

A statistical comparison of vocabulary in the *Pervigilium Veneris* and the eclogues of Vergil

Dietmar NAJOCK

The main motives of the *Pervigilium Veneris*, love and nature, play an important part in the Eclogues, too, although other motives – such as song and the longing for peace – are equally important there (1). The dates of composition differ probably by about four centuries (2), and the difference of genre is obvious (3), but a comparative vocabulary study may yield no result or unspecific results, if the texts compared are either too similar or too different.

The data

Frequency lists were available for both texts, they could not be used, however, in the given form. Our own lists for the *Pervigilium Veneris* (4) were modified according to the new edition of L. Catlow (2). His arguments on special readings are rather convincing, except perhaps for *nudae* in verse 22; this reading, however, is at least not worse than *ut udae* or *totae*, so that Catlow's text was followed throughout. For the single Eclogues new frequency lists were extracted from the word index of Lecrompe (5), who gives only the overall frequencies. Homograph distinction – Lecrompe used an unpractical system of numbers – was made explicit, and word classes were assigned to all lemmata. Finally the lemmatization principles had to be harmonized in some cases.

Vocabulary richness

Vocabulary sizes were compared by means of a standardized random variable Z calculated on the hypothesis that a given short text is extracted at random from a long text – here the collection of the Eclogues as a whole. The variable

$$Z = \frac{V - E(V)}{\sqrt{\text{Var}(V)}} ,$$

where V stands for the vocabulary size of the given text, can be regarded to be approximately distributed as $N(0,1)$. $E(V)$ was calculated as suggested by Ch. Muller (6), i.e.

$$E(V) = v - \sum_i \left(1 - \frac{m}{n}\right)^i v_i ,$$

where m and n are the text-lengths of the short and the long text respectively, v the vocabulary size of the long text, and v_i the number of words (lemmata) occurring exactly i times in the long text. In our case we have

| | | | | | | | | | | | | | | | | |
|--------|-----|-----|-----|----|----|----|----|----|----|----|----|-----|-----|-----|----|----|
| $i:$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| $v_i:$ | 826 | 286 | 137 | 93 | 58 | 37 | 33 | 18 | 16 | 10 | 12 | 5 | 9 | 2 | 11 | 4 |
| $i:$ | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 29 | 31 | 33 | 35 | 38 |
| $v_i:$ | 3 | 2 | 1 | 2 | 4 | 3 | 2 | 2 | 2 | 2 | 1 | 3 | 1 | 1 | 1 | 1 |
| $i:$ | 40 | 45 | 47 | 49 | 52 | 54 | 61 | 64 | 69 | 73 | 96 | 107 | 111 | 140 | | |
| $v_i:$ | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | |

$\text{Var}(V)$ was calculated as described in an earlier article (7), i.e.

$$\text{Var}(V) = \sum_i \left(\frac{n-m}{n}\right)^i \left\{1 - \left(\frac{n-m}{n}\right)^i\right\} v_i .$$

This formula for the variance can only be used reasonably, if the short given text is considerably shorter than the long reference text. Now the *Pervigilium Veneris* and the single Eclogues are about 1/10 of the united Eclogues, and bootstrap experiments (8) have shown that the above formula indeed yields a sufficient approximation of the variance in this region.

A first run showed extremely low Z -values for the *Pervigilium Veneris* (- 6.05) and for Eclogue 8 (- 5.28), obviously caused by word repetition in the refrains of these poems. Removal of the refrains (9) gave the following results :

Table 1

| | TL | V | E(V) | $\sqrt{\text{Var}(V)}$ | Z | Type |
|-----|-----|-----|-------|------------------------|--------|---------------------------------|
| E02 | 499 | 333 | 336.1 | 14.49 | - 0.22 | monologue |
| E06 | 571 | 367 | 373.1 | 15.00 | - 0.41 | monologue |
| E01 | 551 | 356 | 363.0 | 14.87 | - 0.47 | dialogue |
| E04 | 418 | 276 | 292.4 | 13.80 | - 1.19 | monologue |
| E05 | 609 | 372 | 391.9 | 15.24 | - 1.31 | dialogue/certatio (two songs) |
| E07 | 471 | 300 | 321.3 | 14.27 | - 1.49 | certatio (alternating) |
| E09 | 472 | 296 | 321.8 | 14.28 | - 1.81 | dialogue |
| E08 | 651 | 380 | 412.3 | 15.49 | - 2.08 | certatio (two songs) |
| E03 | 795 | 443 | 478.6 | 16.18 | - 2.20 | dialogue/certatio (alternating) |
| E10 | 530 | 318 | 352.3 | 14.72 | - 2.33 | monologue |
| PER | 565 | 324 | 370.1 | 14.96 | - 3.08 | |

(TL is the text length in words)

We see that the *Pervigilium Veneris* has a poorer vocabulary than all Eclogues. The significance level of this deviation differs from that of the poorest Eclogue by a factor of about 10. On the other hand a great variation of vocabulary richness can be observed within the collection of the Eclogues. The significance levels of the Z-values of the poorest and of the richest Eclogue differ by a factor of about 20. This variation seems to be mainly an effect of difference in form: if the Eclogues are roughly classified as monologue, dialogue or *certatio* (10), and if they are arranged according to decreasing vocabulary richness, as in Table 1, the monologues, with the richest vocabulary, gather at the top of the list, the dialogues occupy a medium region, with a certain overlap with both monologues and *certationes*, and the *certationes*, with the poorest vocabulary, gather at the end of the list. The only striking exception is Eclogue 10, a monologue, which should be expected just at the other end of the list. One reason for the vocabulary poorness of Eclogue 10 is the relatively frequent occurrence of single words, such as *Gallus* in the initial and ending parts of the poem, in verses 72/73 even with anadiplosis (... *Gallo, Gallo* ...) or *ego* - which occurs with high frequencies in *certationes* and laments for rejected love - in the embedded verses of Gallus; furthermore of a few prosaic words (11) such as *neque, sive, inquit, esse, and venire*. Another reason is the frequent repetition of words in rhetorical figures,

especially in the numerous, perhaps too numerous cases of anaphora, but also in some cases of antistrophe (e.g. vv. 75/76 ... *gravis ... umbra, ... gravis umbra; ... umbrae*) and other forms of repetition. It would be interesting to find out, why Vergil emphasized these features in Eclogue 10, but this question is beyond the scope of the present article; the date of composition alone seems to be an insufficient explanation (12).

