

# A Statistical Analysis of the Vocabulary of Medical Research Articles (3): Technical and Subtechnical Vocabulary

Faculty of Social Information Science, Kure University  
Simon A. Fraser

---

**Abstract** This paper completes my investigation into a word frequency count derived from a computer corpus of 16 medical journal articles. It is generally agreed that a command of both “technical” and “subtechnical” vocabulary is important for learners of Medical English. The paper looks at two types of subtechnical vocabulary: anaphoric nouns (a category of discourse-organizing words) and cryptotechnical words (common words which can also have a specialized meaning). The distribution of these words across the IMRAD sections of the articles is investigated. Anaphoric nouns are found to occur with the highest frequency in Introductions and Discussions, with cryptotechnical words most frequent in Methods and Results. Examples are given of how these words are used in context. Two different types of vocabulary found in limited contexts – “technical” and “lay-technical” words – are identified. The implications of these findings for both teachers and learners of Medical English are discussed.

**Key Words:** Medical English, Computer corpus, Frequency count, “IMRAD” structure, Anaphoric nouns, Cryptotechnical words

## 1 INTRODUCTION

This is the third and final paper in a series of articles investigating a computer corpus of medical research articles. In the first paper, (Fraser<sup>1)</sup>) a comparison was made between the most frequent items in the medical corpus with the corresponding totals in the Cobuild corpus of general English. The second paper, (Fraser<sup>2)</sup>), in the previous issue of this journal, investigated the distribution of the most frequent words across the “IMRAD” (Introductions, Methods, Results, Discussion) structure of medical research articles.

It is accepted that ESP (English for Specific Purposes) learners need a command of both technical vocabulary (context-dependent words which occur with high frequency in a given

specialty) and subtechnical vocabulary (common words which occur with special meanings in specific scientific and technical fields). Here, I would like to investigate both of these categories of words. Subtechnical vocabulary will be divided into “discourse-organizing” words and “crypto-technical” words, and examples will be given of how these words are typically used, along with their distribution patterns. A further category of words, “lay-technical” vocabulary, will also be described. (See Fraser<sup>3)</sup> for a more detailed discussion of these terms.)

## 2 THE ANALYSIS

I used the corpus of sixteen medical articles (21,000 words) compiled at the Institute for Applied Language Studies, Edinburgh University.

---

サイモン・A・フレイザー

〒737-0182 呉市郷原学びの丘1-1-1 呉大学社会情報学部

The articles covered a wide range of medical fields, and were taken from both British and American journals. (Appendix 1 lists the articles in the IMRAD corpus. For a more detailed description of the analysis, its limitations, and the frequency word list produced by the computer, see Fraser<sup>1)</sup>).

### 3 SUBTECHNICAL VOCABULARY

#### 3.1 Anaphoric Nouns

Discourse-organizing words have an important role to play in the organization and structure of the argument of a text (see Fraser<sup>3)</sup>). It seems likely that the learner who is unfamiliar with the function of these words will have difficulty in interpreting the whole text efficiently. I decided to take a closer look at a particular category of discourse-organizing

words — “anaphoric nouns”. These are words that frequently occur to refer back to chunks of text, e.g. *hypothesis, position, viewpoint*.

#### 3.1.1 Findings

The following words were selected from the frequency list as having the potential of behaving both anaphorically (referring back) and cataphorically (referring forwards). For each rhetorical section, the frequency of occurrence of the word is given, both as a “raw” figure and expressed as a percentage of occurrence in the entire article. The figures in the second row of each entry show the occurrence of the word expressed as a percentage of the total number of words in a particular section. “Range” shows the number of texts in which the word was found (out of a total of 16).

TABLE 1

	Total Freq.	I	M	R	D	Range
1. STUDY	57	10(17.5%) 0.35	12(21.1%) 0.19	1( 1.7%) 0.02	34(59.7%) 0.42	11
2. CASES	42	4( 9.5%) 0.14	5(11.9%) 0.08	13(31.0%) 0.30	20(47.6%) 0.25	12
3. DIAGNOSIS	31	5(16.1%) 0.17	10(32.3%) 0.16	6(19.4%) 0.14	10(32.2%) 0.1	28
4. RESULTS	30	2( 6.7%) 0.07	6(20.0%) 0.09	5(16.7%) 0.02	17(56.6%) 0.21	11
5. FINDING(S)	27	4(14.8%) 0.14	6(22.2%) 0.09	1( 3.7%) 0.02	16(59.3%) 0.20	9
6. STUDIES	17	6(35.3%) 0.21	4(23.5%) 0.06	0( 0%) 0	7(41.2%) 0.09	9
7. EFFECT	14	5(35.7%) 0.17	1( 7.1%) 0.02	2(14.3%) 0.05	6(42.9%) 0.07	9
8. EVIDENCE	13	5(38.5%) 0.17	1( 7.7%) 0.02	0( 0%) 0	7(53.8%) 0.09	9
9. RESULT	12	1( 8.3%) 0.03	3(25.0%) 0.05	2(16.7%) 0.05	6(50.0%) 0.07	8
10. PROBLEM(S)	11	3(27.3%) 0.10	1( 9.1%) 0.02	1( 9.1%) 0.02	6(54.5%) 0.07	6
Total%:		1.55	0.78	0.62	1.59	6

Mean range = 9.2

It was then necessary to refer to the articles and see whether these words are actually being used in their discourse-structuring capacity. It was found that in the majority of cases they are in fact used either to refer back to a preceding stretch of discourse, or forward to a following section of text. In the following examples (arranged in alphabetical order) we can see that

anaphoric nouns can stand in place of segments of text, and that these segments may consist of a sentence or sentences, a whole paragraph, or even more.

#### Diagnosis

1. Diagnosing appendicitis before operation is difficult, especially in children. Thus 15-

20% of appendicectomies are unnecessary on clinical grounds alone. Various tests have been suggested to increase the accuracy of *diagnosis*, including the white cell count and thermography. (Paper 1, p.722, I)

2. Spontaneous rupture of the Achilles tendon has become more common, probably due to increased participation in active sports by a generally unconditioned and susceptible population. The correct *diagnosis* is not infrequently missed when the patient is first seen. (Paper 11, p.1099, I)

In both of these examples, the interpretation of *diagnosis* relies on information contained in the previous sentence.

