

Tactics and Terminology in Information Retrieval: a Summary of Recent Work

THESE TENTATIVE remarks on terminology have been stimulated by conversations overheard at meetings¹ of persons interested in the organization and retrieval of information and by the chapter on terminology in a recent book.² It has become clear that when talking to people in the field one should be careful to use the appropriate terms, not so much to facilitate communication as to establish status. Use of the right words in the right way early in a conversation can mean the difference between being one up or one down.³

Aside from the landmark articles by Chadwick⁴ little formal work has been done in this area to date. The outlines of the field are, however, emerging. The writer has had the opportunity of hearing in action, as it were, a number of persons who, though unschooled, show remarkable natural ability. The promise they display may perhaps justify publish-

ing this rough outline of their primitive but ingenious work, which is submitted merely as a basis for future study.

Before discussing individual terms it seems desirable to make the point that one must not only use the right words but must use them in the right way. Certain guiding principles of conduct in their use can be tentatively advanced at this stage with, however, the warning that future work may be expected to alter their meaning and application. The reader is expected to employ these principles judiciously, recognizing that all possible situations cannot be anticipated in advance of their actually occurring. They are merely generalizations drawn from the, as yet, pitifully small literature and the few opportunities for observation which have come to the writer's attention. As the number of observed and recorded situations grows we must expect to amend these generalizations to conform with the facts. It is hoped that in this way a truly scientific body of doctrine can be assembled and perfected.

Thus understood these generalizations may be useful. We begin with the most basic of all, the Competitive Principle, more commonly called Chadwick's Law.⁵ It is difficult to overstate the importance of this discovery, which must be considered one of the great theoretical advances of the age. Before Chad-

¹The most recent of these was the Symposium on Systems for Information Retrieval held in Cleveland on April 15-17, 1957, under the auspices of the Council of Documentation Research and a number of cooperating organizations.

²James D. Mack and Robert S. Taylor, "A System of Documentation Terminology," in J. H. Shera, Allen Kent and J. W. Perry, eds., *Documentation in Action* (New York: Interscience Press, 1956), p. 15.

³A colleague has suggested that the work of a British chap with some such name as Podder or Pother or Putter, on cheating at games, is relevant to the subject of this article. Perhaps. On the surface the suggestion seems bizarre. The writer has found no references to Podder (Pother? Putter?) in the information retrieval literature.

⁴A. John Chadwick, "The Competitive Principle," *Occasional Papers of the Bombay Bicycle Club*, vol. 51 (4th series), no. 17 (1954); "Lying in Wait," in *Berichte und Schriften zum 60. Geburtstag*, R. S. Nahtanaghar, p. 425 (Leipzig, 1955).

⁵First revealed to the world in Chadwick's famous paper, "The Competitive Principle," *op. cit. supra*. The reader may already be familiar with the reaction of the audience, when Chadwick read it to them in the Main Hall of the Club's headquarters. Some stalked out in anger, others tried to shout him down, but the wise few listened and, at the end, stood in respect.

Mr. Merryman is Associate Professor and Librarian, School of Law, Stanford University.

wick's work, it is fair to say, all was chaos. Simply stated, in non-mathematical form,⁶ it means this: in any conversation in the information retrieval field one must regard the person to whom one speaks as an adversary and the conversation as a contest in which, inevitably, one will gain prestige at the expense of the other. (This formulation refers only to the two-party situation. The beginner should avoid more complex conversations until he has a sure grasp of the fundamentals. Many a talented novice has been spoiled trying to move along too fast. Obviously the greater the number of adversaries the greater the odds against the individual—and the larger the number of witnesses.)⁷

Failure to apply Chadwick's law is a common error of the most drastic sort. An obvious example is the tendency among beginners to ask questions which indicate ignorance. Thus Chadwick tells⁸ of an occasion on which he let drop the relatively common term "superimposed random coding" in the first sentence of a conversation with a person of unknown ability. Imagine Chadwick's reaction when, as he put it, "Do you know what the fellow did, Merryman? He asked me what 'superimposed random coding' meant." Chadwick naturally turned on his heel in contempt and walked off without answering, not, as some spectators thought, because of the poor chap's ignorance, but because his ready display of that ignorance showed him to be an inexperienced and unworthy adversary. He had failed to apply this elementary principle and had sought

information, rather than tactical advantage. (The reader is warned that this example does not prove that questions should never be asked. There are a number of situations in which carefully devised questions have tactical value. See especially Chadwick's paper on "Lying in Wait.")⁹

Perhaps Chadwick's Law, lying as it does at the root of the thing, can support one more example. This is the case of the invitation to cooperate. As Chadwick records it,¹⁰ A and B have begun a conversation and appear to be evenly matched. Neither has gained any appreciable advantage when A pauses in the middle of a sentence, obviously groping for the appropriate term. The possibilities are two: either he really is searching for the term or he is not. In the former case if B supplies the term, say "interfix," he has (1) helped his opponent and (2) laid himself open to a *riposte*. If A is a conversationalist of any ability he will instinctively decline "interfix" with scorn and substitute "conceptual linkage" which, he will imply, is the only possibly appropriate term in the context. The consequences to B would be even more drastic if A's groping for the term were feigned (as would normally be the case), since he then would be lying in wait and could be expected to have a carefully prepared *riposte* ready should B be so foolish as to cooperate. Thus the corollary "never cooperate" is an appropriate extension of Chadwick's Law. Pseudo-cooperation, or merely appearing to cooperate, however, is quite another matter, especially when there are spectators. Thus B can create the impression that he is quite a good fellow and at the same time score well by offering to lend A a recent article (or an advance copy of a book not yet on the stalls) that perhaps A has not seen and which A might

⁶ The mathematical derivation of Chadwick's Law is quite complex and will not be repeated here. The qualified reader will find an adequate discussion in Chadwick's articles, *op. cit. supra*, n. 4.