The vocabulary poorness of the *certationes* results from their improvisatory features, which include the exchange of keywords between the competitors. In Eclogue 3 even a complete verse of Menalcas (v. 43) is repeated by Damoetas (v. 47). Another reason is the frequent occurrence of personal pronouns of the first person.

Although the *Pervigilium Veneris* is rather free from improvisatory and prosaic elements, it is, as mentioned above, considerably poorer in vocabulary than Eclogue 10 and the *certationes*. This results from the fact that the poet, or rather the poetess (13), uses a number of favorite words closely related to the contents and that Vergil is obviously superior in the art of variation. The words favored in the *Pervigilium* will be given in the section on characteristic vocabulary.

It should be noted that collections such as the book of the Eclogues, on account of thematic variation, tend to have relatively richer vocabularies than the single texts of the collections. This effect has also been observed in the Plays of Shakespeare (14). In our case even the richest single text, Ecl. 2, has a poorer vocabulary than the collection of the Eclogues as a whole. This observation supports the view of Vergil as a master of variation in the book of the Eclogues.

Word class distribution

The frequencies of word classes in the *Pervigilium Veneris* and in the single Eclogues are given in table 2. Altogether eight word classes are differentiated : nouns (SU), adjectives (AJ), pronouns (PN), verbs (VB), adverbs (AV), conjunctions (CJ), prepositions (PP) and interjections (IJ). The X^2 test for homogeneity shows that, with error probability 0.0000 (rounded), the differences in word class distribution cannot be regarded as a chance effect ($X^2 = 167.144$, 70 degrees of freedom). The

Table 2
Word-class-distribution

| | SU | AJ | PN | VB | AV | CJ | PP | IJ |
|-----|-----|----|-----|-----|----|----|----|----|
| PER | 212 | 61 | 46 | 137 | 28 | 42 | 36 | 3 |
| E01 | 171 | 70 | 72 | 110 | 67 | 35 | 21 | 5 |
| E02 | 151 | 67 | 70 | 105 | 43 | 41 | 12 | 10 |
| E03 | 227 | 61 | 136 | 189 | 70 | 77 | 28 | 7 |
| E04 | 154 | 56 | 41 | 88 | 36 | 36 | 5 | 2 |
| E05 | 220 | 67 | 84 | 105 | 43 | 67 | 20 | 3 |
| E06 | 191 | 63 | 62 | 131 | 48 | 57 | 15 | 4 |
| E07 | 170 | 62 | 55 | 84 | 31 | 47 | 21 | 1 |
| E08 | 214 | 74 | 89 | 139 | 40 | 57 | 26 | 3 |
| E09 | 127 | 43 | 64 | 116 | 58 | 41 | 17 | 6 |
| E10 | 168 | 62 | 75 | 107 | 39 | 54 | 21 | 4 |

contributions of the single table entries to the X^2 value are given in Table 3, multiplied by 10, rounded off to integers and preceded by the minus symbol, if and only if the observed frequency is below the expected frequency.

Table 3
Word-class-distribution : contributions to X^2

| | SU | AJ | PN | VB | AV | CJ | PP | IJ |
|-----|-------------|-------------|--------------|-------------|-------------|-------------|-------------|-----------|
| PER | <u>39</u> | - 1 | - <u>102</u> | 21 | - <u>74</u> | - 17 | <u>117</u> | - 5 |
| E01 | - 5 | 11 | 0 | - 6 | <u>104</u> | - <u>45</u> | 0 | 1 |
| E02 | - 10 | 22 | 4 | - 1 | 0 | - 4 | - 21 | <u>94</u> |
| E03 | - <u>43</u> | - <u>89</u> | <u>105</u> | 20 | 3 | 3 | - 1 | 0 |
| E04 | 21 | 17 | - 33 | - 1 | 0 | - 1 | - <u>69</u> | - 5 |
| E05 | 21 | - 1 | 3 | - <u>50</u> | - 10 | 25 | - 2 | - 7 |
| E06 | 0 | - 1 | - 20 | 6 | 0 | 5 | - 16 | - 1 |
| E07 | 16 | 16 | - 7 | - 29 | - 16 | 4 | 9 | - 20 |
| E08 | 0 | 0 | 3 | 0 | - 31 | - 1 | 3 | - 9 |
| E09 | - <u>50</u> | - 19 | 1 | 22 | <u>95</u> | - 1 | - 1 | 14 |
| E10 | - 2 | 1 | 5 | - 4 | - 5 | 7 | 1 | - 1 |

Contributions significant on the 5 %, 1 % and 0.1 % levels are underlined with one, two or three strokes.

We see that the *Pervigilium Veneris* is characterized by a highly significant preference for prepositions, by a significant avoidance of pronouns and adverbs and by a less evident preference for nouns. In these features the *Pervigilium* differs from all Eclogues. There are, however, considerable differences between the Eclogues, too, a variation which can be assumed as non-fortuitous with very low error probability. The most outstanding peculiarities are the high frequency of pronouns, at the cost of nouns and adjectives, in Ecl. 3 and the high rates of adverbs in Ecl. 1 and Ecl. 9 and of interjections in Ecl. 2; furthermore the low rates of prepositions in Ecl. 4, of verbs in Ecl. 5, of conjunctions in Ecl. 1 and of nouns in Ecl. 9. The low rates are difficult to explain, unless they are balanced, as in Ecl. 3, by high rates of another word class. We must concentrate therefore on the high rates, and fortunately these are much more significant in our case.

We have already mentioned the frequent occurrence of personal pronouns of the first person in *certationes*. In Ecl. 3 these pronouns, mainly *ego*, are complemented by a considerable number of occurrences of *tu*, *ille*, *qui* and *quis*, a feature which may be characteristic of this special type of *certatio*. Similarly the relatively high rate of interjections in the lament of Corydon, in Ecl. 2, needs no explanation. It remains therefore to comment on the high rates of adverbs in Ecl. 1 and Ecl. 9.

