

LEVELS OF TECHNICALITY IN SCIENTIFIC COMMUNICATION

RENATA TAGLIACCOZZO

Mental Health Research Institute, The University of Michigan, Ann Arbor, MI 48104, U.S.A.

(Received September 1975)

Abstract—The large number of specialized languages which have resulted from the rapid development of modern science and technology, although meant to facilitate communication within each area of specialization, may create barriers to understanding between different levels of competence and specialization. Ways have to be found for identifying and describing the fundamental features of technical and scientific writing and for clarifying its differences from what is called “common” language.

In this paper, the style of scientific writing has been discussed from the point of view of its lexical, semantic, and use characteristics. On the assumption that the level of technicality of a scientific publication is responding to the degree of training and specialization of its readership, a comparison was made between the language of a group of articles taken from the *Scientific American* and the language of articles given as references by them. Distribution and overlap of vocabulary in the two groups of articles were analyzed. Some differences between the writings in two areas of science were found.

How can we identify a “technical” or “scientific” language, as distinct from “ordinary” language, that is, from the language that is commonly used to deal with the basic aspects of everyday life? Can we describe distinctive features that will enable us to recognize a fragment of written text as belonging to one rather than to the other type of language? Better still, can we formulate rules for content analysis so precise that a computer can be instructed to assign a text to the correct category?

The problem of specifying and explaining technical languages is in some way equivalent to the problem of translation. Technical languages are analogous to foreign languages: they are used by particular groups of people, who share some common activity or interest; their elements, although different in form, are equivalent in meaning and can be translated from one into the other language; and they can be learned with appropriate training and practice. But perhaps it is more appropriate to compare the technical language to a dialect, that is, a language-variety spoken by a subgroup in the larger language community of people using a common language[1]. As in the case of dialects, some technical languages may be only slightly different from the ordinary language, and therefore be intelligible outside the group of habitual users, while others are so different as to sound like foreign languages.

Different levels of technicality are found not only in spoken languages but also—and even more—in written text. A “highly technical” text can usually be recognized at first sight as different from a text written on the same subject for a lay audience or for an audience less proficient in the use of the particular specialized language.

Languages exhibiting a very high degree of technicality are sometimes labeled “jargons”[2]. The term, when used by outsiders, has a pejorative connotation, implying not only that the language in question presents difficulty of understanding to the noninitiated, but also that it is somewhat ugly-sounding and clumsy. In addition, the technicality of jargons is often suspected of being misused, or of being unnecessary and misleading. When the label “jargon” is used by insiders, however, its connotation seems to be positive. In this case the term makes an appeal to the sense of camaraderie, the understanding and the bond of common interests of the group.

The problems of communicating across different languages are today important for their implications which have extended from interpersonal communication to man-machine communication. Language can be an obstacle to effective communication even among people who speak the same language; it certainly is when the parties attempting to communicate speak different sublanguages. Doctor-patient and lawyer-client communication are examples of verbal exchanges in which no understanding occurs unless the technical language of the person trained in the particular discipline is translated into the common language of the untrained[3]. Modern societies, with their increasing emphasis on consumerism and accountability to the public from

the various professions, demand that pursuers of the various sciences, arts and crafts be skilled in communicating technical matters in nontechnical language. This demand seems to run counter to the need for preserving that clarity, precision, and conciseness which traditionally are associated with scientific language.

The interactive information systems now being designed for a user population with different levels of scientific and technical competence (e.g. doctors, technicians, paramedical personnel, patients) require that translation of the language from one level to another be possible. The question of compatibility and flexibility of languages in the interaction between user and information system is one which can mean the difference between success and failure of innovative technology.

The term "technical language", often used interchangeably with the term "scientific language", would seem to indicate that the use of this special language is confined to science or technology. Actually there are numerous technical languages which belong neither to the field of science nor to that of technology. The language of legal documents or the language of religious sermons, for instance, may be as "technical" and as different from the common language as the language of physics[4]. Even areas of human endeavor that cannot be counted among the professions, trades, and occupations to which the term "technological", in its original meaning, applies (Gr. "techne" = art, craft, practical skill) have developed their own technical languages (see, for example, the writings in various fields of sport or entertainment, which may be as obscure to people lacking the proper training or interest as those of any scientific discipline). We will therefore define a technical language as a special sublanguage or dialect developed by speakers who have particular competence in some area of knowledge, to communicate about matters belonging to that area.

Technical languages are probably as old as the areas of specialized knowledge or practical skill to which they apply. As a consequence of the rapid development of science in the past 300 yr. the number of scientific and technical languages has increased at an accelerated pace. We can look at the proliferation of special languages as an answer to the need to facilitate communication within special subgroups of the population. That this process has at the same time created barriers to understanding between groups and in the general population is well evident.

From the expressions "the language of science" or "the language of technology" one would be led to conclude that there exists a universal language used in all the branches of science and technology. Actually the somewhat nebulous concept "the language of science" refers to the way scientific thinking and scientific processes are organized rather than to the language used to communicate about specific subject matters. But undoubtedly, when it comes to *writing* about scientific matters, conventions and norms dictated by publishers and editors aim at a certain uniformity of writing, as attested by those numerous tools of scientific writing, i.e. the manuals of style and instructions for authors. A considerable body of literature exists about the writing of science, mostly devoted to criticizing the excesses and distortions to which too great a zeal in applying the conventions of scientific writing may lead[5, 6].

1. LEXICAL ANALYSIS OF TEXTS

When one tries to identify the features of scientific and technical prose, one should consider the argument of whether a technical language is simply one which is rich in technical terms. This argument is based on the notion that each specialized field of knowledge uses, in addition to the terminology of the common language, a special terminology characteristic of that field, and that the totality of its special terms differentiates the technical language from common language or from other technical languages. Leaving aside for the moment the question of what is the minimum level of technical terminology qualifying a text as "technical", the problem of recognizing a technical language would then consist in differentiating technical terms from common terms.

To establish criteria for identifying technical terms, one could experiment with different approaches, alone or in combination. At least three come to mind which I shall discuss briefly in the following paragraphs.

(a) *Morphological characteristics*

It would be unrealistic to expect that precise rules could be stated which would identify all the technical terms of a specialty by their formal characteristics. But if a substantial portion of them

could be so identified, the problem would be in part solved. This type of approach looks attractive from the point of view of empirical testing. It does not involve analyzing the meaning of the words and it is promising in regard to computerization.

We know that in the past, with the expansion of man's knowledge of the universe around him, many words had to be created to name new things, concepts and events. This continuous process has been accelerated in recent times in accord with the increasing pace of scientific discovery and technologic development. A multitude of technical terms had to be invented almost overnight. The artificiality and recency of origin of these neologisms is often betrayed by the fact that they have retained a distinctive morphology [7]. The natural process of change and assimilation into the common language, in fact, did not have time to take place. For example, (a) Latin terms used in the taxonomies of the biological sciences and in descriptive anatomy of medicine (e.g. *Quercus alba* = white oak; *Felis domesticus* = cat; *musculus dilator pupillae* = the muscle which dilates the pupil). (b) words containing characteristic chemical radicals (e.g. -phenyl-, -hydroxy-, -amide). (c) words containing Greek affixes or stems which have kept their original form (e.g. -hyper- = over; macro- = large; -odont- = tooth). (d) eponyms, such as those used in the medical field (e.g. Bright's disease, Parkinson's disease). (e) long words (the probability of technical words being long, especially in the case of composite neologisms, is high).

