Communicating scientific experiments in journal articles

W. John Hutchins

1. Introduction

It has been common practice to describe texts and their structures independently of authors and readers, as if they have ‘meaning’ in themselves. In such a perspective it would seem legitimate to describe the communicative function of texts as something existing without communicating individuals. The practice has been common among linguists investigating discourse to describe the interpretation of texts as a process with universal validity, as a process that all readers do, with greater or less success.

In human linguistics, however, the focus is on people and their structure, not on texts and their presumed structure. The question then arises as to how to reconstitute in human linguistics the valid linguistic insights already won but expressed in terms of a linguistics focused on text. We will investigate here what scientists do linguistically during the conduct and reporting of their scientific research.

2. Basic features of the scientific article

The motivation for conducting a scientific investigation or experiment is frequently the desire of a scientist to resolve a problem or anomaly in what is currently known about a particular set of facts, or in what is commonly regarded as an explanation or hypothesis for particular phenomena, events, effects, or processes. In order to resolve the problem or anomaly, the scientist will typically formulate a new hypothesis about the ‘state of affairs’ and will undertake tests of the hypothesis. These tests will provide evidence for or against the hypothesis, which may consequently alter the perception or understanding of the phenomenon in question.

Typically, scientific explorations are concerned (a) with the investigation of a particular problematic area and proposals for resolving difficulties, (b) with the examination of alternative hypotheses regarding some problem of interpretation, or (c) with the development of methods for experimentation. There are thus three different types of scientific article: (a) Problem-Solution, (b) Hypothesis-Testing, and (c) Methodological. In each type there is typically an introductory section describing the current ‘state of affairs’, the area of problem or deficiency which is the ‘topic’ of the article, and then statements about what is to be tested (practically or theoretically).

In the Problem-Solution paper (a), the author will describe the current situation and the problem he is tackling, and then the methods he will use. In the Hypothesis-Testing paper (b), the author will state the ‘problematic’ area and describe the various proposals that have been suggested, and then examine each for their advantages and disadvantages. In the Methodological paper (c), the author will state the deficiencies of current methods and propose a ‘new’ method, which he will then demonstrate by tests as an improvement (or substitution) for other methods.

---

1 Here following the convention of using the masculine pronoun to refer to individuals of any gender.
There will follow in each type of article, a section of tests, results and evaluations (or judgements), and a statement of the solution, hypothesis, or method the results indicate (although this section may be implied only). Finally, there will usually (but not invariably) be the author’s observations about what his findings may have for other areas of the science (often labelled as “Implications”), and what further investigations might be undertaken (“Further Work”). Schematically, the typical Problem-Solution type – the type discussed in this paper – has the following overall structure:

- **Situation**: description of common knowledge about field X; statement of current hypothesis
- **Problem**: anomaly in some area of X; drawbacks of current hypothesis
- **Hypothesis(es)**: statement of proposed solutions, new hypotheses
- **Tests (Basis for evaluations)**: experiments/testing of hypotheses
- **Evaluation**: results of tests/experiments; proof of new hypothesis
- **Solution**: adoption of one hypothesis
- **Implications**: what the ‘solution’ may mean for other areas of X
- **Further Work**: how the ‘solution’ could be developed in the future

In practice, authors provide scientific articles with more simple section headings. We find typically an “Introduction”, embracing the first three sections (“Situation”, “Problem”, “Hypothesis”) and including a statement of the overall ‘topic’ of the paper; it will be followed by “Experiments and Results”, covering the three sections (“Tests”, “Evaluation”, “Solution”), and a final “Conclusions” (with sections of “Implications” and “Further Work”).

### 3. The scientist as author

In the Situation section, the author triggers in the reader (or brings to the foreground of his attention) a particular ‘domain of reference’ (area of knowledge). This is achieved by referring to phenomena (real or mental) by means of anaphoric and deictic expressions, i.e. to things presupposed by the author to be already known by the reader. Referring can also be indirect by citing previous publications – which may or may not be known to the reader, but which he can consult. The domain is described or alluded to by referring to particular relevant parts of the domain – those parts sufficient to trigger the whole of the reader’s area of knowledge. As Yngve describes it, mention of a train brings into the field of interest also known features of ‘trains’ such as engines, wheels, drivers (engineers), etc. The effect of this section is therefore to change the reader’s initial state (precontext) where there is no focus on the subject area to a state (postcontext) where the reader is attending to a particular domain of reference.

