SCIENTIFIC ENGLISH: AN OVERSIGHT IN STYLISTICS?

DONALD PORTER

University of Warsaw

1. The work which forms the basis for this paper originated in an attempt to fill a clear need in language-teaching. I had been teaching English to students of scientific subjects for a number of years, and had been struck by the number of courses, and of teachers, purporting to teach the English relevant to such students, but almost without exception failing to produce *any* evidence that this teaching was ever based on any systematic description of the language as encountered, and needed, by students. To my knowledge, only a handful of courses are based on a linguistic analysis, and in spite of the inestimable value these analyses would have for others engaged in constructing teaching materials in this important field, they have never been published. I suspect this is partly for reasons of linguistic inadequacy, partly because of questionable principles underlying the selection of materials for analysis.

2. This is not to say that the 'Language of Science' has not received linguistic attention; merely that the results have not, so far as can be seen, been taken into account in the construction of materials.

The attempts which have been made at a linguistic characterization have been surprisingly careless. As far back as 1938, Bloomfield wrote a section in *The international encyclopedia of unified science* called "Linguistic aspects of science". In spite of the fact that the section is quite lengthy — more than 40 pages — Bloomfield gives explicit examples of what he is talking about only once (1938:261) when he says that fundamental scientific processes (I interpret: processes common to all branches of science) produce in scientific language frequent:

(a) expressions of exclusion: e. g. *not*, the verb *exclude*, the sentence-structure *if* ... , *then*...;
(b) words or phrases of existence or predication: e. g. *there exists*...; ...*is*...;
(c) equational sentences: ...*means*...; ...*equals*...;

— for Bloomfield, these are surprisingly informal, ad hoc classifications, but this is the nearest he comes to making actual syntactic description. He also makes a claim about sentence-connection in scientific discourse, but this is simply left as an assertion, with no explicit information, no illustration of what he has in mind being given.

Many writers have written about 'the Language of Science', but refer only to scientific vocabulary (e. g. Grove 1949, Savory 1953, Jespersen 1958, Baugh 1959). In particular, Savory's book: *The language of science, it's growth, character and usage* is often mentioned.[1] I can see no reason why anyone could conceive of taking this book seriously; quite apart from equating 'language' with 'vocabulary', Savory is often prescriptive, and full of subjective vagueness, e. g. (1953: 67):

> Invention of new words should aim at 3 qualities: brevity, euphony, and purity.

and:

> It almost seems as if scientists preferred ugly words.

This sort of attitude permeats the book. Where Savory does seem to at least attempt objectivity, he is found hopelessly wanting: in writing of the comparative frequency of words of Greek and Latin origin in scientific texts, he bases his conclusions on "thirty consecutive words taken from the index of one of the most popular and widely used textbooks of biology of the day" (1953:25). And so he goes on.

The most frequently cited early serious attempt to define the characteristics of scientific English is that by Barber (1962). Barber's purpose, like mine, was "to provide teachers of English as a foreign language with quantitative information on the language used in science, in particular those features which are common to scientific texts". (1962). However, this initial assumption that there *are* features common to scientific texts in unfortunate, as it leads Barber to economize in the texts he investigates: he makes 3 texts represent 6 subject areas, by choosing texts which 'straddle' two subject areas each — e. g. he makes a text in biochemistry suffice for both biology and chemistry, although it remains to be shown whether the characteristics of biochemistry are in fact the same as either or both.

Moreover, in spite of his declared aim of providing quantitative information on features common to scientific texts, comparison of texts is minimal. Barber's interest appears to be solely in overall frequency totals for his corpus:

[1] For example, the British Council lists it in its current ETIC *Information guide on English for special purposes.*

thus the totals with which his analysis provides us may obscure wide divergencies between texts in the frequency of occurrence of syntactic categories: such divergencies *can* only be brought to light by systematic and explicit comparison of texts.