Unfortunately, the number of technical terms that can be identified on the basis of their morphology is quite limited. Many technical terms have in fact been assimilated into the common language and have lost their distinctive morphological structure (e.g. therapy from Gr. *therapeia*; diet from Gr. *diaita*). On the other hand, words which have kept their original morphology—and can be easily identified as transliterations of a single or multiple Greek word—are so commonly used in the ordinary English language, that one would hesitate to classify them unequivocally as technical terms (e.g. phobia, schizophrenia). And then there are words that have been taken bodily from the common language and have been given a special technical meaning (e.g. ring, conditioning, feedback). There is nothing, of course, in the morphological structure of the latter that could give a clue to their being technical terms.

(b) *Semantic characteristics*

To identify the semantic characteristics of technical terms one must abandon the analysis of structure and form for the more complex and arduous task of understanding how words perform their main function, that is, the function of conveying meaning [8]. Ideally, scientific and technical languages should be performing this function more efficiently than common language can, especially in terms of conciseness, precision, and clarity.

That technical terms contribute to the *conciseness* of a text can be clearly seen when the need for translating technical prose into common prose arises. Where one technical word may be sufficient to express a concept, a long expression is often required to convey its meaning in ordinary language. In this respect technical terms can be thought of as convenient short-hand symbols. But this is not always the case. Many technical terms have their equivalents, or near-equivalents, in the common language and in some cases it happens that the common term is the shorter of the two (for example, in the case of the descriptive structural terminology of chemistry).

A much more fundamental quality of language that technical terms are supposed to optimize is *precision*. Technical terms usually define and identify the things or concepts for which they stand more exactly than common words do. While the expression "a broken shoulder" may be adequate to describe an unfortunate state of one's own body and may even be sufficient for admission to a hospital, a more precise localization of the fracture, better expressed by one of the three anatomical terms "scapula", "clavicle", and "humerus", would probably be needed by the physician in charge of the treatment. Or, to take another example, brand names of chemical compounds seem to serve well in the exchanges of everyday life, but to the chemist the true nature of a chemical compound is better revealed by the technical name, i.e. the structural name showing the molecular composition (e.g. "acetylsalicylic acid" for "aspirin"; "17-hydroxy-11-dehydrocorticosteronacetate" for "cortisone").

The concept of terminological precision does not lend itself to easy definition or measurement. For some authors it overlaps to some extent with the concept of terminological consistency which, for a given technical language, is the degree of conformity displayed by the scientific or technical community in using agreed-upon (standardized) terms [9]. For these authors

the fewer the synonyms contained in the language of a scientific discipline, the more terminologically consistent is the language of that discipline.

Precision of technical terminology is better defined as lack of ambiguity. It is important that technical terms mean only one thing, and always the same thing, that is, their meaning does not have to depend on context. We must add, though, that this statement is true only within the boundaries of a particular technical or scientific sublanguage. The word "ring" for instance can be said to be a technical term and therefore to have only one meaning (out of the 25 or more listed in Webster's dictionary) when it is made clear that we are referring to the field of chemistry (or physics, or mathematics). In its idealized form, the language of science is conceived by some authors as being completely devoid of homonyms but rich in synonyms [10]. It is contrasted to the language of poetry, which presents the opposite features: no two words have the same meaning, and each word has an infinite number of meanings depending on context (*ibid.*).

It is doubtful whether *clarity*, the third equality claimed—or aspired to—by technical and scientific writing, is a property of single terms or of the syntactic and stylistic structure of the text. Obscure language can be generated even in the absence of technical terms, as the numerous books of instruction on scientific writing and the heated controversy on the matter amply testify [5]. Clarity goes hand in hand with lack of ambiguity. But again we must emphasize that this is true only in a given language community, and that even there overly specialized terms or rarely used technical terms may baffle everybody, including the specialist.

(c) *Use characteristics*

A third criterion for lexical analysis consists in comparing technical words to common words from the viewpoint of relative *frequency of use*. It is to be expected that while the frequency distribution of common words will be approximately the same in different types of texts, technical words will have a higher concentration in writings belonging to scientific or technical disciplines. In his research aimed at the automatic construction of a technical vocabulary, D. C. Stone has tried several statistical measures which could be applied to a text to discriminate "specialty" from "non-specialty" words on the basis of their relative frequency of use [11].

A comparison of technical vocabularies to common-use vocabularies would also show that most technical and scientific terms have a recent origin. To quote a comment on language change in the 1964 edition of the Webster-Merriam English Dictionary, the "...astounding developments in science, medicine, space and other areas have produced floods of new words". We don't know which portion of the 100,000 new entries added to the new edition of the Dictionary are in fact scientific or technical words. But, if we extrapolate from the often reported observation that 90% of the scientists who ever existed are alive today, it is easy to conclude that many of the words now used to communicate in science and technology did not exist one or two generations ago.

2. STYLISTIC ANALYSIS OF TEXT

The problem of identifying the features of technical and scientific prose may be approached from the point of view of stylistic analysis. This type of analysis disregards the terminology of texts, and investigates instead their syntactical patterns and other features of sentence construction.

One definition of stylistic analysis characterizes style as a "... pattern of recurrent selections from the inventory of optional features of a language" [12, p. 3]. Under this aspect a style has strong resemblance to a dialect, in particular a social dialect. In the same way as patterns of use differentiate the language spoken by different social strata within the same linguistic community, stylistic features differentiate the language of poetry, for instance, from the language of science [1].

There appears to be general consensus, within the scientific and technical community, that scientific writing has its own style, which differs, for better or for worse, from other forms of writing. Conventions have imposed on scientific writing stylistic rules which are believed to be conducive to clear and concise exposition, while at the same time conforming to the social norms of science. The requirement that expository prose, in addition to being clear and concise, be impersonal, objective, and somewhat detached is reflected in the abundant use of the passive

form, in the avoidance of personal pronouns and adjectives, and in the nominalization of verbs. The excesses induced by too eager an adoption of these conventions have been the source of heated debates [5, 6]. In his paper "Sounder thinking through clearer writing" F. P. Woodford gives a merciless expose of the worst products of "scientific" style. The author, Woodford writes, "... takes what should be lively, inspiring and beautiful and, in an attempt to make it seem dignified, chokes it to death with stately abstract nouns; next, in the name of scientific impartiality, he fits it with a complete set of passive constructions to drain away any remaining life's blood or excitement; then he embalms the remains in molasses of polysyllable, wraps the corpse in an impenetrable veil of vogue words, and buries the stiff old mummy with much pomp and circumstance in the most distinguished journal that will take it" [13, p. 743]. Woodford maintains also that the present standards of scientific writing have a corrupting influence on the development of the young scientist, as they impair "... his ability to write, to read and to think" [ibid. p. 745].

