

DAVID CRYSTAL

English and the Communication of Science

David Crystal is Honorary Professor of Linguistics at the University of Wales, Bangor, Patron of the International Association of Teachers of English as a Foreign Language (IATEFL), Fellow of the British Academy, and a prolific writer and broadcaster in the field of language. He currently divides his time between work on language and work on general reference publishing.

In 2005 I was asked to contribute a paper to a conference with the above title.¹ I had difficulty, because I found the title three-ways ambiguous. Were we talking about ENGLISH and the communication of science? That would be a discussion focused on the role of English as opposed to other languages in the way science is communicated. Or were we talking about English and the communication of SCIENCE? That would be a discussion focused on what it is that makes scientific language different from other domains of intellectual endeavour, such as religion or the arts. Or were we talking about English and the COMMUNICATION of science? That would be a discussion of how scientists get their message across, in an information-conscious age.

We need to be talking about all three, but I believe the most useful focus, in 2006, is the third. The first two, after all, have been well rehearsed and the issues they raise do not change much. But in relation to the third topic, everything has changed in the last ten years, and there are some fresh and fundamental issues to be addressed. As the media people say, we are faced with a whole new ball-game. And the name of that ball-game is the World Wide Web.

But let me begin at the beginning, and discuss the first word in the title: ENGLISH. The point that English is the language of science can be traced back to the Industrial Revolution. Two-thirds of the scientists and technologists who made that revolution possible had English (at first British and later American) as a mother-tongue, so that anyone who wished to learn about the latest advances had no option but to acquire some competence in it, either directly or through translation. Two hundred years on, and we find people beginning to talk about the preeminence of English, but even in the 1960s observers were being cautious.

Take Theodore Savory, whose influential *The Language of Science* was first published in 1953. It was revised in 1967, yet even at that late date Savory says no more than 'English shows signs of becoming the language of science' (p.151). This caution was justified, because case studies had

shown that English was being used in only around 40 per cent of cases. For example, an analysis of the *Zoological Record* from 1865 to 1955 showed that five languages had been used for about 90 per cent of the publications: English had never fallen below 40 per cent, but German, French, Spanish, Portuguese, and Russian were all important. German, for example, was 31 per cent in 1880, and there was a time when scientists all over the world were being advised to learn German if they wanted to get on. But the figure for German had fallen to 9 per cent by 1948, for reasons perhaps too obvious to need recapitulating. Yet, old habits die hard, and in the decade after 1955 Savory repeated the exercise and found German up to nearly 15 per cent with English down to 38 per cent, and new languages – Italian and Japanese, in particular – then having some presence. So he was right to be cautious.

But he was perhaps being overcautious, for bibliographical studies of other scientific domains were also taking place, and these were showing higher results for English. A study by Sandra Ellen (1979) of seven abstracting and indexing journals in 1977 showed the following results: the lowest was chemistry with 47 per cent English, then maths with 65 per cent, geology with 72 per cent, medicine with 76 per cent, biology with 79 per cent, engineering with 83 per cent, and physics with 87 per cent. The average is 73 per cent. A similar study by Andrew Large (1983) of four databases in 1980 found chemistry up to 62 per cent, medicine 73 per cent, biology 88 per cent, and physics 82 per cent. The average is 76 per cent. The trend is steadily upwards, and so it is not surprising to find in the 1980s the figure we most often encounter today – that some 80 per cent of the world's science is expressed in English. It has to be interpreted cautiously, of course, but it has an intuitive recognition. I think most of us would say that it is about right, for written material. But my point is that it has not changed, as an estimate for the written language, in the past 25 years. It does not seem to be increasing, even though English is steadily growing as a global linguistic presence, because most of the new users of English are not scientists but lay people. Nor is it particularly decreasing, for no other language has yet increased its world status in the scientific domain to provide a viable alternative global lingua franca. So for this first topic, ENGLISH, there is nothing really new to report.

