

SÕNALÕPU JA SÕNAPIKKUSE VAHEKORRAST RESTITUTSIOONILISES TEKSTIS

=====

Ü. Kaasik, J. Tuldava

Teatavate rakenduslingvistiliste ja ka keeletüpoloogiliste ülesannete lahendamisel pakub huvi sõna foneetilise struktuuri seos sõna mõningate muude formaalsete tunnustega. Selle probleemi üks alaküsimusi on sõnalõpu ja sõnapikkuse vahetõrge. Käesolevas töös vaatleme eesti keele pöördagedussõnastiku materjalide põhjal tekstis esinevaid sõnavorme nende lõpptähtede ja lõpptäheühendite ning pikkuse järgi.* Vaadeldava teksti maht on 93.420 sõnet, valim on võetud tänapäeva eesti ilukirjandusproosa autorikõne tekstidest (19 osavaliitist) ja ca 5000 sõnet).

1. Lõpptähe järgi on vaadeldavas tekstis kõige enam g-lõpuliselt sõnavorme - arvult 19.690 ehk 21,08%. Sageduselt järgmised lõpptähed on e (13,58%), s (13,43%), i (12,92%), d (10,96%). Nimetatud viis sagedamat lõpptähte katavad tekstis kokku 72% kõigist sõnalõppudest.

Võrdluseks võib tuua lõpptähtede sagedused sõnastikus (pöördagedussõnastiku alusel, sõnastiku maht on 26.025 sõnavormi). Viis sagedamat lõpptähte on e (17,03%), d (14,94%), s (14,52%), a (14,05%), t (13,0%), mis kokku moodustavad 73,5% kõigist sõnalõppudest sõnastikus. Võrreldes tekstisagedustega võib märkida, et viie sagedama tähe hulgas on mõlemal juhul a, e, s, d; tekstis on suure sagedusega ka i, sõnastikus aga t (i on sõnastikus kuuendal kohal sagedusega 9,1%).

Andmed ülejäänud tähtede kohta leiab lugeja tabelist 2.

2. Tekstis esineb kõige rohkem 4-tähelisi sõnavorme (15.715 ehk 16,82%), seejärel 5-tähelisi (14,98%), 6-tähelisi (14,08%), 2-tähelisi (10,9%) ja 3-tähelisi (10,53%). Ülejäänud moodustavad kokku 32,8% (vt. tabeli 1 lõppkokkuvõtet).

* Mõningad erinevused vormide tekstisagedustes (ja seetõttu erinevus ka teksti üldpikkuses) on tingitud üksikute vaadeldavate valimite mittetäpsest kokkulangemisest käesolevas eksperimentis ja pöördagedussõnastikus.

Keskmine sõnavormi pikkus (sõnapikkus) ilukirjandusproosa autorikõne tekstis on 5,8 tähte.

Sõnalõpp ja sõnapikkus kombineeritult annavad meile iselaadse vormitüübi. Tekstis on kõige sagedam a-lõpuline kahtäheline sõnavorm (mida tähistame -a/2). Sellist tüüpi sõnavormide tekstikatvus on 6,33%, seega keskmiselt iga viieteistkümnes sõne eestikeelses tekstis on -a/2-tüüpi sõnavorm. Need on põhiliselt suure sagedusega ja seega suure funktsionaalse koormusega sõnavormid ja, ta, ka, ma, sa. Esitame andmed kahekümne sagedama sõnavormitüübi esinemissageduse kohta:

<u>-a/2</u>	6,33 %	<u>-a/5</u>	2,90 %	<u>-a/4</u>	1,94%	<u>-e/7</u>	1,56%
<u>-a/4</u>	3,83	<u>-a/3</u>	2,36	<u>-s/5</u>	1,83	<u>-s/7</u>	1,54
<u>-s/6</u>	3,75	<u>-i/5</u>	2,33	<u>-d/8</u>	1,80	<u>-s/3</u>	1,52
<u>-e/5</u>	3,38	<u>-d/6</u>	2,14	<u>-i/6</u>	1,66	<u>-e/3</u>	1,49
<u>-i/3</u>	3,37	<u>-d/4</u>	1,99	<u>-d/7</u>	1,63	<u>-t/7</u>	1,48

Kümme sagedamat tüüpi moodustavad tekstist 32,38%, kakskümmend sagedamat - 48,83%. Kõigi võimalike ja tekstis tegelikult esinenud lõpptähe ja sõnapikkuse kombinatsioonide koguarv läheneb kolmesajale (vt. tabel 1).