It is not my intent to discuss the case of unskilled or awkward use of the language in scientific and technical writing. There are enough good writers among scientists to show convincingly that clarity and elegance of expression are not exclusive prerogatives of literary prose. For the purpose of this study, I will assume that on the average scientific and technical writing is handled with an adequate degree of competence.

"Specialized" languages have been recognized as having stylistic rules of their own. Crystal and Davy, for instance, identify a language of religion, a language of newspaper reporting, and a language of legal documents, among the written languages [4]. Bross, Shapiro and Anderson have studied the language of written surgical reports [2]. According to these authors, the structure of specialized languages has evolved in response to the function for which they are intended. While the function of some specialized languages is to persuade, or to convince (e.g. the language of the courtroom lawyer or the advertiser), scientific sublanguages, such as the language of written surgical reports, are used solely to transmit information. This function is facilitated by some of the restrictive features of the particular sublanguage, as, for instance, the almost total elimination of semantic ambiguity, the scarcity of synonymous terms and the simplification of syntax. The authors do not hesitate to call the language of surgical reports "jargon", making it clear that it is intended for subject specialists only. Although they deplore the dangers of language insularity, they do not offer any suggestion on how to break the communication barriers created by specialization.

Funkhouser and Maccoby, on the other hand, are concerned with research aimed at identifying the textual variables of scientific writing that favor the communication of scientific information to a lay audience [14]. They are interested in those stylistic and lexical features which make for easier understanding of scientific or technical material, such as the use of short sentences, examples and concrete words.

One of the stylistic features independent from special rules and conventions of particular sublanguages, and therefore reflecting more fundamental properties of different languages, is the frequency distribution of the vocabulary. Statistical analysis of text vocabularies has been used in literary analysis to solve problems of authorship attribution [15]. From the pioneer work of YULE [16] to more recent investigations by MOSTELLER and WALLACE [17] attempts have been made at identifying the characteristic traits of style by quantitative measures of vocabulary distribution. One such measure is the type-token ratio.

The type-token ratio (TTR) reflects the richness of the vocabulary of a text, or, said another way, its verbal diversification. It is obtained by dividing the number of different words appearing in the text (types) by the total number of words in the text (tokens). A text in which the ratio is high, that is, in which the number of types is large in relation to the total number of tokens, is a text in which there is a large proportion of different words, while a low type-token ratio will indicate abundance of repetitive occurrences of words.

The type-token ratio varies with the length of the text on which it is measured. When the length of the text increases, the TTR decreases. To make this measure independent from length of text, some computational modifications have been devised at different times, such as Yule's characteristic K [16], or the logarithmic type-token ratio introduced by HERDAN [18]. Both techniques have not been exempt from criticism (see Herdan for criticism of Yule, and Weitzman for criticism of Herdan [19]).

3. AN INVESTIGATION OF TYPE-TOKEN RATIOS AND VOCABULARY DISTRIBUTION

As a starting point for our investigation we used data from a computational analysis of different genres of present-day American English published by KUČERA and FRANCIS[20]. In addition to other data, they give the word-frequency distribution of 50 samples, each consisting of approx. 2000 tokens, taken from different kinds of writing. Eight of the 50 samples were placed in a category that they called "scientific and learned writing".

We compared the TTRs of these 8 samples with those of two other categories of writing, i.e. newspaper writing and literature. It appears that on the basis of type-token distribution it is possible to discriminate scientific and learned writing from newspaper writing and, to a lesser extent, from literary writing (see Fig. 1). The 8 samples of "scientific and learned writing" have in fact TTRs in the lower range of the scores, while the 9 samples of newspaper reporting and editorials have TTRs in the higher range[‡]. The 8 samples of the third group (literary writing) are somewhat closer and partially overlap the first group.

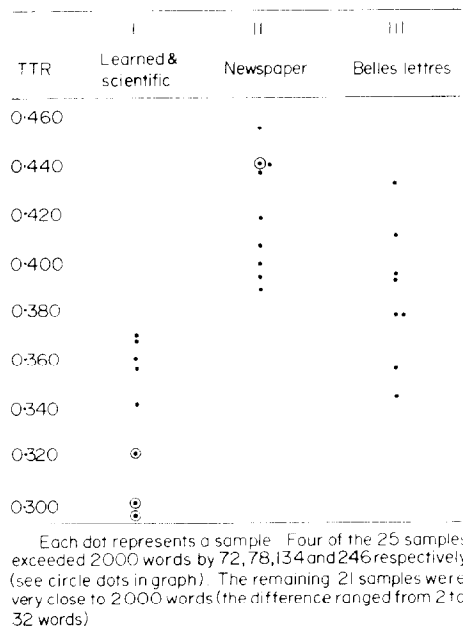


Fig. 1. Distribution of type-token ratios (TTR) in 25 samples of approx. 2000 words each, from three categories of writing (data from Kučera & Francis, 1967: pp. 331-355) (Samples of writing).

The above data would suggest that a scientific and scholarly style of writing is on the whole more repetitive than the writing of other forms of prose. Given the heterogeneity of the corpus from which the samples were selected and in the absence of a clear definition of "scientific and learned writing", however, no general conclusion is warranted. In addition, the references of the sources give the impression that the samples of the second group (press writing) are an aggregate of different items, rather than continuous texts. This could explain the higher TTRs (lower degree of repetitiveness) in this group. Also there is no clear-cut differentiation between the sources of Group I and the sources of Group III: some of the samples qualified as "Belles Lettres" could have been included in the group of "Learned Writings" and vice versa.

It is evident that, to compare the type-token ratios of different kinds of prose, more stringent rules for the selection and the categorization of the material should be applied. This is even more of a requirement when the two groups of texts that one wants to compare are much more similar than, let's say, the groups of scientific and literary texts.

It appears that for any field, there are at least 3 types of scientific and technical literature, one addressed to a restricted audience characterized by a high degree of special training and competence, the second addressed to a more general audience with interest and some basic

[‡]The excess number of words in three of the eight samples of learned and scientific writing may have resulted in a lowering of the TTR, but not enough to change the relative position of the three groups on the TTR scale.

training in science, but lacking the high degree of specialization of the first group, and the third addressed to the public at large. A publication of the first type will usually be written in a more technical language than a publication of the second type, and this, in turn, in a more technical language than a publication of the third type.

The hypothesis that we wish to test is that the type-token ratio (TTR) of scientific and technical writings varies in relation to the degree of technicality of the text and, more specifically, that highly technical writings have lower TTRs than writings of lower technicality. Publications belonging to the first of three groups outlined above will therefore have lower TTRs than publications of the second group, and these in turn will have lower TTRs than those of the third group.

