

Željko Bujas

A "Time" Magazine Vocabulary Study

A. INTRODUCTION

1.1. This paper is intended as a contribution to quantitative research into the English vocabulary, for several years now one of the principal interests of this author (Bujas, 1966, 1968a, 1968b, 1971).

1.2. The object of the present analysis is "magazine English", as exemplified by *Time* magazine, and effort will be concentrated on two sectors: word frequency and etymological composition of the vocabulary.

1.3. Since all procedures in an analysis of this type (counting, tab-keeping, etymology look-ups, tabulation and computation of absolute and relative frequencies) were to be done manually, the present effort had to be quantitatively restricted. Consequently, only one magazine issue was chosen and the word count was limited to 5,000 tokens.

1.4. The choice of *Time* magazine was made for several reasons. Best known of all US "magazines of opinion", the weekly *Time* has for five decades now been in the forefront of linguistic innovation. Its unique style (the "Timesese", in the mouths of the non-*aficionados*) owes considerably to a breathlessly up-to-date vocabulary — one of the best sources of new words for the lexicographer. *Time* is never above coining a word itself, and some of its coinages have become part of modern English vocabulary.

1.5. After a recent issue of *Time* (November 15, 1968)¹ was selected for analysis, a sampling technique was devised. If only

¹ This four-year gap is due to the pressure of other research obligations, after the *Time* issue in question had already been sampled and the samples marked for etymological origin.

5,000 tokens (running words) were to be processed out of an estimated total of 45,000 running words for that issue's total text (minus captions and advertisements), a proper dispersion of samples had to be secured.

1.5.1. To achieve this, 100-word samples were counted off on each textual page, starting (arbitrarily) from the top of the left-hand column. The total of 5,000 words thus obtained from 50 pages covered the full span of *Time's* sections (from "Letters", p. 13, to "Books", p. 70).

1.6. Etymology look-ups were made from the Concise Oxford Dictionary (Fourth Edition, reprinted 1958). The following principles were observed in recording etymological sources:

- a) The ultimate rather than the immediate etyma was recorded (except, obviously, for Romance languages).
- b) Probable etymologies were counted as definite.
- c) Frequency values for words of mixed etymologies were split.
- d) Proper names, figures and abbreviations were disregarded.

B. FREQUENCY ANALYSIS

2.1. The immediate result of this word-count were two lists:

- a) Alphabetical List (with frequencies)
- b) Rank List

2.2. While the first of these lists is basically a look-up aid, the Rank List is a direct statement about the vocabulary structure of a text and, ultimately, of a language. The next table compares the top 20 words in the *Time* Rank List with the corresponding top section of the Brown Corpus Rank List:²

² The Brown Corpus, so called after Brown University (Providence, R. I., U.S.A.) where, between 1963 and 1966, H. Kučera and W. N. Francis assembled and analysed by computer a representative "corpus of present-day edited American English". Its size (1,014,232 running words), range (500 two-thousand-word samples on ten style levels) and simultaneity (all samples from texts published in 1961) make it the most reliable corpus of English for quantitative study of that language's vocabulary.