In the Problem section, the author alerts the reader to anomalies or problems in this domain (‘state of affairs’). The problem may be one already familiar to the reader or it may be ‘new’ in some sense – perhaps not completely new, in that the reader may have had a feeling of unease, disquiet or dissatisfaction about some aspect, but could not formulate where. The author’s description of the ‘problem’ may itself refer to other aspects of the domain which the author assumes the reader has some previous knowledge of, but primarily the intention is to inform the reader that a problem exists that requires solution. In the reader there is thus set up an expectation that the author will offer some solution.

In the Hypothesis section, the author suggests one or more ways in which the problem may be solved. These may include ‘solutions’ already familiar to the reader,
or which he may infer from what he knows already, or which the author may
demonstrate can be derived (logically) from what is already known. Some ‘solutions’
may, on the other hand, be completely new to the reader, and the author will be
aware that he has to do some persuasion. The most effective means of persuasion are
to describe some experiments or tests.

In the Tests section, the author brings to the foreground other areas of
knowledge, namely those concerning methods of testing, evaluating, proving the
truth or falsity of a particular event (or rather the author’s explanation for the
phenomenon). Again the type of test proposed may be already known to the reader,
at least in essence, and the author may perhaps assume familiarity with tests
employed in other (previously undertaken) experiments. Tests can be of many kinds,
not just physical experiments but also mental arguments (Gedankenexperimente).
Referring to such already known (or presumed known) tests may also be explicit by
citing previous publications or by quoting from them.

In the Evaluation section, the author provides his opinion on the success or
failure of the tests undertaken. The background for such assessments will be
generally assumed to be part of the knowledge of any reader concerning the conduct
of research in the particular area of science. Evaluations may range from simple
assertions that the tests were successful (e.g. that they showed the results and values
expected) to more extensive discussions pointing to further tests (which may be
described), and perhaps referring to previous experiments by the author or other
scientists (by citation.)

In the Solutions section the author informs the reader of the solution, based
on evaluations of the tests. Its effectiveness (persuasiveness) depends on the reader’s
satisfaction with the author’s evidence and argument. Here, the author will refer
back to preceding parts of the text (which can now be assumed to be known), and
only rarely to external domains not previously mentioned. (If there are such
references then this may represent a weakness in the author’s presentation.)

In the Implications section, the author will probably again refer to the initial
domain of reference (the Situation), since his aim is probably to point out what other
parts of this domain could be affected by his solution. He may refer to other
problematic areas (known or unknown to the reader) which could also be solved by
the proposal he has made. Alternatively, he may point out that his solution produces
further (new) anomalies in the ‘state of affairs’, which in turn demand resolution.

Finally in a Further Work section (which is not always present, or it may
occur instead of an Implications section), the author will suggest how the proposed
Solution may be improved or how it may be applied in other areas of the subject
field. Again, the author will refer implicitly or explicitly to the Situation and to other
of its problematic aspects, some of which may not have been known to the reader
before reading the article.

For each of the sections (paragraphs) there are overt linguistic signals for the
functions and relationships of segments within the total text. These signals should be
regarded within a human linguistics as arousing conditions or triggering procedures
that cause appropriate changes in the reader’s plex.

The signals have long been studied by discourse analysts, and their findings
are of obvious relevance, but they need to be placed in the context of human
linguistics. While discourse analysts would interpret linguistic elements (lexical
items, anaphors, etc.) as signs of particular textual relationships or functions, in the
present context we would understand them as triggering particular conditions and
networks (plexes) for the knowledge and awareness of individual writers and readers.
In this framework, texts are interpreted not as entities in themselves, i.e. as objects that ‘contain’ messages, but as epiphenomena of authors communicating to readers.

4. Text understanding: scientist as reader

In order to interpret a text the reader must bring with him (i.e. have already present in his plex of properties) the same or a similar realm of common knowledge about the state of affairs as the author presupposes his readers will have when they start to read his text. He must also, of course, know the same language. What this means in human linguistics terms is the subject of intensive research. And he must have familiarity with the ways in which knowledge and information is spoken and written about by others, i.e. an ability to recognise general patterns (such as those outlined in sections 2 and 3).