Arguments in favor of this hypothesis can be given. (a) Highly technical writings are often confined to specialized areas and are more restricted in subject matter than more general writings. The vocabulary of specialized technical writings will be more limited and the language more repetitive; consequently, the type-token ratio will be lower. (b) The requirement that scientific writing be clear and precise necessarily induces the avoidance of those stylistic features of literary writing which could lead to ambiguity—for instance, the use of anaphoric terms. As a result the frequent repetition of the same words or expressions will occur. (It must be added also that the rules of literary style are often disregarded not only for the sake of precision, but also because highly technical writing is by many considered outside the realm of literary writing.) Articles directed to less specialized audiences, that rely for their acceptance not only on the interest of the material, but also on some elegance of form, may on the other hand try to avoid a too repetitive style. (c) Some people challenge the commonly held belief that the writing of research work has the primary goal of communicating information. Ziman, for instance, is convinced that a scientific paper is “. . . a cunningly contrived piece of rhetoric. It has only one purpose; it must persuade the reader of the veracity of the observer, his disinterestedness, his logical infallibility, and the complete necessity of his conclusion” [21, p. 319]. If it is true that writings addressed to the inner circle of experts—the highly technical articles—have the main purpose of persuading, while those addressed to a more general readership are merely intended to convey information, it is not inconceivable that the former have a more repetitive style (and therefore a lower TTR) than the latter.

To test our hypothesis, we compared a group of articles from the *Scientific American* with the group of reference articles listed in their bibliographies. The articles published in the *Scientific American* cover a broad range of scientific fields and are written for a readership that may have technical knowledge in some areas, but not in others. To appeal to the journal's heterogeneous public, the articles have to be written in a language which is understandable by a non-specialized audience. The reference articles, on the other hand, for the most part have appeared in specialized journals, which are organs of scientific societies or represent particular fields of research. They, therefore, do not have the same editorial constraints and can use the language appropriate to their more restricted and specialized readership. Our assumption is that the reference articles are written in a more technical language than the articles of the *Scientific American*. The TTRs of the former should therefore be lower than those of the latter. Following is a description of the procedure that we used.

From the 1968–1973 volumes of the *Scientific American* we selected a sample of 5 articles in the area of psychophysics (perceptual and motor behavior in man) and a sample of four articles in the area of genetics. The conditions for the selection were that the articles should be 2000 words long or more and that they should have at least three usable references in their bibliographies. (A usable reference was defined as a research paper 2000 words long or more with the exclusion of review papers.) The 9 articles of the *Scientific American* constituted our *Group I*. *Group II* was formed by 27 reference articles (3 for each Group-I article) selected from the bibliographies of the *Scientific American* articles.† One of the reference articles had subsequently to be eliminated, reducing the total to 26. Most of these reference articles are evidently the source material on which the *Scientific American* articles are based. (The whole set of articles is listed in the Appendix.)

To avoid the complications involved in analyzing articles of different length—due to the fact

†The bibliographies of the articles of the *Scientific American* are printed separately from the articles, in the back of each issue. The majority of them consist of 3–4 articles.

that the type-token ratio of a corpus varies with the length of the corpus—we decided to take article samples of equal length, namely 1000 word samples. We decided also to disregard function words (articles, prepositions, conjunctions, etc.), so that the 1000 word samples would consist entirely of *content* words. To extract from the articles the 1000 word samples, we followed Yule's technique of *spread sampling*. We took sections of text 100 words long spread out over the entire article, so that each page of the text (with a few exceptions in the case of pages taken over by tables or illustrations) would be sampled. The number of sections needed to reach the total of 1000 *content* words varied slightly from article to article, depending on the ratio of content words to function words in the entire article. If the ratio had been an even 50/50, it would have taken, of course, 20 samples per article. With a proportion of content words close to 60%, as it was the case, the number of samples was between 18 and 19.†

The intervals between the selected samples had to be adjusted to the length of the article. For instance, in an article estimated to contain 2000 content words, or twice as many words as necessary, the 100 word sections were placed at a distance of 100 words from each other, that is, the number of sampled lines was equal to the number of skipped lines. But if the article was four times as long as necessary, the 100 word sections were placed at intervals of 300 words from each other. A slightly more complicated adjustment had to be made in intermediate cases.

In preparation for our analysis, we punched the set of 100 word sections from each article into IBM cards and the cards were run under the Spitbol program, which is similar to the Snobol 4 language processor program. In the Spitbol printout the cards were typed out and numbered in the order they had been keypunched and submitted. In addition, the program produced a concordance of the text, that is, a complete alphabetic listing of the individual words found in the set of 100 word sections, along with their locations on the cards. After we deleted from the printout the function words (articles, prepositions, conjunctions, auxiliary verbs, etc.) we counted by hand the number of card locations for each content word, that is, the number of occurrences of that word in the sampled text. By grouping together words that had occurred only once, words that had occurred twice, three times, and so on, we obtained a frequency distribution of word *types* by number of occurrences. Their sum gives the total vocabulary for the 1000 word corpus. If we multiply the frequencies of types by the number of their occurrences, we obtain another frequency distribution, this one consisting of the number of *tokens* produced by each category of word-types. For instance, if three different words (types) appeared each eleven times in the corpus, the tokens produced by those types were 33; if 16 types appeared each four times, the result was 64 tokens. The sum of these products amounts to 1000, which is the total number of content words in the corpus. By dividing the number of types by the number of tokens we obtain the TTR for each of the articles under analysis.

For illustration purposes, we show the distribution of types and tokens in one of the articles that we analyzed (see Table 1). The first column lists the occurrence categories, that is, the categories of words which occurred once, twice, three times, and so on. The figures of the second column (f_x) denote the frequencies in each category; their sum is evidently equal to the number of different words (types) and is equivalent to the vocabulary of the sample. Column 4 ($f_x X$) was obtained by multiplying the X values of Col. 1 by the corresponding f_x values of Col. 2. These products represent the number of words (tokens) contributed to the text by each occurrence category; their sum is equal to the total number of words in the sample.‡

Columns 3 and 5 are the cumulative percentages of f_x and $f_x X$. They are helpful in illustrating the relative weight of the different categories in building the vocabulary and the text. We notice that the words which occurred only once (*hapax legomena*) in the sampled text were responsible for over 64% of the entire vocabulary of content words. The *hapax legomena* together with words that had a low rate of repetition (between 2 and 5 occurrences) formed the bulk of the vocabulary (93.28%). Words belonging to the high-occurrence categories—11–39 occurrences—constituted a minute portion of the vocabulary (2.6%).

The distribution of tokens follows a different curve. Here the *hapax legomena* contribute only 29.20% of the text. Words occurring from one to 5 times provide 63.72%, and words of higher rate of occurrence (11–39) are responsible for a respectable portion of the text (22.42%).

† In 23 out of the 35 articles analyzed the sample represented more than 40% of the corpus; in 10 of the remaining articles it represented more than 20%.

‡ This should be 1000 words, but small variations were acceptable, up to a maximum of 20 words (or 2%).

Table 1. Distribution of types and tokens in a 1017-word sample from a scientific article in the area of Psychophysics¹

1	2	3	4	5
X	f_x	cum % f_x	$f_x X$	cum % $f_x X$
1	297	64.43	297	29.20
2	79	81.56	158	44.74
3	32	88.50	96	54.18
4	13	91.32	52	59.29
5	9	93.28	45	63.72
6	6	94.58	36	67.26
7	5	95.66	35	70.70
8	4	96.53	32	73.84
9	2	96.96	18	75.61
10	2	97.40	20	77.58
11	1	97.61	11	78.66
13	1	97.83	13	79.94
14	1	98.05	14	81.32
15	2	98.48	30	84.27
16	3	99.13	48	88.99
23	1	99.35	23	91.25
25	2	99.78	50	96.17
39	<u>1</u>	100.00	<u>39</u>	100.00
	461		1017	

¹Perrott, R. & Nelson, A. M. "Limits for the Detection of Binaural Beats" J. Acoust. Soc. Am. 46:1477-1481, 1969.