What about the last word in the title, SCIENCE? The study of what makes scientific language distinctive was one of the major developments in language study during the second half of the twentieth century. I have attended two conferences on scientific language, and both focused on the stylistic issues. The books that appeared during this period did the same. Savory's was one of the earliest, notable for its breadth of subject-matter, and it is interesting to see how it dealt with the topic, for this was typical of the period. It begins with a general reflection on scientific language, contrasting

¹A conference held at the English Speaking Union, London, 10 March 2005.

it with other varieties, and especially with varieties at the supposed opposite extreme, such as poetry and religion. Thus we find two characterizations of 'man': Shakespeare's 'What a piece of work is a man! How noble in reason! how infinite in faculty! in form and moving how express and admirable! in action how like an angel! in apprehension how like a god! the beauty of the world! the paragon of animals!' (*Hamlet*) And a zoologist's account in 1912 (Lancelot Borradaile): 'Man is metazoan, triploblastic, chordate, vertebrate, pentadactyle, mammalian, eutherian, primate'. The aim was to demonstrate the contrast between the two ways of thinking. It is a linguistic reflection of the 'two cultures' theme which C. P. Snow and others were expounding at the time (1959).

One consequence of this was a focus on vocabulary, always the most noticeable dimension of a linguistic variety because there is so much of it. In English we have a language which has 26 letters, typically 44 spoken vowels and consonants, and some 3,500 grammatical constructions, but a vocabulary which is well over a million items – 90 per cent of which is terminology. The identity of scientific English, as Derek Davy and I helped to demonstrate in 1969, in *Investigating English Style*, is a consequence of all these factors, and when it comes to intelligibility the grammar and graphic presentation is just as important as the lexicon. But for a long time the discussion of scientific English focused on the vocabulary alone. The reason is obvious: vocabulary is easy to talk about; grammar is not. We can all produce a list of scientific vocabulary at the drop of a hat. We need no especial linguistic skill. But if I were to ask you to give me a list of the distinctive grammatical constructions in scientific English, you would be hard-pressed, unless your ability to parse sentences was part of the same dropped hat. The issue is a general one, by no means restricted to science. Contemporary linguistic amateurs who give accounts of English – I am thinking of radio series such as *Word of Mouth* or *Routes of English* – usually do little more than talk about the lexicon. Melvyn Bragg's recent television *Adventure of English* is actually only an adventure in the history of English vocabulary.

So, in most discussions of scientific English, we find a focus on the words, and this is what Savory does – and does very well. Note his chapter headings: Chapter 2 'The Words of Science'; Chapter 4 'On word formation'; Chapter 5 'A Science Word Museum'; Chapter 6 'The Vocabularies of the Sciences'. And his Chapter 3, 'The Growth of the Language of Science' is entirely devoted to a historical account of which terms arrived in which centuries. All the themes of later discussion of scientific English are found here. He emphasises the Latin and Greek origins of most scientific terminology, something which dates from the Middle Ages, and which gives English – as I have recently argued in my book, *The Stories of English* (2004) – its diglossic character, manifesting 'low'

and 'high' levels of style. Ordinary people say that something is *shapeless*; scientists (or people trying to impress others with their scientific know-how) would say that it was *amorphous*. Ordinary people talk about *sleeplessness*; scientists about *insomnia*. There are hundreds of such cases where a simple translation between scientific term and everyday concept is possible. Alongside these we find the scientific terms which do not have such ready word-for-word translatability, where to gloss them adequately would require a phrase, a sentence, a paragraph – or even a book. Savory distinguishes three kinds. First, there are words which have been borrowed from ordinary speech and given a scientific application, such as *salt*, *atom*, and *force*. Second, there are words which have been borrowed from other languages – chiefly, as already mentioned, Latin and Greek – such as *cortex*, *nucleus*, and *vertebra*. And third, there are the words built up out of other elements of English – prefixes and suffixes, as in such cases as *ultraviolet*, *univalve*, and all the words beginning with *hyper-* and *hypo-* – or invented words, as in many drug names.

