

## RECHERCHES STATISTIQUES POUR L'ENSEIGNEMENT DU LATIN.

### Theme and Purpose of the Thesis

A statistical analysis of the texts prescribed by the curriculum has become necessary. After having completed a three or four years' course in Latin (3-3-3-3 or 5-4-4 lessons a week) pupils are expected to be able to translate an original text into German, to have gained some knowledge about languages in general, and to be able to find out the meaning of the more usual words of Latin origin, in German at least. There is a special branch of Erweiterte Oberschule (EOS - approximately = Grammar School) for pupils who want to specialize in Latin and Greek, but the two courses mentioned above are what the majority of pupils, who attend the Natural Sciences (B) or the Modern Languages (A) Branch have to expect.

We have some difficulties in realizing these aims in the given time, and these cannot be solved by trying to find better methods of instruction, at least not only in that way alone. It is necessary at first to answer the following questions: are we really teaching the pupils the vitally important things in respect to grammar and vocabulary? Or is perhaps too much stress put on matters of secondary importance, while indispensable things are undervalued? Does the textbook really prepare them in the best possible manner for those original texts which they are going to read?

It is desirable that the mistrust against the use of computers and similar machines in linguistic research, which is still existent in the G. D. R., should be overcome quickly. This thesis was begun in the hope of proving that these technical means can be applied with profit.

The texts analysed are those read by the pupils of the B and A-branches of EOS. As Latin is no longer taught with the intention of enabling the pupils to master the language actively, it is quite legal to concentrate on these texts,

which were chosen for their historic and aesthetic values but not for belonging to a special (the "golden") era of Latinity. If the curriculum were to be altered, it would still be worth while to have worked out a method of analysis which could easily be applied to other texts.

As the work had to be done in a relatively short time, there was only the choice left between analysing a small portion of the texts thoroughly or all of them partially. I decided on the latter course, choosing the verbal forms for analysis, because in this way the results may at once be used for school purposes (the verb being one of the most important classes of words) while the other variant would only have served as a technical exercise.

The material came from the following sources:

1. Caesar, Bellum Gallicum: 2678 forms (total of words: in these texts: 11747)
  2. Short passages from various technical authors, as Vitruvius, Columella, Cato: 545 forms (total: 2501 words)
  3. Livy : 2261 forms (total: 9522 words)
  4. Catullus: 217 forms (total: 801 words)
  5. Vergil: 158 forms (total: 777 words)
  6. Claudius Quadrigarius: 22 forms (total: 85 words)
- Pupils in the B-Branch read texts 1 and 2, pupils in the A-Branch texts 3 -6 as well.

The following problems are treated:

A. Vocabulary

- a) Which verbs occur at all?
- b) What is the relative frequency of these different verbs?

B. Grammar

- a) Which conjugations, which conjugated forms, which nominal forms, which nominal structures, what sorts of subordinate clauses are to be found?

b) What is their relative frequency respectively?  
c) The results of a) and b) are compared to those of an analysis of a second group of verbs from Cicero and Ovidius (1963 in all: Cicero, De Lege Manilia 4626 words, 947 verbal forms; Ovidius, Metamorphoses, Fasti, Tristia 4031 words, 1016 verbal forms); for the conjugated forms only, also with a group of verbs from Vergil, that had been treated before by E. Chubb (1), in order to find out (with the help of some statistic calculations) in what degree generalization might be possible.

C. Then the material provided by several Latin textbooks and wordlists is compared to the results of A and B in order to ascertain their effectivity in preparing pupils for the texts of the curriculum (2).

Basing on the result of the points A -C the questions

- a) Is it possible to improve the textbook used in the G. D. R. (Wolff), and in what ways?
- b) Is the use of 80-column-punched cards, which can be sorted and printed out mechanically to be recommended for linguistic research of a grammatical and didactical tendency? are answered.