When graphed on log-normal paper, the distribution of types and tokens in the low and intermediate ranges of occurrence—words occurring up to 10–11 times in the sample—fit a straight line fairly well (see Fig. 2). This type of graph is useful for illustrating vocabulary distributions in texts and can be used as an adjunct to TTR measures for comparing different kinds of writing.

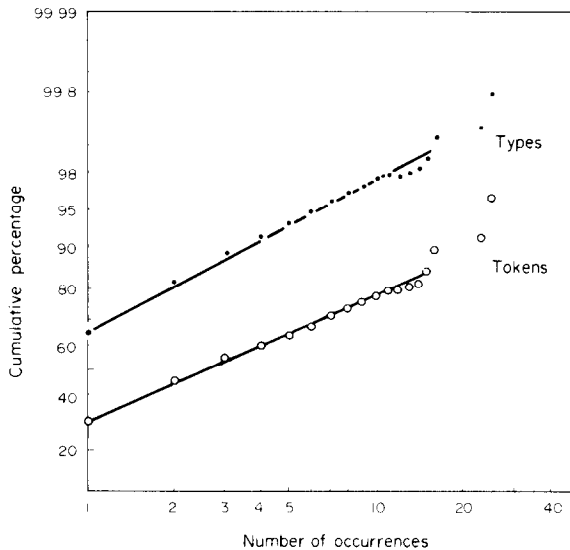


Fig. 2. Distribution of types and tokens in a sample of 1017 content words from article on Table 1.

4. RESULTS

The type-token values for the two groups of articles that we analyzed are shown in Table 2. The figures in the first line (Group I) are the type-token ratios of the *Scientific American* articles, while those in the second, third, and fourth line (Group II) are the type-token ratios of the reference articles listed in their bibliographies.

The table shows that, in the case of the five sets of articles from the field of Psychophysics,

Table 2. Type-token ratios in articles of two different levels of technicality

	Psychophysics					Genetics			
	A	B	C	D	E	F	G	H	I
Group I, Articles from <i>Scientific American</i>	0.5134	0.4811	0.5235	0.4999	0.5473	0.4755	0.4711	0.4737	0.5039
Group II, Reference Articles	0.4932	0.5005	0.5199	0.5034	0.5065	0.5351	0.5130	0.5242	0.4780
	0.5134	0.4826	0.5197	0.4976	0.5073	0.4755	0.4920	0.4833	0.5034
	0.6551	0.5941	0.5970	0.4734	0.6997	0.5440	0.4780	0.4665	0.5034

* One of the three selected reference articles of set "I" was, on reconsideration, discarded as it appeared to fit in both the categories of "research" papers and "review" papers.

the overall trend is in agreement with our hypothesis that writings of higher technicality would have smaller TTRs than writings of lower technicality. Thirteen of the 15 reference articles (higher technicality) have in fact lower TTRs than the *Scientific American* articles which make reference to them. It should be pointed out, however, that the differences are small and in two cases (the second reference article from the top in the *C* and *D* columns) almost non-existent. The difference between the means of the two groups is not statistically significant.

If the five sets of articles in Psychophysics are examined individually, it is evident that three of them (*A*, *B* and *E*) show a clear-cut and consistent relationship between the TTR of the Group I article and the TTRs of Group II articles—with the latter all being lower than the former—while in the two remaining sets no consistent relationship can be found (only in two of the reference articles is the TTR lower than in the *Scientific American* article, and in one case there is almost no difference). In the four sets of articles from the field of Genetics no specific trend can be found. Out of 11 reference articles, in fact, 6 have higher TTRs and 5 lower than the corresponding *Scientific American* article.

In the complete set of data (Psychophysics and Genetics together) the mean TTR is 0.5006 for the *Scientific American* articles and 0.4753 for the reference articles. The difference is not statistically significant ($F = 3.055$; $df = 34$; $p = 0.09$). In conclusion no significant difference can be detected between the two kinds of writing that we analyzed on the basis of their type-token ratios. In a subset of writings we noticed a trend toward a higher TTR in Group I than in Group II, but this trend is too inconsistent to form the basis for a reliable differentiation.

Hoping that a comparison of the vocabulary distribution in the two groups of writings would give us better insight into the differences, if any, of their styles, we graphed the frequency of words of different rate of occurrence both in the *Scientific American* articles, and in the sets of three reference articles. Rather than making three separate comparisons for each set of articles, we took the average of the three values for the reference articles for each point of the distribution and compared this "average" vocabulary distribution to that of the *Scientific American* article. In view of the fact that the *Scientific American* articles are usually a synthesis and condensation of the articles given as references, this procedure seemed appropriate. Our comparison had to be confined to word types occurring between one and 12 times in the sampled texts. Content words with a higher number of occurrences were too few in number—and their occurrence rate too widely spread out—to be suitable for comparison (see, for instance the frequencies of types occurring 13 times and above in the first and second column of Table 1).

In Fig. 3 we show, plotted on log-normal paper, the frequency distributions of word-types from 4 sets of articles, using solid circles to indicate the word-frequencies of the *Scientific American* articles, and open circles to indicate the average frequencies of the three reference articles of each set. The 2 graphs on top (sets *A* and *B*) show a clear-cut separation of the two distribution curves, with all the points of the first (*Scientific American* article) above the points of the second (reference articles). In the 2 graphs at the bottom of the figure (sets *C* and *G*) no clear separation can be seen between the two frequency curves, which at times cross each other and at times overlap.

These results are analogous to those obtained with the type-token ratios: the sets of articles in which the TTR of the Group-I article is consistently higher than the TTRs of the Group-II articles also show a differentiation between the frequency distributions of the word-types in the range of

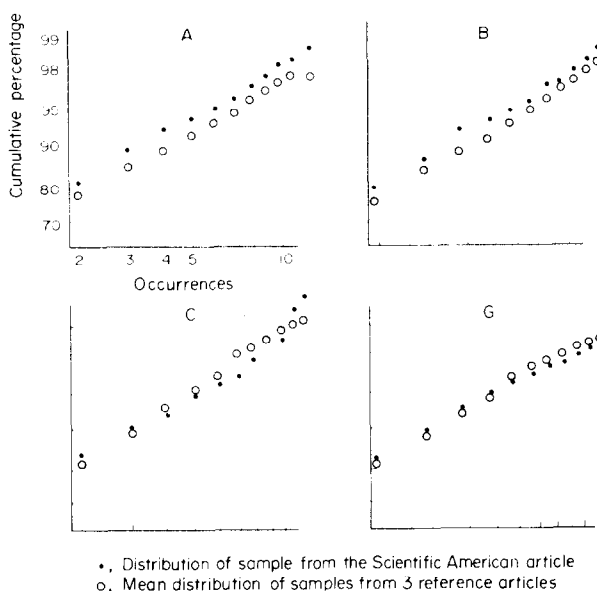


Fig. 3. Distribution of word-types occurring 2–12 times in samples of 1000 content words.

low and medium rate of occurrence. When this relationship between the TTRs of the 2 groups of articles is missing, no distinct curves for the frequency distributions of the two groups of articles can be identified.