Vocabulary is the only area of scientific language investigated in detail in Savory's book, and that was typical of the time. It would be another two decades before linguists such as Michael Halliday would pay proper attention to the features of scientific grammar and discourse (see, for example, Halliday 1993). In Savory, references to grammar and other general properties of scientific language are incidental and tend to be vague. Some of his observations are worth quoting, nonetheless, for they identify a series of stereotypes about the nature of scientific English which have lasted until the present day.

The chief myths are that scientific language is emotionless, factual, objective, and stable. Here is Savory:

in the mouths and in the books of the students of science, emotive language is rare and informative language is common (p 17)
 the scientist does not write in metaphors (p.116)
 the language of science makes no provision for the slightest gleam of humour (p.116)
 scientific words do not change their meanings in the course of centuries, as many ordinary words do (p. 44)

Having met a fair few scientists over the years, I can personally vouch for the fact that they use emotive language a lot, write in metaphors a lot, do have a sense of humour, and often change their mind. But for purposes of illustration, let me offer the following. First, emotive language:

Suddenly Rosy came from behind the lab bench that separated us and began moving towards me. Fearing that in her hot anger she might strike me, I grabbed up the Pauling manuscript and hastily retreated to the open door. My escape was blocked by Maurice, who, searching for me, had just then stuck his head through. While Maurice and Rosy looked at each other over

my slouching figure, I lamely told Maurice that the conversation between Rosy and me was over and that I had been about to look for him in the tea room. Simultaneously I was inching my body from between them, leaving Maurice face to face with Rosy.

This is not from a crime novel, though such words as *anger*, *hastily*, *escape*, *slouching*, and *lamely* might lead you to think so. It is from James Watson's book recounting the discovery of the structure of DNA, *The Double Helix* (p. 131). He calls it 'a personal account', but that does not disqualify him as a scientist, or his book as popular science.

Second, no metaphor? I need perhaps do no more than quote some book titles from one of the most well-known scientists of our age, Richard Dawkins:

The Blind Watchmaker
The Selfish Gene
River out of Eden

Or this paragraph from his *River out of Eden*. Note the metaphorical way (italicized) of expressing the content:

In every one of your cells, half your mother's genes *rub shoulders* with half your father's genes. Your maternal genes and your paternal genes *conspire* with one another *most intimately* to make you the *subtle and indivisible amalgam* you are. But the genes do not blend. Only their effects do. The genes themselves have a *flintlike integrity*.

And so it goes on, in a river of metaphor.

And if I had to sum up in single words the evidence of humour in scientists, I would simply look at physics and recite: *quark*, *hedgehog*, *boojum*. This last, for example, became the subject of a whole book, by David Mermin, *Boojums all the Way Through: Communicating Science in a Prosaic Age* (1990). He needed a term for the phenomenon where an observed symmetry in the behaviour of the superfluid helium-3 was seen to disappear, and a different pattern to emerge. He found the term he needed at the end of Lewis Carroll's poem 'The Hunting of the Snark', where the last lines are:

He had softly and suddenly vanished away
For the Snark was a Boojum, you see.

Thereby following in the tradition of stealing words from humorous literature – *quark* (from James Joyce's *Finnegan's Wake*) being its illustrious predecessor.

Fourth, stability of meaning. The only scientific subject I know well enough here is my own, linguistics, often referred to as 'linguistic science' – or even 'linguistic sciences' when phonetics is seen as separate from

linguistics, or 'speech sciences' when the focus is on the contributing disciplines to the study of spoken language – physics (acoustics), anatomy, physiology, and neurology. The first edition of my *Dictionary of Linguistics and Phonetics* came out in 1980. It has since gone through five editions. In each edition it has proved necessary to make changes to about 15 per cent of the book, to keep pace with changing terms and senses. This is partly a reflection of the relative immaturity of linguistics as a science. As linguist Dwight Bolinger once put it (in *Aspects of Language*, p. 554):

One sign of immaturity [in a science] is the endless flow of terminology. The critical reader begins to wonder if some strange naming taboo attaches to the terms that a linguist uses, whereby when he dies they must be buried with him.

