(1)   All words with a total frequency of 100 or more were registered and
      the division into the several groups was recorded.
(2)   All instances were recorded of zero frequency in either PROSE or
      POETRY and details were given of the frequency of occurrence in the
      other group. In this case, MEDIEVAL LITERATURE was not taken
      into consideration.

My main intention is to show how far the relative frequency of the 20 most
frequent words in the pieces of poetry mentioned above corresponds with the
average frequency of these words as shown by the overall frequency list. This
could guide us - with all the necessary caution - in assessing the ways of four
poets in the use of, or the deviation from, the average vocabulary.

Furthermore, the occurrence in the several groups has been examined of some
words with a view to define them as mainly "poetic", "prosaic" or "late".

A few other aspects of vocabulary have also been glanced at. This paper can
be no more than a first attempt to draw attention to a complex of problems
it may be rewarding to deal with. If this study helps to promote research in
this field with the aid of Computers, it will have achieved its main purpose.

We shall first consider the relation of frequency of the 20 most frequent
words in the four collections of poems mentioned above to their occurrence
in the overall frequency list. *TABLE 1* shows these 20 words in each collec-
tion, but in the order of frequency according to Diederich's list (6).

2

TABLE 1 :   THE 20 MOST FREQUENT WORDS ACCORDING TO DECREASING
             FREQUENCY IN OVERALL FREQUENCY LIST

Serial Number	Eclogues		Odes		Tibullus		Properce	
1	dico (2)	(I)[7]	dico (2)	(I)	manus (20)	(XIX)	nox (13)	....
2	facio (12)	(III)	video (15)	(IV)	amor (5)	....	malus (15)	....
3	video (3)	(IV)	deus (1)	(VI)	carmen (19)	....	amor (2)	....
4	deus (15)	(VI)	do (4)	(VIII)	puella (1)	....	carmen (7)	....
5	venio (4)	(XI)	dies (7)	(IX)	iuvenis (18)	....	eo (12)	....
6	magnus (11)	(XII)	multus (5)	....	pes (7)	....	amo (9)	....
7	cano (10)	....	magnus (9)	(XII)	aqua (13)	....	verbum (16)	....
8	domus (17)	....	fero (8)	(XVI)	sto (14)	....	os, oris (18)	....
9	amor (6)	....	pater (19)	(XVII)	sacer (17)	....	puella (1)	....
10	puer (7)	....	terra (18)	....	cano (6)	....	caput (17)	....
11	carmen (1)	....	domus (6)	....	fleo (11)	....	aqua (11)	....
12	duco (5)	....	audio (12)	....	tener (2)[8]	....	equus (20)	....
13	cano (10)	....	bonus (16)	....	coma (3)	....	cano (14)	....
14 [9]	mons (20)	....	puer (3)	....	fores (4)	....	fatum (19)	....
15	silva (8)	....	pono (14)	....	durus (8)	....	fleo (8)	....
16	incipio (9)	....	duco (13)	....	herba (9)	....	os, ossis (3)	....
17	pecus (13)	....	mare (11)	....	precor (10)	....	formosus (4).	....
18	canto (14)	....	dulcis (10)	....	sinus (12)	....	durus (5)	....
19	formosus (16)	....	gratus (17)	....	turba (15)	....	mollis (6)	....
20	herba (19)	....	virgo (20)[8]	....	parco (16)	....	lectus (10)	....

The 6 most frequent words in the *Eclogues* are among the 20 most frequent
words in Latin literature as represented by the overall frequency list; 8 others
occur more than 100 times in the overall list, while 5 have a frequency of
less than 100. One of these 20 words - *canto* - appears in poetry only.

In the *Odes* no less than 8 among the 20 are to be found among the 20 most

3

frequent words in the overall list, 9 others have a frequency of over 100, while 2 only occur less than 100 times. One - *virgo* - is found in poetry only.