Table 1

TIME Magazine						Brown Corpus					
Rank	Type	f	% of total	cum f	cum %	Type	f	% of total	cum f	cum %	
1	<i>the</i>	376	7.52	376	7.52	<i>the</i>	69,971	6.90	69,971	6.90	
2	<i>of</i>	178	3.56	554	11.08	<i>of</i>	36,411	3.59	106,382	10.49	
3	<i>and</i>	136	2.72	690	13.80	<i>and</i>	28,852	2.85	135,234	13.34	
4	<i>a</i>	130	2.60	820	16.40	<i>to</i>	26,149	2.58	161,383	15.92	
5	<i>to</i>	113	2.26	933	18.66	<i>a</i>	23,237	2.29	184,620	18.21	
6	<i>in</i>	96	1.92	1,029	20.58	<i>in</i>	21,341	2.10	205,961	20.31	
7	<i>his</i>	70	1.40	1,099	21.98	<i>that</i>	10,595	1.05	216,556	21.36	
8	<i>he</i>	69	1.38	1,137	23.36	<i>is</i>	10,099	0.99	226,655	22.35	
9	<i>that</i>	59	1.18	1,196	24.54	<i>was</i>	9,816	0.97	236,471	23.32	
10	<i>is</i>	55	1.16	1,251	25.64	<i>he</i>	9,543	0.94	246,014	24.26	
11	<i>for</i>	49	0.98	1,300	26.62	<i>for</i>	9,489	0.94	255,503	25.20	
12	<i>by</i>	40	0.80	1,340	27.42	<i>it</i>	8,756	0.86	264,259	26.06	
13	<i>but</i>	36	0.72	1,376	28.14	<i>with</i>	7,289	0.72	271,548	26.78	
14	<i>not</i>	34	0.68	1,410	28.82	<i>as</i>	7,250	0.72	278,798	27.50	
15	<i>was</i>	34	0.68	1,444	29.50	<i>his</i>	6,997	0.69	285,795	28.19	
16	<i>with</i>	34	0.68	1,478	30.18	<i>on</i>	6,742	0.67	292,537	28.86	
17	<i>as</i>	33	0.66	1,511	30.84	<i>be</i>	6,377	0.63	298,914	29.49	
18	<i>from</i>	30	0.60	1,541	31.44	<i>at</i>	5,378	0.53	304,292	30.02	
19	<i>on</i>	30	0.60	1,571	32.04	<i>by</i>	5,305	0.52	309,597	30.54	
20	<i>it</i>	27	0.54	1,598	32.58	<i>I</i>	5,173	0.51	314,770	31.05	

Comment on Table 1:

2.2.1. Close correspondence is obvious between the frequency values (*f*) for most of the first ten word-types from either list.

2.2.2. There is also considerable parallelism between the positions for the first ten types (4 pairs of identical positions, 2 off by one notch, 3 by two notches). Comparing under the broader criterion of presence or absence among the first ten (irrespective of rank position), the correspondence is as high as 18 out of 20 — the unpaired two being *his* from *Time* and *was* from the Brown Corpus.

2.2.3. There is still significant parallelism between the two lists in terms of presence or absence for the second ten types (12 out of 20). Correspondence by rank position, however, becomes rather meaningless from this bracket on, due to the appearance and increasing presence of types with an identical number of tokens which are then assigned rank by the arbitrary criterion of alphabetic order.

2.3. However, for vocabulary study purposes, it is useful to supply a list of both *Time* and (in italics) Brown Corpus items (Kučera & Francis, 1967: 5) between the 21st and 100th rank-list positions.³

	21—30	31—40	41—50	51—60	
1	be	this	which	we	out
2	*states	had	one	him	so
3	have	not	you	been	*said
4	were	are	more	what	what
5	been	but	had	her	up
6	has	from	an	all	its
7	most	or	could	she	about
8	only	have	out	there	into
9	at	an	their	would	than
10	who	they	there	their	them
	61—70	71—80	81—90	91—100	
1	they	can	than	are	must
2	about	only	do	always	**F ⁴
3	before	other	first	does	through
4	can	*new	any	down	back
5	great	some	my	into	*years
6	though	could	now	*nation's	where
7	through	*time	where	*never	much
8	when	*good	will	off	your
9	while	*homes	*apportion	*party	*way
10	over	less	after	*president	*well
		*life	all		
			before		

Even a cursory look at the above list will reveal the following:

2.3.1. Nearly all the items belong to what may be termed "functional words" (articles, auxiliaries, modals, conjunctions, prepositions, pronouns and a few other "grammatical" words).

³ Asterisks mark the unpaired items within the range presented (21st to 100th rank-list positions).

2.3.2. The few "principal" words (nouns, full verbs, adjectives, adverbs) occur much less frequently (20% for *Time*, 8% for the Brown Corpus), and then in the lower rank positions.

2.3.3. The larger proportion of "principal" words for *Time*, as well as their higher rank position, is due to the restricted size of the sample, as a consequence of which the word-type list is shorter and its purely "functional" portion gives out sooner.

