_lognostics
tools for vocabulary research

Technical report pm96y.
Lexical Loadings in English Language Examinations.

Paul Meara
Swansea University

Introduction
The brief for this study was to produce a preliminary analysis of the lexical demands that ULEAC English language examinations make on people who take them. Two basic questions arise from this brief:

a) is there are suitable progression of difficulty across the six levels of the examination, and
b) are the lexical demands made by a particular examination appropriate to the level of the candidates' knowledge?

The raw data for this study was a set of examination papers set as part of the ULEAC profile. One examination paper at each level was provided. The tests vary considerably in style and length. However, it was possible to extract from the material I received a set of texts which were roughly comparable across all six levels. This material -- essentially parts one and two of each test -- makes up the analysis reported in Part One of this report. Part Two looks in more detail at some of the other test parts. Part Three reports a new type of analysis that was specially developed for this project. In all three parts, the main question that we are asking is whether the texts at the different levels were making significantly different lexical demands of the candidates.

This question turns out to be rather more difficult to answer then it appears at first glance. However, the analysis shows that there is a progression of difficulty across the test levels, though the progression is not always as clear-cut as it might be. The analysis also raises some important questions about the appropriateness of the lexical demands made in the low-level tests.

Part One

The analysis reported in the section is based on Parts One and Two of the six examinations that I received. The material analysed included the written examination paper, and the taped material that accompanies them.

The texts varied considerably in length, as shown in Table 1 and Figure 1. Not surprisingly, the lower level tests were generally shorter than the higher-level tests. However, the highest level, 6-7a, was shorter than all the other tests, except for Level One.

Table 1: ***Raw text length (in words)***

text	*1-6A*	*2-6A*	*3-6A*	*4-8A*	*5-4B*	*6-7A*
Running words	1552	1827	1871	2063	2149	1692
Lemma count	240	283	263	324	453	400

Our basic method of comparing the lexical loads of these texts is straightforward, but time-consuming. As a first step, each text is lemmatised -- that is, we reduce each word form in the text to a basic word form which ignores morphological inflections and other similar variants. The output from this analysis is a lemma count -- a list of all the word families which are represented in the text. For obvious reasons, the lemma count is considerably smaller than the word count: most texts contain words which are repeated a number of times, and for normal texts, a lemma count is typically around one-third the size of the word count. The texts here are abnormal in that respect: a rather large proportion of the words are repeated several times.

The lemma counts for this material are also reproduced in Table 1 and Figure 2. The data shows that the lower level examinations tend to have a smaller lemma count than the more advanced examinations. However, the growth in lemmas is not uniform. The Level Two test has more lemmas than the more difficult Level Three text, and the Level Five text contains more lemmas than the more difficult Level Six text.

Figure 1: running words in texts

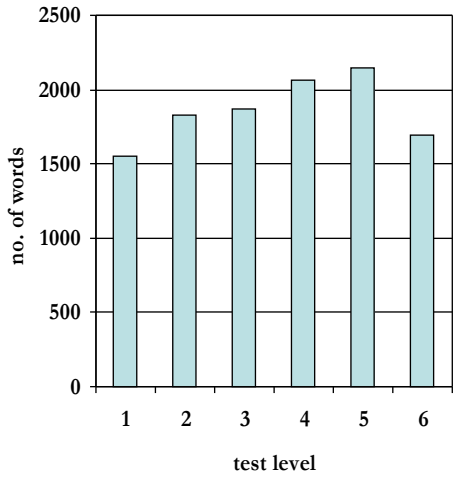
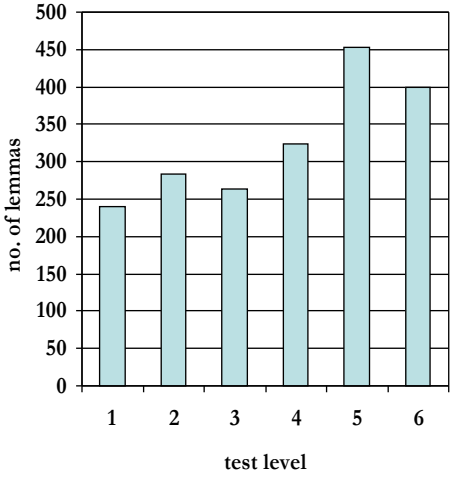


Figure 2: lemmas in texts



The second step of the analysis takes the raw lemma files and classifies each lemma according to its frequency. We do this by comparing the lemma list with a word frequency count: the count used in this report is Nation's 1984 *Word Lists*. This list is an adaptation of the standard frequency counts and specifically aimed at non-native speakers of English. It lists some 2500 word families, broadly grouped into frequency ranges. The analysis produces a breakdown of the lemmas in each text, classifies each of them into a frequency band, and produces a list of 'unusual' words contained in the texts. 'Unusual' in this context means a lemma which is not found in Nation's lists. Since Nation's list covers approximately two and a half thousand common words in English, items which fall outside this range are relatively infrequent.

Tables 2A to 2F provide the basic raw material for the analyses that follow. Each table consists of two sets of figures and a list of words. The first set of figures reports the number of different **word types** in the text: the top line report the raw numbers, while the second line reports the percentage figures. In each case, the total is divided into four categories. Category NAT0 consists of a very high frequency closed class words (prepositions, days of the week, proper names and their derivatives, numbers, and words which are of special significance to EFL learners like *noun* and *verb*). NAT1 includes words which appear in Nation's 1000 word list; NAT2 includes words which appear in Nation's 2000 word list; NAT3 words include anything which the frequency analyser has failed to recognise, and a list of these words is printed out at the foot of each table. In most cases, these items are genuinely unusual, though occasionally a compound word made up of high frequency parts is included in these listings. The type counts are included for the sake of thoroughness, but we will mainly be dealing with the second set of figures in the analyses that follow.

This second set of figures records the total number of **lemmas** found in the text. The top line reports the absolute number of lemmas in the text, while the second line reports the same data as a percentage figure. The lemma counts are typically a little smaller than the type counts, and they weight unusual words slightly more heavily than the type counts do. This is because in the type counts AM, IS, ARE, WAS, WERE, etc. are all treated as separate entries. In the lemma count, all these forms would be reduced to a single word family: BE.

A number of points emerge from these analyses.

Firstly, the lowest level test, Level One, makes substantial lexical demands on the testees. 13% of the lemmas identified in this test have been classed as NAT2 or NAT3 words, and do not occur in the 1000 most frequent word families in English. This represents a considerable level of difficulty for beginners: the figure of 13% corresponds to one word in eight. Whether this is an acceptable figure depends on what levels of vocabulary knowledge ULEAC expects of candidates at this level. My hunch is that the level of difficult words in this text is probably too high, and there may be a case for limiting this first level examination to a basic vocabulary of 1000 words.

The second point to emerge from these analyses is that there does appear to be a general progression through the six levels, but the details of this progression are not straightforward. Table 3 and Figure 3 below show the main indices.