#### Technical Experiences

The linguistic centre at the Deutsche Akademie der Wissenschaften in Berlin was not yet there when I began my work, therefore a part of this chapter is consecrated to hints for beginners who have to work with an industrial centre where they do not use machines which can punch and before all print letters as well as

some special points of codification arise if a text is not analysed word-for-word:

A. Alphabetic punching ("Alphalochung") practical, but its drawback is the fact that more space is taken up this way,

therefore it is worth while to give the words code-numbers, if the number of words which can possibly occur is not too great: double cards are rather difficult to sort, one ought to confine oneself to one card per word if at all possible.

B. It is very important to make sure that the exact place of a given form in the texts can easily be found again, for it is often necessary to have recourse to the whole sentence. Mechanical counting is only possible, when every single word of a given text is being analysed.

C. An interaction takes place between the numeric code adopted and the result of the analysis- elements which have not been codified can not be taken into account later. Therefore

a) a code ought to be tried upon a material of no less than 1500 items before it is finally adopted, because it is next to impossible to foresee every possibility -especially if different authors are being analysed.

b) The code ought to contain some elements of interpretation (for didactic purposes at any rate), e. g. not only Present tense, but: "normal" Present or historic Present, not only Subjunctive, but also "final sense, consecutive sense" a. s. o. On the other hand it must not become so complicated that a lot of mistakes are bound to arise in the act of codifying.

c) For some categories the code can easily be completed before the beginning of work (Tenses, Participles, Moods). But in other cases it is advisable to work with an 'open' code, dividing the numbers roughly on the letters of the alphabet and filling the single numbers as need arises (e. g. verbum regens of the A. c. I. or of certain conjunctions. If three columns are reserved for them; "agere, ut" will get 003, "committere, ut" 025, it does not matter if some places are left free in the end. But to give codenumbers to all possible verbis regentibus beforehand would be very hard indeed, if it were possible at all).

DO It is very difficult to find a way to codify words with different meanings. As many of these in the case of the verbs are short words (ago, facio) it helps to put the ob-

ject into the space reserved for letters and taken up fully by longer words (15 columns must be kept free, in order to be able to put up words like *communefacere*, *animadvertere*). For scholars with a grammatical purpose it might be best to confine themselves to a few words of this kind, which can be decided upon beforehand. Lexicologists will have to resort to double cards with more context.

D. If some categories can never occur at the same time, such as Person and Case (a verbal form can either be conjugated or nominal) it is possible to use one column for two purposes (That is the card model of the LASLA!)

These principles are illustrated with examples in the text, there is also a sample of codification. I have not used printed cards like those of Prof. Delatte (3), but big lists with 80 columns, so that one line contained the data for one card. These lists had a printed headline in order to make them more easy to handle, and fat vertical lines divided complexes of columns belonging together, from the next group. As codification was rather complicated and the numbers had to be verified and the total number of cards was not too great, the system proved quite satisfying.

#### Results concerning vocabulary

The 5881 verbal forms of the texts read in school belong to 1178 different verbs.

Group a more than 20 occurrences: 34 verbs

Group b 11 to 20 occurrences: 74 verbs

Group c 6 to 10 occurrences: 132 verbs

Group d 4 to 5 occurrences: 134 verbs

Group e 2 to 3 occurrences: 336 verbs

Group f one occurrence only: 468 verbs

More than two thirds of the forms belong to the verbs of groups a, b, c, these verbs ought be treated with extra care.

In considering the different wordlists and textbooks it is therefore essential to ascertain whether they contain a relatively high percentage of verbs from the upper frequency-groups, and as few as possible verbs that do not occur at all. The comparison shows that the efficiency does not necessarily increase proportionally to the total number of words. With comparatively few, well chosen words a high degree of efficiency can be reached (Cp. Table 1) Wolff is rather good, but contains 97 verbs which do not occur in the texts.