2.3.4. Also, for the same reason — small size of the sample — specific vocabulary items (those reflecting topics) occur out of their turn. So, *Time* has *states* and *united* topping its "principal" words (22nd and 42nd positions). Also, since the magazine issue analysed discusses at large Nixon's victory in the presidential election of 1968, other "principal" words included such nouns as: *campaign* (rank position: 75), *election* (76), *party* (99) and *president* (100).

2.4. The entire span of the top 100 word-types in the two word-counts analysed is compared and presented in the following numerical table:

Table 2
Correspondence Survey (Top 100 Rank List Items)

Bracket	Time and B. C. types		Eventually unpaired types (within 1-100 span)	Time cum %	B. C. cum %	Time margin in %
	paired within bracket	unpaired within bracket				
1-10	18	2	—	25.64	24.26	+ 1.38
11-20	12	8	—	32.58	31.05	+ 1.53
21-30	2	18	1	36.14	35.33	+ 0.81
31-40	4	16	3	39.14	38.33	+ 0.81
41-50	2	18	4	41.40	40.66	+ 0.74
51-60	2	18	4	43.28	42.52	+ 0.76
67-70	2	18	7	44.30	44.10	+ 0.80
71-80	—	20	11	46.38	45.40	+ 0.98
81-90	4	16	8	47.72	46.49	+ 1.23
91-100	—	20	14	48.92	47.43	+ 1.49
1-20	30	10	—	32.58	31.05	+ 1.53
21-40	6	34	4	39.14	38.33	+ 0.81
41-60	4	36	8	43.28	42.52	+ 0.76
61-80	2	38	18	46.38	45.40	+ 0.98
81-100	4	36	22	48.92	47.43	+ 1.49
1-50	38	62	8	41.40	40.66	+ 0.74
51-100	8	92	44	48.92	47.43	+ 1.49
1-100	46	154	52	48.92	47.43	+ 1.49

Comment on Table 2:

2.4.1. As was to be expected, the high rate of correspondence in the top two brackets (90% and 60% of items paired within the bracket) drops off sharply for the remaining eight brackets (to 20% or, predominantly, 10%), reaching zero in two of them. The number of items unpaired within the bracket rises, naturally, in inverse proportion.

2.4.2. Eventually unpaired items (for the top 100), completely absent from the first two brackets, rise to 14 in the 91—100 bracket. The uneven distribution of this parameter is best seen in the ratio 8 : 44 (or 15.3% : 84.7%) for the two halves (1—50 and 51—100) of the top 100. Of all unpaired items for the top 100, one-third (33.8%) remain eventually unpaired within the 1—100 rank-list range.

2.4.3. The comparison of cumulative frequencies for brackets reveals no clear trend for the top 100 items in the two lists compared, except for the consistently higher values for *Time* (reaching + 1.49% for the 91—100 bracket, and with the 41—80 bracket lowest: + 0.74%). The trend, however, becomes considerably clearer if the first 1,000 items on the parallel rank lists (by 100) are examined:⁵

2.5. Table 3

	<i>Time</i>		Brown Corpus		<i>Time</i> margin in %
	cum f	cum %	cum f	cum %	
1—100	2,446	48.92	481,055	47.43	+ 1.49
101—200	2,816	56.32	543,495	53.59	+ 2.73
201—300	3,147	62.94	580,235	57.21	+ 5.73
301—400	3,347	66.94	591,903	58.35	+ 8.59
401—500	3,547	70.94	628,081	61.93	+ 9.01
501—600	3,672	73.44	645,833	63.68	+ 7.76
601—700	3,772	75.44	661,153	65.19	+ 10.25
701—800	3,872	77.44	674,735	66.49	+ 10.95
801—900	3,972	79.44	687,011	67.71	+ 11.73
901—1,000	4,072	81.44	698,079	68.83	+ 12.61

⁴ One of 27 coding symbols used in preparing the Brown Corpus for computer processing. All these symbols (plus figures, proper names, foreign words and abbreviations) were counted among the 1,014,232 words of the corpus and presented in the Rank List.