Table 2A *Level One Test*

Type count	310 wds	**Nat Class**	**0**	**1**	**2**	**3**
			107	168	22	13
		%	35	54	7	4
Lemma Count	240 wds	**Nat Class**	**0**	**1**	**2**	**3**
			67	40	20	13
		%	28	58	8	5

NLEM has identified the following unusual words:
CRICKET ALBUM CLASSICAL MATHS GEOLOGY GRANDFATHER CONCERT PHOTO DEER BIRTHDAY BOOKLET LOVELY

Table 2B *Level Two Test*

Type count	360 wds	**Nat Class**	**0**	**1**	**2**	**3**
			107	194	28	31
		%	30	54	8	9
Lemma Count	240 wds	**Nat Class**	**0**	**1**	**2**	**3**
			69	161	23	30
		%	24	57	8	11

NLEM has identified the following unusual words:
HINGES EXHIBITION PITCH MOMENT BESIDE BELOW REFRESHMENTS SHED BLANKS STAFF PRINCIPAL SENIORS REMIND ALLOW PUPILS LUNCH HALL MATHS ACTUALLY SMARTY COUPLE LITTER SENIOR TIMETABLE REFECTORY MESS HOCKEY AWFUL POOL FACILITIES BOOKLET

Table 2C *Level Three Test*

Type count	343 wds	**Nat Class**	**0**	**1**	**2**	**3**
			113	167	35	28
		%	33	49	10	8
Lemma Count	240 wds	**Nat Class**	**0**	**1**	**2**	**3**
			69	134	32	28
		%	26	51	12	11

NLEM has identified the following unusual words:
BELOW BREAKDOWN HOVERCRAFT BROCHURE FURTHER BABYBUS TOURIST DEPOSIT REMIND DEPARTURE KIT AWFULLY INFORMATION RUCKSACK LUGAGE FERRY ADVICE SERVICE MINIBUS CODE PER HIGHWAY LEAFLETS MAXIBUS EMERGENCY BOOKLET PLUS INSURANCE

Table 2D *Level Four Test*

Type count	343 wds	**Nat Class**	**0**	**1**	**2**	**3**
			112	191	66	39
		%	27	47	16	10
Lemma Count	240 wds	**Nat Class**	**0**	**1**	**2**	**3**
			66	162	61	35
		%	20	50	19	11

NLEM has identified the following unusual words:
ENTHUSIASM ACCESSIBLE DOUBLE INTERMEDIATE LODGINGS ACCOMMODATE ACCOMMODATION EUROLANGUAGE MUSEUM QUALIFICATION EXTENSIVE QUALIFICATIONS CONCERNED HALLS CONCERT BASIS SPECIFIC LOCATED CHOICE ADVANTAGE STAFF APPLY REMIND QUALIFIED MOTOR ABTROAD CAMPUS SELF-SERVICE CATERS CONCERTS REASONABLY EXCEED SERVICE WORLD-FAMOUS COMMERCE FACILITIES BOOKLET COUNTRYSIDE THRIVING

Table 2E *Level Five Test*

Type count	578 wds	**Nat Class**	**0**	**1**	**2**	**3**
			130	249	118	81
		%	22	43	20	14
Lemma Count	453 wds	**Nat Class**	**0**	**1**	**2**	**3**
			77	199	102	75
		%	17	44	23	17

NLEM has identified the following unusual words:
OBSESSED CHARGEDS EMPHASIS HIGH-TECH AIRLINE MINIMUM LANDSCAPE OVERSEAS BELOW PRICE-CONSCIOUS MOTORISTS WEEDS MODERN PLANES CHARTER LAST-MINUTE PERIMETER COMFORTABLE TREND EXTREMELY ALTERNATIVE EYESORE FUTURE SURVIVING ECOLOGY TOURS DISCOURAGE UNDERBOOKED HAND-OUT TYOUR GUARANTEED MOUNDS SOPHISTICATED OFFEND DAUNTING ADVANTAGE MOWN INTERVIEW COLLEAGUE BRIEFING TREMENDOUS THOUGHTLESSLY SURCHARGES WHEREAS DELIBERATELY TOURIST ACCORDING COMFORT TACKLE CONCERNS CONCERN AIRCRAFT RUNWAYS DOUBTLESS ALLOW INFORMATION HAZARD STUBBLY VEGETATION ATTITUDES ACTUALLY OPERATING DE-ICING TUNNELS CLEANSE NEVERTHELESS LONG-TERM CHAIRM,AN EXCESSIVE INSIST DISAPPOINTMENT HOLIDAY-MAKER HOLIDAY-MAKERS CLEANSED AWARDED PRICE-CUTTING INCHES DESTINATIONS FACILITIES BOOKLET TAXI-WAYS

Table 2F *Level Six Test*

Type count	521 wds	**Nat Class**	**0**	**1**	**2**	**3**
			122	230	109	60
		%	23	44	21	12
Lemma Count	400 wds	**Nat Class**	**0**	**1**	**2**	**3**
			67	186	93	54
		%	17	47	23	14

NLEM has identified the following unusual words:
OUTLINES ABILITIES MOMENTS APPLICANTS APPLICANT BELOW PERSONNEL ALTOGETHER DRAMA PROS APPLIED NON-ACADEMIC QUALIFICATIONS CONCERNED INITIATIVE COLLEAGUES INTERVIEWS EXTREMELY ROUTINE GRUMBLES INFALLIBLE FUTURE REPUTATION RECRUITING IDENTIFIED POLITICS WELL-ROUNDED INTERVIEWING STAFF INTERVIEW BRIEFING PRINCIPAL TABLEWARE BRANDS HOBBIES GOODS CRITERIA INTELLECTUALLY HONOURS REASONABLY NEVERTHELESS SENIOR DELIGHTED ISSUES CONS EXCELLENT RELATIONSHIPS DEVOTED EXTRA-CURRICULAR AWARDED BOOKLET RECRUITMENT GIFTWARE DISRUPTIVE

Table 3 Proportion of items allocated to NAT classes

text	***1-6a***	***2-6a***	***3-6a***	***4-8a***	***5-4b***	***6-7a***
% NAT0 lemmas	28	24	26	20	17	17
%NAT1 lemmas	58	57	51	50	44	46
%NAT2 lemmas	8	8	12	19	22	23
%NAT3 lemmas	5	11	11	11	17	14
%NAT0+1 lemmas	86	80	76	69	61	63
%NAT2+3 lemmas	14	20	24	31	39	37
Raw NAT0 lemmas	67	69	69	66	77	67
Raw NAT1 lemmas	140	161	134	162	199	186
Raw NAT2 lemmas	20	23	32	61	102	93
Raw NAT3 lemmas	13	30	28	35	75	54
Raw NAT0+1 lemmas	207	230	203	228	276	252
Raw NAT2+3 lemmas	33	53	60	96	177	147

Figure 1 has already shown that texts increase in length as the tests get harder. Figure 2a shows that this increase in length is generally accompanied by an increase in the complexity of the