It is interesting to note how close the *Eclogues* and the *Odes* are in this respect. The *Elegies* of Tibullus and of Properce, much like between themselves, are however different. Only *manus* in Tibullus occurs on the overall frequency list (10) and none of the 20 most frequent words in Properce is there.

Among the 20 words in Tibullus 8 occur less than 100 times in the overall list, while one - *tener* - is used in poetry only. 5 of the 20 words in Properce have a frequency of less than 100 in the overall list.

This seems to show that Horace in his *Odes* and, to a smaller extent, Vergil in the *Eclogues*, prefer words frequent in the literary language, while Tibullus and Properce employ words less frequent (11).

The next Table shows two frequency lists prepared on the base of the overall frequency list : *LIST A* contains all word categories, while *LIST B* comprises Substantives, Adjectives and Verbs only, to make possible a comparison with the lists given in *TABLE 1.*

4

TABLE 2 :   LIST OF MOST FREQUENT WORDS IN LATIN LITERATURE

LIST A :   All Words (Frequencies in brackets)

1	*-que*	6771	15	*suus*	1111(M)	29	*tuus*	625(P)
2	*qui/quis*	6553	16	*dico*	1068(M)	30	*e/ex*	624
3	*et*	6412(M)[12]	17	*nec*	1056(P)	31	*se*	610
4	*sum*	4853	18	*de*	935(M)	32	*deus*	600
5	*in*	3481(M)	19	*sed*	924	33	*te*	591(P)
6	*is*	1998(M)	20	*ipse*	876	34	*me*	590
7	*hic*[13]	1995	21	*si*	825	35	*res*	578(PR)
8	*non*	1827	22	*per*	820	36	*iam*	572(P)
9	*cum*[14]	1567(M)	23	*possum*	789	37	*do*	566(P)
10	*ad*	1563(M)	24	*facio*	785(M)	38	*atque*	565
11	*ille*	1461	25	*video*	768	39	*aut*	544
12	*ut*	1450	26	*meus*	756(M)	40	*alius*	536(M)
13	*omnis*	1184	27	*habeo*	654(M)			
14	*a/ab*	1128	28	*mihi*	626(P)			

LIST B :   Substantives, Adjectives, Verbs

1	*dico*	V (1) , H (1) [15]	11	*venio*	V (5)
2	*possum*		12	*magnus*	V (6) , H (7)
3	*facio*	V (2)	13	*homo*	
4	*video*	V (3)	14	*tempus*	
5	*habeo*		15	*primus*	
6	*deus*	V (4) , H (1)	16	*fero*	H (8)
7	*res*		17	*pater*	
8	*do*		18	*dominus*	
9	*dies*		19	*manus*	T (20)
10	*rex*		20	*modus*	

With regard to grammatical categories the list is made up as follows :

Substantives	10
Adjectives	2
Verbs	8

5

The most frequent words, apart from indeclinables and pronouns, are verbs expressing basic actions : *dico, possum, facio, video, habeo.* The most frequent substantive is *deus* (16); next come *res, dies, rex.*

The verb *dico* which heads this list, is second in both the *Eclogues* and the *Odes.*

So far we have been dealing with words of higher frequency. We shall now consider another part of the word list, namely those words which are missing altogether in one of the two main groups, prose or poetry. The third group, medieval literature, will be excluded.

With a certain measure of justification we can classify as "poetic" or "prosaic", respectively, a word which occurs in one of these groups only (17).

TABLE 3 shows details with regard to frequency range, number of words and number of occurrences. Proper Names are included (18).

