

# *The Vocabulary of the English Coursebooks: An Analysis*

---

Antonio Miranda  
Universidad de Málaga

In this paper, the vocabulary of some English coursebooks used in Spanish Secondary Education [BUP] is analyzed to find out the size of the common lexical core. The results indicate that there does not exist a regular lexical distribution in them, which shows that different selection criteria, if any, were adopted.

## 1. Hypothesis

Assuming that the lexical aims to be reached in L2 learning must be similar for all the students in the same level, it may seem logical to suppose that the coursebooks<sup>1</sup> for the referred levels should at least share a minimum lexical core (MLC). However, we suspect that the MLC will not be quantitatively meaningful and we will try to prove this hypothesis empirically<sup>2</sup>.

## 2. Method

To find out the size of the lexical core common to sixteen three-level coursebooks<sup>3</sup>, a comparative analysis has been performed, and in order to get valid, reliable and representative results, the sixteen coursebook series<sup>4</sup> were selected having in mind the following criteria:

- a) Coursebooks had to contain either a list with the words used in each unit, or a vocabulary summary at the end of the book, or both.
- b) They were used as textbooks in BUP in the eighties.
- c) Textbooks which different methodological approaches were included in the study.
- e) We tried to find a balance between textbooks published by British and Spanish publishers.

With the 9159 different words<sup>5</sup> appearing in the coursebooks a corpus was produced which was later transferred to a database file, called Total Lexical Set (TLS), where each record contained the following fields and contents:

entry:	word <sup>6</sup> or phrase (used as the identifier of the record).
level <sup>7</sup> :	level of the textbook where the entry appears for the first time.
class <sup>8</sup> :	grammar category with which the word may function.
range:	number of series <sup>9</sup> where the entry is used.

The entries in the TLS have been sorted to calculate the number of words of each coursebook in the sixteen series, taking into account the values of level and range, or the grammatical class. The percentage rating the number of entries of each series and the

TLS has been also calculated.

### 3. Proof

#### 3.1. Counting of the entries of each series

The number of entries of each series<sup>10</sup> making up the TLS is indicated in the table below.

A 3931	F*	1438	O	1780	T	1830
B 3028	L*	2304	P	2249	U	2205
D* 1052	M	2339	R	1308	V	2066
E 3086	N	2522	S*	1072	W	2032

The contribution of each series to the TLS is unequal, ranging from 3931 to 1052, and some of them are clearly below the minimum standards recommended by the Education Authorities which is, namely, a «productive» lexical inventory of about 3000 English words<sup>11</sup> at the end of the Secondary Education Period. If we assume that the textbook is the main source of lexical input<sup>12</sup>, the achievement of this goal will be quite hard, as some series do not even reach that level. Consequently, teachers are then supposed to complement the textbook and to develop their students' lexicon with the help of other materials such as simplified readers, vocabulary builders, magazines, etc.

#### 3.2. Analysis of the TLS according to range

Once the TLS entries have been counted and sorted according to range, we obtain the results shown in the table below.

R16	153	R12	133	R8	223	R4	491	
R15	135	R11	158	R7	244	R3	770	
R14	182	R10	177	R6	309	R2	1332	
R13	127	R9	198	R5	403	R1	4124	
RR4	597	RR3	666	RR2	1179	RR1	6717	
RR4+RR3+RR2						2442	RR1	6717
RR4+RR3+RR2						26.7%	RR1	73.3%

The distribution of 26.7% of the items is spread all over the values of range (5 - 16) while 73.3% of the TLS is piled up on range values 1 to 4; in other words, the entries appearing in one/two/three/four series amount up to 73.3% of the TLS. Furthermore, the entries used exclusively by only one series account for 45% of it, while the entries common to all the series barely cover 2% of the TLS which make up the Common Lexical Core (CLC). This sort of distribution is known as hyperbolic, representing the function  $y=k/x$  which means that the distribution would not change<sup>13</sup> if new more data were to be added. As the CLC seems to be too small, we are forced to include the entries ranging from 4 to 16 to constitute the MLC which amounts up to 2718 words<sup>14</sup>, and which differs from what is generally referred to as pedagogical vocabulary lists. The main difference consists in the criteria adopted for the selection of words. Whilst most wordlists (WL) have been prepared by frequency-sorting the words of a written corpus, the MCL is produced sorting the TLS words according to range, which seems to embody the values of frequency, availability, learnability: in the former, a word is classified as a high/low frequency word depending on the number of times it appears in the corpus; in the latter a top-range word is a word which has proven to be necessary for sixteen British/ Spanish textbook-authors when writing their books. Furthermore, the MCL can update the WL removing «obsolete» words and/or including «new» ones or those connected with the class-room context which turn out to be necessary in everyday tasks, though they were not listed due to their low-frequency index, i.e. «chalk» or «blackboard».

