

## The Lexis of Fundamental Medical English: Classificatory Framework and Rhetorical Function (A Statistical Approach)

Françoise Salager

*University of the Andes, Mérida, Venezuela*

A statistical study of a corpus of 100,000 words from medical English literature has been carried out in order to determine the core lexis of medicine, i.e., those items which are homogeneously distributed across the medical spectrum, whatever the medical speciality. The results of this investigation show that: (a) the verbal roots are made up of those items which describe the general methodology of scientific inquiry (description, analysis, comparison, cause-effect relationships) as well as those which indicate the evolution of diseases and/or the patient's clinical state; (b) the noun roots describe medical procedure, and express measurement; (c) the adjectival roots describe illness or injury, and the quality and/or timing of treatment; and (d) the function words mainly express causality, opposition and purpose. It is contended that applying these findings to vocabulary development in teaching the reading of medical English for professional purposes will result in more appropriate teaching materials and a more viable overall program. It is also suggested that the research procedure described is equally of relevance for vocabulary input to teaching-reading materials in other LSP fields:

### INTRODUCTION: THE IMPORTANCE OF A VOCABULARY INVENTORY FOR TEACHING PURPOSES

In the last two decades there has been in certain countries (particularly at upper secondary and tertiary levels) a movement away from general English courses to those that are designed to give students an English competence that is narrower in function and more selective in language. In such specialized courses (generally referred to as ESP, English for Specific Purposes), it is commonly thought that a knowledge of scientific terminology will provide the non-native readers with what they need. However, experience has shown that a knowledge of specialized terms alone is not a sufficient condition for successful reading of specialized material. Indeed, investigations into the reading problems of non-natives (e.g. Edwards 1974, Martin 1976, Anderson 1980, Hoffman 1981, Pica 1981, Qvistgaard 1981) demonstrate that a more serious impediment to fluent comprehension are what Cowan (1974) calls *sub-technical words* and Martin (1976) *academic vocabulary*, i.e. those context-dependent words which are used across different scientific disciplines, but which tend to occur infrequently in general word-frequency counts.

---

Françoise SALAGER holds a BA and MA in Russian (University of Lyons, France) and a PhD in Foreign Language Education (University of Texas at Austin). She is currently teaching ESP and Russian at the University of the Andes, Mérida, Venezuela, and has published several articles in the *ESP/EST Newsletter* and in the *Russian Language Journal*. In 1982 she contributed a paper at the international symposium on LSP: *Reading for Professional Purposes in Native and Foreign Languages*, Eindhoven, Holland. She may be contacted at Apartado 253, MERIDA 5101, Venezuela.

Nowadays, with the current emphasis on written discourse, vocabulary may be considered an unfashionable, neglected area. However, Elliman (1980: 40) reminds us that when the proportion of unknown words rises above the 10% level, reading comprehension sinks to frustration level. Thus it would be sensible to consider vocabulary as a key element in teaching ESP. Moreover, according to Hoffmann (1981: 114) 'The peculiarities of LSP (Languages for Special Purposes) are first and foremost of a quantitative nature. It is the significantly frequent occurrence of certain speech elements, forms or structures that characterizes scientific writings... As a consequence, statistical methods play an important role in selecting an inventory for teaching purposes... It is the word and phrase levels that yield the best results, i.e. lists of typical lexical and syntactical items which may serve as a highly effective teaching/learning minimum'. Other investigators (e.g. Greenhall 1981, Qvistgaard 1981) also agree that frequency lists provide useful information for the course writer as long as they are used with circumspection and discretion. In particular they have a useful role '... as aides-mémoire, reminding us of language items or linguistic features which we might otherwise overlook' (Swales 1976), and they provide a framework within which pedagogical selection in reading, teaching and testing can be made.