TABLE 3 :   *WORDS OCCURRING IN ONE GROUP ONLY*

a) *Poetry - (Prose Frequency : Nil)*

Frequency	N° of Words	N° of Occurrences	Remarks
86 - 64	4	288	
53 - 31	13	514	
30 - 21	25	563	
20 - 16	26	538	
15 - 10	105	1271	
9 - 1	540	2758	
Total	713	5932	

6

b) Prose - (Poetry Frequency : Nil)

Frequency	N° of Words	N° of Occurrences	Remarks
32 - 21	9	237	
20 - 15	9	156	
14 - 10	24	283	
9 - 1	371	1196	
Total	413	1872	
GRAND TOTAL	1126	7804	

There is a rather striking difference between the two groups : Words occurring in poetry only and missing in prose are more in quantity in number of occurrences and in range of frequencies. What is the meaning of the fact that a much larger number of words are excluded from prose and used in poetry only, while a comparatively small number of words is used in prose only, but not in poetry ? First we have to bear in mind that the list is based on two selections comprising different quantities of words. (Prose 50.000 - Poetry 75.000). It should therefore not be surprising that the numbers of poetry are larger than those for prose, but this cannot explain all these differences. The exclusion of certain words from poetry may certainly be due to their sound structure which precludes their occurrence in certain metres (19).

But, as the following Table shows, the meaning of most of these words in addition to their sound form, quite explains their exclusion from poetry.

7

TABLE 4 :   *WORDS USED IN PROSE ONLY (Frequency 10 and more)*

> *memoria* (32); *multitudo, existimo* (30); *magnitudo* (29); *dignitas,
> interficio* (25); *voluntas* (23); *superior* (22); *imperator* (21); *consue-
> tudo, iudicium, Marcus* (19); *militaris, universus* (18); *equitatus* (17);
> *excito* (16); *magistratus, opinio* (15); *aliquando, animadverto, familia*
> (14); *conicio, omitto* [20], *perterreo, quomodo* (13); *antea, castellum,
> Hispania, incredibilis, legatus, quies, valetudo* (12); *avunculus, centu-
> rio, familiaris, praesum, utilitas* (11); *auctoritas, celeritas, consularis,
> contio, saburra* (10).

Of the words occurring in poetry only a few examples will suffice.
The most frequent of these "poetic" words is *unda* (86), next comes
tener (70) and *vates* (68), followed by *virgo* (64) and *Phoebus* (53).
It is not surprising to find here words like : *aequor* (46), *rosa* (35),
ales (32), *aether* (31), *antrum* (23), *caeruleus* (23), but *cras* (31),
aratrum (16), *patrius* (17) may be there by chance only.

Interesting is *nil* (36) which is not found in prose in Diederich's list,
while *nihil* has 128 occurrences in prose, 82 in medieval literature,
but *32* only in prose.

To conclude this study, I should like to quote a few examples of
characterization of words by means of their frequency in one of the
three groups : PROSE, POETRY, MEDIEVAL LITERATURE. Only
a very small selection can be given here, but this seems to be the most
promising aspect of this kind of research and it is in this direction that
the application of the computer could yield interesting results. In order
to exclude as far as possible sources of uncertainty only such examples
will be mentioned in the following Table, where divergences in frequency

8

in the various groups are so great as to make up for inaccuracies in counting and for possible lack of representation of Diederich's selections.

TABLE 5 :   *CHARACTERIZATION OF WORDS BY MEANS OF FREQUENCY*

a) *Prosaic*

| Word | Frequency | | | Remarks |
	Prose	Poetry	Medieval	
aliquis	117	27	64	
apud	61	5	50	
autem	101	7	304	
civitas	46	2	68	
enim	166	37	166	
etiam	182	95	113	
homo	149	72	175	
hostis	104	35	34	
is	731	89	1178	
neque	177	88	26	compare *nec* !
nihil	128	32	82	compare *nil* !
publicus	101	13	16	
quasi	46	6	73	
res	321	149	108	
virtus	75	38	35	
itaque	68	3	54	

b) *Poetic*

| Word | Frequency | | | Remarks |
	Prose	Poetry	Medieval	
aqua	18	80	52	
caelum	19	162	58	
cano	4	99	15	
carmen	12	183	29	
cura	37	99	42	
deus	75	231	294	
do	90	277	199	
eo, ire	18	115	70	

9

TABLE 5 - b) (continued)