Again, Barber concerns himself with very few features of his texts: vocabulary, tense-forms, sentence-lengths and clause-types. And the last two of these are considered for one text only, which of course makes them valid for that text only, so that not even tentative conclusions can be drawn about sentence-length and clause-type characteristics common to scientific texts. There is little or no definition of the syntactic categories considered, which makes the findings at least suspect, and at worst useless: if we do not know what Barber includes under 'sentence', and what he excludes, then his discussion of the number of words per sentence cannot begin to be appraised.

Barber's study is thus of interest not so much for the information which it provides or the light it sheds on a Scientific English style, as for its attempt at a statistical approach, and its illustration of what should and what should not be done in such an investigation.

By far the largest and most thorough study of Scientific English is that carried out by Huddleston, Hudson, Winter and Henrici, and reported in *Sentence and clause in scientific English* (1968). The study is based on a large corpus consisting of 27 texts, each 5000 words long. The texts are grouped into the three subject-areas of physics, chemistry and biology, and also into three levels of 'brow' — 'high-stratum' texts from specialist journals, 'mid-stratum' texts from undergraduate text-books, and 'low-stratum' texts from popular science addressed to the educated layman. The aim of the study is stated very generally to be "to give an account of selected areas of the grammar of written scientific English" (1968), this account being a statistical appraisal of the occurrence in the corpus of carefully defined syntactic features. The study concentrates on the clause and its constituents, and is said to be 'influenced by' Hallidayan systemic grammar.

3. In view of the existence of this massive study, I decided to break new ground in my own investigations, and to look into the characteristics of Scientific English as revealed in the syntactic devices of *sentence-connection*. That particular types of sentence-connection are characteristic of Scientific English was claimed by Bloomfield in the above-mentioned article. He says, cryptically, that from "fundamental scientific processes" arise the rules of calculation, and these rules govern the sequence of sentences. Then (1938: 262):

> *Formal, scientific discourse uses a rigidly limited vocabulary and syntax, and moves from sentence to sentence only within the range of conventional rules* (my italics).

I take this as meaning conventional rules for Scientific English, not for English as a whole — or the sentence would be pointless. The claim is a strong one, but Bloomfield unfortunately does not justify it. I decided to try to justify it for him.

4. For want of an integrated linguistic model for dealing with intersentential relationships, I decided to take a purely surface approach, basing my analysis on the relevant sections of Quirk, Greenbaum, Leech and Svartvik (1972) and Greenbaum (1969).

For the purpose of this work, the notion of Scientific English was restricted to the language used in university undergraduate textbooks. Thirty-six different aspects of sentence-connection were considered from various points of view, as manifested in ten texts of 100 sentences each, two texts being taken from each of five subject areas: physics, chemistry, botany, zoology and mathematics. The limitation to a hundred sentences per text was not considered to be over-restrictive, as it was felt that, if it was indeed possible to characterize Scientific English in terms of sentence-connection, the characteristics should be evident in a text of this length, and — for the title 'Scientific English' to be valid at this level — should be evident in any text from the various subject-areas which are normally considered to be part of 'science'.

5. In attempting to state explicitly some characteristics of Scientific English, I was clearly attempting the partial delineation of a style. As the notion of *a style* means so many different things to so many different people, a few words are in order here in partial explanation of what I consider to be at least necessary — if not sufficient — to an adequate definition of *a style* in a non-literary sense.

The notion of *a style* is essentially two-sided: it has a linguistic and a non-linguistic face. If either is missing, the notion becomes vacuous. Thus it would be pointless to talk of some sort of recurring configuration of features in language unless we also stated in relation to what these recurring configurations are found. It would be rather like saying: "certain configurations of nouns sometimes occur". So when we define a style, we must make an explicit statement of its linguistic characteristics, and an explicit statement of the non-linguistic variable(s) with which the particular configuration of linguistic characterics exclusively occurs.

Over the years, considerable effort has been expended in elaborating sets of relevant non-linguistic variables. Thus Firth (1950) proposed a set of parameters for describing the 'context of situation' with which language varies. Then Catford (1961), and Halliday (1964) developed the notion of 'register', proposing a new set of non-linguistic variables with which language was said to covary. This register-notion underwent minor modifications in

the works of Spencer and Gregory (1964) and Leech (1966). Enkvist (1964) and Crystal and Davy (1969) proposed rather more elaborate schemes of situational variables, while Lakoff (1972) and Fillmore (1973) suggest still further refinements and additions.