#### *Evidence*

1. The *evidence* for pathogenic and protective roles of Vi is as follows: (Paper 1, p.1101, I)
2. The preponderance of *evidence* from many sources finds no familial aggregation of cases of EHBA and lends no support to a genetic etiology. (Paper 5, p.752, D)

In the first example, *evidence* is being used to stand in place of a large segment of text which is to follow. In (2), *evidence* represents whole texts – the work done by other researchers in the field.

#### *Findings*

1. Tejani showed that during follow-up ranging from 5 to 18 years renal death occurred in 25%; an actuarial survival curve showed that at a mean of 20 years' disease duration all those with morphological transition would probably reach end-stage renal status. Tejani's *findings* are very different from those in two other reports. (Paper 10, p.256, I)
2. We report here the *findings* of a prospective trial comparing routine DCBE alone with either colonoscopy alone or a combination of flexible sigmoidoscopy and routine DCBE as initial investigation of patients

with colorectal symptoms...

(Paper 13, p.549, I)

In (1), *findings* is standing in place of the previous, quite lengthy, sentence. In (2), *findings* represents a much larger segment of text – the entire paper, in fact. This time the word is referring forwards.

#### *Problem(s)*

1. Two other children had a haematemesis, and one required treatment for burns. Despite these *problems* none of the children in the study appeared to develop any long term physical *problems*. (Paper 12, p.521, R)
2. A report from Newcastle highlighted an increasing *problem* with intoxication among adolescent boys in that city. Our study does not suggest a similar trend in Glasgow or Nottingham.... Nevertheless, the pronounced difference in the apparent size of the *problem* in our two cities, where hospital admission policies are identical, is intriguing. (Paper 12, p.521, D)

In (1), the first *problems* is simply standing in place of the previous sentence. In (2), the second *problem* refers back to the previous sentence, which in turn refers to another medical article.

#### *Results*

1. We reported that the immunogenicity was increased and T-dependent properties were conferred on capsular polysaccharides injected into humans after their covalent binding to proteins. These *results* confirmed observations made many years ago by Avery and Goebel in laboratory animals. (Paper 1, p.1103, D)
2. A problem in interpreting these *results* is that each food is not of equivalent importance in children's diets. (Paper 2, p.1698, D)

*Results* was frequently found in collocation with *these*. In (1), *These results* represents the

previous sentence. In (2), *these results* is standing in place of a large segment of text – the preceding Results section.

*Study*

1. Although the association of alcohol ingestion and hypoglycaemia in young children is well known, our *study* suggests that it occurs more commonly than was previously appreciated. (Paper 10, p.521, D)
2. I now intend to carry out a larger *study*, using the Feverscan thermometer to assess all patients presenting with acute abdominal pain. (Paper 8, p.722, D)

In the first example, *study* is referring to the findings which had been described previously in the paper. In example (2), the word is referring forward to work which the author elaborates on later in the sentence. The fact that *study* is preceded by *larger* tells us something about the author’s designs – we know that he sees this future work as being very important.

3.1.2 *Comment*

It is interesting to note that several of the words on the list appear on Baker’s<sup>4)</sup> list of rhetorical / functional words: *results*, *study*, *evidence*, *diagnosis*, and *findings*. There is also overlap with Yang’s<sup>5)</sup> list of subtechnical terms: *effect*, *problem*, and *result*. *Diagnosis*, *analysis*

and *finding* appear on one of Francis’s lists of “A-nouns” (Francis<sup>6)</sup>)

We can also see that there is high distribution of words. With the exception of *problem(s)*, they all appear in at least half of the articles, and the “average” anaphoric noun appears in nine articles. So not only do anaphoric nouns occur with a relatively high frequency, they are of high distribution too.

It is apparent from Table 1 that anaphoric nouns occur most frequently in Discussions, closely followed by Introductions. They occur least frequently in Results (with only about half the frequency of occurrence in either Introductions or Discussions). Since the Introduction and Discussion sections are where the author is developing his argument, we would expect text-structuring words such as anaphoric nouns to be found mainly in these sections. The more procedural Methods and Results sections will have less need for these “signposts”, which signal the interactive structure of the text.

3.2 **Cryptotechnical Words**

3.2.1 *Findings*

The term “cryptotechnical” was coined by Howard<sup>7)</sup> to define words which, in addition to their common, general meanings, have a specialized meaning in a particular discipline. The following words in the frequency list looked as though they would fall into this category:

TABLE 2

	Total Freq.	I	M	R	D	Range
1. CASES	42	4( 9.5%) 0.14	5(11.9%) 0.08	13(31.0%) 0.30	20(47.6%) 0.25	12
2. INTOLERANCE	36	9(25.0%) 0.31	11(30.6%) 0.17	4(11.1%) 0.09	12(33.3%) 0.15	1
3. STROKE	35	7(20.0%) 0.24	14(40.0%) 0.22	4(11.4%) 0.09	10(28.6%) 0.12	2
4. TREATMENT	34	2( 5.9%) 0.07	11(32.3%) 0.17	11(32.3%) 0.25	10(29.5%) 0.12	8
5. POSITIVE	31	4(12.9%) 0.14	8(25.8%) 0.13	7(22.6%) 0.16	12(38.7%) 0.15	5
6. TABLE	31	0( 0%) 0	5(16.1%) 0.08	26(83.9%) 0.60	0( 0%) 0	13
7. CELLS	27	5(18.5%) 0.17	8(29.6%) 0.13	3(11.1%) 0.07	11(40.7%) 0.14	4
8. ADMITTED	25	3(12.0%) 0.10	7(28.0%) 0.11	9(36.0%) 0.21	6(24.0%) 0.07	5