Word	Frequency Prose	Poetry	Medieval	Remarks
fatum	5	108	4	
fero	67	228	72	
ferus	14	85	10	
fleo	11	70	22	
gravis	37	78	33	
ignis	23	113	38	
mens	18	101	56	
miser	18	74	31	
mors	59	110	76	
nec	194	651	211	compare *neque* !
o	9	143	66	
puella	1	123	50	
puer	19	110	114	
rapio	9	65	28	
sacer	13	101	25	compare *sanctus* !
semper	29	109	60	
vinco	42	110	23	
vir	75	140	85	
vita	55	150	77	
vivo	40	90	45	

c) Frequency increasing towards medieval Literature

Word	Frequency Prose	Poetry	Medieval	Remarks
audio	59	45	183	
bibo	2	24	76	
de	176	231	622	
dominus	30	43	267	
et	1278	1900	3234	
filius	20	11	199	
gratia	29	22	74	
ibi	32	13	102	
invenio	25	19	106	
magister	6	20	74	
post	43	54	116	

10

TABLE 5 - c) (continued)

| Word | Frequency | | | Remarks |
	Prose	Poetry	Medieval	
rex	60	67	367	
sanctus	5	40	139	compare *sacer* !
sicut	18	4	117	
volo, velle	98	93	163	

The data shown in Part a) of *TABLE 5* are necessarily the most reliable; this on account of the fact that the number of words of prose passages is smaller than that of the poetic passages. In Part b) I should like to mention *deus,* one of the most frequent nouns; it seems to occur less often in prose, but understandably its frequency increases in Medieval Literature. Worth mentioning are also *fatum* and - strangely - *ignis*; *miser, o* and *puella* do not surprise us.

Three pairs of words were pointed out especially ! While *neque* and *nihil* are mainly prosaic, *nec* and *nil* are used predominantly in poetry. *Sacer* seems to be essentially a poetic word, but less frequent in Medieval Literature, while *sanctus,* although occurring in Poetry, is used mainly in Medieval Literature.

Part c) of *TABLE 5* lists a few words the frequency of which seems to increase very much in later and medieval literature; while this steep increase is easily explained in words like *dominus, filius, sanctus* on the one hand and in *de* - instead of the simple genitive - as well as in *et* (increase in parataxis) on the other, I have no explanation whatsoever in other cases. Why are *audio, bibo, invenio* so much more frequent in later literature ? This may be due to chance. Until further examinations are carried out,

11

preferably with the aid of computers, no answer can be given.

At this point I should like to summarize the results of this study, but before doing so I shall again draw attention to the fact that a research of this kind cannot claim accuracy and the measure of reliability of its results cannot be assessed exactly.

1) While Vergil in his *Eclogues* and Horace in his *Odes* use most frequently words which are frequent also in the average language of the literature, Tibullus and Properce prefer words less frequent.

2) The *Eclogues* and the *Odes* are similar in this respect but are different from the *Elegies*.

3) Among the most frequent words are the conjunctions *-que* and *et*, the pronouns *qui/quis, is* and *ille*, the prepositions *in* and *ad* as well as *sum*. The most frequent verb is *dico*, the most frequent noun is *deus*.

4) More than 770 words in about 6000 occurrences are found in poetry, but are missing in prose; the number of words found in prose only, but not in poetry, is much smaller : some 400 in less than 2000 occurrences.

5) On account of the fact that a word is not occurring in one of these two groups or on account of its prevalence in one of them, words may be characterized as "poetic" or "prosaic". This characterization can be carried out more thoroughly with the aid of a computer.

6) There are certain words which show a marked increase in frequency in later literature; such increases may partly be explained by the

12

content and meaning of the words, but in other cases may be due to chance.

7) It seems that research of this kind could be promoted by the use of computers and this study should serve as a stimulus to further studies, but first of all a frequency list should be prepared.

TEL AVIV University   E.D. KOLLMANN
December, 1973

13

NOTES

(1) *A Study of the vocabulary of Vergil's Eclogues,* in REVUE, 1973, n° 3, pp. 1-24.