Figure 2a: frequency of lemmas in texts

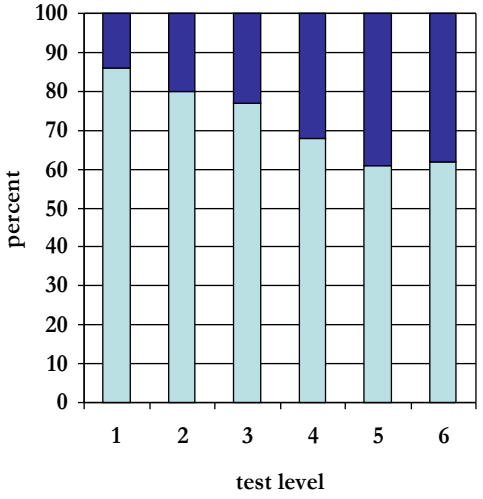
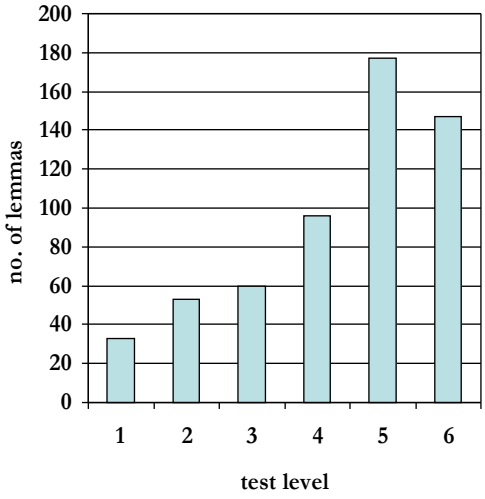


Figure 2b: number of Nat2/3 words

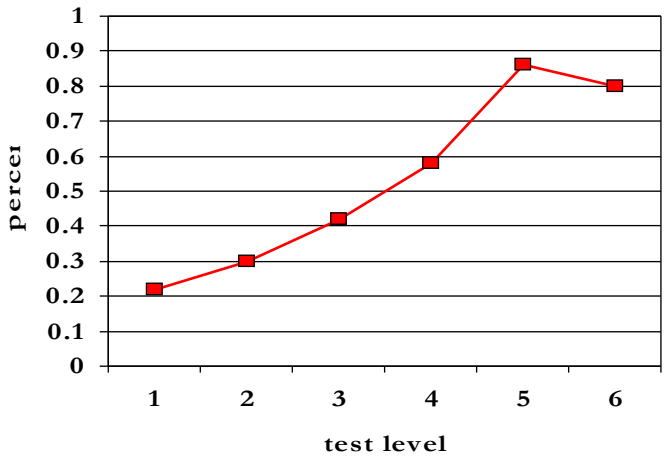


vocabulary used. The proportion of very common words in the texts falls in line with the difficulty of the examination, while the corresponding proportions of unusual words in the texts rise as the examinations get more difficult. The percentage of very unusual words (NAT3) ranges from 5 to 17. The corresponding figure for moderately unusual words (NAT2 and NAT3 combined) ranges from 14 to 39%. In percentage terms, these figures mean that the Level Six test contains almost three times the rate of difficult words than the level one test. Figure 3 also suggests that the Level Five test analysed here may actually be more difficult than the Level Six test. The Level Five text contains the same proportion of NAT2 and NAT3 words as the Level Six test, but it is almost 20 percent longer than the Level Six test. These differences in text length produce a real difference of 30 words in the NAT2+3 categories. The analysis therefore suggests that from a narrow lexical point of view, the quality control on these tests is less than ideal. The Level Five test, contains almost six times as many hard lemmas as the level one test.

A useful way of comparing texts which takes some account of their length, is to look at the relative proportions of the text which come from the different frequency bands. We do this by calculating for each text with a number of words in category NAT2 and NAT3, and the number of words in category NAT1, and comparing the two figures (cf. Laufer and Nation 1995). The raw figures for this calculation are shown in Table 3, and the result will be found in Figure 3. Again, the figures suggest that the test paper for Level Five might be out of line. The scores range from .25 for the simplest text, to .80 for the Level Six text, but again, the Level Five text produces an anomalously high score. The graph in figure 4 suggests that the Nat2Nat3 ratio for the Level Five test ought to

be around .710. This figure implies that the levels get progressively more difficult, but that after Level Four, the rate of increase slows down a little.

Figure 3: ratio of low frequency to high frequency words



It is actually hard to talk sensibly about lexical difficulty when we have only a small corpus of text to deal with. In the data analysed here, we have only a single instance of a text at each level and we don't know whether the texts included in the sample are really typical or not. With the exception of the Level Five test, which looks anomalously difficult, the tests **seem** to show a steady increase in the level of demands that they make on the testee. However, it is possible that we just struck lucky with our tests, and that a bigger set of samples would show that there was a high degree of overlap between the levels. This is obviously something that needs to be monitored in future.

Part Two

The analysis reported in this section is based on the reading comprehension passages in Sections Four and Five of each of the six pilot papers. Direct comparisons between the papers are difficult, since the questions are not directly parallel with each other, and the reading material is used in very different ways at the different test levels. However, it is possible to draw a broad comparison between some of the texts, but the reader needs to bear in mind that the parallels are weaker here than they were in the previous section.

The questions included in this analysis were:
Level One question 5, Level Two question 5, Level Three question 4, Level Four question 4, Level Five question 4 and Level Six leaflet.

Fairly close parallels exist between the Level One and Level Two material, and between the material from Levels Three to Five. The Level Six material does not relate easily to any of the other levels. All the material involves authentic (or apparently authentic) reading texts. Candidates are required to demonstrate their understanding of this material in a variety of ways, but these tests of comprehension will not concern us here.

As before, the texts differ considerably in length. Here however, the increase in length does not correlate well with the different levels (see table 4). The Level Two text is close to double the length of the Level One text. Similarly, the Level Four text is more than twice as long as the Level Three text. Levels Four and Five are effectively identical in length, while the Level Six text is considerably longer than these two levels. The anomalies come in the Level Three text, which is more than 300 words **shorter** than the Level Two text, and 700 words shorter than the Lev el Four text (see Figure 4.) Again, it is difficult to assess the significance of this variation with only a single text at each level, but the data reported here suggest that the Level Three test is seriously out of line by being too short, while both the Level Two test and the Level Four test may be too long.