A second aspect of vocabulary distribution that we investigated in our two groups of articles (higher-technicality and lower-technicality articles) was the extent of their vocabulary commonality. This part of the study leads us back to the concern with *kind* of terminology used in addressing different types of readers. The question to be asked at this point was: how many word-types used in the *Scientific American* articles appear also in the reference articles (or vice versa)?

The extent of overlap between vocabularies of the two groups, that is, the percentage of common word-types over the total number of word-types of the two articles to be compared is shown in Table 3. The table lists the percentage of vocabulary overlap in pairs of articles formed each by one *Scientific American* article and one reference article from the same set. The overlap is obtained by computing the ratio of the number of *types* that the two articles have in common to the total number of types of both articles. If we denote by V_{sa} the number of types appearing only in the *Scientific American*, by V_r the number of types appearing only in the reference article, and by V_c the number of types appearing in both articles, the overlap is given by: $V_c / (V_{sa} + V_r - V_c)$.

Table 3 shows that the overlap ranges from 15.5 to 23.7% for the pairs in the field of Psychophysics, with a mean overlap of 19.0%, while in the field of Genetics the range is from 14.3

Table 3. Vocabulary overlap in pairs of articles in two subject areas

PSYCHOPHYSICS		GENETICS	
A	15.5 %	F	15.3 %
	17.4 %		16.5 %
	15.7 %		15.9 %
B	18.6 %	G	18.4 %
	22.3 %		19.1 %
	20.6 %		17.1 %
C	23.7 %	H	14.3 %
	20.3 %		17.8 %
	18.5 %		16.8 %
D	18.6 %	I	20.0 %
	18.1 %		19.9 %
	19.0 %		
E	18.6 %		
	22.4 %		
	15.5 %		
Mean: 19.0 %		17.4 %	

to 20.0%, with a mean of 17.4%. On the basis of these data, then, it would seem that the amount of vocabulary commonality in pairs of articles on the same subject but at different levels of technicality may be, to some extent, dependent on the field to which the articles belong. An analysis of variance indicated, however, that the difference between the two groups of scores does not reach statistical significance ($F = 3.2$; $df = 25$, $p < 0.10$).

Another factor which might affect the percentage of overlap between the vocabularies of the pairs could be—we suspected—the authorship of the two articles. Many of the *Scientific American* articles of our sample were in fact authored by the same person (or persons) who wrote some of the reference articles. It seemed natural to expect that pairs of articles written by the same authors would exhibit a more similar vocabulary than pairs of articles written by different authors. The data show that this is the case (see Table 4). The mean overlap is 19.1 for pairs of articles with the same authors, and 17.2 for pairs of articles with different authors. The analysis of variance in this case showed that the difference is statistically significant ($F = 5.19$; $df = 25$; $p < 0.05$).

Table 4. Vocabulary overlap in pairs of articles by same or different authors

	SAME AUTHOR	DIFFERENT AUTHOR
A		15.5 % 17.4 % 15.7 %
B	22.3 % 20.6 %	18.6 %
C	23.7 % 20.3 %	18.5 %
D	18.6 % 18.1 % 19.0 %	
E	22.4 % 15.5 %	18.6 %
F	15.9 %	15.3 % 16.5 %
G	19.1 % 17.1 %	18.4 %
H	17.8 % 16.8 %	14.3 %
I	19.9 %	20.0 %
Mean:	19.1 %	17.2 %

5. SUMMARY AND CONCLUSIONS

The complex and protean nature of scientific and technical languages poses severe problems to those interested in analyzing the basic characteristics of such languages. Scientific and technical languages have been shaped by several different and sometimes conflicting forces. In contrast to the "common" language, which has slowly evolved through centuries of continuous changes, scientific and technical languages have often arisen rapidly, in response to the sudden need to name new things and concepts that were discovered. While the common language is relatively independent of constraints, scientific and technical languages are subjected to the norms of the scientific community that, although subtle, are very powerful in prescribing a particular etiquette of writing. An additional constraint impinging on scientific writing derives from the nature of scientific thought. The user of common language may be free to follow all the vagaries of fantasy, of metaphoric and poetic thinking, but the user—and especially the writer—of scientific and technical language is forced to stay within the narrow path of logical and rational exposition. And finally, a further complication for the scientific writer is that he is expected to communicate with an audience that may or may not have the same competence in his special language as he has.

Well aware of the difficulties of extricating the distinctive features of scientific and technical languages from those of common language, we have focused this study on a comparison between scientific writings of different levels of technicality. To define the level of technicality of a publication we used the degree of competence and specialization of the particular readership for

which that publication is intended. Our first step was to explore the possibility that the difference in writings of different levels of technicality would merely consist of their relative amount of technical terms, the highly technical article being richer in technical terms than the article at a lower level of technicality. We discussed the morphological, semantic and use characteristics of technical terms and pointed out the extreme difficulty of finding clear demarcation lines between the class of technical terms and the class of non-technical terms.

Using some of the methodology of computational linguistics we made a comparison between the type-token ratios (TTR) of two groups of articles, one consisting of articles that could be defined as highly technical, the other of articles to be considered less technical. The type-token ratio reflects the diversification, or richness, of the terminology of a corpus. Our hypothesis was that the highly technical articles—in our case scientific research papers—would use a less diversified vocabulary, and therefore would display a higher rate of repetition, than articles treating the same subject matter but written for a less specialized readership (articles in the *Scientific American*). There were arguments in favor of this hypothesis, such as the fact that the highly technical research paper usually covers a restricted, specialized area, and therefore uses a limited vocabulary; that its preoccupation with precision and avoidance of ambiguity favors the repetitive use of the same terms; and that its function—which is not only to inform, but also to convince—is conducive to a more repetitive style than that of the general, less technical paper.

In a sample of articles from the field of Psychophysics the overall trend appeared to be in agreement with our hypothesis. But the differences between the TTRs of the two groups of articles were neither large nor consistent throughout the five sets of articles analyzed. The trend did not hold in the second sample (articles in the field of Genetics) where the results show, in half the cases, a reverse effect (i.e. higher TTRs in the research papers than in the *Scientific American* articles). No apparent explanation could be found for this discrepancy between the two subject areas.

On the same set of data we have looked at the distribution curves of word-types of low to medium rate of occurrence (word-types, that is, occurring between a single time and 12 times in the 1000 content-word samples that we analyzed). In five out of the nine sets of articles a clear-cut separation between the word distribution curves of articles belonging to the highly technical group and those of articles of lower level of technicality was apparent. In the remaining four sets this pattern could not be detected.