If we assume that range implies frequency, then the MLC must contain a great deal of the words listed in the frequency studies<sup>15</sup>. To prove this, the MLC entries were checked against the words of the below VL and the level of coincidence was also found out, as follows:

DIXON<sup>16</sup>. The MLC contains 457 words out of the first group, 433 words out of the second and 762 out of the third. As a summary, the 1652 coincidences represent 83% of Dixon's list.

EATON<sup>17</sup>. The MLC contains 590 entries included in the groups (2a & 2b) which represent 59%. As a summary, the MLC covers 65%.

HINDMARSH<sup>18</sup>. 87% of the entries in the CLP can be found in Hindmarsh's CEL.

KUCERA<sup>19</sup>. 1545 entries in the MLC are included in Kucera's work, which represents about 62%. Nevertheless, this percentage can increase if proper nouns not taken into account.

LONGMAN<sup>20</sup>. The MLC contains a 84% out of the Longman word list.

PALMER<sup>21</sup>. 95% of the words out of Palmer's list are included in the MLC.

RICHARDS & OGDEN<sup>22</sup>. 711 Basic English words are contained in the MLC, which represent a 84%.

THRESHOLD<sup>23</sup>. Out of the 1500 words of The Threshold Level, 1137 words are included in the MLC, rising up to 76%.

WEST<sup>24</sup>. The MLC comprises 1751 entries out of those studied by West.

LUND<sup>25</sup>. Just the first one hundred words of the Lund corpus have been analysed and they are comprised in the MLC.

Once the items of the MLC have been compared with those of the studies above, and notwithstanding the already mentioned differences, we find that the MLC accounts for about 80% of the entries in them. Then, to know which items are present in all the other studies we filter the MLC, which becomes MLC1 and which comprises 485 words. If the entries of Basic English are not included to filter the MLC, then it contains 1037 words which we refer to as MLC2<sup>26</sup>.

RANGES	MLC	MLC2
R4 - R6	1.040	129
R7 - R11	925	359
R12- R16	743	549
Total	2.718	1.037

If we compare the distribution of entries in MLC y MLC2 (see table above) we notice that, after the filter, the higher the ranges are, the greater the concentration becomes, and the opposite is also satisfied, which seems to prove that range must be considered a valid criterium for selection and classification purposes.

#### 4. Conclusions

1. The results show that the vocabulary of some coursebooks could be considered quantitatively insufficient and not very carefully selected.
2. The MLC has been validated and RANGE has proven to be a valid criterium for the selection of vocabulary.
3. The words making up the MLC can be used to rate the vocabulary of the coursebooks, with respect to range<sup>27</sup>, in two ways:
  - a. By studying the distribution of the entries of each series with respect to range, we consider that the ideal series should accomplish the following conditions:

$$\boxed{RR4 > RR3 > RR2 > RR1}$$

which can be read as follows: the number of higher range items should be greater than those of lower range.

VOCABULARY OF ENGLISH COURSEBOOKS

- b. By comparing the distribution of the entries of the different series with respect to range, we consider that the best series would be the one closer to the following formula:

$$\boxed{\text{RR4} + \text{RR3} + \text{RR2} \parallel}$$

which can be understood like this: greater number of RR4 items plus greater number of RR3 plus greater number of RR2.