Table 4: Raw text length (in words) and lemma counts

test	*L1_5*	*L2_5*	*L3_4*	*L4_4*	*L5_4*	*L6_L*
Running words	469	839	514	1136	1133	1622
lemmas	158	276	201	364	423	507

Figure 4a: running words in texts

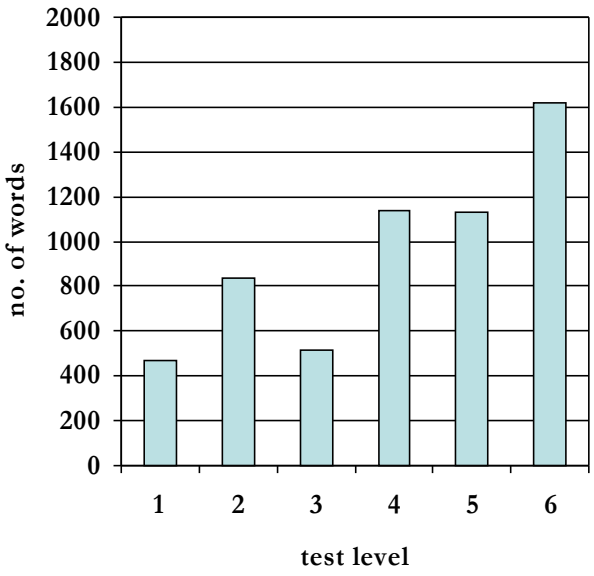
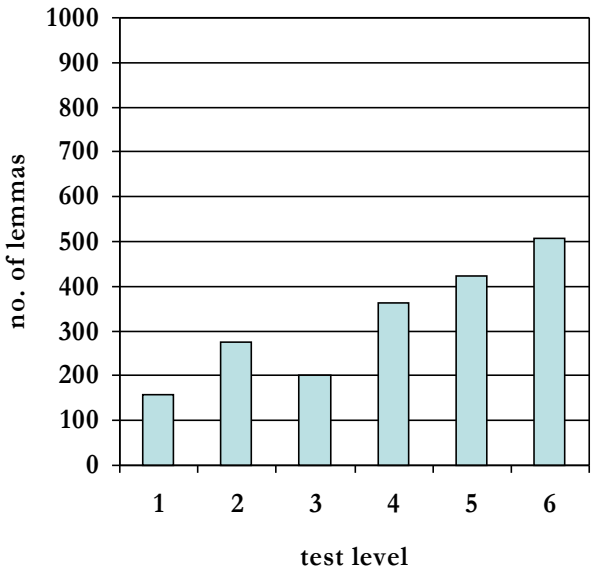
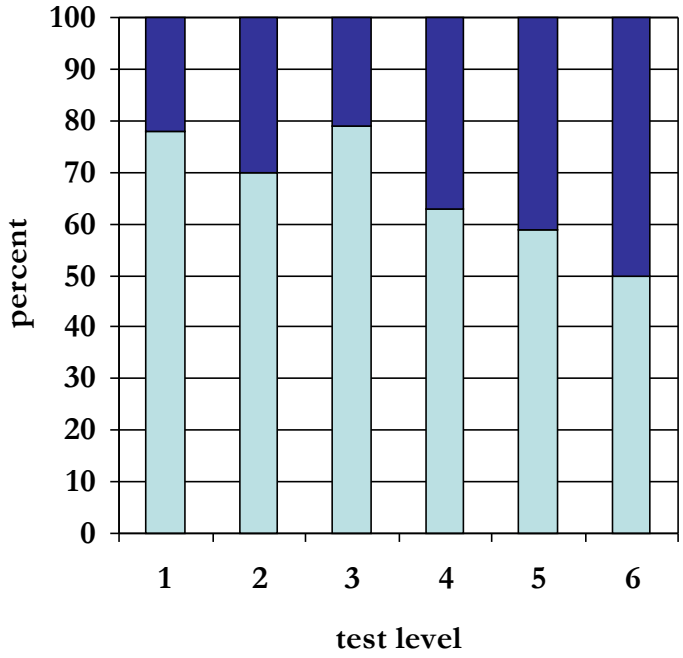


Figure 4b: lemmas in texts



The raw lemma counts, also shown in table 4, show a similar story. The Level Two text appears to contain a considerably richer vocabulary than the level three text. (See figure 5). In this figure, NAT2/3 words are shown in dark shading, and NAT0/1 words are shown in light shading.

Figure 5: frequency of lemmas in texts



Tables 5-A TO 5-E on the following pages show the results of the more detailed analysis of these texts. The procedures followed were identical to those reported in section 1. These detailed analysis make a number of obvious points.

Firstly, the lexical demands of the Level One tests once again appear to be high. 22% of Level Two and Level Three lemmas is an unusually high proportion, and must present considerable difficulties for testees at this level. (See figure 5 a)

Secondly, the detailed analyses confirm that the Level Three test is out of line with the other materials. Level Three, on these counts, is not greatly different from level one, and considerably easier than the Level Two test. (See figure 5b)

Thirdly, and partly in consequence of point 2, the rise in difficulty between levels three and four is much greater than the equivalent rise between the other levels. Levels Four and Five show the smallest increase, while the differences between levels one and two, and levels five and six are substantial.

Finally, the level of difficulty in the two hardest levels is very high indeed. Level Five has a total of 41% NAT2 words and NAT3 words, while for Level Six, this figure is exactly 50%. These proportions are high, but not excessively so. For comparison, table 6 shows a set of equivalent data from a study of lexical richness in Cambridge Proficiency examinations. The Cambridge material is slightly shorter than the texts analysed here, and this might be expected to affect the overall

distribution of the word classes. Nonetheless, the figures suggest that the top level tests are broadly in line with the practices of other examination boards, but may be rather more difficult. (See also figure 6. In this figure, the stacked bars show the % of lemmas in each of the NAT categories – NAT0 wds at the bottom of the stack, NAT4 words at the top of the stack.)

Figure 6: CPE vs ULEAC Lv5 & Lv6

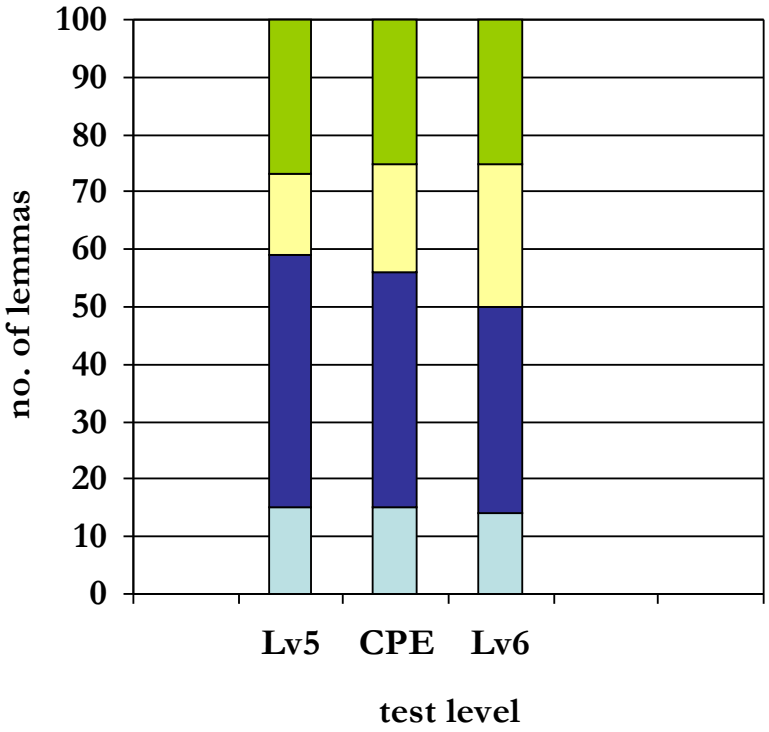


Table 6 *Distribution of words in a sample CPE listening comprehension test*

Type count	551wds	**Nat Class**	**0**	**1**	**2**	**3**
			116	225	94	116
		%	21	41	17	21
Lemma Count	460 wds	**Nat Class**	**0**	**1**	**2**	**3**
			70	188	88	114
		%	15	41	19	25