**A Statistical Analysis of the Vocabulary of Medical Research Articles (3)**

	Total Freq.	I	M	R	D	Range
9. ACTIVITY	24	1 ( 4.2%) 0.03	3(12.5%) 0.05	13(54.1%) 0.30	7(29.2%) 0.09	6
10. INTOLERANT	23	1 ( 4.3%) 0.03	4(17.4%) 0.06	1(4.3%) 0.02	17(74.0%) 0.21	1
11. CONTROL	22	1 ( 4.5%) 0.03	10(45.5%) 0.16	6(27.3%) 0.14	5(22.7%) 0.06	7
12. SUBJECTS	22	0( 0%) 0	10(45.5%) 0.16	5(22.7%) 0.11	7(31.8%) 0.09	4
13. RISK	21	3(14.3%) 0.10	2( 9.5%) 0.03	0( 0%) 0	16(76.2%) 0.20	7
14. NEGATIVE	19	1 ( 5.3%) 0.03	6(31.6%) 0.09	3(15.8%) 0.07	9(47.4%) 0.11	9
15. POPULATION	18	2(11.1%) 0.07	7(38.9%) 0.16	4(22.2%) 0.02	2(11.1%) 0.06	7
16. HISTORY	16	1 ( 6.3%) 0.03	11(68.8%) 0.17	1( 6.3%) 0.02	3(18.8%) 0.04	8
17. RESPONSE	15	3(20.0%) 0.10	6(40.0%) 0.09	2(13.3%) 0.05	4(26.7%) 0.05	6
18. SIGNIFICANT	15	2(13.3%) 0.07	2(13.3%) 0.03	4(26.7%) 0.09	7(46.7%) 0.09	8
19. MEAN	14	1 ( 7.1%) 0.03	5(35.7%) 0.08	4(28.6%) 0.09	4(28.6%) 0.05	6
20. CULTURE(S)	14	2(14.3%) 0.07	5(35.7%) 0.08	7(50.0%) 0.16	0( 0%) 0	2
21. TREATED	14	0( 0%) 0	9(64.3%) 0.14	1 ( 7.1%) 0.02	4(28.6%) 0.05	6
22. CONTROLLED	13	6(46.1%) 0.21	1 ( 7.7%) 0.02	1 ( 7.7%) 0.02	5(38.5%) 0.06	4
23. EXAMINATION	12	1 ( 8.3%) 0.03	6(50.0%) 0.09	0( 0%) 0	5(41.7%) 0.06	3
24. ONSET	12	1 ( 8.3%) 0.03	2(16.7%) 0.03	4(33.3%) 0.09	5(41.7%) 0.06	3
25. ACUTE	11	(18.2%) 0.07	7(63.6%) 0.11	1 ( 9.1%) 0.02	1 ( 9.1%) 0.01	11
26. ADMISSION	11	0( 0%) 0	2(18.2%) 0.03	7(63.6%) 0.16	2(18.2%) 0.02	4
27. COURSE	11	2(18.2%) 0.07	3(27.3%) 0.05	1(9.1%) 0.02	5(45.5%) 0.06	11
28. RANGE	11	0( 0%) 0	5(45.4%) 0.08	4(36.4%) 0.09	2(18.2%) 0.02	6
29. REFERRED	11	0( 0%) 0	3(27.3%) 0.05	0( 0%) 0	8(72.7%) 0.10	6
30. REACTIONS	10	2(20.0%) 0.03	4(40.0%) 0.06	0( 0%) 0.05	4(40.0%) 0.04	3
31. SAMPLE	10	0( 0%) 0	4(40.0%) 0.06	1(10.0%) 0.02	5(50.0%) 0.06	6
Total%		2.32	2.50	3.10	2.26	

Mean Range=6.1

The actual use of these words in the medical articles was then investigated. In nearly all instances, it was found that the word was being used with its technical meaning. Some examples, arranged in alphabetical order, are given below:

*Activity*

1. IL-1 *activity* was observed in most BWs (nine of 12 harvested on day 4; on day 7 no

IL-1 *activity* was detected).

(Paper 6, p.478, R)

2. Since the peak incidence in both rural and urban populations was for both births in the three month period of August, September, and October, the peak *activity* period for the presumed etiologic agent in these term infants would be from the middle of March to the middle of June. Although the etiologic agent is unknown, these are the months of

intense agricultural *activity* in north Texas.  
(Paper 5, p.752, D)

*Activity* is a word which was found to occur with both its general and technical meanings. The first example illustrates how it is normally used in medical articles – to mean *biological* activity. The second example shows both this use and its general meaning “state of being busy”.

#### *Acute*

1. All patients with a diagnosis of *acute* stroke were investigated. (Paper 14, p.11, M)
2. ...with a primary diagnosis of *acute* alcohol intoxication... (Paper 12, p.520, M)

The examples clearly show how *acute* is used – to describe a disease which comes to a crisis.

#### *Admitted / Admission*

1. Each year roughly 1000 children aged under 15 are *admitted* to hospital in England and Wales suffering from acute alcohol intoxication. (Paper 12, p.519, I)
2. All of the ten patients were *admitted* under the care of a staff orthopaedic surgeon to Dr. Steevens Hospital, Dublin, Ireland, between January and July 1983, with a suspected spontaneous rupture of the Achilles tendon. (Paper 11, p.1099, M)
3. Seasonal variation in the rate of *admission* was evident. (Paper 12, p.520, I)

All instances of *admitted* were used with the sense “entered hospital” rather than with the general meaning of “concede”.

#### *Cases*

1. The attack rate (*cases* per 1000 vaccinated persons) of typhoid fever with positive blood cultures was 2.60 in the Vi group... (Paper 1, p.1102, R)
2. ...we examined the distribution of *cases* of EHBA in North Texas from 1972 to 1980. (Paper 5, p.749, I)

In most instances, *cases* was used in the sense of “instances of a disease”.

#### *Complications*

1. Colonoscopy is reported to carry a higher risk of *complications* than DCBE... (Paper 13, p.551, D)
2. Secondary physical *complications* of intoxication were also seen. (Paper 12, p.521, R)

*Complications* was found to occur with its medical sense, as shown in the above examples.

#### *Control*

1. The ratio of the number of cases in the Vi group to that in the *control* group changed little throughout the study. (Paper 1, p.1103, R)
2. *Control* cells were incubated in the presence of 0.1 ml of non-infected allantoic fluid. (Paper 6, p.477, M)
3. *Control* of the disease requires uncontaminated water, effective sewage disposal, and prompt diagnosis and treatment of patients and asymptomatic carriers. (Paper 1, p.1101, I)

The first two examples illustrate how *control* is normally used in medical articles. However, the word was also found to occur with its general meaning “regulation”, as we can see in (3).