⁷ The problem of quantifying the effect of non-participants (spectators) on the flow of prestige remains to be solved, as does the infinitely more complex problem of quantifying the interactions of non-isolated groups operating simultaneously (the well-known foyer-phenomenon). Perhaps, as Chadwick suggests, the solution may lie in the use of three-dimensional matrices.

⁸ "The Competitive Principle," *op. cit. supra*, n. 4, p. 22.

⁹ *Op. cit. supra*, n. 4.

¹⁰ "The Competitive Principle," *op. cit. supra*, n. 4, p. 47.

find helpful as an elementary introduction to recent work in the field.

A second fundamental contribution comes to us from the work of Forsch. Forsch has not done any research in the information retrieval field, but his investigations of the use of language by the behavioral scientists are now recognized as classics.¹¹ As the reader will recall, Forsch found and described three personality types, each with its own characteristic habits of language utilization and affectation. These are the *ectolinguistic* (adapts jargon from other fields, usually the pure sciences; favors polysyllabic constructions), the *endolinguistic* (invents jargon specially for the field and works it hard; frequently earnest to the point of being intense), and the *mesolinguistic* (impatient with jargon and wishes that people would use plain words and say precisely what they mean). For the sake of brevity we will adopt Forsch's short titles for these classes: *Ectol*, *Endol* and *Mesol*.

The writer's own investigations tend to confirm his earlier hypothesis that these types exist in the information retrieval field in almost as clearly defined form as in the behavioral sciences. This discovery has important consequences in the development of terminology strategy, since it enables one to classify one's adversary early in the conversation. It then becomes possible to adopt the appropriate tactic with confidence. Clearly, if the adversary has revealed himself as, say, an *Ectol*, it becomes much simpler to choose among the available courses of action. Consequently the second basic principle of conduct in the use of information retrieval terminology can be very simply stated in this way: Know Your Adversary. Get him to speak up so as to help you classify him; draw him out until he commits himself; make him talk

¹¹ Pablo Forsch, *The Scientific Study of Language Behavior Among the Behavioral Scientists: A Report and an Analysis* (Buenos Aires, 1952).

first and the conversation is half won.

A few examples may illustrate this principle in such a way as to show its utility. The writer has been extremely fortunate in having the opportunity of working with Chadwick in the field and has recorded a number of actual conversations which show the master's technique.¹² Only the relevant parts of these conversations are reproduced here.

Example 1.

Adversary. . . . You might be interested in our problem. We have 100,000,000 bits of information in a high entropy state. At present we are trying to get an O.R. team together to structure a system for us. We think a coordinate system, with suitable parameters, might be worth trying on a pilot plant basis but we are worried about the noise problem.

Chadwick (who always enjoys meeting an *Ectol*, particularly one as far gone as this). Sounds to me like a job any decent librarian could handle. What precisely do you expect these operations research chaps to do?

Example 2.

Adversary. . . . As a documentalist you must have followed the dispute between Perry's semantic factoring school and Taube's Uniterm descriptors system with some interest. Which do you think holds the greater promise for solution of the problems of information organization, storage, and retrieval?

Chadwick (who immediately recognizes his man as an *Endol*). Neither. The work of the Bureau of Standards people in using analog computers to search steroids, relying on some simple notions from topology, is the only significant advance in documentation to date. Possibly the Minicard, Filmorex, and Rapid

¹² The work was carried on under a grant from the Jas. Joyce Foundation for Research in Linguistic Behavior. A full report will be published in 1958 by the Foundation. All responsibility for the conclusions expressed in this article and the forthcoming book is of course taken by Chadwick and the writer.

Selector experiments may, in time, reduce manipulation and storage problems. But as to the theory of the thing look to the Patent Office and Bureau of Standards groups, not to these gadgets.

Example 3.

Adversary. . . . All this rot one hears about documentation and information retrieval and all that. A lot of jargon. Bunch of computer-happy adolescents. Using words to hide a dearth of ideas. Etc.

Chadwick (smiling inwardly at the prospect of dealing with a genuine *Mesol*). I suppose it's true that one who is unfamiliar with the field occasionally does bog down in the terminology.

The reader will agree that in each case Chadwick has scored well. What may not be so obvious is the fact that each of these corking good shots is an example not only of virtuosity but of careful analysis and cool, deliberate planning. Here are classic illustrations of the three types identified by Forsch, together with the three major categories of manoeuvre properly executed.¹³ The reader may profit from some explanation of these manoeuvres in the context of the examples given.