⁵ This, is, naturally due to the much shorter list of word-types in the *Time* sample (1,865), as opposed to the Brown Corpus (50,406). As a result of this, the cumulative relative (in %) value of each rank position in the smaller samples represents a higher proportion of the total word-type list. The disparity between these proportions (i.e. the positive *Time* margin in %) grows as the list unfolds. A good measure of disparity between the two corpora is provided by their type-token ratio. The value of this important parameter decreases as the sample grows. Thus, for *Time* it is 0.373 (1,865 : 5,000) and for the Brown Corpus only 0.050 (50,400 : 1,014,232), or 7.5 times less.

Comment on Table 3:

2.5.1. The trend noticed in Table 2 (of consistently higher values for the parallel rank-list items in *Time*) is now very pronounced, reaching as high as +12.61% in the 901—1,000 bracket. The slight slow-up in the middle (decline in the absolute value of the increase for the 501—600 bracket) reflects the relatively sharp reduction of increase in the *Time* rank list (two-token types cease at the 525th position, one-token types taking over until the end).

2.5.2 The steady rise in the higher value of *Time*'s parallel rank-list positions (which reaches +18.44% at the 1,500th position and +24.54% at the 1,865th, *Time*'s last, position) is due to the fact that from the 525th position on the *Time* rank list has run exclusively through the inventory of types with $f = 1$. We have, in other words, been dealing with the sample portion where tokens equal types, and their cumulative-frequency values start behaving as type cumulative values, leaving behind the same values of the Brown Corpus rank list. The latter list will have to run through nearly 28,000 items before it reaches the 27,864th position from which its $f = 1$ types commence (Kučera & Francis, 1967: 300—307).

2.5.3. All this will be even more evident if we take a look at the following brief comparison of the increase in cum % values for *Time* and the Brown Corpus word-type lists:

	<i>Time</i>	Brown Corpus
top 10 items	0.54%	0.02%
top 100 items	5.36%	0.20%
top 1,000 items	53.62%	1.98%

As demonstrated, while the first 100 *Time* types account for 5.36% of the total number of different words used in the sample, the top 100 types in Brown Corpus cover only one-fifth of one per cent (Kučera & Francis, 1967: 301). Even sharper is the absolute disparity between these values for the top 1,000 word-types: over one-half in *Time* as compared to under two per cent in Brown Corpus (Kučera & Francis, 1967: 306). The stress was on "absolute" since relatively the two values keep their ratio (27.8 : 1) at all list levels.

2.6. While Table 2 dealt with the aspect of correspondence (parallelism), the next survey shows the degree of non-correspondence (shift in brackets) observed:

Table 4

Bracket Shift Survey (top 100 Rank List Items)

Shift in bracket	Total of pairs	Higher-bracket item found in:	
		<i>Time</i>	Brown Corpus
none	23	—	—
1 off	22	13	9
2	13	10	3
3	6	3	3
4	6	2	4
5	1	—	1
6	1	1	—
7	1	—	1
8	—	—	—
9	—	—	—
Total	73	29	: 21

Comment on Table 4:

2.6.1. Complementing the observations on the high rate of correspondence from comments on Table 2, the above survey provides us with some additional parameters. Thus, for instance, to the previously established total of 23 same-bracket word-type pairs we can now add 22 pairs which shift merely to the neighbouring bracket. We shall now find that as many as 55 pairs (or 110 items) may be termed as closely parallel out of 100 pairs (200 items) ideally possible.

2.6.2. Data on the affiliation of the higher-placed item within the pair reveal a slight bias in favour of the *Time* sample which is normal in view of its smaller word-type inventory, the items of which (even on the top-100 list) will tend to bunch up closer to the top than those of the Brown Corpus. A look at the following vocabulary lists should offer additional useful insights, though basic observations can be said to have been made in comments at 2.3.1.—2.3.4.