#### *Course*

1. In view of these conflicting reports we have examined the clinical *course* and long-term prognosis of a well defined group of children with steroid-sensitive, biopsy-proven, minimal-change nephritic syndrome. (Paper 10, p.256, I)
2. 34 of these 44 received a single *course* of cyclophosphamide and 3 had two *courses*, 1 patient received azathioprine only, and the remaining 6 received more than one cytotoxic agent at different times. (Paper 10, p.256, M)

This word was used in two different ways in the articles, as the above examples show. In (1), *course* is used to mean the progress of an illness, and in (2) it has the meaning of “prescribed treatment”.

#### *Culture(s)*

1. A survey of blood *cultures* at the Infectious Disease Hospital at Teku and at clinics in Kathmandu... (Paper 1, p.1102, I)
2. Exposure of alveolar macrophages... induced the production of IL-1 activity in *culture* supernatants. (Paper 6, p.478, R)

*Culture* and *cultures* were found to occur exclusively with their biological meanings, as shown in the examples above.

#### *Examination*

1. After a routine physical *examination* the calf-squeeze test of Thompson and Doherty was performed. (Paper 11, p.1099, M)
2. After a full physical *examination*, including digital rectal *examination* and rigid sigmoidoscopy... (Paper 13, p.549, M)

The above two examples clearly show how *examination* is generally used in medical research articles.

#### *History*

1. Patients with a dyspeptic *history* of greater than three months prior to perforation... (Paper 9, p.198, M)
2. Ten children, however, had a *history* of chronic behaviour problems... (Paper 12, p.520, R)

*History* was found always, as would be expected, in its sense of *medical* history.

#### *Intolerance / Intolerant*

1. Determining the prevalence of food *intolerance* and its severity is important. (Paper 2, p.1696, I)
2. A child was classified as food *intolerant* if

the answer was “yes” to both questions. (Paper 2, p.1697, M)

*Intolerance* and *intolerant* were only found in one paper, where they are used when discussing hypersensitivity to a particular food.

#### *Mean*

1. ...the estimated *mean* difference in height standard deviation scores was... (Paper 2, p.1697, I)
2. 213 patients had a *mean* follow-up of 17.1 years. (Paper 10, p.258, D)

*Mean* was found only with its mathematical / statistical meaning “the average of a set of numbers”.

#### *Negative*

1. *Negative* results were obtained for all the controls. (Paper 8, p.722, R)
2. Two patients had a positive Thompson and Doherty test but a *negative* needle test. (Paper 11, p.1100, R)

Both *negative* and *positive* appear to be used in a similar way in medical research articles, usually when reporting the results of a test.

#### *Onset*

1. It takes into account the nature of *onset* of stroke... (Paper 14, p.11, M)
2. The modal age of *onset* was 2 years... (Paper 10, p.257, R)

*Onset* was found to occur with the sense “start of a disease”.

#### *Population*

1. The target *population* was Nepalese inhabitants of five villages outside Kathmandu, 5 through 44 years of age. (Paper 1, p.1102, M)
2. First, as a tertiary referral centre we could have been seeing a highly selected *population*. (Paper 10, p.258, D)

3. Northeast Texas is a forested area with a *population* of 1,633,500.

(Paper 5, p.749, M)

4. With *population* data from the 1981 census the mortality from head injury per 100,000 children in the ward group was calculated for each of the 10 groups of wards.

(Paper 7, p.1193, M)

As we saw with *control*, *population* was also found to occur with both its usual meaning and its technical, statistical meaning (“a group of persons, items, etc., considered statistically”).

#### *Positive*

1. ...all the patients in whom the response was *positive* had appendicitis.

(Paper 8, p.722, R)

2. Five patients had both a *positive* Thompson and Doherty test and a *positive* needle test.

(Paper 11, p.1100, R)

The meaning of *positive* in both of these examples is that something is present in a medical test.

#### *Range*

1. 213 patients had a mean follow-up of 17.1 years (*range* 10-21 years).

(Paper 10, p.258, D)

2. Those in the *range* of probability between 30% and 70% for either stroke type...

(Paper 14, p.11, M)

*Range* was found to be used exclusively in its statistical sense.

#### *Reactions*

1. Local *reactions*, observed in 25 subjects, were minimal and diminished 24 hours later.

(Paper 1, p.1102, M)

2. That they are non-invasive and carry no risk of adverse *reactions* or side-effects are additional attractions.

(Paper 3, p.327, D)

These examples show that *reactions* is typically used with its biological meaning.

#### *Referred*

1. This review of all new cases of EHBA *referred* to UTHSCD from 1972 to 1980 demonstrates such a time-space clustering.

(Paper 5, p.751, D)

2. We studied 20 consecutive patients with chronic schizophrenia (17 inpatients and three outpatients) who had been *referred* for non-pharmacological treatment of persistent auditory hallucinations.

(Paper 3, p.327, M)

*Referred* was found only to be used with the sense of “hand over for the consideration of a medical expert”.

#### *Response*

1. The *response* of thymocytes to BW collected at day 7 was lower than that of thymocytes in the presence of medium.

(Paper 6, p.478, R)

2. There was significant antibody *response* in all groups ( $p=0.0001$ ).

(Paper 1, p.1102, M)

*Response* was found to occur with its specialized meaning of “biological reaction”, as we can see in the above examples.

#### *Risk*

1. *Risk* factors for a cerebrovascular accident were common in the 8 epileptics presenting with their first stroke.

(Paper 14, p.12, D)

2. ...the *risk* of relapse in the subsequent 5 years is 22.6%.

(Paper 10, p.257, R)

*Risk* was used only with its meaning of “probability of illness developing”.

#### *Sample*

1. In this *sample* of children aged 5 to 11 attending primary schools in Britain...

(Paper 2, p.1698, D)

2. We think it is unlikely, therefore, that our control group was a biased *sample*.

(Paper 14, p.12, D)



*Sample* is another example of a word which has both general and statistical senses. The statistical meaning was found in the articles.