Table 5A Level One Test

Type count	205 wds	**Nat Class**	**0**	**1**	**2**	**3**
			71	99	13	22
		%	35	48	6	11
Lemma Count	158 wds	**Nat Class**	**0**	**1**	**2**	**3**
			38	85	13	22
		%	24	54	8	14

NLEM has identified the following unusual words:
CRICKET PLACEMENT FAVOURITE FOND CLASSICAL DRAMA A_LEVEL COMPUTER CHEMISTRY SQUEAK LITERATURE MANCHESTER_UNITED RECIPES COMICS MOUSE INTERNET GUITAR KEEN CHESS BIRTHDAY CHOIR APPLE_PC

Table 5B Level Two test

Type count	327 wds	**Nat Class**	**0**	**1**	**2**	**3**
			73	166	44	44
		%	22	51	13	13
Lemma Count	276 wds	**Nat Class**	**0**	**1**	**2**	**3**
			53	141	39	43
		%	19	51	14	16

NLEM has identified the following unusual words:
PERMITTED SELF-ACCESS RECEIPT ACCOMMODATION TEENAGE RELIABLE SUPERVISOR DENTIST SEATBELT DISRUPTS LOCATED RUDE DIRECTOR SOLUTION PRINCIPAL AUDIO-VISUAL HOTLINE OVERCHARGED CANTEEN HAIL RECOMMEND BINS MEDICAL WELFARE DESIGNATED LOCATION LITTER TIMETABLE SCHEULED TIPS FIST PM RESIDENCE ACCOUNTS EXIT ETC. COUNTER EMERGENCY FACILITIES BOKLET FASTEN LIRA INSURANCE HOST

Table 5C Level Three Test

Type count	254 wds	**Nat Class**	**0**	**1**	**2**	**3**
			83	117	26	28
		%	33	46	10	11
Lemma Count	201 wds	**Nat Class**	**0**	**1**	**2**	**3**
			50	102	25	24
		%	25	51	12	12

NLEM has identified the following unusual words:
CYCLING POLLUTION SLIM OVERWEIGHT CYCLE FEDERATION VENTS MENDICAL HEALTHIER HALVE LIFESTYLES CYCLISTS JUNIOR CYCLIST HILLMAN EN_ROUTE BREATHE ADVICE PAVEMENT MINIBUS CAMPAIGN PER_CENT VEHICLE STATISTICALLY INTAKE EVIDENCE

Table 5D Level Four Test

Type count	452wds	**Nat Class**	**0**	**1**	**2**	**3**
			105	199	63	85
		%	23	44	14	19
Lemma Count	364wds	**Nat Class**	**0**	**1**	**2**	**3**
			63	162	58	81
		%	17	45	16	22

NLEM has identified the following unusual words:
AA PRESERVED ELABORATE BIRTHPLACE DECLINED BESIDE PUPIL BELOW WORKMEN BRONZE FOUNTAIN DEDICATED WOOL PROMINENCE MUSEUM UNDERLINING FOOTPATHS LIFE-STYLE HERBS BENEATH BC RURAL UNKNOWN RESTORED POET RETAIN RELICS GODDESS CHOICE MEMORIAL COTTAGE IMPLEMENTS CRESCENT CITIZENS DELIGHTFUL DECORATIONS TOURISM TICK DELIBERATE BROAD TOURIST RUBBISH UNDERLINE GALLONS ACCORDING EXHIBITS FAME GRAVE EASTER PORTRAITS SCHOOLBOY STATUE ARC APPLICATION MEADOW ACTUALLY FLANKED MUSEUMS RIVERSIDE OUTBUILDINGS SETTLEMENT PRESERVE FURNISHED DRAMATIST CANAL FURNISHINGS INSPRIATION COMMUNITY IDENTITY GUILDHALL HEALING RECALL MEMORABLE INN REGAIN MAYOR CRAFTSMEN LEGEND GLOVER FLEETING ERA DOCUMENTS LINGER SNATCH COUNTRYSIDE

Table 5E Level Five Test

Type count	512 wds	**Nat Class**	**0**	**1**	**2**	**3**
			106	230	61	115
		%	21	45	12	22
Lemma Count	423 wds	**Nat Class**	**0**	**1**	**2**	**3**
			64	188	58	113
		%	15	44	14	27

NLEM has identified the following unusual words:
MIDST REVELLED APOLOGETICALLY LUDIRCOUSLY BOW ENTHUSIASM FABRICS OVERCROWDED URGED CRUMBLING SPICY DIMLY AUNTIE DELIGHT CLUSTERED FELLOW PETS NERVOUS WESTERN REPELS SCHEDULE COLLEAGUES DETERMINATION ZIPPED AGONY ACCOMPANIED ROUTINE UNHYGIENIC AWFULNESS BLOCK-PRINTED BLOUSE HEAVEN SWORE CRAFTS AFGTERNOONS ESCORT AMBLED KINGDOM ASUNDER SIGHTSEEING STICKINESS RUDE TAILOR VOLUNTARY LICK BAZAAR INTRACTABLE APATHY LACK-LUSTRE COLLEAGUE IN-SERVICE TRANSFROMED FUNDS TOURIST LANES HOSTESS REGARDLESS TOES CREATURE CLOTHS RAMSHACKLE INEFFECTIVE GAZED PYJAMAS SCOOTER PUPILS TRIP PERCEIVE COUPLE SUMMONED CHARGED UNDERFUNDED REMARKABLY BROWSED SNEEZED CORRUPTION IMMUNISED RELATIONSHIP INDEFINITE ECONOMIC WHISKED GARLANDED CUFF TIKPS REMORSE ELEPHANT PHOETICS DESPERATE EUPHORIC CRAFTSMEN TRUNK BILLETED FLASHED THRUST SACRED RAPACITY AWFUL MONKEYS ROOTLING CHAT NICHE INSTANT SOMEHOW SEMINAR RUPEES SUBVERSIVELY RUSTLED

Table 5F Level Six Test

Type count	656 wds	**Nat Class**	**0**	**1**	**2**	**3**
			122	234	159	141
		%	19	36	24	21
Lemma Count	507 wds	**Nat Class**	**0**	**1**	**2**	**3**
			69	184	127	127
		%	14	36	25	25