A crucial contact-point for the reader of a scientific article is the introductory section (or paragraph) – the ‘base section’ – where the author reviews the research (the current state of knowledge), points out lacuna, inconsistencies and anomalies, and then states as clearly as possible what he has discovered or concluded from the work being reported. In this section (typically labelled “Introduction”), it is quite common for the author to state explicitly what he considers to be its ‘topic’ as a whole, what he contends it to be ‘about’.

Successful understanding of a scientific text presupposes also that the reader knows the scientific field. He must know (have learned or experienced) most, if not all, of those aspects, features, objects, concepts, events, etc. that the author writes about as presupposed elements (i.e. the elements of the initial sections where the author invokes background knowledge.) He ought also to know as much as possible about any other items of special knowledge that the author assumes his readers will already know. Of course, not all readers will have all this knowledge — indeed, perhaps very few, particularly in the leading edges of science — and many readers may have to refer to other scientific articles (some by the same author) on the same subject in order to learn more and understand more. This is one of the reasons why authors refer to the writings of others. On the other hand, there may well be some readers for whom not only will the author’s presupposed knowledge be fully familiar but even some (or much) of what the author believes or assumes to be new. This may be because the experiment and the ideas are not as original as he may have supposed. More likely, however, it will be new information because he is writing not for the experts who know most of it already, but for others who do not. In any case, even for the experts there may be something that is new; in particular, some of the results and some of the conclusions. Since it is obvious that readers of a scientific article come from many different backgrounds and many different levels and spheres of knowledge, the relevance of a particular article to their own information needs will be highly variable.

In this respect, the reader will clearly make his own judgement of what the text is ‘about’. It will quite possibly differ in some (perhaps many) respects from what the author states its ‘topic’ to be. The reader may well identify various elements of the text as ‘topics’; they are likely to include subjects of particular interest to the reader at the time when he is reading it, e.g. in the case of a scientific article, it might be a particular method of chemical analysis – for such a reader the article may be mainly ‘about’ this method. It follows also that ‘topics’ can change over time; what may interest the reader on one occasion may be of no interest to him on a later occasion. And vice versa, what may have seemed irrelevant (or not understood, and therefore not given topic status) when first read may later become of greater interest
on later reading. The individual reader’s state of knowledge (his plex of conditional properties) change over time – from experience, from learning, and from reading.

In general, we may say that from the perspective of human linguistics, the reader’s task combines an attempt to ‘reconstruct’ the associative knowledge network that the writer has constituted and an attempt to integrate this network (or part of it) into his own associative knowledge network. The nature of this network is of course unknown. It may or may not be similar to the kinds of networks conceived by linguists and researchers in artificial intelligence, i.e. ‘semantic networks’ of hierarchical (hyponymic, etc...) relationships among the senses (meanings) of lexical items, where ‘meanings’ are defined in terms of ‘primitive’ (‘atomic’) sense elements and logical relations. It may also be that ‘reconstruction’ of networks operates by processes activated by various strategies at different levels: linguistic, cognitive, emotional, inferential, etc. (as described by Van Dijk and Kintsch 1983.)

The difference will be, of course, that the networks are conceived not as interconnected atoms of meaning in the logical domain but as interconnected nodes of plex structures of people, dispositions, intentions and conditions with the potentialities of producing texts and of understanding texts.

5. Tasks and subtasks

Within the human linguistics framework we may describe the activities of authors and readers as sets of tasks and subtasks, sequential, parallel and overlapping. We may also describe the particular activities of the scientist-author as a specific set of tasks and subtasks within his general activity as a scientist, as a researcher, as a human being, etc.

The primary task of the scientist-author is to communicate the outcomes of his investigations (mental and physical) to his fellow scientists or to the general public. As we have described above, there are different types of scientific papers depending on the nature of the investigation or the way in which the author chooses to present his findings. In the Problem-Solution type of paper, the author has two basic subtasks: to determine the nature of the ‘problem’ and to describe a ‘solution’. Setting up the ‘problem’ may itself involve various subtasks: describing the ‘state of affairs’, citing previous relevant work, highlighting anomalies and difficulties, etc. Likewise, the ‘solution’ subtask involves other lower-order subtasks: stating a hypothesis, describing the tests and their results, determining whether they constitute solutions or not, and so forth. Some of these subtasks may be common to other writing tasks, e.g. the citation of previous work; others may be specific to this type of scientific paper, e.g. the evaluation of possible solutions to scientific problems.