#### *Significant*

1. A *significant* difference in height between the food intolerant and non-food intolerant children ( $p < 0.01$ ) remained after adjusting for the following factors in a multiple regression analysis... (Paper 2, p.1697, R)
2. Similarly, Table IV suggests clustering by season since 23 of the 30 cases of EHBA occurred during the last six months during the nine years, and analysis shows this impression to be statistically *significant*.  
(Paper 5, p.751, R)

As with *mean*, *range*, and *sample*, *significant* was found to be used exclusively with a statistical sense.

#### *Subjects*

1. ...the value was adjusted to reflect the value that would be obtained if there was an equal distribution of *subjects* in each level of category. (Paper 1 p.1102, M)
2. There were 6907 vaccinated *subjects*, of whom 6438 were members of the target population... (Paper 1, p.1102, M)

As the examples show, *subjects* was found to be used with its meaning "sufferers of a disease".

#### *Table*

1. *Table 1* gives the patients' demographic details. (Paper 4, p.789, R)
2. The *table* shows the results.  
(Paper 8, p.722, I)

Unsurprisingly, this word was always found with its meaning of "a set of facts and figures".

#### *Treatment / Treated*

1. Cimetidine *treatment* did confer significant benefit... (Paper 9, p.198, M)

2. All were receiving maintenance *treatment* with antipsychotic drugs.  
(Paper 3, p.327, M)
3. This decrease is most pronounced in those *treated* with cytotoxic drugs and steroids.  
(Paper 10, p.258, D)

*Treatment* was always found with its medical meaning of "medical care and attention".

#### 3.2.2 *Comment*

Our examination of how the words in Table 1 are used in the medical research articles has shown that they generally occur with their specialized meanings. However, it appears that there is a continuum of such words, ranging from words like *sample*, in which the technical sense is merely an extension of the general meaning, to words such as *culture*, in which the technical sense is truly hidden.

A large number of cryptotechnical words appear in Salager's<sup>8)</sup> list of the lexis of fundamental medical English. These include *subject*, *onset*, *risk*, *data*, *range*, *mean*, *acute*, *positive* and *negative*.

Cryptotechnical words occur most frequently in the Results section, followed by Methods. This is probably what we would expect – the greatest use of these words, in their technical senses, when experimental procedures are being described and reported.

Several cryptotechnical words were found to occur in half or more of the texts, and it may be worthwhile focusing on these: TABLE (13 texts), HISTORY (eight texts), SIGNIFICANT (eight texts), EFFECT (nine texts), CASES (12 texts) and CLINICAL (11 texts). These, in common with anaphoric nouns, fulfill the criteria for "subtechnical" according to Cowan<sup>9)</sup>, in that they occur with a relatively high frequency in a wide range of articles. Such words must have an important role to play if they are found in so many papers concerned with a variety of topics within the medical field.

4 TECHNICAL VOCABULARY

as those context-dependent words which occur with high frequency in a given medical specialty.

4.1 Findings

“Technical” vocabulary was defined earlier

The following table lists the technical terms found in the medical corpus:

TABLE 3

	Total Freq.	I	M	R	D	Range
1*. DCBE	43	5(11.6%) 0.17	4( 9.3%) 0.06	14(32.6%) 0.32	20(46.5%) 0.25	1
2. COLONOSCOPY	33	3( 9.1%) 0.10	5(15.2%) 0.08	10(30.3%) 0.23	15(45.4%) 0.18	1
3*. EHBA	33	0( 0%) 0	6(18.2%) 0.09	4(12.1%) 0.09	23(69.7%) 0.28	1
4. RELAPSE	27	1( 3.7%) 0.05	3(11.1%) 0.05	2( 7.4%) 0.05	21(77.8%) 0.26	2
5*. TNF	25	2( 8.0%) 0.07	4(16.0%) 0.06	10(40.0%) 0.23	9(36.0%) 0.11	1
6. TENDON	23	5(21.7%) 0.17	4(17.4%) 0.06	11(47.8%) 0.25	3(13.0%) 0.04	1
7. ISCHAEMIC	21	2( 9.5%) 0.07	10(47.6%) 0.16	2( 9.5%) 0.05	7(33.3%) 0.09	2
8. EPILEPSY	20	5(25.0%) 0.17	2(10.0%) 0.03	1( 5.0%) 0.02	12(60.0%) 0.15	1
9. RUPTURE	19	7(36.8%) 0.24	2(10.5%) 0.03	4(21.1%) 0.09	6(31.6%) 0.07	1
10. SIGMOIDOSCOPY	19	2(10.5%) 0.07	3(15.8%) 0.05	6(31.6%) 0.14	8(42.1%) 0.10	1
11. ACHILLES	16	4(25.0%) 0.14	3(18.8%) 0.05	6(37.5%) 0.14	3(18.8%) 0.04	1
12. INTOXICATION	14	2(14.3%) 0.07	1( 7.1%) 0.02	6( 42.7%) 0.14	5(35.7%) 0.06	2
13. VACCINES	12	3(25.0%) 0.10	4(33.3%) 0.06	0( 0%) 0	5(41.7%) 0.06	1
14. ANTIBODIES	11	1( 9.1%) 0.03	6(54.5%) 0.09	0( 0%) 0	4(36.4%) 0.05	1
15. BOWEL	11	3(27.3%) 0.10	2(18.2%) 0.03	2(18.2%) 0.05	4(36.4%) 0.05	1
16. CIMETIDINE	11	3(27.3%) 0.10	4(36.4%) 0.06	3(27.3%) 0.07	1( 9.1%) 0.01	1
17. COLON	11	3(27.3%) 0.10	0( 0%) 0	3(27.3%) 0.07	5(45.5%) 0.06	1
18. HYPERTENSION	11	3(27.3%) 0.10	5(45.4%) 0.08	0( 0%) 0	3(27.3%) 0.04	2
19. SERUM	11	1( 9.1%) 0.03	3(27.3%) 0.05	1( 9.1%) 0.02	6(54.5%) 0.07	4
20. VARIABLES	11	0( 0%) 0	9(81.8%) 0.14	2(18.2%) 0.05	0( 0%) 0	3
21. COHORT	10	1(10.0%) 0.03	1(10.0%) 0.02	0( 0%) 0	8(80.0%) 0.10	1
22. MACROPHAGES	10	4(40.0%) 0.14	1(20.0%) 0.02	5(50.0%) 0.11	0( 0%) 0	1
23. MORPHINE	10	0( 0%) 0	3(30.0%) 0.05	4(40.0%) 0.09	3(30.0%) 0.04	1
24. RELAPSES	10	0( 0%) 0	2(20.0%) 0.03	3(30.0%) 0.07	5(50.0%) 0.06	1
25. TYPHI	10	6(60.0%) 0.21	4(40.0%) 0.06	0( 0%) 0	0( 0%) 0	1
Total%		2.26	1.43	2.28	2.17	