A comparison of the vocabularies of pairs of articles of different level of technicality—one *Scientific American* article and one reference article—showed that the extent of overlap between content words was on the average higher in the Psychophysics than in the Genetics set of articles. Although not reaching statistical significance ($p = 0.08$), this difference may suggest that in certain areas of science the language used in highly technical articles is more similar to the language used in less technical writings than in other areas.

It is evident that the results of our study did not give a conclusive answer to the question of how to measure the level of technicality of written texts. Nevertheless the study has brought to the foreground interesting aspects of the languages used in technical and scientific prose and has suggested ways for exploring their characteristic features.

It is possible that the type-token ratio is not a sensitive enough measure for detecting stylistic differences in genres of writing as closely related as those that we have analyzed. The type-token ratio is, after all, a global measure which may hide subtle differences in vocabulary distribution. Some authors have pointed out that the vocabulary distribution cannot be represented by a single curve. For instance, the bilogarithmic straight line relation between frequency and rank, known as the Zipf law, is said to fit the intermediate frequency distribution, but not to hold for high-frequency and for low-frequency words [18, pp. 35–38]. Yule maintains that a distribution of the Poisson type is a good approximation for the great majority of words, but admits that the approximation is less close for words used with great frequency [16, pp. 45–46]. Looking at our data, we realize that in order to separate low-frequency words from intermediate frequency words an arbitrary demarcation line would have to be used, but that the high-frequency section of the distribution (words with a rate of occurrence above 10–12) could be easily identified, being discontinuous and showing large jumps in frequency. This part of the distribution clearly includes words which describe central issues in the article and therefore form the pool of terms from

which keywords or index terms are taken†. The individual analysis of the high, as well as of the intermediate and low range of the frequency distribution, may provide more fertile ground for comparing writings than the measurement of a type-token ratio did.

In the present study we have confined our analysis of vocabulary distribution to content words. We excluded the function words fearing that their high rate of repetitiveness may in some cases obscure differences in type-token ratio of what we considered the most significant part of the vocabulary, i.e. the content words. The distribution of function words, however, is by itself a factor in determining style of writing. We have some data indicating that the mean percentage of function words was higher in our Group I articles than in Group II. The difference was not large, but was statistically significant (44.4–42.0%; $p < 0.01$). These figures, which we based on an estimate of about 10% of the corpus of the 9 *Scientific American* articles and 26 reference articles, agree in principle with some data reported by Funkhouser, who found that the ratio of function words to content words was lower in a sample of technical articles than in “ordinary prose” [14, pp. 61–62]. If it can be conclusively proven that technical and scientific writing includes a lower proportion of function words than non-technical writing, then it may be interesting to test the hypothesis that the proportion of function words in a corpus varies inversely to its level of assumed technicality. An analysis of the role of function words in building the vocabulary and the text of scientific and technical articles may contribute to clarify the characteristics of scientific prose.

The two groups of articles that we have analyzed show a difference not only in the mean percentage of function words, but also in the variance of such percentage. It is perhaps not surprising that Group I, consisting of articles belonging to the same journal, have a more uniform style than Group II, as indicated in this case by a smaller variance of the percent of function words. The direct or indirect influence that journal editors have on the style of their contributors may be another variable to be added to the list of forces that shape technical and scientific writing, that we have outlined at the beginning of this section. The effects produced by such influences or pressures may well be worth analyzing.

Our analysis of the overlap between the vocabulary of the *Scientific American* articles and the vocabulary of the reference articles suggests that the extent of such overlap may be dependent on the specialty to which the writing belongs. In other words, there may be more similarity between the languages used in articles of different levels of technicality in some areas of science than in others. It seems that this should be the case. It is a common observation that articles in the field of psychology (the area in which we observed the larger overlap) appear more comprehensible to the non-trained person than articles in the field of biology, and these in turn may appear easier to read than articles in the field of physics. It would be interesting to follow this line of investigation to see whether the greater similarity between the languages of two different levels of technicality is a general fact in the “softer” sciences as compared to the “harder” sciences.

If the conclusions suggested by our data were supported by the analysis of a larger and more varied sample, the measure of vocabulary overlap may be used as an indirect method for estimating the distance between levels of technicality of articles in the same area. This method would circumvent the difficulty, that we discussed at the beginning of this article, of estimating the level of technicality of scientific articles by comparing the numbers of technical words that they contain. If we assume that the overlapping portion of the vocabulary consists primarily of common terms and that the technical terms are concentrated mainly in the non-overlapping portion, then a comparative analysis of extent of overlap could lead to an estimate of level of technicality. Comparing two articles—or two groups of articles—against a standard, one could in fact conclude that the one having the smaller overlap is the more technical. This line of investigation, too, may be interesting to pursue.

In conclusion, testing some of the hypotheses suggested above may contribute to a clarification of what we mean when we qualify a language as “technical” or “scientific”, and in particular it may help in estimating the level of technicality of such languages. No claim is made at this moment that a method for identifying the stylistic characteristics of technical prose will be

†We found that on the average nearly 40% of the high-frequency words of each article did in fact appear in the title of the article.

found solely on the basis of the statistical properties of its vocabulary, although there are some indications that in samples of overspecialized writing, i.e. in writing sufficiently removed from "common" or "literary" prose this may be the case. It is possible that in order to clarify the basic problems involved in the communication of technical information across different levels of specialization and competence, other properties of scientific languages will have to be analyzed, namely those properties, such as precision and clarity, that we have listed under the heading *semantic characteristics*. Combined efforts from various research directions will be needed to reach an understanding of the essential features of the special sublanguages that scientists use to communicate among themselves and with others.

Acknowledgement—The valuable assistance of ELIZABETH PODOLSKA and ANITE IMMELE in collecting and analyzing the data is gratefully acknowledged.