The ordering of the series could be different depending on whether the percentage items/TLS is calculated with all the entries of each series, or only with those words making up the MLC, as you can see in the next chart:

ORDERING			
ABSOLUTE		RELATIVE	
Series	%	Series	%
A	42.95	A	22.11
E	33.72	B	20.14
B	33.09	E	19.74
N	27.56	N	18.41
M	25.56	L	18.39
*L	25.18	U	18.31
P	24.58	W	17.76
U	24.09	O	15.94
V	22.58	V	15.83
W	22.20	T	15.36
T	20.00	P	14.16
*O	19.45	F	12.53
*F	15.71	M	12.52
R	14.27	R	10.82
*S	11.71	S	9.71
*D	11.50	D	9.16

If we compare both ordered lists, we realize that the selection of entries on the right (qualitative aspect) is not necessarily conditioned by the number of entries on the left (quantitative aspect), being the series referred to as O the most obvious case. This fact denotes that some series have in common some basic criteria in the vocabulary selection

while others seem not to have followed any.

## Notes

- 1 «Different books which offer themselves as covering similar ground show widely different treatment of vocabulary» [Sinclair & Renouf, 1988:142]
- 2 Renouf (1984) studied nine level-one coursebooks [Sinclair & Renouf, 1988:142]
- 3 At the moment of writing this paper, only some of them are still being used.
- 4 Text-book author(s), titles, publishers, & series identifier.  
 ABBS, B. & I. FREEBAIRN, *Opening Strategies & Building Strategies*. Longman (S). ALEXANDER, L.G., *Follow me 1/2*. Longman (F). BLUNDELL, J. *English Visa 1/2/3*. OUP (V). CAPELLE, G. & D. GIRARD, *It's up to you 1/2/3*. SGEL (U). COLES, M. & B. LORD, *Starting out, Getting on & Open Road\**. OUP (O). EASTWOOD, J. et al., *Network 1/2/3*. OUP (W). ELTDO., *Teens 1/2/3*. Omnivox (T). ESTRANY, M., *Inglés 1/2/3*. Anaya (A). FENTE, R. & E. WULFF, *English Course 1/2/3*. Edelvives (E). FOWLER, W.S. et al., *New Incentive 1/2/3*. Nelson (N). HARTLEY, B. & P. VINEY, *Streamline Departures, Streamline Connections & Streamline Destinations\**. OUP (L). LUTTIKHUIZEN, F. et al., *Ready, Steady & Go*. Vicens Vives (R). PUJALS, M. E. et al., *Inglés 1/2/3*. Magisterio Español (M). RAMSEY, R.M., *Inglés 1/2/3\**. Teide (D). WARSHAWSKY, D. & D. R. BYRD, *Spectrum 1/2/3*. Regents (P). WULFF, E. & N. McLAREN, *England 1/2/3*. Bruño (B).
- 5 As a rule, all the words have been picked up, excepting proper nouns.
- 6 The different lexical items of homographs have been grouped under the same entry since some series do not supply the grammatical class or the meaning in the vocabulary lists. The sorting and counting of English homographs involve a great deal of difficulty as it was pointed out by W. N. Francis (1982:21), and by Bradley (1983:2).
- 7 '1' to refer to Level-1 textbooks (TB1) used in «Primero de BUP».
- '2' to refer to Level-2 textbooks (TB2) used in «Segundo de BUP».
- '3' to refer to Level-3 textbooks (TB3) used in «Tercero de BUP».
- 8 The values are: 1 for nouns; 2 for verbs; 3 for adjectives; 4 for adverbs; 5 for function words, & 6 for phrases.
- 9 The values range from 1 (bottom-range) up to 16 (top-range), so a top-range word appears in the sixteen series, while a bottom-range word just appears in one.
- 10 From the series marked with an asterisk, only TB1 and TB2 were studied.
- 11 For the distinction word/lexical item, see Wallace [1982], Nation [1983:1], Howatt [1983:254] & Carter [1987:26].  
 «... a graded reader scheme, which produces word lists at several levels. The school syllabuses in several countries are similarly organized» [Sinclair & Renouf, 1988: 141]
- 12 Textbooks are firmly thought to be the most important vocabulary input device in the learning process of lexis in L2, notwithstanding the criticisms issued by Brumfit [1980:30], Greenall [1984:14], and Sheldon [1988: 237].
- 13 A similar hyperbolic distribution results when the TLS is divided into subsets containing the entries sorted taking into account the grammatical class.
- 14 The 2933 words with range values 4 to 16 are reduced to 2718 by removing compounds and noun-phrases from it.
- 15 Although there are many vocabulary studies available at the moment, however there do exist differences among them, due mainly to the sources used, the criteria employed, and the updating.
- 16 Dixon lists the 500 most frequent words in the first group, the next 500 in the second, and finally in the third those ranging from 1001 to 2000.
- 17 Eaton classifies the entries according to the number of frequency. As some words were not found inside these groups, the groups 4a/b a 5a/b were also considered.
- 18 Roland Hindmarsh collects 4470 lexical items in the CEL.
- 19 Kucera's work lists 2500 words.
- 20 The Longman dictionary lists the 2000 words which have been used to define the words in the dictionary.