But more mature sciences evince significant lexical change too. All one needs to do to demonstrate this is compare the terms in a present-day dictionary of a science with the corresponding terms in a dictionary of a century ago. One could write a book on the historical lexicology of, say, *gravity* and the associated terms in its semantic field (*gravitation*, *gravities*, etc) over the past 400 years. Peter Medawar, in *The Art of the Soluble*, spends some pages discussing the varying meanings of *induction* and *experiment* over the centuries (p. 150ff). And if further evidence were needed, one would need to look only at the preamble to many a scientific paper, in which the authors take care to distinguish the meanings of the critical terms they use from those used by their predecessors.

Words, words, words. Whether stereotype or reality, the discussion of scientific English has been almost exclusively concerned with words. I suppose it is the sheer quantity of the terminology that makes this unavoidable. Scientific nomenclature comprises most of the vocabulary of English. We do not have a 'super-dictionary' of the whole of the English language, in this respect. The nearest, the *Oxford English Dictionary*, consists of something over half a million lexical items. But we know, for example, that there are nearly a million species of insects: the order Coleoptera (beetles) alone has some 350,000 species. Only a tiny fraction of these names are in the *OED*. Many of them will of course be Latinate phrasal names, but that is beside the point. If entomologists want to talk about such things in English then they must use the appropriate terms, and whether they are in Latin or not, ipso facto these then become part of the language.

But the weight of subject-specific nomenclature should not blind us to what is in many respects a much more important feature of scientific vocabulary – namely, the subject-neutral terms without which no science could begin to express itself at all. These are the words which scientists need in order to give instructions to act in a certain way, or to report on the consequences of having so acted. They include, for example:

- verbs of exposition: *ascertain, assume, determine, estimate, label, record, verify*
- verbs of warning and advising: *avoid, check, prevent, remember, not drop, not spill*
- verbs of manipulation: *adjust, assemble, boil, clamp, connect, dilute, mix, weigh*
- adjectives/adverbs of method: *careful, continuously, finally, gradual, secure, vertical*
- adverbs of uncertainty: *approximately, usually, often, almost certainly*

Here is a short 100-word example, taken at random from a recent paper in acoustic phonetics, to show just how important these terms are:

Pitch contours were calculated, using Praat, for the entire set of 270 analysis tokens and 102 test tokens. Pitch can be difficult to measure in breathy and creaky voice due to the amount of noise that is superimposed on the periodic signal in breathy phonation, and the irregular nature of vibratory cycles in creaky phonation. Since the objective was to use curve-fitting software to estimate equations for each contour, it was important to ensure that the measurements were accurate, and to have sufficient measurement points for curve fitting. Pitch was estimated every 0.01 seconds. Several methods were used to ensure the accuracy of the results... (*Journal of the International Phonetic Association*, Dec 2004, p. 128)

Let us analyse the lexical content of this passage. Of the 105 words, just under half are grammatical words such as *the* and *of*. If we omit these we are left with the following 59 words:

Pitch contours calculated, using Praat, entire set 270 analysis tokens 102 test tokens. Pitch difficult measure breathy creaky voice amount noise superimposed periodic signal breathy phonation, irregular nature vibratory cycles creaky phonation. objective use curve-fitting software estimate equations contour, important ensure measurements accurate, have sufficient measurement points curve fitting. Pitch estimated 0.01 seconds. Several methods used ensure accuracy results

Let us now identify the words that are specific to acoustic phonetics. There are only 18 of them:

pitch contours Praat pitch breathy creaky voice noise periodic signal breathy phonation vibratory cycles creaky phonation contour pitch