Because of this, and because of the fact that up to now the selection of linguistic materials for this kind of ESP teaching has been mainly governed by empiricism and intuition, we decided to define quantitatively, through a statistical frequency-count analysis, the fundamental vocabulary or nucleus language (what Hoffmann (1981: 125) calls a 'dynamic minimum') of ME (medical English) literature as well as its rhetorical function. Our reasons were that, first, teaching vocabulary involves selecting a corpus of lexical items to suit the learners' immediate needs and presenting a number of classificatory frameworks or language events within which the learners can pattern their new vocabulary. And second, the learning of science involves the acquiring of superimposed knowledge of certain universal concepts and methods which are general to scientific writing (Stevens 1973, Mackay and Mounford 1978, Widdowson 1979). As Trimble (1979) states: 'The objective of the whole ESP text is realized by general rhetorical functions at the level of parts of the texts (e.g. stating the purpose of an experiment) which in turn are realized by a limited number of specific rhetorical functions, namely, to describe, to define, to clarify, to give instructions'.

## **CORPUS**

The corpus on which this research is based is a body of 100,000 words of natural language text. It comprises 50 sample texts (drawn from 50 different textbooks, journals and reviews) of approximately 2,000 words each. The corpus is both synchronic and representative. Synchronicity is ensured by confining the data to texts printed since 1960. To ensure representativeness, the 50 sample texts are drawn from 12 specialities (Internal Medicine, Pediatrics, Obstetrics/Gynecology, Cardiology, Neurology, Nephrology, Ophthalmology, Endocrinology, Surgery, Hematology, Traumatology/Orthopedics, and Psychiatry). The texts characterize the kind of literature likely to be consulted by graduate students or researchers in the field of medicine. In all cases, the material in a sample is one continuous passage from a single source. In the case of reviews and articles, the selection starts at the beginning of the article. In the case of longer items (e.g. textbooks), a page number was taken at random. In either case, the selection runs on continuously (omitting such things as

footnotes, table or figure captions) and ends at the first sentence after a count of 2,000 words.

The present study defines the word as a continuous string of symbols uninterrupted by space and having the root as its basic unit. Taking the root as the basic unit precluded the use of a computer because of the great number of words derived from a single root. (e.g. a computer would not be able to recognize that *mean*, *mean-ing*, *mean-ing-ful*, *mean-ing-less*, etc. are derived from the same verbal root *mean*.) But the advantage of basing our analysis on the root was that the incidence of very low-frequency lexical items was considerably reduced; otherwise, those items would have been excluded on the grounds that their frequency of occurrence was not statistically significant. (e.g. the word *remainder* might occur with a very low frequency, but the root *remain* might be very frequent in medical discourse). Moreover, as every word was recorded by hand, homographs were given separate entries; a computer would not have been able to differentiate between *to mean* (as a verb), *a means* (as a noun) and *mean* (as an adjective, synonymous with *average*).

### STATISTICAL PROCEDURE

For each root, the following data were recorded:

- average frequency per text
- average frequency per 1,000 words
- standard deviation per text
- standard deviation per 1,000 words
- range, i.e. the number of texts (expressed as a percentage) in which a given root occurs

Once these data had been recorded, the next step was to ascertain whether each root belonged to Basic English (BE), Fundamental Medical English (FME), or Specialized Medical English (SME). For purposes of comparison, we used the data from Kučera and Francis's (1967) study, which comprises one million words and covers fifteen literary genres. We first computed the average frequency of roots per 1,000 words in Kučera and Francis's study, and then computed their true mean of the root (MT), estimated by the average frequency 'm' ('m' being calculated as the sum of all word-types listed by Kučera and Francis which derive from a single root). An ANOVA (analysis of variance) test compared Kučera and Francis's MT with the average frequency found in the ME corpus. The ANOVA test indicates:

1. if  $\bar{f}\% < m \rightarrow$  the root belongs to BE
2. if  $\bar{f}\% > m$ , then
  - 2.1 if  $\bar{f}\% < 1.5 \times m \rightarrow$  the root belongs to FME
  - 2.2 if  $\bar{f}\% > 1.5 \times m \rightarrow$  the root belongs to SME