(2) P.B. DIEDERICH, *The Frequency of Latin Words and their Endings,* Thesis, University of Chicago Press, 1939. Cf.
L. DELATTE, *Key-words and poetic themes in Propertius and Tibullus,* in REVUE, 1967, n° 3, pp. 31-79, especially p. 32. Delatte mentions also M. MATHY, *Vocabulaire de base du latin,* Paris, 1952, but prefers Diederich's list.

(3) This list is based on three anthologies : The Oxford Book of Latin Verse, Avery's Latin Prose Literature and Beeson's Primer of Medieval Literature. These selections comprise passages of all important authors from Ennius to Erasmus, but - unfortunately for our research - Plautus and Terence are excluded. This is indeed a very serious setbeck.

(4) About 80 poets are represented as well as 29 anonymous collections of poems, altogether *75,323 words.* The prose selections include, according to Diederich, 5 - 40 pages of everyone of the following : Cato, Varro, Cicero, Caesar, Sallustius, Nepos, Livy, Seneca Rhetor, Seneca Philosophus, Velleius Paterculus, Curtius, Petronius, Quintilianus, Tacitus, Plinius, Suetonius, altogether *49,363 words.* The anthology of medieval literature includes prose and poetry from the fall

14

of Rome to the Renaissance, a total of *77,472 words*. The total numbers of words included is *202,158*, the total number of authors exceeds *200.*

Apart from the fact that Comedy is not represented, there are some other shortcomings; in medieval literature there is no separation of prose and poetry; furthermore, sometimes different words are put together, for instance *cum* (preposition *and* conjunction), *hic* (pronoun *and* adverb). In spite of this the list is usable.

(5)   Among the total of 202,158 words there are 7771 Proper Names, about 4 % in the average. Diederich tried to check the reliability of his list. (*Op. cit.,* p. 12). In order to prove that "the frequency of these words in our three anthologies would correspond closely with their frequency in a random sampling of Latin authors" he selected 10 passages of 100 words each which had not been included in the anthologies used. The results of this examination tallied almost exactly with those obtained from the anthologies.

(6)   For the original lists see my *Study,* in REVUE 1973, n° 3, p. 8.

(7)   The Roman numbers mark the relative frequency according to Diederich's list; see below, *TABLE 2, List B,* p. 5.

(8)   Not found in prose in Diederich's list.

(9)   Under this line are words which occur in Diederich's list less than 100 times.

(10)   See below, *TABLE 2, List B*; *manus* has about the same overall
        frequency (19) as in Tibullus (20).

(11)   Out of the most frequent words in *List B* there are in

Eclogues	Odes	Tibullus	Properce
6	8	1	0

(12)   The letter in brackets shows the predominant literary genre, if the
        word occurs in one of them in more than 50 % of the total occur-
        rences.
        P = Poetry, PR = Prose, M = Medieval Literature.

(13)   These data cannot be used because they contain both pronoun and
        adverb.

(14)   See Note 13. The data contain preposition and conjunction.

(15)   V = Vergil's Eclogues, H = Horace's Odes, T = Tibullus. The num-
        bers mark the relative frequency.

(16)   *deus* is the most frequent word in the *Odes*.

(17)   Although the number of occurrences is comparatively small - about
        8000 only out of more than 200,000, it can nevertheless be consi-
        dered as representative because it contains all cases of zero frequency
        in one of the two groups.

(18)   In poetry only occur 132 names (1102 occurrences), in prose only
        26 names (171 occurrences). This seems to strengthen the assumption

16

by the present author that Proper Names are rather important in Poetry. Compare *A Study of Proper Names in Vergil's Eclogues,* to be published shortly in CLASSICAL WORLD.

(19)   The fact that Plautus and Terence are excluded may also blur the results.

(20)   But see Horace, *Ars Poetica* 44 : "praesens in tempus omittat.".

17