Improvement is possible by trying to avoid the unnecessary words and by concentrating on the important ones. For it is as much work to learn a hundred necessary verbs as to learn a hundred unnecessary ones, and the difference will make itself felt as soon as the pupils are confronted with original texts and have to look up the unknown words. In former times it did not matter so much, because there was time enough.

In building up a minimum vocabulary we can not rely on frequency alone. On the other hand the criteria used by the experts who select words for a "lexic minimum" ("lexikalisches Minimum" "français élémentaire") (4) must not be simply adopted, because in their case the ability to express oneself in the foreign language is the first aim of instruction, in Latin it is the translation into German.

Frequency is of great importance for a Latin vocabulary, too. But whereas in the modern languages it is desirable to make the basis as broad as possible, some hundred thousand words as a rule, in Latin, and specially in our case, we had better keep to the comparatively few texts read in school. Absolutely frequent words (e. g. agere, facere, capere) will occur simply anywhere therefore be included in our vocabulary as a matter of course. It is the next group, the verbs of

middle frequency, where we have to select, as the minimum must not be too abundant, and it is preferable that the pupils should learn what they can afterwards use. Thence range (répartition, Reichweite) is not absolutely decisive for us, but only in cases of lesser frequency. What occurs above six times in the texts ought to be learnt in any case. But a word that occurs thrice is more valuable if it is used by three different authors than if the three occurrences are confined to, say, one chapter of Livy. The fact that a word has different meanings (Mehrdeutigkeit) is important, and so is fertility in producing derivatives (Fruchtbarkeit). The latter word must, however, be taken in a wider sense than in the modern languages, including derivatives in non-Latin languages.

The amount of "thematic words" (mots disponibles, thematischer Wortschatz) (these are words which are not necessarily very frequent, yet indispensable for leading a conversation on certain themes of everyday life, on politics, literature a. s. o.) will be smaller in Latin than in the modern languages because the range of themes to be treated is more restricted. Lastly, it is not possible to pay much attention to the criterium of morphological and grammatical regularity ("Mustergültigkeit"), because anomalous words cannot be shirked in original texts: and adaptations are anathema in Germany!

The share of verbs in a minimum vocabulary is limited on the outside by the share that the verbs have in the most frequent words in general: 39,5% judging from Mathy. If the size of the minimum vocabulary is small, the share of the verbs goes down because the number of functional words (prepositions, pronouns and the like) is constant and cannot be cut down under 250.

For a minimum of 1500 words, which is about the amount that is learned by pupils in two years in the modern languages, somewhat less than what can be expected in England or France, because there are not so many Latin derivatives in Ger-

man which could save as a prop of memory, I proposed 384 verbs.

of group a ((34): 34 = 100%

of group b (74): 74 = 100%

of group c (132): 132 = 100%

of group d (134): 64 = 47,8%

of group e (336): 62 = 18,5%

of group f (468): 18 = 3,8%

The criteria of frequency, range, existence of various meanings have been considered. Some more verbs of the groups d, e, f will have to be included for the sake of their derivatives or because they are necessary for thematic reasons, but that cannot be the work of the statistician.

328 of these verbs are in Wolff's vocabulary, 323 of his verbs were omitted, 56 were not contained in his vocabulary. It must be noted, that neither Wolff's book nor any of the others was written with a special view to the texts of the curriculum, so the results of the request must not be taken as derogative to any of these books, they only show that it is better to have a book that is specially designed for them, as Sweet f. i. is for reading Vergil as first author.

#### Results concerning Grammar

A comparison of the four textbooks (Wolff, Schneider, Jarcho, Sweet, see note 2) shows the following basic differences: Wolff and Schneider start with the a- and e-conjugation and with esse, developing all the tenses, moods and voices in comparatively few lessons, the other conjugations follow later. Jarcho and Sweet, on the contrary, start with all the conjugations, first introducing only the indicative active, leaving the subjunctive to the end. Sweet even limits himself to the 3rd person singular and plural for rather a long time. The nominal forms and nominal structures are divided and treated in the course of more than 20 lessons by Wolff (who takes other matter in between), on a good few lessons by Schneider, while they are concentrated by Sweet and Jar-



cho. (Differences in the didactic approach to the material are treated at length in the text, because the American as well as the Soviet textbook differ widely from all that is usual in Germany).