NLEM has identified the following unusual words:
COMPREHENSION SEEKERS JOURNALS FERRET ALBEIT ANTICIPATE GRADUATE FOCUS MEDIA RELEVANT EMPHASIS UNSUITABLE CONGLOMERATE BELOW PERSONNEL ENTRANTS INSTITUTEARTY INSTITUTION INSTITUTIONS RELIABLE TRADITIONALLY CIVIL QUALI-FICATIONS GRADUATES BRILLIANT MEDIUM TYPICALLY SUPERVISION JUNIOR SPANS EXPERTISE POST-SCHOOL FURTHER MAJORITY MANPOWER PUBLICISED ALTERNATIVELY UNDERKILL UNKNOWN DISILLUSION DISTINGUISHES FUTURE RECRUITING PREMATURE BASIS EMPATHY CHARACTERISTICS ENORMOUSLY ACADEMIC PENNY-PINCHING CIRCUMSTANCES VARIATIONS LOSSES SOPHISTICATED NOMINALLY REALISTIC RECRUIT CRUCIAL DETRACTS WHOSE STATURE STAFF CURRICULUM BROADER RECOGNITION ENTRANT CONTENTION PRINCIPAL TABLEWARE BRANDS ASSESSING ACQUISITION MUTUAL NON-EXISTENT PERCEPTION PERCEPTIONS FUNDS CONTRADICT APPLIED EG. CONTENT TACKLE OPTIONS CONCERN SALESPEOPLE ENTRUSTED RELYING COUNTERED INTAKES UNSATISFACTORY THUS ABOVE IDENTIFYING PROMOTED SIGNIFICANT INFORMATION QUOTE RECRUITERS ACTUALLY COMPUTER ENTITLED AMATEURS METHODOLOGIES RECRUITED TERTIARY ENDEMIC EFFECTGI VE DISCLOSURE DISASTROUS REVIEW LOCATION RECRUITER CADIDATE INVIGILATOR VISUAL COMMUNITY CULPRITS SERVICE ISSUES SOUGHT ALAS EXCELLENT RECALL DEVOTED DISCLOSED ORGANISATIONS CREDIBLE ORGANISATIONAL INOUT FANCY SECTORS SUPPLEMENTARY UNREALISTIC MATURITY INSERTION INTAKE RECRUITMENT GIFTWARE CROSS-BORDER AMAZING

Part 3

While I was working on these analyses, I became increasingly concerned that the patterns produced by the computer programs were being distorted by the length of the source text. The problem here is that it is unreasonable to expect short texts to have the same lexical characteristics as long ones, and this problem is exacerbated when we base our counts on lemmas, rather than word tokens. The reason for this is that highly frequent words like THE or A car only counted once in our lemma counts, no matter how many times they occur in a text. In longer texts, therefore, the contribution of high frequency words tends to diminish, and low frequency words come to dominate the overall distribution of the profiles.

Trying to get round this problem, I devised a way of looking at the texts, which appears to be relatively independent of text length. The approach is basically very simple. Each text is divided up into 10 word segments, disregarding punctuation, sentence structure, or any other factors other than lexical ones. The texts are then reanalysed, by counting the number of unusual words occurring in each 10 word segment, where 'unusual' is defined as anything not included in the NAT0 and NAT1 vocabulary levels. We can then calculate for each text, the probability that a randomly chosen segment will contain 0,1,2,3,4...10 unusual words. Our assumption is that easy tests ought to have a higher probability of segments containing 0 unusual words. For more difficult test, the chances of getting 0 unusual words would be relatively low, and the chances of a large number of difficult words correspondingly higher. Ideally, we would like to find a steady progression across the six levels of difficulty.

Table 8 shows an analysis of this sort for parts one and two of the test set. For each test, the table shows the probability (to three decimal places) of getting N unusual words in successive 10 word sequences. The data is also illustrated in figure 8 which is discussed in more detailed below.

Table 8: probability of finding N unusual words in a segment
parts one and two

text	*N=0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
1_6a	609	292	054	036	006	000	000	000	000	000
2_6a	367	414	186	026	025	000	000	000	000	000
3_6a	358	353	179	051	035	020	000	000	000	000
4_8a	230	355	254	100	052	004	000	000	000	000
5_4b	160	311	302	151	064	009	000	000	000	000
6_7a	257	314	234	120	045	022	005	000	000	000

At first sight, the data look more than a little confusing. The different data sets seem to vary quite substantially -- for instance, the highest point of the level one texts corresponds to 0 unusual

words, whereas for the remaining texts, the highest point in the figures corresponds to 1 unusual word per 10 running words. The 'tail' of the figures -- the chance of getting more than 0 or 1 unusual words in a set of 10 -- increases as the test levels get harder. The curves for the level two and level three tests look very similar, and the curves for the level four and level six tests also resemble each other closely. Fortunately, the data can be significantly tidied up by fitting this empirical data to set of curves called a *Poisson distribution*. Poisson distributions are statistical tools which are very useful in describing the occurrence of rare events. They were originally used to examine the pattern of Prussian cavalry officers being kicked to death by their horses in any one day, but may have more general applications where the phenomena under investigation is relatively rare, and they are widely used in quality control situations. Poisson distributions are defined by equation 1:

$$P(N)= (\lambda^{N} (e^{-\lambda}))/ N! \qquad \textbf{Equation 1}$$

where P(N) is the probability of of getting N items in an interval and λ is a parameter which describes the shape of the probability distribution.

The point of this is that complex curves like the ones reported in table 8 can be economically described in terms of the single parameter λ. Figures 8a to 8f show how this is done. Each of these figures contains two thirds: the red line with square points graphs the data reported in table 8, while the blue line with the circle points graphs the Poisson curve that most closely matches the empirical data. The λ parameter is reported in the title of each figure.

It will be immediately apparent that the Poisson curves are uncannily close matches for the empirical data in almost all cases. The worse match is the level two text, but in all the other cases, the Poisson curve is practically indistinguishable from the real data. This suggests that we can compare the characteristics of the six texts by simply plotting their λ values on a single chart, and this has been done in figure 9 and figure 10.

The analysis here broadly confirms our earlier interpretation. For the listening tests, the texts at level two and level three are virtually indistinguishable, and the texts that levels four and 6 are similarly close. The graph suggests that there is a clear progression across the tests, but the level two test may be slightly to difficult and the level six test may be too easy. There is also a suggestion in the data that the level one test may be too difficult too, that conclusion that echoes our earlier analysis.

Table 9 summarises the distribution data for these tests, and also presents a parallel analysis of the reading comprehension data analysed in part two what this report.

Table 9: lambda values for texts

Level	***1***	***2***	***3***	***4***	***5***	***6***
Listening Texts	0.48	0.98	1.02	1.44	1.47	1.42
Reading Texts	0.87	1.53	1.52	1.76	1.76	2.66