That leaves 41 which fall into the category of terms of general science:

calculated using entire set 270 analysis tokens 102 test tokens difficult measure amount superimposed irregular nature objective use curve-fitting software estimate equations important ensure measurements accurate have sufficient measurement points curve fitting estimated 0.01 seconds several methods used ensure accuracy results

That is 70 per cent. It is a typical figure. And it would enable me to boast that I understand perfectly some 70 per cent of what, for example, Sir Roger Penrose has written – without any physics training whatsoever!

But there is something else about the vocabulary of science which we must not ignore. Much of the intelligibility of a scientific text is due to the way a relatively small number of terms is manipulated to control the direction of the discourse. By 'discourse' I mean the connectedness of what is being spoken or written. It is what enables us to sense that an exposition has a beginning, a middle, and an end, or that someone has made a series of distinct points, or has made a contrast between point A and point B. All coherent discourse contains signposts which help the writer (or speaker) organize thoughts and the reader (or listener) to follow them. A good piece of writing will have many of them, and if the subject matter is especially difficult, they prove to be life-savers. Here is an example: one of the longest paragraphs in Stephen Hawking's *A Brief History of Time*, in Chapter 9 (p. 150) is 27 lines long, and it immediately follows the sentence 'People in the contracting phase [of the universe] would live their lives backward: they would die before they were born and get younger as the universe contracted'. The next paragraph, we might imagine, will be a test case as to whether we ever reach Chapter 10. But Hawking leads us very gently through it, as you will sense if I quote just the first sentence and the subsequent discourse signposts:

This idea is attractive because it would mean a nice symmetry between the expanding and contracting phases. However... The question is: ... As I said... However... Further...

Before you know where you are, you have reached the end of the paragraph, and been fooled, by the sophisticated clarity of expression, into thinking that a degree in cosmological physics wouldn't be such a big deal after all.

It's all in the discourse markers. These items, like *however, furthermore*, and *nevertheless*, are part of a complex system of several hundred words and phrases in English that express the movement of thought. They occur in speech as well as writing, and in all languages, and are so subtle that we often do not notice they are there. They are, incidentally, among the most difficult things to pick up in a foreign language. I once had to give a lecture in French, and spent a great deal of time preparing what I wanted to say, ensuring that my sentences were terminologically spot-on. But when I gave the lecture it was a disaster, as I had omitted to prepare the discourse markers to a comparable level, and found myself making several hundred isolated points. I simply could not remember the French for *furthermore, incidentally, as a matter of fact, frankly*, and a host of other crucial linking words and phrases. I pitied my audience more than I pitied myself.

This leads to the third topic in the trilogy: the COMMUNICATION of science. Here too, Theodore Savory is unequivocal. In a chapter on the

nature of scientific prose, talking about its apparent incomprehensibility to outsiders and its clarity to insiders, he says this:

Science books are only exceptionally written in prose of the highest quality, and more often they are written in prose that must be described in contrasting terms, yet the students who use these books do not in fact find them hard to understand. (p.133)

And he concludes:

There are plenty of obstacles to scientific research, but the problem of language is not, as a rule, the most formidable (p.142)

I have to disagree. It does not square with experience at any of the three main levels at which scientific language is encountered: by scientists communicating with each other; by scientists communicating their subject to their students; or by scientists communicating their subject to a general public. It does not square, firstly, with the way scientists communicate with each other, for many papers begin by the authors expressing their uncertainty or disquiet over the way previous authors have expressed themselves. And in relation to the second, pedagogical, level, I have unequivocal – albeit anecdotal – evidence that Savory is wrong. ‘The students who use these books do not in fact find them hard to understand’?