This statistical procedure thus permitted us to split the ME lexicon into three superimposed strata of sizeable extent: (1) Basic English (BE) defined as that part of the lexicon common to all kinds of writing, whether it be scientific, technical or literary; (2) fundamental medical English (FME) - the object of the present study - made up of those roots whose frequency of occurrence is homogeneously distributed over the whole ME corpus but in a significantly different way than in BE; and (3) specialized medical English (SME) composed of specialized terms particular to a given

discipline and known only to the small group of workers in that field. This lexical stratification has also been noted by other researchers (Salager 1977a in Science; Cohen et al 1979 in Genetics; Coulon 1981 in Social Sciences; Anderson 1980 in Medicine; Osuna 1982 in Biology), but their investigations (with the exception of Salager and Osuna) are either empirical and/or based on a small linguistic sample. As Anderson (1980: 4) suggests: 'The results of this small survey need to be followed up with similar studies on a larger scale for this kind of research has important implications for vocabulary selection in ESP courses.'

**RESULTS AND DISCUSSION**

A total of 1,425 roots above the level of statistical significance have been identified in the ME corpus. These break down into BE 31.84% (462 roots), FME 23.36% (339), and SME 44.79% (650). Table 1 compares the ME data with data from Engineering (Salager 1977a) and Biology (Osuna 1982).

Table 1 Comparison of statistical data with Engineering (Salager 1977a) and Biology (Osuna 1982)

FIELD	TOTAL recorded roots	BASIC	FUNDAMENTAL	SPECIALIZED
Medicine	1425	462	339	650
Engineering	1476	538	320	618
Biology	1245	438	299	508
Medicine	100%	31.84%	23.36%	44.79%
Engineering	100%	36.00%	22.00%	42.00%
Biology	100%	35.18%	24.02%	40.80%

As the the table shows, the total number of recorded roots and the percentages of roots forming each stratum are very similar. A t-test proves that the difference is not statistically significant: (p < 0.05). The remainder of this article analyzes and discusses the FME lexicon,<sup>1</sup> the roots of which may be broadly classified according to grammatical category.

**A. Verbal Roots**

The verbal roots (and a few noun roots) mainly express the taxonomy of scientific acts. Among the highest frequency verbs we find the 'organization of knowledge' verbs, i.e. those procedural verbs the primary function of which is to articulate the development of an idea in a discourse. They bear an explanatory and illustrative function likely to be encountered whenever a research project is involved and whatever the branch of science (engineering, medicine, etc.). This part of the FME lexicon could be labelled *discursive reasoning* vocabulary, and may be sub-categorized as follows.

<sup>1</sup> A copy of the 339-word FME lexicon may be obtained from the editors of *Reading in a Foreign Language*. The lexicon is presented in the form:

(root)	FME			BE (Kučera and Francis)	
	frequency	standard deviation	range	frequency	standard deviation
agent	.42	1.09	32.5	.08	.26

### A.1 Description of Process and Methodology of Investigation ( $4.32 \geq f \geq .80$ )<sup>1</sup>

These lexical items are generally encountered in the introduction or the summary of the text, e.g.:

to study	4.32	to describe	.95
to present	1.95	system	.93
to report	1.64	to perform	.90
to observe	1.35	procedure	.90
to determine	1.31	to investigate	.86
to evaluate	1.21	to examine	.86
to develop	1.21	to define	.80
technique	.95		

### A.2 Classification ( $2.84 \geq f \geq .62$ )

These are mainly found in a description of method, and include:

to differ	2.84	to select	.81
to compare	1.85	to consist of	.80
to vary	1.34	to contain	.62
group (e.g. age group)	1.34		
to distribute	.82		

### A.3 Cause and Effect ( $2.35 \geq f \geq .35$ )

These lexical items are mainly found in a description/discussion of results of medical investigation, e.g.

to result from	2.35	to require	.89
to relate	2.20	to cause	.89
to find	1.94	to indicate	.85
to follow	1.78	to suggest	.85
to produce	1.10	to deduce	.65
to obtain	.92	to conclude	.35

### A.4 Analysis ( $.90 \geq f \geq .05$ )

At lower frequencies we find those FME verb-roots which are characteristic of ME literature. They mainly describe the evolution of the disease and/or the patient's clinical state, e.g.:

<sup>1</sup> = frequency range, i.e. from 432 to 80 occurrences per 100,000 words

to perform	.90	to injure	.19
to remain	.86	to recur	.17
to improve	.81	to predict	.17
to induce	.64	to persist	.16
to exhibit	.55	to transmit	.16
to prevent (from)	.55	to prescribe	.09
to be involved	.55	to explore	.09
to administer	.44	to regress	.09
to assess	.44	to enhance	.09
to affect	.42	to heal	.07
to reveal	.31	to constrict	.07
to alter	.30	to deteriorate	.07
to confirm	.29	to remit	.05
to manifest	.22		
to dilute	.19		
to suspect	.19		

**B Noun Roots**

The majority of the FME noun roots can be divided into two broad categories: (B.1) those which describe medical procedure, and (B.2) those which express measurement, not only in medicine but in scientific inquiry in general.

**B.1 Medical Procedure ( $1.95 \geq f \geq .09$ )**

presence	1.95	agent	.42
subject (= patient)	1.76	contrast (e.g. contrast medium)	.36
report (e.g. case report)	1.64	pattern	.36
appearance (= aspect)	1.62	mortality	.26
failure (e.g. heart failure)	1.51	edema	.26
onset (e.g. onset of a disease)	1.50	conduct	.25
risk	.71	clot (e.g. blood clot)	.15
probe	.56	clamp	.15
complication	.56	morbidity	.15
diagnosis	.52	syndrome	.09
dosis	.51		

**B.2 Measurement ( $2 \geq f \geq .12$ )**

This sub-category includes:

level (e.g. peak level)	2.00	range	.46
test	1.57	grade	.45
factor	.70	mechanism	.37
data	.68	loss	.35
fact	.61	amount	.35
series	.61	proportion	.35
limit	.59	index	.35
view	.55	maximum	.32
distribution	.52	sum	.32
weight	.52	degree	.31

critterion	.29	curve	.20
length	.25	monitoring	.20
hypothesis	.24	segment	.20
quantity	.22	peak	.15
graph	.22	deviation	.12
		logarithm	.12

### C Adjectival Roots

These may be grouped into those which (C.1) describe illness or injury, and (C.2) express the quality and/or the timing of treatment administered to the patient.

#### C.1 Description of illness or injury ( $2.21 \geq f \geq .05$ ) e.g.

mean (= average)	2.21	spontaneous	.27
marked	.91	mild	.25
sharp	.81	early (stage)	.24
severe	.64	moderate	.22
acute	.54	transient (level)	.21
major	.54	diffuse	.11
primary (site)	.46	premature (beats)	.10
secondary	.35	idiopathic	.05
slight	.30	superficial	.05

#### C.2 Quality and/or timing (including duration, sequence, frequency of treatment) ( $.95 \geq f \geq .09$ ) e.g.

initial	.95	poor (prognosis)	.20
standard	.49	simultaneous	.17
recent	.49	negative	.16
gradual	.45	consecutive	.16
rapid	.40	definitive	.14
positive	.37	identical	.14
correct	.32	prompt	.11
accurate	.27	appropriate	.11
subsequent	.24	successful	.11
sufficient	.22	preliminary	.10
stable	.20	concomitant	.09

### D Function Words

The rational and unambiguous character of the technical discourse is also to be felt in the use the technical writer makes of function words, particularly discourse markers. The majority of these inter-clause and inter-sentence connectors have been classified by the statistical procedure as BE. Nevertheless, 12% of them fall into the FME

stratum. Their high frequency in scientific-technical writing expresses the grammatical-conceptual manifestations of advanced and complex thought. Whatever the subject matter, as Strevens (1973) explains, 'the importance of these general concepts of advanced thought is that they state the logic—the rhetoric, the argument—of the text as well as having grammatical consequence'.