From a statistical point of view the verbs of the curriculum texts (Lehrplanverben = L) and the verbs of the second and third group (Kontrollverben = K, Vergilverben = V) form random samples out of a whole which might be called "Literary language of the years 100 A. C. to 100 A. D."

But that is not a homogeneous whole, especially as prose and poetry is mixed in the samples L and K - with a preponderance for the prose texts in L. Yet a comparison between the samples shows, whether a certain phenomenon is typical only for L or of greater importance.

Because the samples (Stichproben) NL, NK, and NV are of different size, differences concerning the various grammatical phenomena cannot be compared directly. Even if we compare the relative frequencies  $f_L$ ,  $f_N$ ,  $f_V$ , not every difference is statistically significant. Statistical significance is tested with the aid of the Bernoulli-formula (two items can be compared at a time)

$$u = \frac{D}{S_D} = \frac{|f_L - f_K|}{\left(\frac{f_L(100-f_L)}{N_L}\right)^2 + \left(\frac{f_K(100-f_K)}{N_K}\right)^2}$$

If  $u$  is bigger than 3,09, the difference is significant (rate of probability 99,9%). Those series of calculations were done by the mathematical centre of the University of Rostock, with the aid of a computer.

Phenomena which are of frequent occurrence have to be treated

ted earlier and "drilled" more intensely than those which occur seldom, because the pupils must be able to handle them when they start on original texts. What is absolutely rare may be left to the commentaries or simply be mentioned in the Grammar, without receiving attention in the textbook, provided that the system is not shattered in the procedure: we would not advise to keep the 2nd person plural from the paradigm because it is not used often. But it does not seem necessary to spend a lot of time on drilling these forms. Though vocabulary is more deeply influenced by the contents of a given text than grammar, there are some significant differences with respect to the use of tenses, moods, nominal structures between the different samples. It would not do to take the average: these differences must be taken into account. See Table 2.

For every complex which is treated the relative frequencies are ascertained then the results are compared to the treatment these phenomena receive in the textbooks, and at last some propositions are ventured concerning possibilities of amendments. The statistic analyses are gathered into tables, of which table 2 is an example.

This is a summary of the proposals:

A. It is desirable that the 3rd conjugation (agere) should be treated earlier than now. Besides it must be noted, that the other conjugations (i-, anomala, verbs like capio) which have been treated in the second year of the course contain a few very important verbs, which ought to be learned as soon as possible (see Table 3)

B. Conjugated Forms

It is possible to limit oneself for a considerable time to the 3rd person singular and plural. The four great tenses (present, perfect, past (Imperfekt) and pluperfect) had better be treated in this order. The passive voice might be postponed until the middle of the preliminary course, the two future tenses need not be introduced before the end. The subjunctive is a frequent mood, in the L group (indirect

speech, lots of cum-and ut-clauses), much more so than in K, let alone V. Therefore we cannot adopt Jarcho's and Sweet's way of leaving it to the last lessons, it must be well drilled and introduced rather early in the course.

#### C. Nominal Forms and Nominal Structures

It is right to treat the infinitives rather early, participles and their syntactic structures had better be concentrated in the middle of the course. The A. c. I. deserves a fuller treatment than it gets now, longer periods ought to be included in the textbook, or the pupils meet with great difficulties when they start on *Bellum Gallicum*. Gerund and gerundive will have to stay at the end of the course, but the construction "ad urbem obsidendam" ought to receive treatment.

Supine and N. c. I. might be left to the Grammar, if "bonum esse videtur" was simply learned like an idiom.