In 2005 I received an email from a mature student in her fifth year of studies at the Open University. The OU is offering a new English language course, E303, called English Grammar in Context, and this student had decided to do this one alongside S282 Astronomy and Cosmology. An unusual combination, and an illuminating one, for when she found the latter course unexpectedly difficult, she used the knowledge from her language course to work out why. This was her conclusion:

Not only did I find the ideas hard to absorb in first reading, but I then found it extremely difficult to navigate around the coursebooks in order to revisit a topic. I came to the conclusion that most of the writers (though not all) were so much in the habit of presenting the subject in the traditional scientific manner that this still took priority over communicating their knowledge. I could see they’d tried to help – some of them included little jokey bits at the beginning, and they were obviously under threat of slow strangulation to ensure that they kept their sentences simple and short. But those sentences often had no linking adverbials (‘however’, ‘so’ etc), and were just sprayed out in short, machine-gun bursts. So it was hard for me to tell how the subject of one sentence was related to the subject of the previous one – was it cause, or effect, or something else?

She identified other problems with discourse expression, and they were all instances where I have encountered problems myself, in reading scientific literature. For example:

- lack of detail within the contents pages
- the practice of creating cross-references to sections (which could cover many pages) instead of page numbers. Or even worse, not even specifying sections, but stating simply ‘earlier’, ‘previously’, or ‘later’.

And she mentioned a difficulty with one of the biggest problem areas of all – the adequacy of the index. This is a well-recognized issue. The journal of the Society of Indexers, *The Indexer*, has a regular section in which book indexes are praised or censured. It is disturbing to find so many science books falling within the latter category. But I am not surprised. As a former president of that Society, I am well aware that an unacceptably high proportion of science authors find indexing an irritating chore, to be avoided if possible (‘let the publisher do it’) and if not, to be done in the shortest possible time and space.

You might think the point about indexing is trivial – the tail of the book, only. And why should the tail wag the dog? All readers of *Moderna Språk* know the answer to that question, for they have all looked something up in an index and not been able to find it, or looked something up and found too many unstructured references, and cursed the author. Some elements of indexing theory and practice ought to be part of the routine training of all scientists. Especially now.

Why now? Because for scientists, as for everyone else, life is about to change dramatically, as we move from a book-based information world to a World Wide Web-based one. And the Web presents us with an indexing problem of an unprecedented scale. A search-engine such as Google indexes every word (bar a few stop-words, such as *the*). You might think this is a good thing, until you carry out a search for a particular topic and find you have a million hits. Life is too short to look at more than the first few dozen. Moreover, most of those hits will turn out to be irrelevant. Let me give you a scenario which you can try out for yourselves next time you log on. Imagine that you are interested in ‘depression’ in the sense of climate. Type *depression* into Google. You will get, as of the week I am writing this paper, 18,700,000 hits. But the first page, and all the sponsored links down the right-hand side of the screen, are to *depression* in the sense of mental health. You will have to scroll down for several pages before you find some sites relevant to your query. Nor is it easy to solve this problem by adding extra search words to your query. We know that depressions are often *deep*, so you might type *deep depression* into Google. You will still get sites offering you tablets to help raise your mood. Eventually – if you have the patience – you will stumble upon a combination of query terms which will give you some relevant sites. But this time-wasting trial-and-error approach to information retrieval is not how scientific enquiry should proceed.

More and more scientific information is being made available on the

Web, but little attention seems to have yet been paid to such questions as how it should be indexed, how cross-references (now called 'hypertext links') should be motivated, or how it should be presented on screen to maximise legibility. The practice of making journals available online is increasing. More publishers are allowing books, or extracts of books, to be available online, through such sites as ebrary. Several of my own books and articles are now 'out there', but all that has happened is that they have been given an electronic presence. No attempt has been made to manipulate the content to suit the new medium. Footnotes are there as footnotes still. You would have to scroll through the pages to find a footnote at the back of a book. Cross-references still say 'see p. 66' and you have to scroll to find p. 66. And the indexes remain indexes at the back of the book, and not (as they could so easily be) dynamic means of collating related information obtained by clicking on a particular word.