None of these writers fully characterise the variables they propose in terms of specific linguistic features; without such characterization, there is no natural limit to the proliferation of such systems, no way of checking whether the variables proposed are in fact genuine, whether they are all distinct or perhaps sometimes merely the same variable under two or more different labels. Such systems are mere constructions of words. A non-linguistic variable can only be justified if it is linked to a characteristic configuration of precisely defined linguistic features.

What would the characterization of a non-linguistic variable in linguistic terms look like? If our thinking concerns a single comprehensive grammar for a language, then any linguistic variation within that overall framework would have to be expressed in terms of restrictions on the grammar covarying with relevant non-linguistic variables. And these restrictions would be in the form of typical frequency-distributions. Of course, it might be that, in a given style, certain features, certain rules never occur. But this would merely be the extreme case on the frequently-scale. As has often been said, the grammar is the same for all linguistic variation, differences between styles being accounted for by relative differences in prominence of various linguistic features.

Again, the characterization of a style would consist not of individual linguistic features, but of the *total configuration* of features regularly cooccurring with the non-linguistic variable(s). Thus, in attempting to set up 'a non-linguistic variable as valid in a style-description, one would have to demonstrate the regular occurrence of certain linguistic features, within a given frequency-range for each, wherever texts representing that non-linguistic variable occurred. The texts would be homogeneous with respect to those features. We would call this an 'internal definition' of the style. Next, to prove that this variable is not itself subsumed within a more general variable, we would have to show that where the variable does not occur, the characteristic configuration of features also does not occur. This distinguishing of the linguistic characteristics of a style from those of other styles we would call an 'external definition'. It is in this 'external definition' that the specification of the style as a *configuration* of characteristics, and not just in terms of isolated features, is important.

It might well be that a style is quite distinct from other styles, while manifesting no single linguistic characteristic which occurs typically and only in that style. For example, in the following diagrammatic representation, five styles are characterized uniquely by three features, none of which is

restricted to a single style:

features	styles 1	2	3	4	5
		a	a	a	
	b	b			b
		c	c		c

It is the overall configuration of features which is unique in each case. Another point: the characterization of a style would not simply concern itself with the most frequently occurring features, but with any features which could be considered to fall within the same range from text to text within the style. The position of Crystal and Davy (1969: 21) that the most frequently occurring features are the most important, seems to me untenable. A regular low-frequency occurrence is surely just as characteristic as a regular high-frequency — and both are quite distinct from purely random occurrence.

6. The test which is normally applied to discover whether a set of frequencies can be considered to be samples from the same population, i.e., to be 'within the same range', is the χ^2 test. There is an important drawback to the operation of this test, however: it cannot be applied where expected frequencies are less than 5. The meaning of this limitation, of course, is that with expected frequencies of less than 5, any variation between the frequencies observed for a range of texts could be the product of chance. I have already required that low *and* high frequencies be considered to characterize a style equally well. It seems that the requirement can exist in stylistic theory, but not in statistical procedure. So where a linguistic feature characteristically occurs with low frequency in scientific texts, there is no means of knowing whether the differences in frequency which we observe from text to text should be considered natural variation within a characteristic range, or whether the differences are so great that the feature is irrelevant to the style.

7. I shall not go into the details of the syntactic analysis here, as these have been presented elsewhere,[2] and it is the overall results which are the point of this paper.

When the ten texts were analysed for the thirty-six features of sentence--connection referred to above, the deficiency of the χ^2 test with respect to expected frequencies of less than 5 was found to affect data for twenty-three of the features. So, statistically, the differences and similarities which I might have observed in my corpus for over two-thirds on the features considered cannot with confidence be said to be anything other than the product of chance. It might be argued that if the corpus had been larger — say double the present size — then the figures for frequency of occurrence of linguistic

[2] Porter (1974).

features would have been correspondingly larger, the χ^2 test could have applied to more of them, and we would have a fuller picture of the characteristics of Scientific English at this level.