2.7. Table 5

List of Paired Items (by Bracket Shift)⁶

Shift in bracket	Paired Word-Types Higher-bracket item found in:	
	<i>Time</i>	Brown Corpus
1 off	<i>bad</i> , but, from, him, his, its, more, not, over, some, were, what, where been, before, do, first,	about, an, at, be, one, or, this, was, which
2	has, <i>new</i> , now, out, who, <i>years</i>	said, <i>time</i> , when
3	could, even, though	I, no, up
4	only, than	into, they, will, would
5	—	all
6	most	—
7	—	are
8	—	—
9	—	—
same bracket (no shift)	a, after, and, as, by, can, for, have, he, if, in, is, it, <i>man</i> , of, on, so, that, the, their, there, to, with	

Comment on Table 5:

2.7.1. *Man*, the only "principal" word among the no-shift items, is evidently a coincidence in view of its low rank position (81st).

2.8. Table 6

List of Eventually Unpaired Items⁶

Bracket	Higher-bracket item in:	
	<i>Time</i>	Brown Corpus
1	—	—
2	—	—
3	<i>states</i>	—
4	—	her, she, you
5	<i>because, national united</i>	we
6	<i>church, last, still</i>	them
7	<i>great</i> , though, while	may, other, these, two
8	<i>campaign, election, good,</i> <i>homes, less, life</i>	any, like, my, our, such
9	<i>apportion, people, since</i>	also, did, <i>mad</i> , many, me
10	always, does, down, <i>nation's</i> , never, off, <i>party, president</i>	back, much, must, <i>way</i> , well, <i>year</i>

⁶ Non-functional words are italicised.

Table 5

⁷ The fractional values result from splitting the odd-figure frequencies of mixed-etymology words (cf. 1.6).

C. ETYMOLOGICAL ANALYSIS

3.1. The compilation of the *Time* Rank List was an important first stage in the analysis of the etymological composition of *Time* vocabulary. Once completed, this rank list considerably facilitated the computation of absolute and relative values for each etymology, using both possible approaches: by the token (counting each occurrence) and by the type (counting each new word only once). Here is first the survey by token:

3.2. Table 7

Time Sample Etymological Proportions (by Token)

Etymology	Frequency	Relative <i>f</i> (in %)
AS	3,344	66.88
F	1,069	21.38
L	354	7.08
ON	92.5 ⁷	1.85
?	40	0.80
Gk.	28.5	0.57
It.	17	0.34
Du.	13.5	0.27
Imit. ⁸	12	0.24
LG	8.5	0.17
Sp.	7	0.14
Norw.	4	0.08
Celt.	2	0.04
Arab., Dan., Gipsy, Hindi, Jap., Scand., Sw., Turk.	1	0.02
Total	5,000	100.00

Comment on Table 7:

3.2.1. The above list demonstrates the fallacy of the usual statement about the etymological composition of English: *One half of the English vocabulary is of Romance origin*. Such a statement is obviously true only when words are counted statically, as word-types (e. g. in dictionaries).

3.2.2. As soon as we count every occurrence of a word, i. e. when tokens are counted, we observe very different proportions. Anglo-Saxon items in our *Time* sample, for instance, account for two-thirds of the text (66.88%), while French takes up

⁸ Roberts (1965: 21) credits these ("echoic") words to Anglo-Saxon.

only about one-fifth (21.38%) (Cf. Bujas, 1968a: 97, 98). This becomes even more obvious when etymologies are grouped under larger headings as illustrated in Table 8.

3.2.3. The first three etymologies (Anglo-Saxon, French and Latin) account for over nine-tenths (95.3%) of the text, the remaining eighteen etymologies for only 4.66%.

3.3. Table 8

Grouped Etymological Proportions of *Time* Sample (by Token)

Germanic	Romance	Greek	Other European	Non-European	Imit.	?
AS 3,344 66.88	F 1,069 21.38	28.5 0.57	Celt. 2 0.04	Arab. 1 0.02	12 0.24	40 0.80
ON 92.5 1.85	L 354 7.08			Gipsy 1 0.02		
Du. 13.5 0.27	It. 17 0.34			Hindi 1 0.02		
LG 8.5 0.17	Sp. 7 0.14			Jap. 1 0.02		
Norw. 4 0.08				Turk. 1 0.02		
Dan. 1 0.02						
Scand. 1 0.02						
Sw. 1 0.02						
Total 3,465.5 69.31%	1,447 28.94%	28.5 0.57%	2 0.04%	5 0.10%	12 0.24%	40 0.80%