REFERENCES

- [1] R. W. BAILEY and J. L. ROBINSON, *Varieties of Present-Day English*. MacMillan, New York (1973).
- [2] I. D. J. BROSS, P. A. SHAPIRO and B. B. ANDERSON, How information is carried in scientific sub-languages. *Science* 1972, **176**, 1303–1307.
- [3] B. M. KORSCH and V. F. NEGRETE, Doctor-patient communication. *Scient. Am.* 1972 (Aug.), **227**, 66–74.
- [4] D. CRYSTAL and D. DAVY, *Investigating English Style*. Indiana Univ. Press, Bloomington, Indiana (1969).
- [5] S. BAKER, Clarity. *Science*, 1969, **166**, 365–366.
- [6] Scientific Writing (letters by various authors), *Science*, 1970, **167**, 932–933.
- [7] O. E. NYBAKKEN, *Greek and Latin in Scientific Terminology*. Iowa State Univ. Press, Ames, Iowa (1968).
- [8] A. G. OETTINGER, The semantic wall. In *Human Communication: A Unified View* (Edited by E. E. DAVID and P. B. DENES), pp. 1–12. McGraw-Hill, New York (1972).
- [9] K. BHATTACHARYYA, The effectiveness of natural language in science indexing and retrieval. *J. Docum.* 1974, **30**, 235–254.
- [10] M. NOWAKOWSKA, *Language of Motivation and Language of Actions*, pp. 130–131. Mouton, La Hague (1973).
- [11] D. C. STONE, Word statistics in the generation of semantic tools for information systems. The Moore School of Electrical Engineering, Univ. of Pennsylvania, Philadelphia Dec. (1967).
- [12] W. WINTER, Styles as dialects. In *Statistics and Style* (Edited by L. DOLEŽEL and R. W. BAILEY). Elsevier, New York (1969).
- [13] F. P. WOODFORD, Sounder thinking through clearer writing. *Science*, 1967, **156**, 743–745.
- [14] G. R. FUNKHOUSER and N. MACCOBY, Communicating specialized science information to a lay audience. *J. Commun.* 1971, **21**, 58–71.
- [15] R. W. BAILEY, Statistics and Style: a historical survey. In *Statistics and Style* (Edited by L. DOLEŽEL and R. W. BAILEY), pp. 217–236. Elsevier, New York (1969).
- [16] U. G. YULE, *The Statistical Study of Literary Vocabulary*. (Cambridge Univ. Press, 1944) Archon (1968).
- [17] F. MOSTELLER and D. L. WALLACE, *Inference and Disputed Authorship: The Federalist*. Addison-Wesley, Reading, Mass. (1964).
- [18] G. HERDAN, *Type-Token Mathematics*, Janua Linguarum, Series Major, IV. Mouton, La Hague (1960).
- [19] M. WEITZMAN, How useful is the logarithmic type-token ratio? *J. Linguist.* 1971, **7**, 237–243.
- [20] H. KUČERA and W. N. FRANCIS, *Computational Analysis of Present-Day American English*. Brown Univ. Press, Providence, R.I. (1967).
- [21] J. M. ZIMAN, Information, communication, knowledge. *Nature*, 1969, **224**, 318–324.

APPENDIX

- A. OSTER C., Auditory beats in the brain. *Scient. Am.* **229**, 94–102 (Oct. 1973).
- JEFFRESS L., H. BLODGETT, T. SANDEL and C. WOOD, III, Masking of tonal signals. *J. Acoust. Soc. Am.* **28**, 416–426 (1956).
- PERROTT R. and A. M. NELSON, Limits for the detection of binaural beats. *J. Acoust. Soc. Am.* **46**, 1477–1481 (1969).
- WERNICK J. and A. STARR, Binaural interaction in the superior olivary complex of the cat. *J. Neurophys.* **31**, 428–441 (1968).
- B. KIMURA D., The asymmetry of the human brain. *Scient. Am.* **228**, 70–78 (March 1973).
- STUDDERT-KENNEDY M. and D. SHANKWEILER, Hemispheric specialization for speech perception. *J. Acoust. Soc. Am.* **48**, 579–593 (1970).
- KIMURA D., Functional asymmetry of the brain in dichotic listening. *Cortex* **3**, 163–178 (1967).
- KIMURA D., Spatial localization in left and right visual fields. *Can. J. Psychol.* **23**, 445–457 (1969).
- C. LIPPOLD O., Physiological tremor. *Scient. Am.* **224**, 65–70 (March 1971).
- LIPPOLD O., Oscillation in the stretch reflex arc and the origin of the rhythmical, 8–12 C/S component of physiological tremor. *J. Physiol.* **206**, 359–382 (1970).
- LIPPOLD O., J. W. T. REDFEARN and J. VACU, The rhythmical activity of groups of motor units in the voluntary contraction of muscles. *J. Physiol.* **137**, 473–487 (1957).
- HALLIDAY A. M. and J. W. T. REDFEARN, An analysis of the frequencies of finger tremor in healthy subjects. *J. Physiol.* **134**, 600–611 (1956).
- D. HABER R., How we remember what we see. *Scient. Am.* **222**, 104–110 (May 1970).
- HABER R. and M. ERDELYI, Emergence and recovery of initially unavailable perceptual material. *J. Verbal Learning Verbal Behav.* **6**, 618–623, 625–628 (1967).
- HABER R. and L. G. STANDING, Direct measures of short-term visual storage. *Q. J. exp. Psychol.* **21**, 43–53 (1969).
- HABER R. and L. NATHANSON, Processing of sequentially presented letters. *Perception and Psychophys.* **5**, 359–361 (1969).
- E. THOMAS E. L., Movements of the eye. *Scient. Am.* **219**, 90–95 (August 1968).
- MACKWORTH J. and N. MACKWORTH, Eye fixations recorded on changing visual scenes by the television eye-marker. *J. opt. Soc. Am.* **48**, 439–445 (1958).

- THOMAS E. L. and E. STASIAK, Eye movements and body images. *Can. Psychiat. Assoc. J.* **9**, 336-344 (1964).
- MACKWORTH N. and E. L. THOMAS, Head-mounted eye-marker camera. *J. opt. Soc. Am.* **52**, 713-716 (1962).
- F. WILLS C., Genetic load. *Scient. Am.* **222**, 98-107 (March 1970).
- HUBBY J. L. and R. C. LEWONTIN, A molecular approach to the study of genic heterozygosity in natural populations. *Genetics* **54**, 577-609 (1966).
- VANN E., The fate of X-ray induced chromosomal rearrangements introduced into laboratory populations of *Drosophila melanogaster*. *The American Naturalist* **100**, 425-449 (1966).
- WILLS C., Three kinds of genetic variability in yeast populations. *Proc. of the Nat. Academy of Sci. of the U.S.A.* **61**, 937-944 (1968).
- G. PTASHNE M. and W. GILBERT, Genetic repressors. *Scient. Am.* **222**, 36-40 (June 1970).
- WATSON J. D., Isolation of the lac repressor. *Proc. of the Nat. Academy of Sci. of the U.S.A.* **56**, 1891-1898 (1966).
- GILBERT W. and B. MULLER-HILL, The lac operator is DNA. *Proc. of the Nat. Academy of Sci. of the U.S.A.* **58**, 2415-2421 (1967).
- PTASHNE M., Isolation of the X-phage repressor. *Proc. of the Nat. Academy of Sci. of the U.S.A.* **57**, 306-313 (1967).
- H. BROWN D., The isolation of genes. *Scient. Am.* **229**, 20-29 (August 1973).
- WALLACE H. and M. BIRNSTIEL, Ribosomal cistrons and the nucleolar organizer. *Biochim. et Biophys. Acta* **114**, 296-310 (1966).
- BROWN D., P. WENSINK and E. JORDAN, Purification and some characteristics of 5S DNA from *Xenopus laevis*. *Proc. of the Nat. Academy of Sci. of the U.S.A.* **68**, 3175-3179 (1971).
- BROWN D., P. WENSINK and E. JORDAN, A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. *J. molec. Biol.* **63**, 57-73 (1972).
- I. BRITTEN R. J. and D. E. KOHNE, Repeated segments of DNA. *Scient. Am.* **222**, 24-31 (April 1970).
- HOYER B. H., B. J. MCCARTHY and E. T. BOLTON, A molecular approach in the systematics of higher organisms. *Science* **144**, 959-967 (1964).
- WARING M. and R. J. BRITTEN, Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science* **154**, 791-794 (1966).