There may be some truth in this. However, although there was not sufficient time to carry out this extra analysis, we are in a position to draw an interim conclusion: if there *is* some overall characteristic for Scientific English with respect to these features, it will be that they tend to occur with very low frequency — less than 5 per hundred sentences.

However, there are two reasons for predicting that, in fact, there would be no linguistic feature, within any frequency range, which would characterize Scientific English beyond the sentence.

Firstly, for the 13 sentence-connecting features whose frequency of occurrence was sufficiently high for the χ^2 test to apply, χ^2 values indicate that variation between observed and expected values could not be regarded as insignificant: that is, statistically, no significant regularity could be observed over the texts of the corpus in the frequency of occurrence of any sentence-connecting feature. Thus, taking all the 36 features of sentence-connection together, we can say that *no* evidence was found of a characteristic style which we could label 'Scientific English'.

Of course, we could criticize the statistical assumptions on which the work is based. *Is* it possible to describe a style statistically? Perhaps we have to say no, in spite of the many assertions to the contrary.[3] Or possibly more sophisticated tests of significance than χ^2 are needed? Perhaps the linguistic base of the analysis was inadequate? Or could it be that the linguistic characteristics of Scientific English style are to be found exclusively below sentence-level?

My second reason, however, for predicting that no feature will be found to characterize Scientific English beyond the sentence, makes these various objections irrelevant, and would imply that both the surface approach of the linguistic description and the statistical treatment are justified.

When the five different subject areas (mathematics, zoology, botany, physics and chemistry) were put in rank-order for each of the sentence-connection features considered, ordering by frequency of occurrence, they showed an overwhelming tendency to rank in the same way each time:

Mathematics — Physics — Chemistry — Botany — Zoology

Thus, mathematics might show a high frequency, zoology a low one, and the other subject would be strung out between the two; for example, con-

[3] For example, Enkvist (1964), Ellegård (1962a and 1962b), Burton (1975); in very similar sociolinguistic matters, Weinreich, Labov and Herzog (1968), Labov (1969), Doležel and Bailey (1969), Bickerton (1971), Cedergren and Sankoff (1973), etc.

nective use of the definite article went:

Math	—	Phys	—	Chem	—	Bot	—	Zoo
72		71		61		40		25

Or, alternatively, mathematics might appear towards the low-frequency end, and the order be reversed; c. g. the overall figures for connective ellipsis:

Phys	—	Math	—	Chem	—	Bot	—	Zoo
1		2		5		6		31

The picture was a little more complex than I have presented here as, for example, chemistry tended to be more 'mobile' than other subject areas, sometimes appearing closer to the maths end of the spectrum, sometimes towards the zoology end. Thus, when the statistics for chemistry were removed from consideration, the overall trend in ranking of Mathematics — Physics — Botany — Zoology was even more striking. The appropriate test for significance here was χ^2_r; the rank order of subject areas for each feature of sentence-connection was placed in a column, the coefficient of concordance of rankings taken as a group was calculated, and the χ^2_r test showed an overwhelming tendency for the subject areas to be ranked in the same way.

My conclusion could only be that the existence of a Scientific English style had not been demonstrated at this level of analysis, but that there was a strong indication that different subject areas within science were distinguished at this level — the 'English of Physics', the 'English of Chemistry', etc.

8. When I mentioned these conclusions to some scientists, their response was something like: "Well, of course!" I interpreted this to mean that they had an intuitive knowledge of styles within science, and that they also had an intuition of some sort of scale of subject areas within science, a scale of abstractness perhaps — mathematics being the ultimate in abstractness and botany and zoology being more tightly linked to living reality. There is no reason to be surprised if the results of the linguistic analysis correspond to the intuitions of native speakers: indeed, any descriptively adequate theory of style must require that a linguistic description of a style correspond to those intuitions, and to that extent a stylistic description discriminating between the 'English of Physics', the 'English of Chemistry', etc., would be descriptively more adequate than one which did not make the distinction. On the other hand, it should by no means be taken for granted that an intuition about the existence of one style rather than another will in fact be substantiated by the objective study of data.