In example 1, Chadwick employed the opposition manoeuvre. That is to say, after identifying his opponent as an *Ectol* he adopted the characteristics of one of the other two language types—in this case the *Mesol*—as a basis for retaliation. One can almost feel the physical impact of his reply. In example 2 he doubled, by which it is meant that he adopted the same game as his adversary but went him one better. (The writer has discussed the matter with Chadwick, who admits that the risk of a redouble exists in this manoeuvre. As he points out, however, *Endols* are normally much too earnest and sincere to pose much of

a threat. To guard against the unusual it is wise to save some ammunition to re-redouble with, as he had done in this case. He was prepared, should the need arise, to move to a discussion of Kirsch's work on a general mathematical theory of information storage and retrieval. Sound man, Chadwick). In example 3, in which Chadwick shows excellent form, the reader will recognize the elegant simplicity of the flanking manoeuvre, which is particularly effective against *Mesols*. Even though he cannot witness the adversaries in action or hear their inflections the reader can well imagine the impressive power Chadwick brought to bear in this flanker. Extraordinarily good show.

These few remarks on the use of terms should be sufficient to convince the reader of the importance of knowing the terminology itself. The best strategies depend, in the last analysis, on a careful choice of the terms to be used. Here again Chadwick's Law operates. The important thing is not the *meaning* of the term. Indeed, nothing could be less relevant. What one needs to keep a firm grasp on is the *use* of the term, its prestige value and the like. Obviously the meaning depends on who is using the term and for what purpose. It is for this reason that the writer feels it necessary to criticize the game try made in a recent publication.¹⁴ The authors, lacking a clear understanding of Chadwick's Law, have botched the whole thing. While driven, perhaps by instinct, to some fairly sound tactical use of terms, they have diluted the effort by trying to convey information. In doing so they have fallen between stools. Their piece fails strategically and it cannot honestly be said to carry much meaning.

An example might make the point clearer. The authors include the following in their table of documentation terminology:

¹³ For a fuller discussion of these manoeuvres see Chadwick's paper, "Lying in Wait," *op. cit. supra*, n. 4, p. 481, *et seq.*

¹⁴ Mack and Taylor, *op. cit. supra*, n. 2.

Index. A systematically arranged list of the names of subjects occurring in a document or group of documents, with an indication of the places in which they occur. One of the stages in the analysis of information. In the sense that it attempts to evaluate the contents, scope or importance of a given publication, an *Annotated Index* approaches an *Abstract*. On the other hand, a *Bibliographical Index* makes no such attempt, but is confined to furnishing verification and tracing data. A *Subject Index* reveals the relationships between subjects by analyzing the given field of knowledge into elementary terms, ordering them systematically, and synthesizing subjects by combinations of these terms.¹⁵

This paragraph shows some promise. "Analysis of information" is good, as are "Verification and tracing data" and "synthesizing subjects." But the rest of it is sheer waste which only weakens the tactical value of the useful part. The following is a clearer and much more useful treatment of the same term:

Index. Origin librarianship, book publishing, etc. Prestige value negative. Occasionally used by Mesols, but otherwise tactical value nil. Alternatives: *coding dictionary; analysis of information; descriptor schedule; etc.*

The superior value of this kind of information is obvious. It gets to the point, rather than puttering around with irrelevancies. One more example may drive the matter home. Mack and Taylor include the following in their tabulation:

Parameter—Computers. In a subroutine, a quantity which may be given different values when the subroutine is used in different parts of one main routine, but which usually remains unchanged throughout any such use. To use a subroutine successfully in many different programs requires that the subroutine be adaptable by changing its parameters.¹⁶

¹⁵ *Ibid.*, p. 21.

¹⁶ *Ibid.*, p. 23-24.

This is very weak. "Computers" is a useful term, but it is merely thrown in here, with no pretense of an attempt to use it tactically. "Subroutine" is only fair, but it might carry slightly more punch than the extremely colorless "program." The total effect is one of insipidity. Contrast the following suggested treatment:

Parameter. Ectol term. Origin mathematics. Prestige value large and positive. Couple with *Computers, systems, etc.*, for maximum effect. Effectively used in questions; e.g. What are the *parameters* of your system?

These examples should make it entirely clear that Chadwick's Law operates in the choice of terms as well as in their use. It is hoped that in the future those who compile annotated lists of terms will keep it more clearly in mind, thus avoiding the mistakes of Mack and Taylor.

In conclusion the writer wishes to remind his readers that much work remains to be done. Chadwick and Forsch were pioneers. They have broken the ground, so to speak. What confronts those of us who follow is the less noble but still rewarding task of working out in detail the application of their discoveries to the multitude of specific problems in the field. There is room for originality here, for inspired creative work, but we also need the work of the dedicated plodder with the sound but prosaic mind. As in the past, science can and will build for the future on the efforts of many such men as they do their parts in applying the theories of the giants.

What has been presented here is merely an outline of the major advances to date. If a few readers are stimulated to apply themselves to the many remaining problems crying out for solution the writer will feel he has been more than adequately repaid for his small effort.