Mean range=1.4

1\*: double contrast barium enema, 3\*: extrahepatic biliary atresia

5\*: tumour necrosis factor

**4.2 Comment**

The distribution of technical words across IMRAD follows a different pattern from that of cryptotechnical words. Technical words are most frequent in Results, followed closely by Introductions and Discussions, and occur considerably less often in Methods than in other sections. This is surprising, as we might expect technical words to occur most often in those sections of the paper most concerned with the details of experimental procedure, techniques and equipment used, drugs given, and so on.

The most striking thing about technical words is that although they occur with similar frequencies to cryptotechnical words, the number of texts in which they are found is far smaller. Most of them in fact occur in only one or two texts, and those which are found in more than one text are found to peak in just one text and have a very low frequency in the others. This accords with the findings of Yang<sup>5)</sup> who showed that technical terms occur with high frequency in texts where they are found, but vary dramatically from one subject area to another.

By looking only at the frequency of occur-

rence, we might arrive at the conclusion that terms like “DCBE” and “colonoscopy” must be useful and important for all medical researchers because they occur with such high frequency. However, if we look at their distribution across texts, we can see that both of these words appear in only one article. The fact that they are to be found in the frequency word list at all is purely a result of particular articles being chosen. A different sample of texts would have thrown up completely different technical words occurring with high frequency. It is therefore important to look not only at frequency but also at the distribution of items if we are to focus on words which are of general relevance to learners.

**5 LAY-TECHNICAL WORDS**

**5.1 Findings**

The following table lists the vocabulary that we might consider to be “lay-technical” – those words which although usually found in a medical context, can be easily understood by the layperson.

**TABLE 4**

	Total Freq.	I	M	R	D	Range
1. PATIENTS	168	19(11.3%) 0.66	38( 9.3%) 0.60	46(27.4%) 1.06	65(38.7%) 0.80	12
2. DISEASE	63	20(31.8%) 0.69	14(22.2%) 0.22	7(11.1%) 0.16	22 0.27	9
3. PATIENT	48	9(18.7%) 0.31	17(35.4%) 0.27	7(14.6%) 0.16	15(31.3%) 0.11	9
4. ALCOHOL	46	2( 4.4%) 0.07	7(15.2%) 0.11	14(30.4%) 0.32	23(50.0%) 0.28	5
5. HOSPITAL	36	6(16.7%) 0.21	17(47.2%) 0.27	6(16.7%) 0.14	7(19.4%) 0.09	10
6. SYMPTOMS	34	4(11.8%) 0.14	14(41.2%) 0.22	2( 5.8%) 0.05	14(41.2%) 0.17	6
7. DIAGNOSIS	31	5(16.1%) 0.17	10(32.3%) 0.16	6(19.4%) 0.14	10(32.3%) 0.12	8
8. VIRUS	24	6(25.0%) 0.21	7(29.2%) 0.11	6(25.0%) 0.14	5(20.8%) 0.06	1
9. FEVER	22	10(45.5%) 0.35	4(18.2%) 0.06	4(18.2%) 0.09	4(18.2%) 0.05	2
10. CLINICAL	19	5(26.3%) 0.17	6(31.6%) 0.09	2(10.5%) 0.05	6(31.6%) 0.07	11
11. HEART	19	2(10.5%) 0.07	9(47.4%) 0.14	2(10.5%) 0.05	6(31.6%) 0.07	2
12. URINE	19	2(10.5%) 0.07	11(57.9%) 0.17	2(10.5%) 0.05	4(21.1%) 0.05	3

	Total Freq.	I	M	R	D	Range
13. INJURY	18	5(27.8%) 0.17	7(38.9%) 0.11	4(22.2%) 0.09	2(11.1%) 0.02	2
14. BLOOD	17	2(11.8%) 0.07	6(35.3%) 0.09	6(35.3%) 0.14	3(17.6%) 0.04	5
15. DIAGNOSTIC	17	10(58.8%) 0.35	2(11.8%) 0.03	2(11.8%) 0.05	3(17.6%) 0.04	6
16. DOSES	14	1( 7.1%) 0.03	8(57.1%) 0.13	2(14.3%) 0.05	3(21.4%) 0.05	6
17. DRUGS	13	3(23.1%) 0.10	5(38.5%) 0.08	0( 0%) 0	5(38.5%) 0.06	6
18. INCIDENCE	12	0( 0%) 0	0( 0%) 0	5(41.7%) 0.11	7(58.3%) 0.09	12
19. THERAPY	11	0( 0%) 0	8(72.7%) 0.13	1( 9.1%) 0.02	2(18.2%) 0.02	3
20. DRUG	10	0( 0%) 0	6(60.0%) 0.09	2(20.0%) 0.05	2(20.0%) 0.02	5
21. INJURIES	10	1(10.0%) 0.03	1(10.0%) 0.02	6(60.0%) 0.14	2(20.0%) 0.02	2
22. WARDS	10	0( 0%) 0	9(90.0%) 0.14	0( 0%) 3.06	1( 10%) 0.01	2
Total		3.87	3.24	3.06	2.50	

Mean range=5.8

## 5.2 Comment

Unlike technical words, lay-technical words are most frequent in the Introduction section. It is interesting that they gradually decrease in frequency as we move across the IMRAD structure: they are least frequent in Discussions. If we look at the range of these items, we see that they are found to occur in a greater number of articles than technical words. We would expect this, since although they are found predominantly in medical contexts, their use is not restricted to any specific field.