9. The conclusions stated are valid for syntactic sentence-connecting devices only. Is it the case that Scientific English as a discrete variety or style can be discerned only in its sentence-internal syntax? This in itself would be interesting — fascinating! — if it were true. The existence of such a style,

with characteristic linguistic features is, after all, implied in the titles of practically all the articles and books written on the subject, and in the content of the writings they introduce.

I have already mentioned the way in which Barber (1962) declared himself interested in "those features which are common to scientific texts", yet ruled out any possibility of finding such common features by his choice of texts for analysis, and by his failure to compare the occurrence of various structures in various frequencies in the texts that he chose. It is clear that Barber assumes the existence of a 'scientific' style from the beginning.

Then there was the study by Huddleston, Hudson, Winter and Henrici (1968) which led to the work discussed here being devoted to sentence-connecting features, as it was felt that the features of Scientific English below sentence level must have been dealt with in great detail. I have already mentioned that the aim of *Sentence and clause in scientific English* was to "give an account of selected areas of the grammar of scientific English". What leads the authors to assume that there *is* a "grammar of scientific English"? In spite of the fact that this was a really extensive research work, taking up the time and efforts of four men over a considerable period, no attempt was made to answer the seemingly absolutely fundamental question: *does Scientific English exist*? The authors write: "Little time was available to look for ways in which the Biology, Chemistry and Physics texts differed" (1968:175). As the quotation indicates, no systematic comparison was made between texts of different subject areas. As a result, all the carefully compiled tables of frequency, evaluations of relative frequency of items, and comparisons between subsections of the corpus defined by their level of 'brow', etc., are in danger of having been a monumental waste of time.

And so it goes on. Gerbert (1970) makes the same error with respect to technical English (a separate style?) by implicitly assuming that "die technische Fachsprache" has a stylistic identity with characteristics which can be defined.

Gopnik (1973) comments on general tendencies in scientific texts in terms of total occurrences observed in her corpus, but still permits herself to refer to "rules which underlie all texts" and to draw conclusions about scientific discourse in general (1973: 26, 27, 28 et passim).

And finally, Beneš makes the same assumption for scientific German, making his position very clear in his 1972 paper "The syntax of scientific German in foreign language teaching". He says:

> Like the standard language itself, even scientific style reveals considerable internal differentiation according to the general purpose of the discourse, the manner of its realization, and its subject. In spite of its many variants, *there exists a relatively homogeneous central area, the impersonal standard style, now commonly used, for example, in textbooks. In a synchronic study of scientific style, it is profitable to start from this central area, the standard scientific style of today* (my italics).

This is just not good enough! Where is his evidence? Let him make an explicit statement of the features common to scientific style, the whole to be distinct from other styles, and let his statement be checked against the data, whether this be texts, or our intuitions *about* texts.

10. So there is the oversight in stylistics mentioned in my title. It was suggested to me recently that stylistics is *really* a matter to be left until the port and cigar stage of a good meal. So long as no attempt is made to make explicit the notion of *a style*, that suggestion will be valid. Indeed, a recognition of the need for explicitness in style-studies is implied in the various statistical treatments which I have mentioned.

My own studies a) do not indicate the existence of a "relatively homogeneous" Scientific English style, and b) bring to light the unfounded assumption of many writers that Scientific English *is* in some sense a homogeneous style — perhaps below sentence-level. It *may* be — but who has taken the trouble to find out?

REFERENCES

Barber, C. L. 1962. "Some measurable characteristics of modern scientific prose". *Gothenburg studies in English 14*. 21 - 43.

Baugh, A. C. 1959. *A history of the English language*. London: Routledge and Kegan.

Beneš, E. 1972. "The syntax of scientific German in foreign language teaching". In Fried, V. (ed.). 1972. 142 - 159.