There is no statistic reason for treating the deponents early or late: their place can be decided upon from purely didactic reasons.

#### D. Subordinate Clauses

Clauses in indirect statements must not be fully neglected. It is important to get the meaning of the structures quite clear, there are some misleading names that can only serve to confuse the pupils (ut "finale" in sentences like "opto ut mihi adsis"),.

Some important types:

##### a) cum

not only cum historicum, cum causale and cum iterativum ought to be treated, but also cum adversativum and cum vere temporale

##### b) quod

for logic reason the treatment ought to be reversed, first the relative pronoun quod, then "faktisches quod" (like *aegre fero, quod*) and then quod causale:

"faktisches quod" has not been treated up to now.

c) ut

There has been enough material about this conjunction, but the treatment might be clearer, so that the different shades become quite manifest.

d) Relative clauses

It would be better if they were treated somewhat earlier and the treatment ought to be continued in a later lesson, so that mood, relative subject and object clauses, place of the clause in the period were fully explained and not simply left to the pupils to discover.

ad III and IV:

In respect to vocabulary we have seen that some items might be left out without doing damage (indeed, there might then be time to work at the rest of the words more thoroughly). In respect to grammar to change in the order of treatment might be desirable. In this case there is no question of reducing, a few things had even better be included again, which have been cut out.

Results

It is possible to make a Latin textbook more efficient by resorting to statistic analysis of the texts for which it is supposed to prepare the pupils. The amendments may not seem great at first sight, but in the given situation every lesson gets valuable.

The use of 80 /-column punched cards for researches of this kind has stood the test. It seems advisable to decentralize the punching (and the subsequent checking of the cards) sorting and printing may then be left to a big centre.

German title: Untersuchungen zum Verbenbestand im Kanon der lateinischen Schulschriftsteller, 142/+16S., 43 Tabellen und Abb.

Effektivität der verschiedenen Wortlisten und Vokabulare

Häufigkeitsgr. :	a	b	c	d	e	f	nie vorgekommen
1178 Vb. Lehrplantexte	34	74	132	134	336	468	-
Wolff (651)	34	70	112	90	165	83	97
(55,4% v.L.) %:	100	94,5	84,8	67,2	49,2	17,7	14,9
Jarcho-Loboda (749)	34	72	118	88	174	109	155
(63,6% v.L.) %:	100	97,4	89,4	65,7	51,8	23,3	20,7
Jarcho-Minimum (178)	31	42	39	21	24	12	9
(15,1% v.L.) %:	91,2	56,8	29,6	15,7	7,1	2,6	5,5
Schneider (601)	32	67	102	87	122	84	107
(51% v.L.) %:	94,1	90,5	77,2	65,0	36,4	18,0	17,8
Mathy (1019)	33	72	132	121	246	185	232
(93,2% v.L.) %:	97,0	97,4	100	90,2	73,2	39,5	22,8
Syllabus (384)	32	67	100	61	69	31	24
(32,6% v.L.) %:	94,1	90,5	75,8	45,5	20,6	6,6	6,2
Schneider Formenlehre (745)	30	58	103	91	176	126	161
(63,2% v.L.) %:	88,2	78,5	78,1	67,9	52,4	26,9	21,6

Zeichenerklärung

a : in den Lehrplantexten über 20 mal vorgekommen  
 b : " " " 11 bis 20mal " "  
 c : " " " 6 bis 10mal " "  
 d : " " " 4 bis 5mal " "  
 e : " " " 2 bis 3mal " "  
 f : " " " 1mal " "

