could also be done with the help of business machines.

Although linguistics is a very old science, its modern experimental phase, which stresses the analysis of unwritten speech, could be called one of the newest. So far as our knowledge goes, the science of linguistics was founded, or put on its present basis, by one Panini in India several centuries before Christ. Its earliest form anticipated its most recent one. Panini was highly algebraic, i.e., pattern symbolic, in his treatment; he used formulas in a very modern way for expressing the obligatory patterns of Sanskrit. It was the Greeks who debased the science. They showed how infinitely inferior they were to the Hindus as scientific thinkers, and the effect of their muddling lasted two thousand years. Modern scientific linguistics dates from the rediscovery of Panini by the Western world in the early Nineteenth Century.

Yet linguistics is still in its infancy so far as concerns wherewithal for its needed equipment, its supply of informants, and the minimum of tools, books, and the like. Money for mechanical aids, such as I referred to above, is at present only a happy dream. Perhaps this condition results from lack of the publicity the other sciences receive and, after all, fairly earn. We all know now that the forces studied by physics, chemistry, and biology are powerful and important. People generally do not yet know that the forces studied by linguistics are powerful and important, that its principles control every sort of agreement and understanding among human beings, and that sooner or later it will have to sit as judge while the other sciences bring their results to its court to inquire into what they mean. When this time comes, there will be great and well-equipped laboratories of linguistics as there are of other exact sciences.

FLYING UP

(*Continued from page 65*)

ground without any forward run and could keep it in the air indefinitely with satisfactory control. It was able to fly backward, sideways, or forward, could make turns at any desired radius, and could be turned completely about in the air while remaining directly over one spot. Enough thus has been done to prove that controlled vertical flight is possible.

Other experimenters have been working along similar lines but with indifferent success. Toward the end of the World War two Austro-Hungarians, Lieutenant Stefan Petroczy and Professor Theodore von Kármán, developed a machine to replace kite balloons as a means of observation. They actually built a helicopter which had a gross weight of about 3,200 pounds and was powered with three 120-horsepower rotary engines. It carried one observer and a machine gun in a tanklike enclosure on top, and enough fuel for a flight of one hour. It was never allowed to fly freely, but was attached to stabilizing ropes. After fifteen successful flights, it was finally damaged badly in a landing and was abandoned. Although the machine had demonstrated a certain degree of value from the military angle, it lacked any means of propelling itself from point to point in the air.

The world was still taking large grains of salt with reports that a couple of bicycle mechanics named Wright had a flying machine that really flew, when another inventor in America began jotting down notes and making sketches on the backs of old (*Continued on page 84*)