Although the function words listed below are characteristic not only of ME literature but of any kind of scientific writing, we will present some of them according to their notional classes:

● causal:	thus	.64
	since... then	.61
	therefore	.56
● adversative:	however	1.31
	in spite of	.30
	nevertheless	.12
● conditional:	except	.21
	unless	.05
● additive:	furthermore	.21
	moreover	.06
● contrastive:	while	.25
	whereas	.24
● purposive:	in order to	.15

### THREE BROAD LINGUISTIC FEATURES OF FME

Cross-cutting the categories of form and function outlined above, three broad linguistic features of FME may be noted. (Such features are equally characteristic of other ESP fields.) The first is the *word-compounding* process. The FME context-bound words describe and report as precisely as possible the phenomena under investigation, and their meaning is unequivocal; this explains why compound words made up of FME lexical elements are so frequently employed to delineate the meaning of general words, e.g., *basement maintenance dosis*, *weight loss*, *patient monitoring*, *contrast medium*, *heart failure*, *basement membrane thickening*, *sinus node recovery time*, *iron-induced anemia*, *blood vessel narrowing*. The compound nominal phrases thus formed are a morphosyntactical feature of Technical English (cf. Salager 1977 b, 1983). They acquire a terminological function and belong to a medical subfield, because words become technical by the compounding of several sub-technical terms.

The second feature may be referred to as the *affixation* process. Unlike the man of letters who strives to find synonyms for denoting one and the same entity in order to avoid repetition, the technical writer frequently resorts to derivatives of the head-word previously used, in order to preclude any possibility of inaccuracy. The following examples taken from the FME medical stratum illustrate this statement.

relate	relationship, cor/relation	unrelated
define	definition	undefined
remain	remainder	remaining
limit	limitation	unlimited
appear	appearance	apparently
mean	meaning, meaningfulness	meaningless, meaningful



improve	improvement	
fail	failure	failed
	in/accuracy	in/accurate/ly
	frequency	in/frequent/ly

The third feature of the FME lexicon may be termed the *doublet phenomenon*. A certain number of FME roots are of Latin origin, the semantic equivalents of which are to be found in the BE category. For example:

BE (Anglo-Saxon)	FME (Latin)
search	investigate
yield	produce
have	obtain
begin	initiate
be left	remain

This doublet phenomenon is also encountered in adjective vs. noun roots, e.g. *kidney function* vs. *renal function*, *heart failure* vs. *cardiac failure*.

Of course, the Latin origin of many FME roots must be taken into account when teaching ME to Spanish-speaking students, since recognition of cognates reduces both the learning and the teaching tasks.

### PEDAGOGIC IMPLICATIONS

The vocabulary element of teaching reading in a foreign language frequently puts the cart before the horse. More specifically, there has for some time been a variety of procedures for teaching vocabulary recognition, i.e. puzzling out the meaning of an unfamiliar word met in a text, by drawing on the word's semantic and syntactic context. Similarly, we have an array of exercise-types for vocabulary development, i.e. the systematic expansion of the learner's stock of words. Thus we know a lot about *how* to teach vocabulary, but far less about what should be a prior consideration - *what* vocabulary to teach.

In our own teaching of medical English at the University Hospital of the Andes, the research described in this article has resulted in our defining quantitatively and functionally the ME core lexicon. As a result, as input to our reading program, we will now be able to give priority to the FME lexical items, in context, and thus optimize our teaching. It is hoped that the research procedure described in this article will be equally valuable in other LSP fields, in providing vocabulary input to programs of teaching reading in a foreign language.

### REFERENCES

- Anderson, J.I. (1980) The lexical difficulties of English medical discourse for Egyptian students. *English for Specific Purposes Newsletter*. 37, 4.
- Cohen, A. et al (1979) Reading English for specialized purposes: discourse analysis and the use of student informants. *TESOL Quarterly*, 13 (4), 551-564.