## 6 CONCLUSION

This study on technical and subtechnical vocabulary concludes my analysis of a medical word list which has the potential of being used to produce more efficient Medical English learning materials. Two types of subtechnical vocabulary were of interest: anaphoric nouns (an important category of discourse-organizing words) and cryptotechnical words. The distribution of these words across the IMRAD sections was investigated. Anaphoric nouns were found to occur with the highest frequency in Introductions and Discussions, and cryptotechnical words were most frequent in Methods and Results.

Examples were given of how these words are typically used in context. Two different types of vocabulary found in limited contexts – “technical” and “lay-technical” words – were identified.

## 6.1 Pedagogical Implications

### 6.1.1 *The Importance of Discourse-structuring Words*

Discourse-structuring words, and anaphoric nouns in particular, have implications for the teaching of both writing and reading skills. We saw that anaphoric nouns have the capacity to function as “structural signposts” – words which enable the reader to perceive the framework of the writer’s argument more easily. An awareness of how anaphoric nouns can signal the author’s intention will help learners to understand the different ways in which the message is being developed. An understanding of such words is vitally important if misinterpretation is to be avoided.

The teacher faces difficulties in teaching discourse-structuring words. It is important that they be taught not in isolation, but in the context of the message. Reading strategies can then be developed which enable the learner to guess unknown words or phrases. Because discourse-oriented items help the reader to interpret large patterns in text, writing tasks

given to students would necessarily require them to produce lengthy stretches of work. The best sources of discourse-structuring words are the learners' own subject material (medical journal articles) coupled with reference material such as the list of anaphoric nouns given in Table 1.

### 6.1.2 Learners Need Help with Cryptotechnical Words

We saw that with a large number of cryptotechnical words, their technical meaning is an extension of their general meaning. This would suggest that techniques which increase not only the students' knowledge of subtechnical vocabulary, but also general vocabulary, would be useful. The more general words the students are familiar with, the better equipped they will be to guess the technical meanings of cryptotechnical words.

In order to make the learning of cryptotechnical words more systematic, a glossary such as that shown in Appendix 2 could be compiled. This would serve as a checklist for the teacher, and an index in which both the general and specialized meanings are explained. The list could be sub-divided into groups of "crypto-medical" words, "crypto-mathematical" words, "crypto-scientific" words, and so on.

Examples of these words in use which make explicit their function in particular contexts could be given.

It may be that students will be able to guess the meaning quite easily of *sample*, for example, but they will almost certainly face problems when they come across a "true" cryptotechnical word which they have never met before. Teachers will need to be aware of such words which have a hidden technical meaning and be prepared to explain them to their students.

### 6.2 Future Research Directions

It would be desirable to carry out similar investigations using a much larger corpus. Obviously, in order to be as representative as possible and secure sufficient evidence for statistical treatment, a corpus should be as large as possible. A larger corpus would also facilitate a study of collocation and phraseology, which is necessary to identify nominal compounds and specialized multi-word items. The use of concordances would be especially useful in providing information about the way vocabulary is used in medical articles, and whether it is used in the same way or differently in different rhetorical sections.

### References

- 1) S. A. Fraser. 2001. A statistical analysis of the vocabulary of medical research articles (1): comparison with the Cobuild frequency count. *Integrated Studies in Nursing Science*. 3 (1): 38-58.
- 2) S. A. Fraser. 2002. A statistical analysis of the vocabulary of medical research articles (2): differences across the "IMRAD" structure. *Integrated Studies in Nursing Science*. 4 (1): 27-34.
- 3) S. A. Fraser. 2001. Vocabulary and the teaching of English for specific purposes. *Integrated Studies in Nursing Science*. 3 (1): 32-37.
- 4) M. Baker. 1988. Sub-technical vocabulary and the ESP teacher: an analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language* 4 (2) 91-105.
- 5) H. Yang. 1986. A new technique for identifying scientific / technical terms and describing science text (an interim report). *Literary and Linguistic Computing* 1 (2): 91-105.
- 6) G. Francis. 1986. *Anaphoric Nouns*. ELR, University of Birmingham.
- 7) R. Howard. 1991. Teaching medical English systematically. *EMP Newsletter* 8: 15-21.
- 8) F. Salager. 1983. The lexis of fundamental English: classificatory framework and rhetorical function (a statistical approach). *Reading in a Foreign Language* 1 (1): 54-64.
- 9) J. R. Cowan. 1974. Lexical and syntactic research for the design of reading material. *TESOL*

*Quarterly* 8 (4): 389-399.

## APPENDIX 1

### The Medical Journal Articles

#### PAPER 1

I. L. Acharya, C. U. Lowe et al. 1987. Prevention of typhoid fever in Nepal with the Vi capsular polysaccharide of salmonella typhi. *The New England Journal of Medicine* 317 (18): 1101-1103.

#### PAPER 2

C. E. Price, R. J. Rona et al. 1988. Height of primary school children and parents' perceptions of food intolerance. *British Medical Journal* 296: 1696-1699.

#### PAPER 3

H. E. Nelson, S. Thrasher, T. R. E. Barnes 1991. Practical ways of alleviating auditory hallucinations. *British Medical Journal* 302: 327.

#### PAPER 4

T. T. C. McLintock, H. Aitken et al. 1990. Postoperative analgesic requirements in patients exposed to positive intraoperative suggestions. *British Medical Journal* 301: 788-790.

#### PAPER 5

A. D. Strickland, K. Shannon. 1982. Studies in the etiology of extrahepatic biliary atresia: Time-space clustering. *The Journal of Pediatrics* 100: 749-753.

#### PAPER 6

F. Vacheron, A. Rudent et al. 1990. Production of interleukin 1 and tumour necrosis factor activities in bronchoalveolar washings following infection of mice by influenza virus. *Journal of General Virology* 71: 477-479.

#### PAPER 7

P. M. Sharples, A. Storey et al. 1990. Causes of fatal childhood accidents involving head injury in Northern region. *British Medical Journal* 301: 1193-1197.

#### PAPER 8

J. E. Hambidge. 1990. Use of skin thermometer to diagnose acute appendicitis. *British Medical Journal* 300: 722.

#### PAPER 9

G. Lamont, C. J. Simpson et al. 1986. Does cimetidine alter the prognosis following perforated duodenal ulcer? *Scottish Medical Journal* 31: 198.