The dialogue of Ecl. 1 reflects a happy past, the still present beauty of the beloved countryside, and an uncertain future for Meliboeus. These themes induce high rates of temporal and local adverbs. *Ante* (3 x), *post* (3 x), *umquam* (2 x) and *saepe* (4 x) occur more often than in any other Eclogue, while *quondam* (1 x), *posthac* (1 x) and *semper* (2 x), although occurring only once or twice, do reach the maximum frequency for a single Eclogue. *Hic* (5 x) is also frequent in Ecl. 7 and Ecl. 9, but *hinc* (4 x) occurs in the other Eclogues at most once. While *hic* is used twice by Meliboeus and three times by Tityrus, the four occurrences of *hinc* are all in the part of Meliboeus, who has to leave his homeland. Only the last of these occurrences (v. 64) is directly associated with his own parting. The first (v. 38) connects the underlying notion of loss and absence with his dialogue partner, the second (v. 53) and third (v. 56) are anaphoric complements to *hic* (v. 51) and mean origin rather than separation, origin of future delight for Tityrus. It is remarkable that *hinc* could well be replaced by *hic* in vv. 53

and 56, that *hic* would appear even more normal, and that the strong position as first word of the verse and the repetition in this position give a special weight to *hinc*. Possibly we may assume that the notion of separation and loss, predominant in the thoughts of Meliboeus on a subconscious level, emerges with these two *hinc* on the surface of a contrasting context. If this interpretation is right, the last *hinc* (v. 64) appears as a final non-transferred utterance of this notion after a prelude of three transferred occurrences. Indeed *hinc* should be a keyword for the deeper feeling of Meliboeus, and vv. 64-78, beginning with the heavy *at nos hinc*, finally express these feelings in full clearness and detail. Also the intensive emotion connected with the first *hinc*, as expressed in

... *ipsae te, Tityre, pinus,*
ipsi te fontes, ipsa haec arbusta vocabant,

is perhaps best explained as a reflex of Meliboeus' own situation.

An usual concentration of adverbs is found in Ecl. 1, 11-13 :

M. Non equidem invideo, miror magis : undique totis
usque adeo turbatur agris. en ipse capellas
protinus aeger ago, hanc etiam vix, Tityre, duco.

Here the accumulated adverbs seem to give, on the whole, an impression of vagueness rather than clear lines, a vagueness which might be interpreted as expressing a certain loss of inner stability, an embarrassment of the disrooted (Meliboeus) who is confronted with a luckier person (Tityrus).

Finally Ecl. 1 has five occurrences of *tamen* (vv. 18, 27, 29, 57, 79), while other Eclogues have one or two occurrences at most. The first and the last occurrence in Ecl. 1 refer directly to the bad luck of Meliboeus, the second and third to the long servitude and late liberty of Tityrus. The experience of Tityrus, including *nec spes libertatis erat* (v. 32), may have had a consolatory effect on Meliboeus, who can hardly hope to return home (vv. 67-71) and who is invited by Tityrus (vv. 79-83) to enjoy – if this is possible – at least the last hours in the old beloved environs. It is possible that *tamen* should be understood as another keyword for the psychic situation of Meliboeus.

Ecl. 9 is almost as rich in adverbs as Ecl. 1, and both poems are inspired by the same motive, the loss of land and home. This motive is connected in Ecl. 9 with an atmosphere of restlessness and discontinuity : fragments of a luckier life and

fragments of poems emerge from the past, contrasted by a threatening present time, but also by an – unfulfilled – attempt to begin bucolic singing *hic et nunc*. It is possible that the oscillating emotions increase the rate of adverbs in general, but it is more difficult to find a direct influence of the contents on special adverbs here. Only *hic* (6 x) and *nunc* (4 x) occur with unusual frequency. Of the four *nunc* (vv. 5, 53, 57, 66) the first and the second are connected with the miserable situation of Moeris, the third with the invitation to bucolic singing (by Tityrus), the fourth with the rejection of that invitation and a more realistic view of the present time (again Moeris). *Hic* occurs in two anaphoric triplets (vv. 40/40/41 and 60/61/62), the latter in the rejected invitation just mentioned, the former, preceded and followed by *huc* (v. 39 *huc ades, o Galatea ... v. 43 huc ades*), in a chant fragment inviting the nymph Galatea to leave the restless sea and to move to the bucolic world. As a part of a myth this invitation has an illusory character, and within the mythical song it must appear as an attempt to reach the impossible. It is probably not by chance that the eight occurrences of *hic* and *huc*, all in invitations and all referring to the bucolic world, appear under the shadow of expected and subsequent denial. So it seems possible to understand *hic*, *huc* and *nunc* as keywords leading very directly to the situation and problems of Moeris. Here it is the immediate context of *hic*, *huc* and *nunc* which contains these problems, in Ecl. 1 it was also the very meaning of *hinc* and *tamen*.

The single texts were also compared with respect to the complete word class distributions: the results of pairwise X^2 comparisons are gathered in Table 4, where the right upper triangle contains the X^2 values (multiplied by 10 and rounded off to integers) and the left lower triangle the corresponding error probabilities (multiplied by 10,000 and rounded off to integers). High X^2 values and low error probabilities indicate a high degree of dissimilarity of the distributions compared.

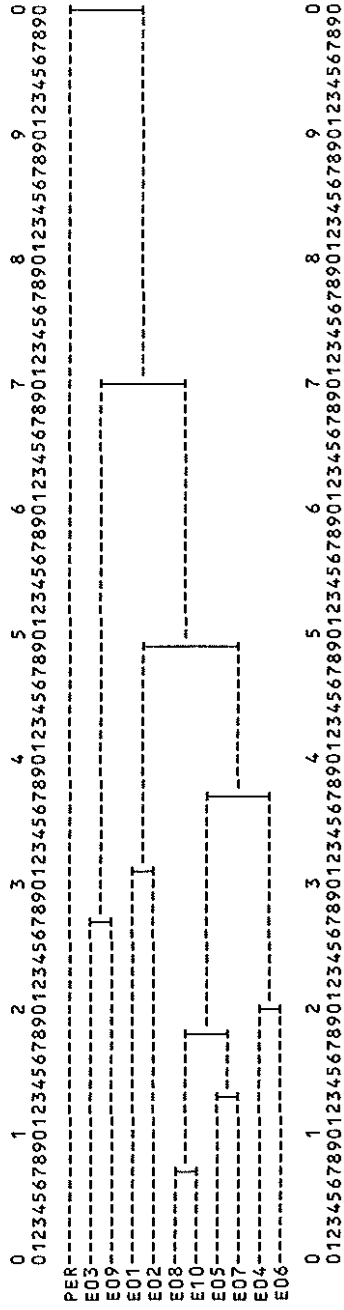
Finally the X^2 values of Table 4 were used as dissimilarity coefficients for a cluster analysis (average-link pair-group method). The resulting dendrogram is given below Table 4.