4. Vaadeldes sõnavorme eri lõpptähtede järgi, võime konstateerida, et kõige sagedama lõpptähe a juurde kuulub, nagu juba nimetasime, kahtäheline sõnavorm (-a/2); need moodustavad 30,1% kõigist a-lõpulistest sõnavormidest tekstis. Järgnevad a-lõpulistest hulgas neljätähelised sõnavormid (a/4) sagedusega 18,2%. Neist on sagedamad (vt. sõnavormide sagedussõnastikku kogumikus "Tõid keelestatistika alalt" I, Tartu, 1976): seda, juba, teda, tema, mida, väga, olla, maha, teha, aega. Artiklile lisatud tabelist 1 võib välja lugeda, et selles rühmas järgnevad sageduselt viie-, kolme-, seitsme- ja kuuetähelised sõnavormid. Sõnavormide sagedussõnastik näitab, et viietähelistest a-lõpulistest sõnavormidest on sagedamad välja, minna, sinna, kolmetähelistest - oma, aga, ära, seitsmetähelistest vaadata, rääkida, niisama, kuuetähelistest - minema, temaga, seisma.

Niisamuti võib analüüsida kõiki teisi vormitüüpe. Vaatleme siinkohal lühidalt veel e-, a-, i-, ja d-lõpulistest sõnavormide sagedamaid tüüpe.

e-lõpulisest sõnavormid, mis on eestikeelses tekstis a-lõpulistest järel sageduselt teisel kohal, annavad kõige sagedamana viietähelise vormi (24,9% kõigist e-lõpulistest), neist on sa-

gedamad: selle, mitte, talle, nende, kõige, poole, jälle, peale, mulle, mille.

s-lõpuliste sõnavormide hulgas on kõige sagedamad kuuetähelised (27,9%): kuidas, hakkas, juures, vaatas, selles, poleks, võttis, rääkis, tundis, seisis jt.

i-lõpuliste seas domineerivad kolmetähelised sõnavormid (26,1%): kui, oli, nii, või, löi, sai, asi, tõi, lõi, uni jt.

d-lõpulistest on sagedamad kuuetähelised sõnavormid (19,5%): polnud, silmad, saanud, läinud, jäänud, läksid, näinud, teinud, tulnud, poisid jt. Kõige sagedamate d/6-tüüpi sõnavormide hulgas on seega peamiselt nud-kesksõnad ja üksikud mitmusevormid.

5. Juurdelisatud tabel 1 võimaldab teostada vormitüüpide analüüsi ka lõpptähepaaride järgi. Võib näiteks nentida, et a-lõpuliste sõnavormide hulgas on kõige sagedamad vormid lõpu-ga (-) ja (kokku 20,7% kõigist a-lõpulistest, kusjuures ja-lõpulistest on kõige sagedamad kahetähelised, seega sõnavorm ja - 80% kõigist ja-lõpulistest; järgnevad viietähelised, näit. välja, neljätähelised, näit. vaja jne.). ja-lõpulistele järgnevad sageduselt (a-lõpuliste hulgas) sõnavormid lõppudega -ta, -ga, -ma, -da.

6. Tabelis 1 on antud sõnalõpu ja sõnapikkuse kombinatsioonide absoluutsagedused. Vaadeldavas tekstis ei esinenud sõnavorme lõpptähtedega š ja ž, mis iseenesest on eestikeelses tekstis võimalikud (näit. nimetava käände vormid tušš, beež). Nende puhul võib nentida, et selliste lõpptähtedega sõnavormid esinevad ilukirjandusproosa autorikõnes üsna harva. Sama kehtib ka ülejäänud võõrtähtede (c, č, g, ģ, y, ž) ning i ja õ kohta; i võib esineda lõpptähena ainult võõrapärastes nimedes (näit. Raj Kapoor) või lühendites (MTJ = masina - traktorijaam), õ esineb lõpptähena hüüdsõnas nõõ ja on võimalik ka võõrnimede või lühendite lõpus.

Tabel 1

Sõnalõpp ja sõnapikkus eestikeelses tekstis
(valimi maht N = 93420 sõnet)

Sõna- lõpp	Sõnavormi pikkus (tähtedes)												Kokku	
	2	3	4	5	6	7	8	9	10	11	12	>12		
AA	0	112	1	2	1	7	19	2	2	2	0	0	148	
BA	0	0	284	28	0	3	3	2	3	2	2	0	327	
CA	1	0	0	0	0	0	0	0	0	0	0	0	1	
DA	0	3	920	525	236	247	177	126	60	7	3	12	2316	
EA	3	196	0	0	10	3	2	3	0	0	1	0	218	
GA	0	889	273	252	177	312	131	179	131	106	69	159	2678	
HA	0	11	374	0	0	8	9	12	1	1	0	0	416	
IA	0	3	84	8	7	10	21	15	5	11	0	2	166	
JA	3266	28	173	340	44	67	31	52	28	16	11	27	4083	
KA	427	0	154	96	33	34	9	13	16	4	5	5	796	
LA	0	12	242	133	39	8	93	11	3	6	6	0	553	
MA	413	624	382	270	352	363	125	94	34	8	0	5	2670	
NA	0	2	434	269	35	38	45	98	44	24	20	39	1048	
OA	1	15	0	0	0	6	1	1	2	0	0	1	27	
PA	0	0	4	43	0	2	5	2	4	3	0	0	63	
RA	0	200	23	89	30	13	21	12	12	4	0	1	405	
SA	78	102	87	152	60	37	10	7	13	2	3	1	552	
TA	1722	0	35	311	196	170	53	78	22	18	9	17	2634	
UA	0	0	72	18	1	6	8	8	1	1	3	0	118	
VA	5	13	12	171	71	51	39	31	19	17	8	9	446	
õA	0	1	0	0	0	0	0	0	0	0	0	0	1	
üA	0	0	25	0	2	0	0	0	0	0	0	0	27	
Lõpp	-A	5916	2211	3579	2707	1294	1385	802	746	400	232	140	278	19690
AB	0	0	196	206	281	94	102	27	0	0	0	0	906	
EB	0	0	66	189	65	83	23	7	1	0	0	0	434	
IB	0	0	119	52	61	14	0	4	0	1	1	0	252	
LB	0	0	10	0	0	0	0	0	0	0	0	0	10	
MB	0	0	1	0	0	0	0	0	0	0	0	0	1	
OB	0	0	35	4	0	0	0	0	0	0	0	0	39	