#### PAPER 10

M. A. Lewis, E. M. Baildom et al. 1989. Nephrotic syndrome: From toddlers to twenties. *The Lancet*, February 4: 255-257.

PAPER 11

T. O' Brien. 1984. The needle test for complete rupture of the Achilles tendon. *The Journal of Bone and Joint Surgery* 66-A: 1099-1101.

PAPER 12

J. O. Beattie, D. Hull, F. Cockburn 1986. Children intoxicated by alcohol in Nottingham and Glasgow, 1973-84. *British Medical Journal* 292: 519-521.

PAPER 13

P. Durdey, P. M. T. Weston, N. S. Williams. 1987. Colonoscopy or barium enema as initial investigation of colonic disease. *The Lancet*, September 5: 549-551.

PAPER 14

R. A. Shinton, J. S. Gill et al. 1987. The frequency of epilepsy preceding stroke: Case control study in 230 patients. *The Lancet*, January 3: 11-12.

PAPER 15

Managing drug dealers who swallow the evidence. *British Medical Journal* (authors and date unknown).

PAPER 16

Snoring as a risk factor for ischaemic heart disease and stroke in men. 1985. *British Medical Journal*, September 7 (authors unknown).

## APPENDIX 2

### Glossary of Cryptotechnical Words

#### **Activity**

General

(noun) a situation in which many things are happening at the same time

Technical

(noun) the situation in which a biological agent is exerting an effect on something and causing change

#### **Acute**

General

(adjective) severe or intense

Technical

(adjective) said of a disease which rapidly develops to a crisis

#### **Admit**

General

(verb) to concede, to acknowledge

**Admission** (noun)

Technical

(verb) if a person is *admitted* to hospital, he is taken to stay there until he recovers.

**Admission** (noun)

**Case**

General

- 1) (noun) a particular situation, event or example
- 2) (noun) a container for something

Technical

- 1) (noun) an instance of a disease
- 2) (noun) a person under medical treatment

**Complications**

General

(noun) circumstances that make a situation more difficult to live with

Technical

(noun, always plural) *complications* arise when a secondary disease develops from original illness

**Control**

General

- 1) (noun) some kind of regulation or check
- 2) (verb) to check, to govern

Technical

(noun) an experiment performed to afford a standard of comparison for other experiments

**Course**

General

- 1) (noun) the path in which anything moves
- 2) (noun) a series of lessons or lectures

Technical

(noun) a prescribed series of medical treatments

**Culture**

General

- 1) (noun) culture refers to the arts and people's appreciation of them
- 2) (noun) the ideas, customs and art of a particular society

Technical

- 1) (noun) a micro-organism, tissue, or organ grown in a laboratory
- 2) (verb) to cultivate such in this way

**Examination**

General

(noun) an official test of capacity and knowledge

Technical

(noun) a physical inspection of someone in order to assess his or her state of health or to diagnose disease

**History**

General

(noun) the study of the past



Technical

(noun) details of the patient obtained by questioning and examination

**Intolerance**

General

(noun) an inability to accept or put up with something which may not be enjoyable or pleasant

**Intolerant** (adjective)

Technical

(noun) an inability to endure unfavorable conditions (e.g. a parasite, a drug) without showing serious effects

**Intolerant** (adjective)

**Mean**

General

1) (verb) to intend or signify

2) (adjective) unkind, cruel; unwilling to spend much money

Technical

(noun) the *mean* is the average of a set of numbers

**Negative**

General

1) (adjective) a negative answer means “no”

2) (adjective) a person who is negative sees only problems or disadvantages

Technical

(adjective) if a medical or scientific test is *negative*, it means that something has not happened or is not present

**Onset**

General

(noun) the beginning of something unpleasant

Technical

(noun) the beginning, or outset, of a disease

**Population**

General

(noun) the number of inhabitants of a particular place

Technical

(noun) a group of persons, objects or items considered statistically

**Positive**

General

1) (adjective) completely sure about something

2) (adjective) confident and hopeful

Technical

(adjective) if a medical or scientific test is *positive*, it means that something has happened or is present

**Range**

General

- 1) (noun) a number of different things of the same kind
- 2) (verb) to vary between two points

Technical

- 1) (noun) a set of values on a scale
- 2) (noun) the difference between the largest and smallest in a set of observations

**Reaction**

General

(noun) a person's *reaction* to something is what he feels, says or does because of it

Technical

(noun) any change in the behavior of an organism in response to a stimulus

**Refer**

General

(verb) to *refer* to something means to mention it

Technical

(verb) when a medical case is *referred* to experts, they are formally asked to deal with it

**Response**

General

(noun) a person's reaction or reply to an event

Technical

(noun) the effect of stimulation, e.g. muscular or glandular *responses*

**Risk**

General

- 1) (noun) a chance that something unpleasant or dangerous might happen
- 2) (verb) to do something knowing that something unpleasant might happen as a result

Technical

(noun) the probability that an illness will develop

**Sample**

General

(noun) a small portion of something to show the quality of the whole

Technical

(noun) a small but representative part of a population, used in a test or survey

**Significant**

General

(adjective) having a meaning, full of meaning

**Significance** (noun)

Technical (adjective) above the threshold value at which the results of a statistical investigation are held to contradict a particular hypothesis

**Significance** (noun)

***Stroke***

General

- 1) (noun) an act or mode of striking
- 2) (verb) to *stroke* something is to move one's hand smoothly and gently over it

Technical

(noun) an apoplectic seizure or sudden attack of paralysis; now usually used to indicate a *cerebro-vascular incident*

***Subject***

General

- 1) (noun) the thing or person being discussed
- 2) (noun) an area of study

Technical

(noun) a sufferer from disease, a patient

***Table***

General

(noun) a piece of furniture

Technical

(noun) a set of facts or figures arranged in rows or columns

***Treat***

General

- 1) (verb) to deal with, handle
- 2) (verb) to behave towards something or someone in a particular way

***Treatment*** (noun)

Technical

(verb) when a doctor *treats* a patient or an illness, he or she gives them medical care and attention in order to deal with the disease

***Treatment*** (noun)

***Ward***

General

- 1) (noun) an administrative division of a town or city
- 2) (verb) to protect from harm

Technical

(noun) a room in a hospital shared by several people who need similar treatment