It is early days. All this will happen eventually. But we must begin to prepare for this brave new world now. If the communication of science is going to be increasingly computer-based, then scientists need training in Web management procedures as well as indexing. They have to anticipate the demands made upon their writing by people working electronically. And they have to recognise the strengths and weaknesses of the Web. The strengths are obvious – immediate access to sources that were previously inaccessible, and rapid in-text word searching. Here are two weaknesses.

For a fast-moving domain such as science, a major weakness is that the Web is comprehensively diachronic – by which I mean that everything is stored, nothing is lost. If I do a search for, say, *gravity*, I will get every document that contains the word *gravity* that has been made available to the Web since it was invented in 1991. In ten years time, the Web will contain documentation covering 25 years. Old ideas will be juxtaposed alongside new ideas, old findings alongside new findings. For some subjects it does not matter. If I want to find the words of a Beatles song, it does not matter if they were put up on the Web in 1995 or 2005. But with science it matters greatly when the document was written. And at the moment there is no easy way of distinguishing old documents from new. It is usually very difficult to find out when a document was written, for there is nothing in the page as downloaded on your screen to tell you. It is a new type of communication problem for students and science journalists, who are perpetually at risk of using out-of-date information. Traditional sources at least would routinely contain a date at the beginning of the article or book.

It is more than science professionals who have a problem. Ordinary people are affected too, as evidenced by the billion or so self-help sites now on the Web. Judging by the hit-rates on some of these sites, behaviour is changing. If people feel ill, their first stop used to be the local surgery. Now it is the Web. A contact I have in a local pharmacy tells me that the

amount of self-diagnosis is rapidly increasing – people coming in to the chemist (a real example, this) asking for some ointment to treat their ringworm, which they have identified through the symptomatology found on a Website. The sites' legal disclaimers do not deter. I do not know how many scientific domains will eventually develop applications in the way that medicine has, but it is going to be a lot. Many areas of technology, such as building science and electrical engineering, are already showing signs of growth. And we know, from recent news reports, that it is possible to build your own bomb using the Web. Self-help indeed.

But back to linguistics, and to a second major weakness of Web searching. One of the things the Web has done is increase the amount of ambiguity in language. If you do a count of all the senses of all the words in a College English dictionary, you will find that the average number of senses per word is 2.5. But in the past few years, every common word in English has been bought by individuals as a potential domain name. There are sites where you can look to see if a word is available, if you wanted to set up a Website of your own, and you would find that it would not be – though you might well find that someone was willing to sell it to you. If you wanted to call your site 'answers.com', for example, it would probably cost you around 10,000 dollars (if the owner was willing to sell). But my point is not the economics of the Web, but the linguistics. By giving all words a domain-name incarnation, the amount of ambiguity in English has been virtually doubled in ten years. The word *answers* means not only what it used to mean, but also, now, the content of the Website going under that name. And as 'answers.com' is different from 'answers.org' and 'answers.co' and the dozen or so other suffixal options currently available, the ambiguity has probably multiplied.

The point affects all of us, in all our worlds and all our languages, but this paper is about science. Scientific vocabulary is traditionally said to be relatively unambiguous, and it is true that most scientific words have just a single meaning. Look up *Tyrannosaurus rex* or *red shift* in a dictionary and you will get one meaning for each. But times have changed. Look them up on the Web and you get a different result. *Red shift* is not only a concept in cosmology, it is the name of a company offering fast Internet services, the name of a film festival, and the name of a theatre company – to take just three of the many additional senses encountered on the first page of a Google search for this item. Admittedly, there are many words which have not yet been affected. *Hydrocortisone* is still only hydrocortisone, on the Web. But for how long? In the old days – pre-1991 – the matter would be insignificant, for the juxtaposition of these senses would never have been perceived. On the Web it cannot be avoided. And it interferes with the communication of science in profound ways.