## VOCABULARY OF ENGLISH COURSEBOOKS

- 21 Palmer lists one thousand words.
- 22 Richards & Ogden list 850 words in their famous Basic English.
- 23 In the Threshold Level a distinction «speaking vocabulary» vs. «reading vocabulary» is made.
- 24 West collects 2000 words which are studied in detail, expressing the value of each entry along with the percentage according to function and meaning.
- 25 We regret we could not have used the whole corpus.
- 26 As far as we are entitled to say that, all the MLC2 words should be included in every course-book, though they have to be supplemented.
- 27 For simplicity reasons we have grouped the 16 range values into four which are to be referred to as RR1 (R1+R2+R3+R4), RR2 (R5+R6+R7+R8), RR3 (R9+R10+R11+R12), and RR4 (R13+R14+R15+R16).

## References

- Bradley, D.C. 1983. *Computational Distinction of Vocabulary Type*. Bloomington, Ind.: Indiana University Linguistics Club.
- Brumfit, C.J. 1980. «Seven last slogans». *Modern English Teacher*, 7/1:30-1.
- Dixon, Robert J. 1956. *The 2000 Most Frequently Used Words in English*. New York: Regents.
- Carter, R.A. 1987. *Vocabulary in Language Teaching*. London: Longman.
- Eaton, H.S. 1940. *An English-French-German-Spanish Word Frequency Dictionary*. New York: Dover Publications, Inc.
- Francis, W. N. 1982. «Problems of assembling and computerizing large corpora» in Stig Johanson, *Computer Corpora in English Language Research*. Bergen: Norwegian Centre for the Humanities.
- Greenall, S. 1984. «The coursebook credibility gap». *EFL Gazette* 53/4:14.
- Hindmarsh, R. 1980. *Cambridge English Lexicon*. Cambridge: C.U.P.
- Howatt, A.R.P. 1983. *A History of English Language Teaching*. London: OUP.
- Kučera, H. & W.N. Francis. 1967. *A Computational Analysis of Present-Day American English*. Providence, R.I.: Brown University Press.
- Longman Dictionary of Contemporary English. 1987. Harlow: Longman.
- Nation, I.S.P. 1983. *Teaching and Learning Vocabulary*. Victoria University.
- Palmer, H.E. 1938. *A Grammar of English Words* (rep. 1971). London: Longman.
- Richards, J.A. 1943. *Basic English and its uses*. London: Kegan Paul.
- Sheldon, L. 1988. «Evaluating ELT textbooks and materials». *ELT Journal*, 42/4: 237-46.
- Sinclair, J.M. & A. Renouf. 1988. «Vocabulary, cloze & discourse» in R.A. Carter & M. McCarthy, *Vocabulary and English Teaching*. London: Longman.
- Svartvik, J. & R. Quirk (eds.). 1980. *A corpus of English conversation*. Lund: Lund Series in English.
- Svartvik, J. et al. 1982. *Survey of Spoken English - Report on Research 1975-81*. Lund: Lund Studies in English.
- van Ek, J.A. & L.G. Alexander. 1980. *Threshold Level English*. London: Pergamon.
- Wallace, Michael. 1982. *Teaching Vocabulary*. London: Heinemann.
- West, Michael. 1953. *A General Service List of English Words*. London: Longman.