Comment on Table 8:

3.3.1. Ratios of grouped etymological proportions are now even more in favour of the correct statement about the predominantly Germanic character of the English vocabulary:

Germanic	69.31%
Romance	28.94%
Others	1.75%

3.4. Using the compiled rank list of the *Time* sample, and counting only word-types, we obtain the following distribution:

Table 9

Time Sample Etymological Proportions (by Type)

Etymology	Frequency	Relative <i>f</i> (in %)
F	762	40.86
AS	693	37.37
L	247	13.25
ON	45	2.42
?	32	1.72
Gk.	24.5	1.31
Du.	13.5	1.26
Imit.	11	0.59
It.	11	0.59
LG	7	0.38
Sp.	5	0.27
Norw.	4	0.22
Celt.	2	0.11
Arab., Dan., Gipsy, Hindi, Jap., Scand., Sw., Turk.	1	0.05
Total	1,865	100.00

Comment on Table 9:

3.4.1. The distributions revealed by Table 9 are very different from those in Table 8, and they once again explain the source of the popular conception of the English vocabulary (being over 50% of Romance origin). This idea has obviously originated in counts based on word-types rather than tokens (Cf. Bujas, 1968a: 97, 98; Bujas, 1968b: 129, 130).

3.4.2 The doubling of the share observable in words of French (from 21.38% to 40.86%), Latin (7.08% to 13.25%) and Greek (0.57% to 1.31%) etymologies is due to the low absolute frequency (total of tokens) of these items, encountered more often as the vocabulary expands.

3.4.3. The relative stability of the Old Norse element (1.85% to 2.42%) is also due to its specific character: relatively few items but with a high absolute frequency as a rule (within their 1,000-word bracket). Thus, the Old Norse element preserves a fairly stable average value of relative frequency (about 0.10% per each word-type in the first ten deciles (Cf. Roberts, 1965: 69—80).

3.5. Table 10

Grouped Etymological Proportions of *Time* Sample (by Type)

Germanic	Romance	Greek	Other European	Non-European	Imit.	?
AS 693 37.37	F 762 40.86	24.5 1.31	Celt. 2 0.11	Arab. 1 0.05	11 0.59	32 1.72
ON 45 2.42	L 247 13.25			Gipsy 1 0.05		
Du. 13.5 1.26	It. 11 0.59			Hindi 1 0.05		
LG 7 0.38	Sp. 5 0.27			Jap. 1 0.05		
Norw. 4 0.22				Turk. 1 0.05		
Dan. 1 0.05						
Scand. 1 0.05						
Sw. 1 0.05						
Total 765.5 41.05%	1,025 54.96%	24.5 1.31%	2 0.11%	5 0.25%	11 0.59%	32 1.72%

Comment on Table 10:

3.5.1. The grouped etymological proportions based on word-types provide an even more striking insight into the make-up of the English vocabulary (in terms of entry-lists rather than of current words):

Germanic	41.05%
Romance	54.96%
Others	3.99%

D. CONCLUSION

4.1. A few years ago, this author compiled a simplified tabular survey of etymological proportions of the English vocabulary on nine style levels (Bujas, 1968a: 97). Based on the results of several authors and one own, fairly extensive, word-count, this survey is a reliable standard for evaluating the "difficulty" of vocabulary as a parameter for particular style levels. Reprinted here, the table includes the proportions (in %) arrived at by this *Time* magazine vocabulary analysis (inserted in italics between Newspaper Reports and Technical-Scientific Text):

Table 11

	Anglo-Saxon	Scandinavian	Total Germanic	French and Latin	Total Romance	All Others
Dialogue	85	1.5	88	10.8	11.1	0.9
Essays	82	1.3		16		
Letters	78.3	2.6	81.7	18.5	18.6	0.3
Poetry	75—80			15—20		
History	76	1.5		16		
Fiction	73—75	2.6	75—78	18—19		1—2
Newsp. leaders	72	2.3	75—76	22—23		1—2
Newsp. reports	66.4	0.7	67.5	31.5	31.8	
Magazines	66.9	2.2	69.3	28.5	28.9	1.8
Technical-Scientific	51—68	1—4	53—74	25—45		1—2

This was to be expected in view of *Time's* make-up: political plus scientific and cultural comment and reports. In this manner, the research results reported in this paper have yielded another statement, covering an additional facet of a not unimportant sector in the field of English Vocabulary.