The activities of the scientist-reader are also divisible into tasks and subtasks. Some are common to all readers: understanding deixic and anaphoric referring, qualification and modification; some are common to readers of the subject: the specific vocabulary of the field, the manner of discussion and argument in the subject, etc.; others are specific to texts of this type, e.g. the structure of ‘problem-solution’ scientific texts; and other tasks are specific to this particular text: i.e. understanding the specific background information presumed by this particular author, following his arguments and integrating the new knowledge into the reader’s own previous knowledge.

Both authors and readers are also operating within a wider context. In this case, as scientists their overall task is to understand scientifically some part of reality or the natural world. Writing and reading are subtasks within this framework. From his reading of scientific papers, the scientist may be stimulated to try new methods,
investigate new areas, gain greater understanding, formulate new ideas and hypotheses, etc. The completing of one task (or subtask) may lead to starting another task (or subtask) before the end of another, or may lead to abandoning a task begun in favour of another one. For example, reading about the methods used by another scientist may inspire him to begin a new investigation even before he has finished the text that he is reading.

6. Linguistic activity of scientists as experimenters

Before he is an author, the scientist is an experimenter, and in this role (which he probably sees as his principal role) he carries out linguistic activity in addition to those described already. When considering whether to undertake a particular experiment, or test a particular hypothesis he will be not just interested in finding information in general in his field of study, he will want to find specific information relevant to what he intends to do. For this purpose he may undertake a search through the literature of his subject using the indexes and abstracts available (as we describe below). Having found articles appearing to be relevant, he will read them with the specific intention of augmenting his knowledge (the plex area related to his focus) so that he is fully up to date with what other scientists have been doing. These other authors will be candidates for the set of bibliographic references he will be making when he writes up his experiment.

The next stage of linguistic activity is likely to be some discussion with colleagues (or, in the case of a student, with his professor or supervisor), either directly or at a distance (by telephone, by email, by letter, etc.) Such discussions will be exploratory, seeking information about aims, methods and, probably, about the prospects of achieving significant results.

Before doing any experiments, it may be necessary to obtain funds, and for this the scientist may well have to submit an application. Typically, fund applications involve descriptions of what is known already in a specific field, what experiments are to be undertaken, and what the expected results may reveal. Evidently, some of the content of grant applications parallels the content of the final scientific article – indeed, as we know, scientists sometimes have the impression that in order to obtain a grant, they have to do all the experiments first! The writing of project proposals and grant applications is clearly another area worthy of separate investigation.

When doing experiments, he will interact linguistically with a number of people, not just colleagues, but also technicians, administrators, secretarial and cleaning staff. Experimental activity is fraught with frustrations, mistakes and failures. These ‘diversions’ will give rise to much linguistic activity, but very little (probably none) of it will be alluded to in the final writing up. The experiments will be idealized, spoken and written about as if they proceeded perfectly. This is the centuries-old scientific tradition.

Before starting to write, the scientist will inevitably (it is presumed) reflect upon the validity of his results, and whether they demonstrate what he intends them to demonstrate. There will at least be an internal monologue, and quite possibly further dialogue with colleagues. Finally, he will formulate (plan) his article, setting out the background, the hypotheses, the tests, the results, and the conclusions somewhat (in broad outline) as described above. The process will involve a further stage of idealization, since he will want to adhere as much as appropriate to the standards and norms of his peers. An illustration of how complex this activity can be, particularly when there is more than one author (which is now quite frequent in science), is provided by Knorr-Cetina (1981). Idealization means that the precise
processes of experimentation are not described, false starts and faulty experiments are omitted, as well as disputes and criticisms within the laboratory. What are also left out are the financial, commercial, academic motivations of the experimenter(s) and the real reasons for undertaking the experiment. In the case examined by Knorr-Cetina, for example, the new technique which is the major finding of the paper was developed in response to difficulties with existing techniques; most of the time in the laboratory was spent on making sure the method worked. But none of this is reported in the paper. Instead, the paper seeks to provide a relevant contextual framework to justify the new technique, and these are not technical justifications but ‘scientific’ ones, describing it in the context of what is known about the current methods and their disadvantages. According to Knorr-Cetina, it is such ‘recontextualization’ in a conventional framework that is the aim of the introductory sections of the paper. Furthermore, this introductory context may well be the last part to be written – as Knorr-Cetina observed, the first sections to be written are the ‘results’ and the ‘conclusions’ (what the scientists regarded as the “core” content of the paper); only later do the authors set down the background and the justifications for doing the work described.