If the refrains are taken into account, the *Pervigilium Veneris* has a high rate of verbs, but a normal rate of adverbs (4 x *amare*, 2 x *cras* and 1 x *numquam* in the refrain), and Ecl. 8 has a high rate of prepositions (1 x *ab* and 1 x *mecum*

Table 4
PAIRWISE X² COMPARISONS

| | PER | E01 | E02 | E03 | E04 | E05 | E06 | E07 | E08 | E09 | E10 |
|-----|-----|------|------|------|------|------|------|------|------|-----|-----|
| PER | 0 | 346 | 347 | 466 | 240 | 276 | 199 | 151 | 161 | 393 | 220 |
| E01 | 0 | 0 | 87 | 227 | 159 | 201 | 132 | 177 | 163 | 91 | 117 |
| E02 | 0 | 2730 | 0 | 183 | 121 | 153 | 91 | 168 | 121 | 114 | 73 |
| E03 | 0 | 23 | 111 | 0 | 325 | 225 | 168 | 297 | 158 | 77 | 112 |
| E04 | 15 | 264 | 952 | 1 | 0 | 129 | 56 | 121 | 141 | 240 | 139 |
| E05 | 4 | 57 | 321 | 25 | 728 | 0 | 90 | 38 | 61 | 262 | 39 |
| E06 | 63 | 678 | 2432 | 188 | 5774 | 2479 | 0 | 98 | 69 | 123 | 56 |
| E07 | 350 | 137 | 190 | 2 | 972 | 7927 | 1986 | 0 | 51 | 280 | 55 |
| E08 | 245 | 226 | 965 | 274 | 494 | 5243 | 5638 | 6402 | 0 | 198 | 20 |
| E09 | 0 | 2414 | 1200 | 3518 | 15 | 7 | 890 | 4 | 64 | 0 | 133 |
| E10 | 30 | 1089 | 3929 | 1269 | 530 | 7822 | 5853 | 5888 | 9567 | 640 | 0 |

SCALE VALUE 100 = CLUSTER DISTANCE 279



in the refrains) and a low rate of adverbs (adverbs do not occur in the refrains, *domum* being assigned to the noun lemma by Lecrompe). The dendrogram separates Ecl. 8, not the *Pervigilium*, from the other texts in this case.

Characteristic vocabulary

The vocabulary characteristic of the *Pervigilium Veneris* – in comparison with the Eclogues – was determined by means of a method described by Ch. Muller (15). For each noun, adjective and verb with a certain minimum frequency (see below), the quantity

$$Z = \frac{F - E(F)}{\sqrt{\text{Var}(F)}}$$

was calculated, where F is the frequency of the word in the *Pervigilium*. $E(F)$ and $\text{Var}(F)$ were calculated on the hypothesis that the word is uniformly distributed over both the *Pervigilium* and the Eclogues : if t is the total frequency of the word in these texts, p the length of the *Pervigilium* (in words), and n the length of both texts, then

$$E(F) = t \frac{p}{n}$$

and

$$\text{Var}(F) = t \frac{p}{n} \frac{n - p}{n} .$$

As standardized version of the binomially distributed variable F , Z can be assumed to be approximately distributed as $N(0,1)$, if the total frequency t is not too low, say about 5. For the present article also those words have been included which occur at least twice in the *Pervigilium*, since many of these have frequency 0 in the Eclogues, and since the proportion 2:0 seems to result from different intentions of the poets in many cases. Although the approximation of the standard normal distribution is rather rough for such words, the resulting values yield a consistent scaling of the words considered. The critical values of $+1.96$ and -1.96 are justified by the more frequent words in the same lists. Z -values greater than 1.96 indicate words which can hardly be regarded as uniformly distributed over the *Pervigilium* and the Eclogues; with error probability 0.05 or less we can assume that their relatively high frequency in the *Pervigilium* is not a chance effect, i.e.