Bickerton, D. 1971. "Inherent variability and variable rules". *Foundations of language 7*. 457 - 492.

Bloomfield, L. 1938. "Linguistic aspects of science". In Neurath, O. et al. (eds). 1938. Vol. 1, Nos. 1 - 5.

Burton, D. 1975. *Style as a mathematical function*. Unpublished paper read to a meeting of the English Institute, University of Warsaw.

Catford, J. C. 1961. *Applied linguistics in the teaching of English as a secondary language*. Unpublished paper read at the Anglo-American Conference on English Teaching Abroad. Cambridge, June 1961.

Cedergren, H. and D. Sankoff. 1973. "Variable rules and performance in a statistical reflection of competence". Forthcoming.

Crystal, D. and D. Davy. 1969. *Investigating English style*. London: Longman.

Doležel, L. and R. W. Bailey. 1969. *Statistics and style*. New York: American Elsevier Publishing Co.

Ellegård, A. 1962a. *Who was Junius?* Stockholm: Almqvist and Wiksell.

Ellegård, A. 1962b. "A statistical method of determining authorship". *Gothenburg studies in English 13*.

Enkvist, N. E. 1964. "On defining style". In Enkvist, N. E., Spencer, J. and M. J. Gregory. 1964. 3 - 56.

Enkvist, N. E., Spencer, J. and M. J. Gregory. 1964. *Linguistics and style*. London: Oxford University Press.

Fillmore, C. J. 1973. "A grammarian looks to sociolinguistics". *Proceedings of the Georgetown Round Table Conference on Sociolinguistics*. 273 - 287.

Firth, J. R. 1950. "Personality in language and society". Reprinted in Firth, J. R. 1957. 177 - 189.

Firth, J. R. 1957. *Papers in linguistics 1934 - 1951*. London: Oxford University Press.

Fried, V. (ed.). 1972. *The Prague school of linguistics and language teaching*. London: Oxford University Press.

Gorbert, M. 1970. *Besonderheiten der Syntax in der technischen Fachsprache des Englischen*. Halle (Saale): Max Niemeyer Verlag.

Gopnik, M. 1973. *Linguistic structures in scientific texts*. The Hague: Mouton.

Greenbaum, S. 1969. *Studies in English adverbial usage*. London: Longman.

Grove, V. 1949. *The language bar*. London: Routledge and Kegan.

Halliday, M. A. K., MacIntosh, A. and P. Strevens. 1964. *The linguistic sciences and language teaching*. London: Longman.

Huddleston, R. D., Hudson, R. A., Winter, E. D. and A. Henrici. 1968. *Sentence and clause in scientific English*. London: Communication Research Centre, Department of General Linguistics, University College London.

Jespersen, O. 1958. (9th edition). *The growth and structure of the English language*. Oxford: Blackwell.

Labov, W. 1969. "Contraction, deletion and inherent variability of the English copula". *Language 45*. 715 - 762.

Lakoff, R. 1972. "Language in context". *Language 48*. 907 - 927.

Leech, G. 1966. *English in advertising*. London: Longman.

Lehman, W. P. and Y. Malkiel. (eds). 1968. *Directions for historical linguistics*. Austin and London: The University of Texas Press.

Neurath, O. et al. (eds). 1938. *The international encyclopedia of unified science*. Chicago: The University of Chicago Press.

Porter, D. 1974. *Scientific English beyond the sentence*. Unpublished M. Phil. dissertation, University of Reading.

Quirk, R., Greenbaum, S., Leech, G. and J. Svartvik. 1972. *A grammar of contemporary English*. London: Longman.

Savory, T. H. 1953. *The language of science*. London: Deutsch.

Spencer, J. and M. J. Gregory. 1964. "An approach to the study of style". In Enkvist, N. E., Spencer, J. and M. J. Gregory. 1964. 57 - 109.

Weinreich, U., Labov, W. and M. L. Herzog. 1968. "Empirical foundations for a theory of language change". In Lehman, W. P. and Y. Malkiel. (eds). 1968. 95 - 198.