Figure 8a: Text 1-6a: λ=0.48

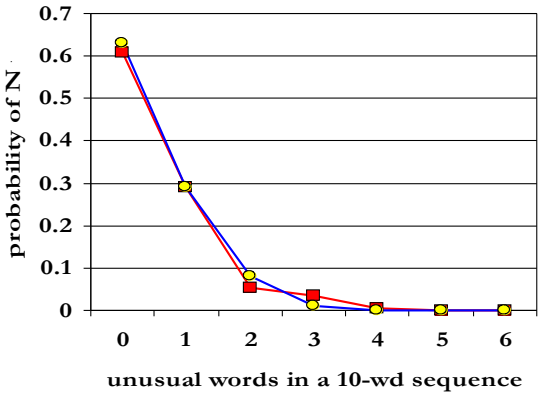


Figure 8b: Text 1-6b: λ=0.98

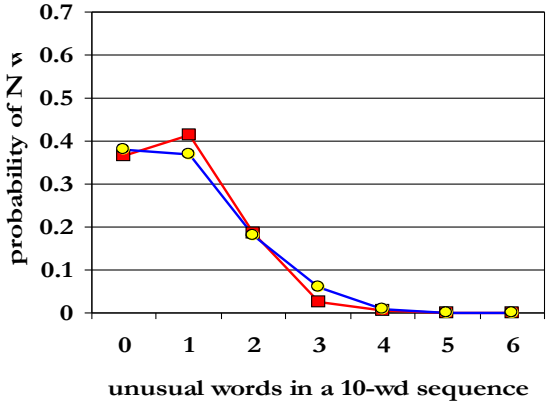


Figure 8c: Text 3-6a: λ=1.02

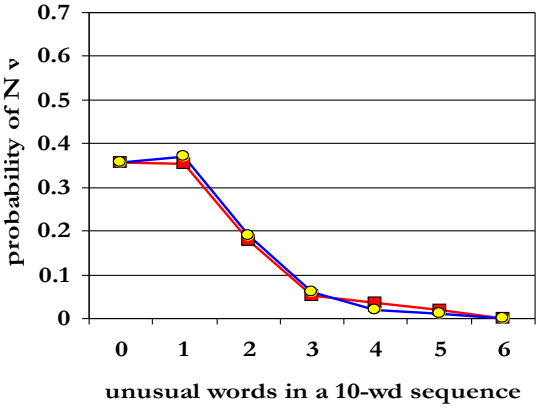


Figure 8d: Text 1-6d: λ=1.44

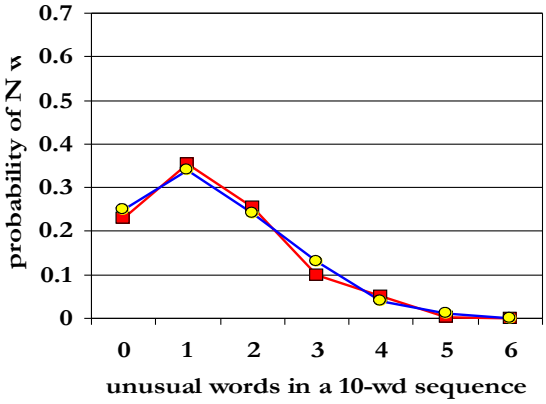


Figure 8e: Text 5-4b: λ=1.77

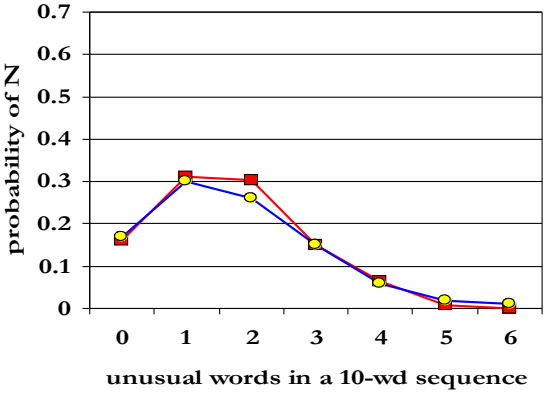


Figure 8f: Text 6-7a: λ=1.42

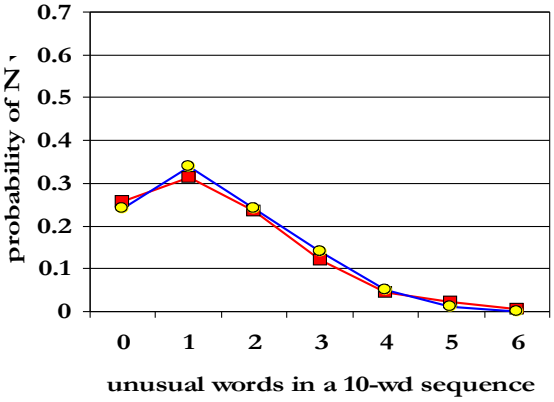


Figure 9: Lambda - parts 1 and 2

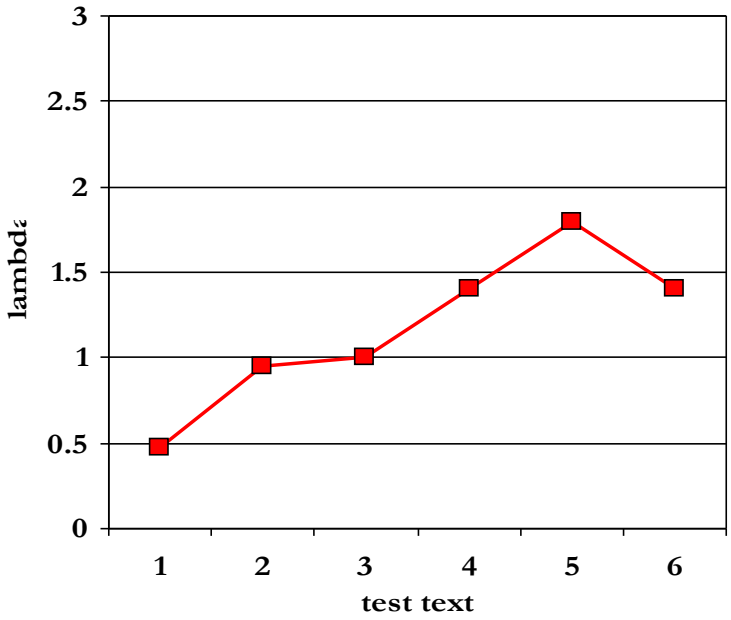


Fig 10: Lambda - reading comprehension

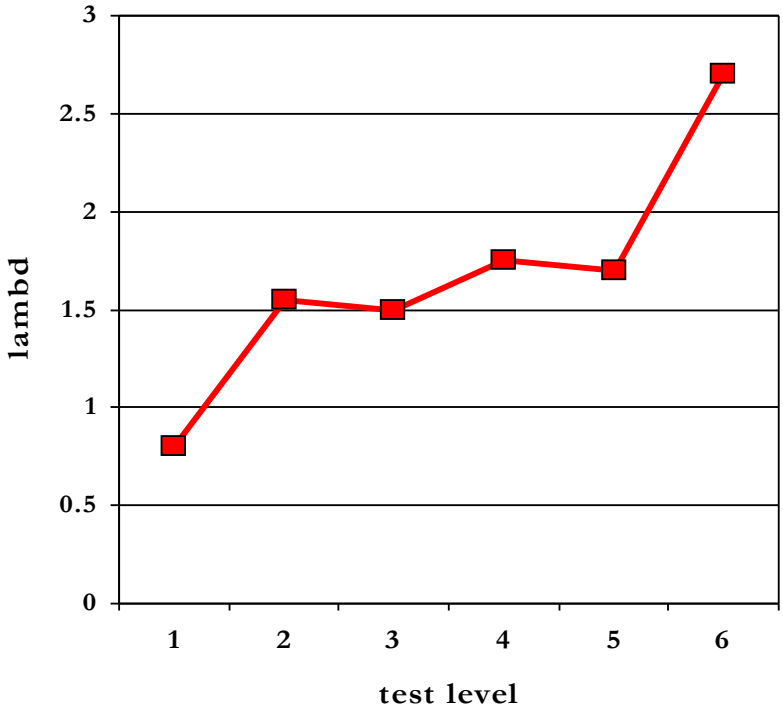
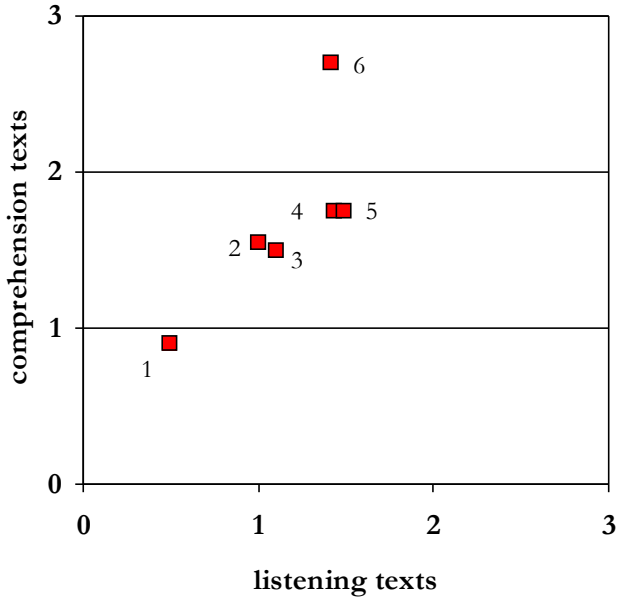


Table 9 shows, not surprisingly, that the lambda figures obtained for the reading texts are generally higher than the equivalent figure for the listening texts. However, the relationship is once again not straightforward (see figure 10). The data suggest that the level two and level three texts are indistinguishable from each other, and that the level four and five texts are similarly indistinguishable. The level six text on the other hand, appears to be very much harder than the other texts, with a marked increase in difficulty between the Level Five and level six texts.

Again, this analysis broadly confirms the analysis put forward in the previous section of this report. The obvious interpretation of these data is that the lexical difficulty of the passages at levels 2 to 5 is better control than was the case with the listening passages, but it is difficult to see how the current texts at levels two to five can be expected to discriminate candidates accurately, except for that a higher level texts tend to be longer: there is no evidence from this analysis that the higher level texts make increasing **lexical** demands on the testees. There is some room for variation within these levels. In particular there is probably a strong case for reducing the difficulty of level two texts in future. There might also be a case for increasing the difficulty of the level five texts relative to the level four texts, although the arguments here are much less clear.

Figure 11 shows all six texts in terms of the lambda values for both parts of the examination. This analysis suggests that there might be some problem with the way lexical loadings vary across the ULEAC suite taken as a whole. There is clearly a marked difference between the level one tests and the others, with a very significant increase in difficulty between the level one and level two tests. The tests at levels two and three clearly form a cluster which are not distinguishable. Similarly, the level four and five tests form a cluster. It is not possible to tell whether these two

Figure 11: Lambda: test parts



clusters actually make up a single large cluster, but it is clear that within the smaller clusters the level of variation is tiny. paper six comes out as a serious anomaly in this graph. As far as the reading tests go, level six represents a huge increase in difficulty from levels four and five, whereas in terms of listening comprehension, the level six is no more difficult than the level four and Level Five tests.

Conclusions and recommendations

The analysis reported here has necessarily been a preliminary one. The amount of text that I had to work with was is relatively small, and included only one test at each of the six ULEAC levels. In spite of this, and number of suggestive ideas have emerged from the analysis.

The examination texts were studied in two different ways. The first analysis was the straightforward count of the proportions of unusual words occurring in each text, with listening and reading texts analysed separately. The tests showed a progression from lower to higher levels in general, but at the detailed level, the progression was not always as clear as we might have expected. The lower level tests make surprisingly high lexical demands on the testees, and the higher level tests were sometimes less demanding than their lower level counterparts.