Table 5
Nouns

| | ECL | PER | Z | | ECL | PER | Z |
|-----------------|-----|-----|--------|-----------|-----|-----|--------|
| CARMEN | 40 | 0 | - 2.01 | DEUS | 16 | 1 | - 0.47 |
| PECUS | 17 | 0 | - 1.31 | MONS | 15 | 1 | - 0.40 |
| HERBA | 15 | 0 | - 1.23 | SILVA | 22 | 2 | - 0.14 |
| OVIS | 15 | 0 | - 1.23 | FONS | 9 | 1 | 0.09 |
| CAPELLA | 13 | 0 | - 1.15 | GREX | 9 | 1 | 0.09 |
| FLUMEN | 11 | 0 | - 1.05 | MUSA | 9 | 1 | 0.09 |
| PASTOR | 11 | 0 | - 1.05 | TERRA | 9 | 1 | 0.09 |
| FRIGUS | 10 | 0 | - 1.00 | AGER | 7 | 1 | 0.33 |
| LAC | 10 | 0 | - 1.00 | ARBOR | 7 | 1 | 0.33 |
| VITIS | 9 | 0 | - 0.95 | CAELUM | 6 | 1 | 0.47 |
| CALAMUS | 8 | 0 | - 0.90 | TAURUS | 6 | 1 | 0.47 |
| HAEDUS | 8 | 0 | - 0.90 | PUER | 22 | 3 | 0.49 |
| LUPUS | 8 | 0 | - 0.90 | ASTRUM | 5 | 1 | 0.64 |
| LAURUS | 7 | 0 | - 0.84 | IGNIS | 5 | 1 | 0.64 |
| MUNUS | 7 | 0 | - 0.84 | POETA | 5 | 1 | 0.64 |
| TEMPUS | 7 | 0 | - 0.84 | MATER | 10 | 2 | 0.90 |
| ANTRUM | 6 | 0 | - 0.78 | AMOR | 22 | 4 | 1.10 |
| CAPER | 6 | 0 | - 0.78 | UMBRA | 15 | 3 | 1.10 |
| CURA | 6 | 0 | - 0.78 | NEMUS | 8 | 2 | 1.19 |
| DIES | 6 | 0 | - 0.78 | NYMPHA | 10 | 3 | 1.74 |
| DOMUS | 6 | 0 | - 0.78 | ORBIS | 5 | 2 | 1.78 |
| FISTULA | 6 | 0 | - 0.78 | LUCUS | 3 | 2 | 2.39 |
| MALUM (A LONGA) | 6 | 0 | - 0.78 | CONIUNX | 2 | 2 | 2.83 |
| PINUS | 6 | 0 | - 0.78 | NODUS | 2 | 2 | 2.83 |
| RUPE | 6 | 0 | - 0.78 | PONTUS | 2 | 2 | 2.83 |
| SIDUS | 6 | 0 | - 0.78 | ARMA | 1 | 2 | 3.45 |
| URBS | 6 | 0 | - 0.78 | CASA | 1 | 2 | 3.45 |
| VERSUS | 6 | 0 | - 0.78 | GEMMA | 1 | 2 | 3.45 |
| AGNUS | 5 | 0 | - 0.71 | NEPOS | 1 | 2 | 3.45 |
| ALTAR | 5 | 0 | - 0.71 | PUDOR | 1 | 2 | 3.45 |
| APER | 5 | 0 | - 0.71 | SINUS | 1 | 2 | 3.45 |
| ARBUSTUM | 5 | 0 | - 0.71 | SPIRITUS | 1 | 2 | 3.45 |
| CANIS | 5 | 0 | - 0.71 | ANNUS | 3 | 3 | 3.47 |
| CAPUT | 5 | 0 | - 0.71 | FLOS | 8 | 5 | 3.66 |
| COLOR | 5 | 0 | - 0.71 | NOX | 5 | 4 | 3.67 |
| CORNU | 5 | 0 | - 0.71 | IMBER | 3 | 4 | 4.40 |
| CORYLUS | 5 | 0 | - 0.71 | RUS | 3 | 4 | 4.40 |
| FAGUS | 5 | 0 | - 0.71 | CATERVA | 0 | 2 | 4.45 |
| HEDERA | 5 | 0 | - 0.71 | CRUOR | 0 | 2 | 4.45 |
| HUMUS | 5 | 0 | - 0.71 | IUS | 0 | 2 | 4.45 |
| IUGUM | 5 | 0 | - 0.71 | NUPTIAE | 0 | 2 | 4.45 |
| IUVENCUS | 5 | 0 | - 0.71 | OS (ORIS) | 0 | 2 | 4.45 |
| LITUS | 5 | 0 | - 0.71 | OSCULUM | 0 | 2 | 4.45 |
| MARE | 5 | 0 | - 0.71 | PAPILLA | 0 | 2 | 4.45 |
| NUMERUS | 5 | 0 | - 0.71 | PURPURA | 0 | 2 | 4.45 |
| POCULUM | 5 | 0 | - 0.71 | SAGITTA | 0 | 2 | 4.45 |
| POMUM | 5 | 0 | - 0.71 | VESTIS | 0 | 2 | 4.45 |
| PRATUM | 5 | 0 | - 0.71 | ALES | 0 | 3 | 5.45 |
| QUERCUS | 5 | 0 | - 0.71 | DIVA | 0 | 3 | 5.45 |
| RIVUS | 5 | 0 | - 0.71 | PUELLA | 4 | 6 | 5.57 |
| SALIX | 5 | 0 | - 0.71 | VIRGO | 2 | 5 | 5.71 |
| UBER | 5 | 0 | - 0.71 | MARITUS | 1 | 5 | 6.30 |
| VIR | 5 | 0 | - 0.71 | VER | 2 | 6 | 6.45 |
| VITULA | 5 | 0 | - 0.71 | | | | |

they may be regarded as positive characteristic of the *Pervigilium* in comparison with the Eclogues. Z-values below - 1.96 indicate the vocabulary negative characteristic of the *Pervigilium* and positive characteristic of the Eclogues. Separate lists of nouns, adjectives and verbs, sorted according to increasing Z-values (last column), are given in the Tables 5, 6 and 7 (the first columns contain the absolute frequencies in the Eclogues and in the *Pervigilium*).

Table 6
Adjectives

| | ECL | PER | Z | | ECL | PER | Z |
|----------|-----|-----|--------|------------|-----|-----|--------|
| FORMOSUS | 16 | 0 | - 1.27 | AUREUS | 4 | 0 | - 0.64 |
| MOLLIS | 13 | 0 | - 1.15 | BONUS | 4 | 0 | - 0.64 |
| TENER | 11 | 0 | - 1.05 | DIVINUS | 4 | 0 | - 0.64 |
| VIRIDIS | 11 | 0 | - 1.05 | EXTER | 4 | 0 | - 0.64 |
| LENTUS | 9 | 0 | - 0.95 | FLOREUS | 4 | 0 | - 0.64 |
| TALIS | 9 | 0 | - 0.95 | FORTUNATUS | 4 | 0 | - 0.64 |
| LONGUS | 8 | 0 | - 0.90 | FRIGIDUS | 4 | 0 | - 0.64 |
| MULTUS | 8 | 0 | - 0.90 | MEDIUS | 4 | 0 | - 0.64 |
| PINGUIS | 8 | 0 | - 0.90 | MISER | 4 | 0 | - 0.64 |
| ALTUS | 7 | 0 | - 0.84 | PALLENS | 4 | 0 | - 0.64 |
| CANDIDUS | 7 | 0 | - 0.84 | QUALIS | 4 | 0 | - 0.64 |
| CRUELIS | 7 | 0 | - 0.84 | SACER | 4 | 0 | - 0.64 |
| DIGNUS | 7 | 0 | - 0.84 | TANTUS | 4 | 0 | - 0.64 |
| AMARUS | 6 | 0 | - 0.78 | OMNIS | 31 | 2 | - 0.62 |
| MALUS | 6 | 0 | - 0.78 | NOVUS | 9 | 1 | 0.09 |
| NIGER | 6 | 0 | - 0.78 | SOLUS | 9 | 1 | 0.09 |
| ALTERNUS | 5 | 0 | - 0.71 | PARVUS | 6 | 1 | 0.47 |
| DULCIS | 5 | 0 | - 0.71 | PRIMUS | 6 | 1 | 0.47 |
| DURUS | 5 | 0 | - 0.71 | TOTUS | 7 | 4 | 3.13 |
| GRAVIS | 5 | 0 | - 0.71 | NUDUS | 3 | 4 | 4.40 |
| INFELIX | 5 | 0 | - 0.71 | CANORUS | 0 | 2 | 4.45 |
| TRISTIS | 5 | 0 | - 0.71 | PUDICUS | 0 | 2 | 4.45 |
| MAGNUS | 19 | 1 | - 0.65 | FERIATUS | 0 | 3 | 5.45 |
| ARGUTUS | 4 | 0 | - 0.64 | MYRTEUS | 0 | 3 | 5.45 |

The most astonishing result is perhaps that only one word, *carmen* (16), is characteristic of the Eclogues, while 12 verbs, 6 adjectives and 32 nouns are characteristic of the *Pervigilium* on the 5 % significance level. On the 20 % level, *carmen* is complemented by *cano* and *pecus*, while the *Pervigilium* has only 3 verbs and 2 nouns in addition. This means that the positive characteristic vocabulary is rather clearly separated from the other words in the *Pervigilium*.