The desire on the part of scientists for their work to be accepted into the scientific consensus encourages them to minimize their criticism (even implicit) of their fellow scientists and of their methods and results. In the case of the Knorr-Cetina example, the authors refrained even from making assertions about the implicit but clear and distinct advantages of the proposed method over existing methods.

7. Indexing and abstracting

Between the author and reader there are commonly also other acting individuals who facilitate the processes of scientific communication. These are the editors of journals, the referees of articles, the publishers, the subscription agents, the workers in the mailing services, the staff in the institutions receiving the journals, etc. In particular, however, there are the individuals who index and abstract articles, who provide the secondary written (printed) texts which scientists (as seekers of information) consult in order to locate articles that may interest them (as readers). These individuals are often helped by authors who provide summaries of their articles, but the processes of indexing and abstracting are separate activities where individuals also pass from initial states (of ignorance) to final states (of some knowledge). They may or may not be familiar with the specific problem area, but their aim is to describe what it is, and (in the case of abstracts) what the author has discovered and what he proposes as an explanation.

There is a familiar distinction made between indicative abstracts and informative abstracts. In indicative abstracts, the abstractor aims to compose a summary (topic paragraph) describing the focal points of the content of the original article in general. Unfortunately, many readers are not satisfied with only statements about the overall topics of articles, they want to know precise pieces of information, e.g. whether a particular chemical was used in an experiment, whether a particular method was adopted, whether a particular effect was observed, etc. This is the function of the informative abstract, to serve as a useful pointer for information embedded in the article outside the introduction and not indicated as topics by the author himself. The distinction between indicative and informative is not rigid or absolute: abstracts commonly combine both functions; as with texts, the abstractor (as author) takes into account the expected needs and background of the reader (or user) of the abstract.
Abstracting (or summarizing) would appear to be primarily a function of generalisation, guided by text features and clues commonly understood by all readers, but directed to the particular end of communicating the essential core. Generalisation itself must be considered a basic operation for all communicating individuals, basic to all learning, apprehension and comprehension, involving (it may be presumed) awareness of similarities and convergences among different sets of (internal) conditions and procedures in the plexes of individuals.

In indicative abstracts, what is summarized is generally only the overall topic of the article, i.e. the subject area and the specific problem being discussed. This may be regarded as a generalisation based on the known elements (mainly in the base sections) and ignoring most of the new information.

Indexing carries indicativeness another stage further. It reduces topic statements to sets of individual independent words or phrases (index terms), which together are intended to cover the content. In effect, index terms are hooks (points of contact) between texts and their potential readers. The task of indexing for a wide range of potential readers is difficult for reasons already given: each reader comes to an article with a different plex of knowledge and understanding. Each will formulate his own idea of what the text is about. All that the indexer can do is to select index terms that capture the topic and perhaps part of the presupposed knowledge of the article for some ideal reader (scientist). In this way, the indexer seeks to provide readers with a common starting point. The author has made presumptions about what readers should already know; the indexer can use these to provide a bridge from a reader’s presumed state of knowledge to the author’s text.

Fortunately, as we saw above, most authors of scientific articles state in the introductory (base) sections of their papers precisely what they consider its topic to be. These statements can be readily located by both indexers and readers. Since the indexer may presume that this topic statement is what the reader would look for when determining whether the article is of interest, the indexer can select it as the basis for producing a standardized set of index words and phrases.

In the indexing (and abstracting) process we have a communicating individual (usually not himself a scientist in the particular subject domain), who functions as a reader of a text (when analysing, understanding or evaluating a scientific article) and simultaneously – apparently – as an author of (a set of) index terms or an abstract. In this authoring function it is possible to envisage the index terms as constituting a ‘language’ analogous to a natural language (of original texts), which indexers use to refer in ways similar to those when using natural language (cf. Hutchins 1975).