A new methodology was developed for the second analysis, which seems to offer some interesting possibilities for quality control in English examinations. The methodology is straightforward and almost fully automated, and could be applied routinely to ULEAC examinations. It also looks as though it might have some interesting applications in evaluating the output of candidates in examinations.

This second analysis broadly confirmed the conclusions of the first analysis. The listening texts show a rather sharp rise in difficulty between levels one and two, a slight rise between level two and level three, and then only a small increase in difficulty level in the higher tests. For the reading comprehension tests, the second analysis suggests that there is very little difference between Levels Two, Three, Four and Five, the very sharp increases in difficulty between Levels One and Two and between Levels Five and Six.

In the absence of any clear descriptions about what levels of lexical competence are expected of candidates at various levels, it is difficult to say whether these results are satisfactory or not. The Level Two test and Level Three test, for example, seem to be much like a higher level tests for written comprehension, but with a simpler listening component. The board might like to consider whether it would be more appropriate to change this emphasis. For example the Board might like the Level Two test to require the same level of listening vocabulary as the Level Three test, but to make less stringent demands in the area of reading comprehension.
The Board may wish to give further consideration to the following points, which strike me as important:

a: Level One tests, both for listening comprehension and reading, contain a surprisingly large number of unusual lexical items, and there might be a case for controlling the vocabulary of the

tests more strictly at this level. There might also be a case for making this level considerably easier overall than it appears to be at the moment.

b: the Level Two tests, similarly, contain a relatively high proportion of unusual words. The Board might consider whether it wishes to draw up policy on this, since the current Level Two tests are not well distinguished from the Level Three tests. As a possible solution to this problem, the Board might wish to continue using texts containing unusual words, but provide some form of gloss for words of this type.

c: a separate issue is whether the listening tests should be allowed to contain **any** unusual words that the candidates may not be familiar with. There may be a case for rigorously control the vocabulary of listening tests at Levels One to Three. At the moment, the chances of a candidate meeting an unknown word in these tests is uncomfortably high.

d: the Level Six tests seem to represent a considerable increase in difficulty as far as the reading comprehension texts are concerned, but only marginal increase in difficulty for the listening comprehension compared to Levels Two and Three. There might be some scope here for developing more demanding listening tests, in order to differentiate between Level Five and Level Six listening skills.

e: the analyses reported here are basically only as good as the word lists on which they are based. The word lists prepared by Nation and used in these analyses are probably as good as any around, but there are still a number of problems with them, and their use is not straightforward. In the long run, it would make sense for ULEAC -- or a consortium including ULEAC -- to consider a large-scale project that aimed to update the work of Michael West. West's *General Service List* underlies almost all the word on word frequency in English as a second language, including Nation's lists. It was first published in its present form in 1953, although it was based on work that had originally appeared in the late 1930s. West's work has never been seriously updated, and a replacement is long overdue.

References

Laufer, B and ISP Nation
Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics* 16,3(1995) and, 307-322.
Nation, ISP
Word Lists. Wellington NZ: Victoria University of Wellington. 1984.
West, M
A General Service List of English Words. London: Longman. 1953.