The solution is obvious, but not yet available. A search for *red shift* can

be made unambiguous in one of two ways. We can do a negative search: get me *red shift* but not in the sense of theatre, film, etc. Or we can do a positive search: get me *red shift* in the sense of cosmology. As there are a potentially infinite number of negative senses, it obviously makes sense to go for the second option. But it is scientists who have to do that. For laymen cannot decide which is the best scientific heading that *red shift* should be assigned to. If I categorise it as simply 'cosmology' this is probably far too general. It would be best categorised as a sub-branch of cosmology, or maybe some other subject. But which? Only the relevant scientists can say. In a field like zoology, finding the appropriate level of classification is a well-recognised issue of nomenclature. All branches of science face it now. And it will radically affect the way we think of our output, as scientists. For example, in providing keywords for Web pages we will need to do far more than we traditionally did when we provided them for paper periodicals. If we want our readers to find our work quickly on the Web, we will need to provide them with electronic beacons in the form of better indexing practices. And the discussion of this issue has hardly begun.²

The World Wide Web is altering our sense of what it means to communicate science in the twenty-first century. It is such a new perspective, for most people, that it lacks systematic investigation. Much of what I have said in this paper has been anecdotal, and I regret that we have so little empirical evidence about the nature of scientific practice. But this regret is nothing new. Peter Medawar wanted the same thing, 40 years ago:

What scientists do has never been the subject of a scientific, that is, an ethnological inquiry. It is no use looking to scientific 'papers', for they not merely conceal but actively misrepresent the reasoning that goes into the work they describe. (p. 169)

We have to examine what scientists actually say, he argues – and what they say, he reports, having listened at keyholes in a biological lab, is things like this:

I don't seem to be getting anywhere...
I'm still at the stage of trying to find out if there is anything to be explained...
That's a very good question...

And the one with which almost all scientific papers conclude, and which is definitely applicable in the case of the present paper:

Obviously a great deal more work has still to be done.

² For one solution to this problem, see the Textonomy 'sense engine' described at www.crystalsemantics.com

References

- Bolinger, Dwight (1968). *Aspects of Language*. New York: Harcourt Brace Jovanovich.
- Bragg, Melvyn (2003). *The Adventure of English*. London: Hodder & Stoughton.
- Crystal, David (1980). *A Dictionary of Linguistics and Phonetics* (5th edn, 2003). Oxford: Blackwell.
- Crystal, David (2004). *The Stories of English*. London: Allen Lane.
- Crystal, David & Davy, Derek (1969). *Investigating English Style*. London: Longman.
- Dawkins, Richard (1995). *River out of Eden*. London: Weidenfeld & Nicholson.
- Ellen, S.R. (1979). Survey of foreign-language problems facing the research worker. *Interlending Review* 7, 31-41.
- Halliday, M.A.K. (1993). 'Some Grammatical Problems in Scientific English' in Halliday, M.A.K. & Martin, J. R. (eds.) *Writing Science: Literacy and Discursive Power*. London: Falmer Press, 69-85.
- Hawking, Stephen (1969). *A Brief History of Time*. London: Bantam.
- Large, J.A. (1983). *The Foreign-Language Barrier: Problems in Scientific Communication*. London: Deutsch.
- Medawar, P.B. (1967). *The Art of the Soluble*. London: Methuen (page reference to the 1969 Penguin edition)
- Mermin, David (1990). *Boojums all the Way Through: Communicating Science in a Prosaic Age*. Cambridge: Cambridge University Press.
- Savory, T.H. (1953). *The Language of Science*. London: Deutsch.
- Snow, C.P. (1959). *The Two Cultures and the Scientific Revolution*. Cambridge: The Rede Lecture.
- Watson, James D. (1968). *The Double Helix*. London: Weidenfeld & Nicholson (page reference to the 1970 Penguin edition.)