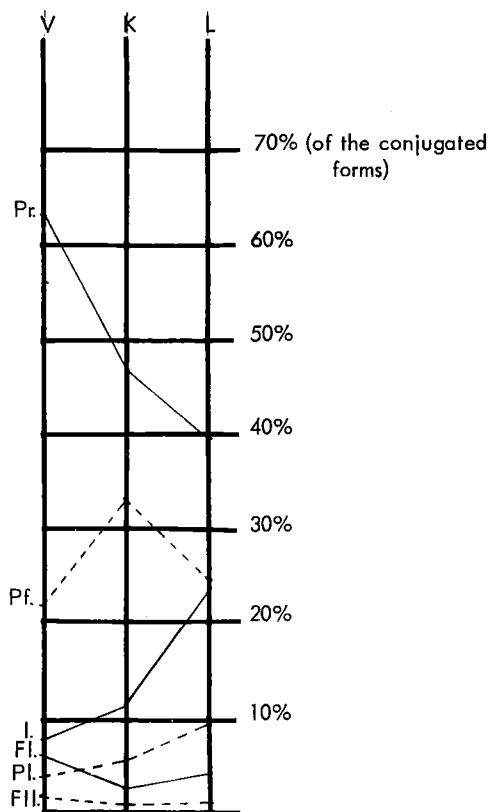
Die Prozentwerte in der letzten Spalte (nie vorgekommen) beziehen sich auf die Anzahl der Verben in der betreffenden Wortliste.

v.L.= von den Verben der Lehrplantexte.

### Tempora

V = Vergil  
 K = Verbs from Ovidius  
 and Cicero  
 L = Verbs from the curriculum-  
 texts

Pr. - Present  
 Pf. - Perfect  
 I. - Imperfect (Post)  
 Pl. - Pluperfect  
 Fl. - Future  
 FII. - Futurum exactum



Verben die in den Lehrplantexten 11mal und öfter vorkommen und nach Wolff in ersten Jahr behandelt werden

Die Zahl gibt das Lehrbuchstück an, in dem das Verb eingeführt wird.

a: das Verb gehört zur höchsten Häufigkeitsgruppe (21mal und öfter).

<u>a- Konj.</u>	<u>e- Konj.</u>	<u>kons. Konj.</u>	<u>anomala</u>
2 pugnare	1 debere	21 adolescere	1 esse a
vocare	3 tenere a	tradere	9 abesse
3 parare	4 habere a	dicere a	praesens
imperare	licet	ducere a	(praesens : 32)
4 privare	5 valere	adducere	12 posse a
(-tus : 10)	8 prohibere	gerere a	
6 postulare	13 placere	intelligere	4
7 rogare	16 videre a	mittere a	
apparare	18 continere	pellere	
9 appellare a		vincere a	<u>defectiva</u>
putare	9	22 alere	3 inquit
10 creare a		cognoscere a	7 futurus
11 dare a		petere a	
(Perfekt : 33)		ponere	
18 iudicare		quaerere	2
20 comparare			
22 existimare a		15	
			Summe : 45
15			davon 16a

Verben die in den Lehrplantexten 11 mal und öfter vorkommen und nach Wolff in zweiten Jahr behandelt werden

<u>a- Konj.</u>	<u>e- Konj.</u>	<u>i- Konj.</u>	<u>kons. Konj.</u>
37 arbitrari conari versari	28 oportere 34 movere obtinere sustinere	29 audire scire 30 aperire reperire venire a pervenire	23 traducere 24 committere dividere 25 accidere reddere addere condere oredere contendere a
3	35 iubere a respondere 37 videri a 40 audere	40 oriri	26 consistere 27 agere a cogere a relinquere considerare 28 constituere instituire vertere animadvertere
<u>i- Erweiter.</u>	<u>anomala</u>	<u>defectiva</u>	38 sequi a uti a 39 proficisci nasci
31 capere a facere a interficere perficere	41 ire transire a 42 ferre a conferre inferre a referre velle a	27 coepisse a 32 fieri a 36 fore	
32 accipere a recipere conficere efficere		3	
39 pati a	7		
9			22
Es kommen bei Wolff <u>nicht</u> vor : rari, convenire, icere, deducere			
			Summe : 59 davon 18 a