4.2. As to the word-type proportions yielded by this paper (cf. Table 9), they too fit usefully among some of the earlier analyses reported by this author (Bujas, 1968a: 87) or computed by him from Roberts' data (Roberts, 1965: 69, 79—80):

Table 12

	First 500 (Grinstead)	First 500 (Cassidy)	First 1,000 (Roberts)	First 10,000 (Roberts)	17—30,000 (Grinstead)	<i>Time</i>
Anglo-Saxon	81.4	61.7	51.8	31.8	23.7	37.4
Total Germanic			56.0	36.0		42.4
French		30.3	35.8	45.0		40.9
Latin	16.6	2.9	7.9	16.7	45.8	13.3
Greek	0.6		0.1	0.5	18.0	1.3
Scandinavian		1.7	3.9			2.8
Others	1.4	2.9	0.3	1.8	12.4	2.1

4.2.1. This table, though incomplete and uneven in its approach (especially in terms of ultimate/immediate etyma), makes it

possible to observe the trend of proportions for major etymologies of the English vocabulary, notably the inverted ratios of Anglo-Saxon (practically: Total Germanic) and French plus Latin (practically: Total Romance). It also presents clearly the steep rise of Greek in the higher vocabulary reaches. If proportions for the 40,000 and 50,000 levels from the Rank List of the Brown Corpus had been computed, they would certainly confirm the continuation of this trend (to, perhaps, 15% for AS, 55% for F and L, and 25% for Gk.).

4.3. It should be pointed out here that the results of this *Time* magazine word-count and etymological analysis bear out the present author's earlier findings, and prove the basic correctness of his simplified statement about the composition of the English vocabulary (Bujas, 1968a: 98):

About one-half of the less usual (the more "difficult") words of English are of Romance origin — if counted only at their first appearance. If, however, all their occurrences are counted, the share of Romance words does not exceed 25%, except in technical and scientific texts.

4.3.1. *Time*, with its ample coverage of domestic and international politics, its interest in culture, science and economy and, last not least, its commitment to linguistic flamboyance and inventiveness — slightly exceeds the 25% level (to 28.9%), though keeping well under one-third.

4.4. In conclusion, and in defence of similar papers, may this author point out that he is aware that efforts of this type are best categorised as basic research. As such, they are in the opinion of some, open to the somewhat uncharitable label: "of general usefulness".

Of the several possible practical uses of our results let us, therefore, mention only one — directly applicable to teaching purposes: Texts can be graded by difficulty, using their established etymological proportions. Precise quotients can be devised, allotting such texts to various well-defined vocabulary levels. It should not be hard to envisage the usefulness of this for the selection of various types of text — examination papers (dictation, translation, précis-writing), vocabulary expansion material, style evaluation passages, and illustrative texts for Growth of English seminars, to list but a few.

REFERENCES

- Ž. Bujas, 1967
Homografski tipovi u engleskom. Filološki pregled Belgrade, 1—4, 1966, pp. 34—43. (English abstract)
- Ž. Bujas, 1968a
Etimološke proporcije engleskog vokabulara. Analize i analizatori. Filološki pregled, Belgrade, 1—2, 1968, pp. 71—98. (English abstract)
- Ž. Bujas, 1968b
Frequency Lists As Aids in Analysing the Etymological Composition of English. SRAZ, Zagreb, 25—26, 1968, pp. 129—148.
- Ž. Bujas, 1971
Chronological and Area Survey of Foreign Element in the English Vocabulary. SRAZ, Zagreb, 29—32, 1970—71, pp. 131—188.
- H. Kučera & W. N. Francis, 1967
Computational Analysis of Present-Day American English. Brown University Press, Providence, R. I., U. S. A., 1967, 424 pp.
- A. H. Roberts, 1965
A Statistical Linguistic Analysis of American English. Mouton, The Hague, 1965, 437 pp.