- Coulon, R. (1981) Français instrumental et compétence de communication. *Recherches et Echanges*, 6 (2), 31-49.
- Cowan, J.R. (1974) Lexical and syntactic research for the design of reading material. *TESOL Quarterly*, 8 (4), 389-399.
- Edwards, P.J. (1974) Teaching specialist English (with special reference to English for nurses and midwives in Nigeria). *ELT Journal*, 28 (3), 247-252.
- Elliman, J. (1980) EST TTI Corvallis 1979. *Al Manakh*, 4 (1), 33-47.
- Greenhall, S. (1981) First experiment in designing ESP courses. *Recherches et Echanges*, 6 (2), 11-20.
- Hoffmann, L. (1981) The linguistic analysis and teaching of LSP in the German Democratic Republic. In J. Høedt and R. Turner (Eds.), *New Bearings in LSP*. Copenhagen: The LSP Centre, UNESCO ALSED LSP Network and Newsletter, Copenhagen School of Economics, 107-130.
- Kučera, H. and N. Francis (1967) *Computational Analysis of Present Day American English*. Providence, Rhode Island: Brown University Press.
- Mackay, R. and A.J. Mountford (1978) The teaching of English for Special Purposes: theory and practice. In Mackay and Mountford (Eds.) *English for Specific Purposes*. London: Longman.
- Martin, A.V. (1976) Teaching academic vocabulary to foreign graduate students. *TESOL Quarterly*, 10 (1), 91-99.
- Osuna, A. (1982) En Busca de un Léxico Básico de Biología. Unpublished PhD dissertation. University of the Andes, Mérida, Venezuela.
- Pica, P. (1981) Quelques remarques à propos des problèmes posés par l'élaboration d'une théorie propre au traitement des langues de spécialité. In J. Høedt and R. Turner (Eds.), *New Bearings in LSP*. Copenhagen: The LSP Centre, UNESCO ALSED LSP Network and Newsletter, Copenhagen School of Economics, 25-46.
- Qvistgaard, J. (1981) La place des langues de spécialité (LSP) dans l'enseignement des langues vivantes. In J. Høedt and R. Turner (Eds.), *New Bearings in LSP*. Copenhagen: The LSP Centre, UNESCO ALSED LSP Network and Newsletter, Copenhagen School of Economics, 143-150.
- Salager, F. (1977a) Optimum Foundations of English Technical Literature and their Application to the Comprehension Training of non-English-Speaking Scientists. Unpublished PhD dissertation. University of Texas at Austin.
- Salager, F. (1977b) Compound words in technical Russian and technical English literature: a comparative study. *The Russian Language Journal*, 34 (118), 45-50.
- Salager, F. (1983) Compound nominal phrases in scientific-technical literature: proportion and rationale. In A.J. Pugh and J.M. Ulijn (Eds.), *Reading for Professional Purposes: Studies and Practices in Native and Foreign Languages*. London: Heinemann. (in press).
- Stevens, P. (1973) Technical, technological and scientific English (TTSE). *ELT Journal*, 27 (3), 223-234.
- Swales, J. (1976) Verb frequencies in scientific English. *ESPMENA Bulletin*, 4, 28-31.
- Trimble, L. (1978) Rhetorical, grammatical and lexical elements in written ESP. In *English for Specific Purposes - an International Seminar*. London: The British Council.

- Widdowson, H.G. (1971) The teaching of rhetoric to students of science and technology. In G. Perren (Ed.) *Science and Technology in a Second Language* (CILT Reports and Papers No. 7). London: The British Council.
- Widdowson, H.G. (1979) *Explorations in Applied Linguistics*. Oxford: OUP.

#### *Editors' Note*

Readers of this article will find useful the vocabulary-teaching approaches in ESP textbooks such as:

- Maclean J. (1978) *English in Basic Medical Science* (English in Focus Series). Oxford: Oxford University Press.
- O'Brien T. (1979) *Nucleus English for Science and Technology — Medicine*. London: Longman.
- Mackin R. and A. Weinberger (1962) *El Inglés Para Médicos y Estudiantes de Medicina*. London: Longman.
- Lachowicz D.J. (1974) *Using Medical English for Students of English as a Second Language, with Drills and Exercises in Medical Terminology and Conversational Usage*. Saigon: Department of Languages, Faculty of Medicine, University of Saigon. (copies may also be obtained from the Project Director, Medical Education Project (Vietnam), The American Medical Association, 535 North Dearborn Street, Chicago, Illinois 60610, USA).