Index entries (and abstracts) are also texts which are read by scientists. As when reading texts, the index user comes with a particular background of knowledge about the field. He formulates his search topic in terms of this background, and when he finds an entry containing these terms he understands it in the light of his own knowledge. What he then expects is that the text (scientific article) to which the index entry refers will in fact be relevant to the topic he is looking for. We should not be surprised that the process can lead to failure, since the background knowledge of the reader (as index user) is unlikely to be close to that of the indexer and that of the original author. In fact, what is surprising is that the process is often successful. It succeeds because authors of scientific articles make reasonably clear what they assume as base knowledge and what they consider their topics to be; indexers learn to interpret the topics of articles and to express them in terms (i.e. index terms) that they expect index users to formulate searches in; and readers (and index users) learn
how to express what they are seeking in terms which they expect authors and indexers to use. Just as authors anticipate in their texts the questions that readers are likely to ask, so indexers anticipate in their texts (index entries) the searches that index users are likely to make.

8. Texts in collections

Indexing and abstracting relate texts to each other by virtue of common index terms (topics). Apart from author-reader linkages there are what we may refer loosely to as text-text associations. Scientific texts are not written in a vacuum, but have relations to other preceding texts. The most obvious of these associations have already been mentioned: the citations made by authors to other scientists, the publication of articles in journals devoted to the same topic, the collection of texts in libraries and databases, and the juxtaposition of bibliographic entries for articles and books in indexes and catalogues. But there are other associations hidden or implicit in the author’s establishment of his base of presupposed knowledge. This is the knowledge that previous scientists have established from previous investigations in the particular scientific field. What is expressed by anaphora, by deixis, and other linguistic cues are indirect references to this familiar background. Since much of this background has been communicated in other texts, these references are also indirect references to ‘associative networks’ set up by authors of other texts – and it is these associations from text to text which constitute the idea of ‘public knowledge’. It has been common practice to describe texts and their structures independently of authors and readers, as if they have meaning in themselves. The practice is particularly tempting in the case of scientific communication whenever there is reference to the scientific literature as constituting ‘public knowledge’ (Ziman 1968).

However, we must not forget that these are indirect relationships, which should be regarded as occurrences within a higher order of linkages embracing the scientific community, i.e. as group participants and as collections of texts as props. References to other texts from a text are made by communicating individuals even if the results are recorded visibly as written forms. The links are activated only when they are triggered by the internal procedures of individual readers (scientists, indexers, etc.)

The situation is obscured by the fact that databases of texts (particularly electronic databases) record relations between texts which do not appear to originate from activities of communicating individuals. Index entries can be, and are now almost invariably, derived automatically from occurrences of words (sequences of symbols) in corpora of texts. It seems that only the computer program which does the indexing is the work of a person; its application to texts is automatic. Collections of interconnected and indexed texts are now essential for all scientific work. The fact that the embedded text relationships do not exist in the mind of any particular scientist – indeed, they are unknown until a search of the literature is undertaken – does not mean that these relationships constitute ‘objective knowledge’ existing independently of any originators.

It is true that written documents seem often to have an existence beyond the life of their creators, e.g. when texts are interpreted in ways contrary to the intention of authors. But the important point is that interpretation has to take place. Popper defines the ‘objective knowledge’ contained in a book (true or false, useful or useless) as “its possibility or potentiality of being understood, its dispositional character of being understood or interpreted, or misunderstood or misinterpreted…”
And this potentiality or disposition may exist without ever being actualized or realized.” (Popper 1972: 116).

The discovery by Swanson (1990) through a search of medical literature that a certain drug might help to cure a certain disease was an act of interpretation, an act of true scientific research, viz. the finding of plausible hypotheses for further investigation, and the investigating of properties of large and complex sets of biomedical data. It was Swanson who recognized the significance and relevance of the connections, not the inert data and documents themselves. Relationships between texts (whether produced automatically or by indexers, etc.) become significant, ‘meaningful’ or relevant only when readers (e.g. scientists) make judgements – just as the effectiveness of an information retrieval system is assessed by its success in retrieving documents that users consider to be relevant to their requests. Texts and text relationships are interpreted by individual readers on the basis of their own conditions and states of knowledge at the particular time when they are read.

9. References


