[0437]   SOME DATA ON WORD FREQUENCY IN THE
SKELTANA RIUCHT

The Old West Frisian *Skeltana Riucht* (from here on referred to as SR) is a law-text that is small in scope and size. It contains 6857 words[1], which belong to 781 lexical units. In this study a lexical unit should be understood to consist of the word(s) that occur(s) as a single entry in Von Richthofen's *Altfriesisches Wörterbuch*. To give an example: *is, se(e), sen, si, sint, wa(e)s, were, weren, wessa* and *wessen* are ten different words, but they compose a single lexical unit[2].

A word frequency count and subsequent arrangement of the data results in the following statistical survey:

A	B	C	D	E	F	G	A	B	C	D	E	F	G
1x	347	347	781	100.-	6857	100.-	30	3	90	37	4.74	4115	60.01
2	126	252	434	55.57	6510	94.94	31	2	62	34	4.35	4025	58.70
3	75	225	308	39.44	6258	91.26	32	1	32	32	4.10	3963	57.79
4	36	144	233	29.83	6033	87.98	35	4	140	31	3.97	3931	57.33
5	22	110	197	25.22	5889	85.88	40	1	40	27	3.46	3791	55.29
6	21	126	175	22.40	5779	84.28	41	1	41	26	3.33	3751	54.70
7	19	133	154	19.72	5653	82.44	42	1	42	25	3.20	3710	54.10
8	9	72	135	17.29	5520	80.50	43	1	43	24	3.07	3668	53.49
9	12	108	126	16.13	5448	79.45	44	1	44	23	2.94	3625	52.87
10	10	100	114	14.60	5340	77.88	45	1	45	22	2.82	3581	52.22
11	12	132	104	13.32	5240	76.42	53	2	106	21	2.69	3536	51.57
12	9	108	92	11.78	5108	74.49	58	1	58	19	2.43	3430	50.02
13	1	13	83	10.63	5000	72.92	61	1	61	18	2.30	3372	49.18
14	7	98	82	10.50	4987	72.73	65	2	130	17	2.18	3311	48.29
15	4	60	75	9.60	4889	71.30	77	1	77	15	1.92	3181	46.39
16	7	112	71	9.09	4829	70.42	78	1	78	14	1.79	3104	45.27
17	4	68	64	8.19	4717	68.79	83	1	83	13	1.67	3026	44.13
18	3	54	60	7.67	4649	67.80	93	2	186	12	1.54	2943	42.92
19	3	57	57	7.29	4595	67.01	103	1	103	10	1.28	2757	40.21
21	2	42	54	6.91	4538	66.18	117	1	117	9	1.15	2654	38.70
22	2	44	52	6.66	4496	65.57	167	1	167	8	1.02	2537	37.-
23	2	46	50	6.40	4452	64.93	168	1	168	7	0.89	2370	34.56
24	2	48	48	6.15	4406	64.26	186	1	186	6	0.77	2202	32.11
25	1	25	46	5.89	4358	63.56	188	1	188	5	0.64	2016	29.40
26	3	78	45	5.76	4333	63.19	200	1	200	4	0.51	1828	26.66
27	1	27	42	5.38	4255	62.05	299	1	299	3	0.38	1628	23.74
28	3	84	41	5.25	4228	61.66	547	1	547	2	0.26	1329	19.34
29	1	29	38	4.87	4144	60.44	782	1	782	1	0.13	782	11.40

A. the frequency (*f*) with which a lexical unit occurs;
B: the rank (*r*) that the lexical unit holds in the frequency-list;
C: the product of the corresponding figures in A and B;
D: the sum of the figures in B from the bottom of the survey upwards;
E: the percentage represented by the figure in D;
F: the sum of the figures in C from the bottom of the survey upwards;
G: the percentage represented by the figure in F.

With a few exceptions, especially at the beginning and end of the list, the product of *f* and *r* varies within relatively narrow limits. This is in accordance with the findings of G.K. Zipf, who observed that the product *fr* is constant. He saw in this symmetry a balance struck between two opposing forces: the greatest ease intended on the part of the speaker, and the greatest clarity desired by the hearer. In other words, the two extremes are: the same word for all concepts that have to be expressed and a separate word for each of these concepts. (For more on these findings, see Bertil Malmberg, *New Trends in Linguistics,* Stockholm/Lund, 1964, p. 192)[3].

It appears that the four most frequently used lexical units (only 1/2 %) account for more than one fourth of the total number of words, and that nineteen of these lexical units (2.43 %) account for half of the total number of words. This confirms one of the findings of Pierre Guiraud in *Les Caractères statistiques du vocabulaire* (Paris, 1954), namely that the greater part of a text is constituted by a very small number of words. (See especially p. 10).

In SR the nineteen most frequently used lexical units in order of frequency are: 1. definite article, demonstrative pronoun[4] (782 x); 2. personal pronoun of 3rd person (547 x); 3. *ende* (299 x); 4. *wessa* (200 x); 5. *to* (188 x); 6. *so(e)* (186 x); 7. *schil* (168 x); 8. conjunction *dat* (167 x); 9. *mit* (117 x); 10. *ief* (103 x); 11. and 12. noun *riucht* (93 x), *sin* (93 x); 13. *o(e)n* (83 x); 14. *aegh* (78 x); 15. *een* (77 x); 16. and 17. relative particle *deer* (65 x), *habba* (65 x); 18. *ma* (61 x); 19. adverb *deer* (58x).

Although conclusions from a text as limited in size as SR are not to be drawn, and this text is not comparable with, e.g., the material on which N.R. French, *The Words and Sounds of Telephone Conversations, Bell System Technical Journal, 9* (1930), pp. 290-324, and G. Dewey, *Relative Frequency of English Speech Sounds* (Cambridge, Mass., 1923), based their research, it is noticeable that the results mentioned before in this study are much closer to the findings of French, whose material consisted of spoken language, than to those of Dewey, who investigated written language. In Dewey's text, sixty-seven lexical units accounted for half the number of words examined; in French's, thirty. This fact might be another indication, not mentioned in Chapter 11 of my *Word-Groups ...,* of the influence on SR of spoken language[5].

Ede   *A. Bor*

NOTES:

1. The edition used for this analysis is the one that is to appear in the series *Oudfriese Taal- en Rechtsbronnen* at some time in the future. Its publication has been delayed by the death of K. Fokkema, its original editor. Part of a proof copy of the forthcoming edition appeared as an auxiliary booklet with my *Word-Groups in the Language of the Skeltana Riucht* (Wageningen, 1971).
2. See G.A. Miller, *Language and Communication* (New York, 1963), p.89.

3. A few other works in the same field are: G.U. Yule, *The Statistical Study of Literary Vocabulary* (Cambridge, 1944); G. Herdan, *Language as Choice and Chance* (Groningen, 1956); G.C. Lepschy, *A Survey of Structural Linguistics* (London, 1970), pp. 140-145.
4. Generally speaking SR has the same word to represent a definite article and a demonstrative pronoun. See my *Word-Groups...,* note 3 to Chapter 1 on p. 219, and p. 28.
5. The various word-classes in SR are represented in the following numbers and percentages: pronouns - 1690 (24.65 %); nouns - 1441 (21.01 %); verbs - 1352 (19.72 %); conjunctions - 752 (10.97 %); prepositions - 604 (8.81 %); adverbs - 562 (8.19 %); numerals - 268 (3.91 %); adjectives - 188 (2.74 %).