The fact that the Eclogues have almost no characteristic vocabulary may partially be explained by variation, both thematic variation between the single Eclogues and variation in the

Table 7

Verbs

| | ECL | PER | Z | | ECL | PER | Z |
|--------------|-----|-----|--------|--------------|-----|-----|--------|
| CANO | 20 | 0 | - 1.42 | PENDEO | 5 | 0 | - 0.71 |
| FERO | 15 | 0 | - 1.23 | RIDEO | 5 | 0 | - 0.71 |
| AMO | 14 | 0 | - 1.19 | VOCO | 5 | 0 | - 0.71 |
| HABEO | 14 | 0 | - 1.19 | CANTO | 16 | 1 | - 0.47 |
| PASCO | 13 | 0 | - 1.15 | DUCO | 16 | 1 | - 0.47 |
| INCIPIO | 12 | 0 | - 1.10 | VENIO | 26 | 2 | - 0.37 |
| LEGO (- ERE) | 11 | 0 | - 1.05 | POSSUM | 12 | 1 | - 0.18 |
| VIDEO | 27 | 1 | - 1.03 | SPARGO | 7 | 1 | 0.33 |
| SEQUOR | 9 | 0 | - 0.95 | DESINO | 6 | 1 | 0.47 |
| CERTO | 8 | 0 | - 0.90 | DICO (- ERE) | 29 | 4 | 0.59 |
| FUGIO | 8 | 0 | - 0.90 | CEDO | 5 | 1 | 0.64 |
| MEMINI | 8 | 0 | - 0.90 | MITTO | 5 | 1 | 0.64 |
| REFERO | 8 | 0 | - 0.90 | RELINQUO | 5 | 1 | 0.64 |
| AGO | 7 | 0 | - 0.84 | FACIO | 17 | 3 | 0.90 |
| ASPICIO | 7 | 0 | - 0.84 | SUM | 96 | 14 | 1.29 |
| ASSUM | 7 | 0 | - 0.84 | CREDO | 6 | 2 | 1.55 |
| CONDO | 7 | 0 | - 0.84 | DO | 6 | 2 | 1.55 |
| CURO | 7 | 0 | - 0.84 | VOLO (VELLE) | 5 | 2 | 1.78 |
| MIROR | 7 | 0 | - 0.84 | SOLVO | 3 | 2 | 2.39 |
| MORIOR | 7 | 0 | - 0.84 | NASCOR | 9 | 4 | 2.70 |
| SOLEO | 7 | 0 | - 0.84 | ROGO | 2 | 2 | 2.83 |
| VINCO | 7 | 0 | - 0.84 | VIREO | 1 | 2 | 3.45 |
| AUDIO | 6 | 0 | - 0.78 | EO | 11 | 6 | 3.73 |
| CADO | 6 | 0 | - 0.78 | ASSIDEO | 0 | 2 | 4.45 |
| ERRO | 6 | 0 | - 0.78 | CREO | 0 | 2 | 4.45 |
| INQUAM | 6 | 0 | - 0.78 | NUBO | 0 | 2 | 4.45 |
| MUTO | 6 | 0 | - 0.78 | PERDO | 0 | 2 | 4.45 |
| SERVO | 6 | 0 | - 0.78 | UMEO | 0 | 2 | 4.45 |
| COGO | 5 | 0 | - 0.71 | TACEO | 0 | 5 | 7.04 |
| IACEO | 5 | 0 | - 0.71 | IUBEO | 4 | 9 | 7.50 |
| PATIOR | 5 | 0 | - 0.71 | | | | |

choice of words within and between the single Eclogues. *Cano* (20 x) and *canto* (16 x), for example, would appear much more characteristic, if counted as the same unit. Another reason is the fact that the most frequent content words (17) of the Eclogues, i.e. *sum* (96 x), *omnis* (31 x), *dico* (29 x), *video* (27 x) and *venio* (26 x), mainly verbs, are frequent in the average language of literature (18) and occur in the *Pervigilium*, too, and, furthermore, that the next frequent content words, i.e. *amor* (22 x), *puer* (22 x) and *silva* (22 x), all nouns, are common to both poets on account of related motives.

Similar observations can be made for the words with frequencies between 10 and 20 in the Eclogues. Some nouns, *deus* (16 x), *mons* (15 x), *umbra* (15 x), *nympha* (10 x) and *mater* (10 x), are used by both poets without marked preference by one of them, but others do not occur in the *Pervigilium* : *pecus*

(17 x), *herba* (15 x), *ovis* (15 x), *capella* (13 x), *pastor* (11 x), *flumen* (11 x), *lac* (10 x) and *frigus* (10 x). Except the four instances of *frigus* in Ecl. 10, these words are typical of the bucolic world, but they are not frequent enough to reach the 10 % significance level (19). Of the adjectives *magnus* (19 x), *formosus* (16 x), *mollis* (13 x), *tener* (11 x) and *viridis* (11 x), the first is rather unspecific and occurs in both poems, while the others do not occur in the *Pervigilium*. *Formosus* and *viridis*, however, correspond to *pulcher* (1 x) and *virens* (2 x) in the *Pervigilium*, so that only *mollis* and *tener* remain unmatched (unless *delicatus* (v. 79) is understood in this sense). It seems that Vergil has a special preference for tenderness and softness (20), while in the *Pervigilium*, with words such as *urgeo* (v. 15), *vis* (v. 64) and perhaps *rigeo* (v. 58) – all absent from the Eclogues – more admiration is expressed for the vigorous creative power of Venus in flowers, human beings and nature in general. The verbs are mainly unspecific, as above : Among *cano* (20 x), *facio* (17 x), *canto* (16 x), *duco* (16 x), *fero* (15 x), *amo* (14 x), *habeo* (14 x), *pasco* (13 x), *incipio* (12 x), *possum* (12 x), *eo* (11 x) and *lego* (11 x), only *pasco*, *amo*, *canto* and *cano* are immediately related to the content of the Eclogues by their meaning. These four verbs, however, *pasco*, *amo*, *canto* and *cano*, can certainly be regarded as keywords for the main events of bucolic life. On the whole, Vergil seems to have coloured his Eclogues much more by the nouns and adjectives used than by the verbs.

Let us turn now to the characteristic vocabulary of the *Pervigilium Veneris*. The *Pervigilium* contains, as mentioned above, a few unspecific verbs which are relatively frequent in both texts, i.e. *sum* (14 x), *dico* (4 x), *facio* (3 x) and *venio* (2 x), but it contains also twelve verbs with significantly higher rates in the *Pervigilium*. These verbs are, in order of increasing significance, *solvo* (2 x), *nascor* (4 x), *rogo* (2 x), *vireo* (2 x), *eo* (6 x), *assideo* (2 x), *creo* (2 x), *nubo* (2 x), *perdo* (2 x), *umeo* (2 x), *taceo* (5 x), *iubeo* (9 x) – and *amo* (44 x) in the refrain. While *amo*, *nubo*, *creo* and *nascor* refer directly to the main motives of the *Pervigilium*, *solvo*, *vireo* and *umeo* refer only to related aspects. *Rogo*, *assideo* and *iubeo* are mainly connected with the spring-celebration in the central part of the poem. The logical subject of *iubeo* is always Dione, i.e. Venus, and three instances of *iubeo* (vv. 22, 67, 84) refer to Venus' power in general. The extent of this power seems to be underlined by the extraordinary frequency of *iubeo*. *Perdo* and *taceo*, finally, are concentrated in the final verses (vv. 89-92),

which express the strong desire of the poetess to find (finally ?) her own spring of love. Love is paralleled by song in these verses, *cf.*

84 *et canoras non tacere diva iussit alites*

87 *ut putes motus amoris ore dici musico,*

and lack of love by silence :

89 *illa cantat, nos tacemus : quando ver venit meum ?*

The close connection of *nos* with the feminine *illa* in this verse, and with the feminine *chelidon* in verse 90 (see below), seems to support the assumption that the *Pervigilium* is composed by a poetess.

The adjectives characteristic of the *Pervigilium* are sharply separated from the non-characteristic ones : while the hypothesis of uniform distribution over both the *Pervigilium* and the Eclogues can be rejected with an error probability less than 0.2 % for *totus*, the error probability would be about 63.8 % for *primus*, which neighbours *totus* in the list sorted for Z-values. *Totus* – in its connections with *annus* (v. 60), *flores* (v. 51), *nox* (v. 46) and *Amor* as the naked child of Venus (v. 35) – may perhaps be understood as expressing an element of passion : the meaning includes the resolution of certain restrictions and the idea of – or striving for – fulfilment and fullness. It is perhaps remarkable that at all occurrences of *totus*, this idea is connected with the future : by imperatives (vv. 34/35, v. 51), by the gerundive (v. 46) and by *cras* (vv. 59/60). This interpretation of *totus* would fit well to verse 89 of the poem (see above). *Nudus* (4 x), *pudicus* (2 x) and *canorus* (2 x) are characteristic of the *Pervigilium* with about 0.001 % error probability, *feriatus* (3 x) and *myrteus* (3 x) with about 0.000002 % error probability. Obviously these adjectives reflect the special contents of the *Pervigilium*, *myrteus* as referring to an attribute of Venus, *feriatus* as referring to her festival.

Most adjectives with medium frequency in the Eclogues do not occur in the *Pervigilium*. Among these a certain group is associated mainly with negative feelings : *crudelis* (7 x), *malus* (6 x), *infelix* (5 x), *tristis* (5 x), *miser* (4 x) and – less clearly – *amarus* (6 x) and *durus* (5 x) (which sometimes refer to plants). Words of similar meanings do not occur in the *Pervigilium*, so that it might appear as a poem without shadows, in accordance with its hymnic character, unless the final verses, especially

- 89 *illa cantat, nos tacemus; quando ver venit meum ?*
 90 *quando fiam uti chelidon, ut tacere desinam ?*
 91 *perdidi musam tacendo, nec me Phoebus respicit,*

would show that, in sharp contrast to verses 1-88, the poetess herself suffers from lack of love and song. But the double *quando* contains an element of hope, and the *Pervigilium* itself may be understood as a successful attempt to end her silence. So *infelix*, *tristis* and *miser* would probably be inappropriate descriptions of her state of mind.

The nouns characteristic of the *Pervigilium* are clearly related to its main motives : love, which is understood essentially as a creative power, nature in its original sense, and spring, the high time of nature and love. Although these motives cannot be clearly separated, love may be associated with *puella* (6 x), *virgo* (5 x), *maritus* (5 x), *arma* (2 x, of Amor), *casa* (2 x, the myrtle bowers), *coniunx* (2 x), *nodus* (2 x), *nuptiae* (2 x), *osculum* (2 x), *papilla* (2 x), *pudor* (2 x), *sagitta* (2 x, of Amor) and *sinus* (2 x), vernal nature with *ver* (6 x), *flos* (5 x), *imber* (4 x), *rus* (4 x), *ales* (3 x), *annus* (3 x), *gemma* (2 x), *lucus* (2 x), *pontus* (2 x), *purpura* (2 x) and *spiritus* (2 x). *Nox* (4 x), *diva* (3 x), *caterva* (2 x) and *ius* (2 x) refer mainly to the festival. Many of these nouns are associated with chastity, matrimony and legality : *virgo*, *pudor*, *nuptiae*, *ius*, *coniunx* and *maritus*. Also *pudicus*, *nubo* and *iubeo* should be seen in this context.

Other words characteristic of the *Pervigilium* may be added, in particular the prepositions *de* (14 x, perhaps an indication of origin in Northern Africa, not before the middle of the fourth century A.D., cf. Catlow, l.l., p. 20) and *subter* (3 x). But it should also be noted that the *Pervigilium* is not free from variation : while *flos* is rather frequent (5 x), *floridus* (v. 13), *florulentus* (v. 19) and *floreus* (v. 44) occur only once.

It is advisable to conclude this section – and the article – with a note on the different lengths of the texts considered. While the Eclogues comprise 5,714 words, or 5,588 words without the refrains, the *Pervigilium* has only 674 or 564 words respectively. This means that the *Pervigilium* was compared with a rather representative sample, and that the vocabulary resulting as characteristic of the *Pervigilium* can be regarded as a well established collection. On the other hand, the *Pervigillum* should be characterized by chance effects more than the Eclogues. A word with frequency 0 in the *Pervigillum*, for

example, might well occur, if the *Pervigilium* were longer; this word would appear less characteristic of the Eclogues in consequence. But other words might still have frequency 0, even if the *Pervigilium* were considerably longer, and such words might emerge as characteristic of the Eclogues on high significance levels. A word like *capella*, for example, with thirteen occurrences in the Eclogues (and none in the *Pervigilium*), would appear as characteristic of the Eclogues with less than 5 % error probability, if the *Pervigilium* had 1,692 words, i.e. three times its actual length (without refrain). In other words, it may be a chance effect that *carmen* was found to be the only word strictly characteristic of the Eclogues in the present comparison.

References

1. E.D. Kollmann, A Study of the Vocabulary of Vergil's Eclogues, *Revue du L.A.S.L.A.*, 1973, 2, pp. 1-24.
2. E.D. Kollmann, Word Frequencies in Latin Literature, *Revue du L.A.S.L.A.*, 1973, 4, pp. 1-17.
3. E.D. Kollmann, The World of the Eclogues, *Revue du L.A.S.L.A.*, 1974, 4, pp. 1-24.

Notes

- (1) Kollmann (reference 1, pp. 9-11) names four main themes : song, love, nature or bucolic life, Rome and reality. He shows that, of the 20 most frequent content words, 5 (incl. *dico* and *incipio*) refer to song, 4 to bucolic life and 3 to love. The present study will show that most of the less frequent content words refer to bucolic life, some also to love, but only few to song. Cf. also Kollmann (reference 2, p. 15). Kollmann (reference 1, pp. 6-8) tends to conclude that "the lively presentation and the emotional style of the Eclogues" are stressed by the frequent use of vocatives, imperatives and demonstrative pronouns, especially *hic* (in comparison with the *Odes* of Horace). This view agrees well with some of the present observations.
- (2) L. Catlow, *Pervigilium Veneris*, edited with a Translation and a Commentary, Bruxelles 1980, pp. 18-25.
- (3) For the main characteristics of the bucolic genre compare, e.g., E.A. Schmidt, *Poetische Reflexion - Vergils Bukolik*, München 1972, pp. 17-57.
- (4) H. Morgenroth - D. Najock, *Concordantiae in Corpus Priapeorum et in Pervigilium Veneris*, Hildesheim 1983.

- (5) R. Lecrompe, *Virgile, Bucoliques - Index Verborum - Relevés Statistiques*, Hildesheim 1970.
- (6) Ch. Muller, *Einführung in die Sprachstatistik*, München 1972, p. 205.
- (7) S. Usher - D. Najock, A Statistical Study of Authorship in the Corpus Lysiacum, *Computers and the Humanities*, 16 (1982), pp. 85-105, especially pp. 101 sq. and the Erratum, p. 271.
- (8) Essentially the bootstrap technique is this : on the base of a single sample, a large number (e.g. 1,000) of pseudo-samples (of equal size) are generated by means of random numbers. This proceeding allows us to calculate the empirical variance of a certain random variable (e.g. vocabulary size) over the pseudo-samples.
- (9) The refrains are, of course, integral parts of the poems, and they were taken into account in a first series of data sets and program runs. In a second series they were left out of account, because it was intended to compare also the vocabulary usage in the non-refrain parts of the poems. Indeed, the results of the second series are more interesting, and they are given throughout this article. The results of the first series are mentioned only where they differ more than slightly from the second series. The modified refrains, which occur only once (Ecl. 8, v. 61 and v. 109), were always taken into account.
- (10) For a similar classification cf. Kollmann (reference 1, pp. 12-13).
- (11) Although *neq*, with regard to the frequencies of occurrences, seems to be almost the poetic equivalent of *neque* (Kollmann, reference 2, pp. 9-11), it would well fit into the list, since I refer "prosaic" to the meaning of the words, here.
- (12) Kollmann (reference 1, p. 14) remarks that Ecl. 10 "stands alone in the collection with regard to content, structure and speaking persons"; this is certainly right, but he also mentions similar differences between other Eclogues. In fact each of the Eclogues seems to be built up in a rather peculiar way, and the special role of Ecl. 10 asks for a more detailed explanation.
- (13) The ascription to a poetess, proposed by P. Boyancé, *Encore le Pervigilium Veneris, Etudes Latines*, 28 (1950), pp. 212-35, in particular pp. 229 sqq., has found substantial support by L. Catlow, l.l., pp. 24 sq. Additional arguments are given below in the section on characteristic vocabulary : the close connection of *nos* with feminines in vv. 89/90 and the stress laid upon words associated with virginity, matrimony and legality. Also vv. 59-67, in particular v. 61, seem to support the ascription to a poetess.
- (14) L. Ule, Recent Progress in Computer Methods of Authorship Determination, *ALLC-Bulletin*, 10, 3 (1983), pp. 73-79.
- (15) Ch. Muller, l.l., p. 243.
- (16) The 9 occurrences of *carmen* in the refrain of Ecl. 8, where it means "love-charms", are not taken into account here. If the refrains are not excluded, *carmen* (49 x) becomes even more characteristic, but it remains the only positive characteristic word of the Eclogues on the 5 % significance level.

- (17) Content words, as opposed to function words, are restricted here to nouns, adjectives and verbs, since many adverbs are pronominal adverbs in our texts. Proper names are also excluded. Numerals are classified as adjectives, pronouns and adverbs, as in the Oxford Latin Dictionary.
- (18) Kollmann (reference 2, pp. 3-4) finds that the *Odes* of Horace resemble the *Eclogues* in this respect, while Tibullus and Propertius prefer words less frequent in the average language of literature.
- (19) In fact, only *pecus* reaches the 20 % significance level.
- (20) L. Delatte, Key-words and poetic themes in Propertius and Tibullus, *Revue du L.A.S.L.A.*, 1967, 3, pp. 31-79, shows that *tener* is a key-word of Tibullus, whereas Propertius prefers *mollis*.