Sõna-
lõpp

Sõnavormi pikkus (tähtedes)

	2	3	4	5	6	7	8	9	10	11	12	>12	Kokku
RB	0	0	13	0	0	0	0	0	0	0	0	0	13
UB	0	0	13	82	143	14	3	3	0	0	0	0	258
ÄB	0	0	39	0	0	0	0	0	0	0	0	0	39
ÖB	0	0	15	0	0	0	0	0	0	0	0	0	15
ÜB	0	0	1	0	0	0	0	0	0	0	0	0	1
Lõpp -B	0	0	508	533	550	205	128	41	1	1	1	0	1968
AD	0	239	113	178	393	269	273	134	103	49	36	40	1827
ED	0	6	226	128	299	91	150	132	108	55	36	74	1305
ID	0	2	900	111	282	448	498	259	141	99	50	103	2893
LD	0	0	20	0	0	1	1	2	0	0	0	0	24
ND	0	124	166	11	19	10	15	15	4	33	3	9	409
OD	0	0	21	6	0	1	1	0	3	1	1	1	35
RD	0	0	98	0	4	46	49	9	1	0	0	0	207
UD	0	0	81	147	1003	658	692	257	185	116	60	95	3294
ÄD	0	0	2	0	0	0	1	0	0	0	0	0	3
ÖD	0	9	29	0	0	2	1	0	0	0	0	0	41
ÜD	0	1	204	0	0	0	1	0	0	0	0	0	206
Lõpp -D	0	381	1860	581	2000	1526	1682	808	545	353	186	322	10 244
AE	0	3	1	0	1	0	1	1	2	0	0	0	9
BE	0	0	14	15	0	0	0	1	0	1	0	0	31
DE	0	18	14	218	231	229	33	32	35	39	37	58	944
EE	0	728	2	3	0	12	3	4	0	0	3	1	756
GE	0	0	67	292	1	3	11	15	6	0	1	3	399
HE	0	64	227	0	6	11	37	10	3	0	2	0	360
IE	0	2	178	0	0	4	6	0	1	0	0	0	191
JE	0	3	0	30	0	0	0	2	0	0	0	0	35
KE	0	0	13	190	100	33	11	27	15	8	1	11	409
LE	0	278	288	1209	354	403	217	217	156	95	85	110	3412
ME	128	2	48	170	63	80	57	48	19	6	3	3	627
NE	0	8	180	270	144	216	121	197	112	77	48	96	1469
OE	0	27	0	0	0	4	1	0	0	0	0	0	32
PE	0	0	2	5	1	0	2	3	1	4	1	0	19
RE	0	1	83	96	2	6	18	5	5	2	2	2	222
SE	2	158	114	290	172	323	295	281	193	109	81	196	2214
TE	15	0	132	293	131	115	221	111	84	45	48	83	1278

Sõna- lõpp	Sõnavormi pikkus (tähtedes)												Kokku
	2	3	4	5	6	7	8	9	10	11	12	>12	
UE	0	38	5	0	0	2	1	0	1	0	0	0	47
VE	0	1	8	77	0	1	4	9	11	4	1	0	116
OE	11	15	0	0	14	4	0	0	0	0	0	0	44
AE	0	47	0	4	1	14	8	1	0	0	0	0	75
Lõpp -E	156	1393	1376	3162	1221	1460	1047	964	644	390	313	563	12 689
AF	0	0	0	0	0	0	0	0	0	1	0	0	1
LF	0	0	0	1	0	0	0	0	0	0	0	0	1
RF	0	0	0	0	0	0	0	7	0	0	0	0	7
Lõpp -F	0	0	0	1	0	0	0	7	0	1	0	0	9
EG	0	39	22	0	2	3	12	5	0	1	1	2	87
IG	0	0	2	0	0	0	0	0	0	0	0	0	2
LG	0	1	21	0	0	2	1	0	0	0	0	0	25
NG	0	0	447	2	0	3	11	6	1	1	3	9	483
OG	0	0	0	0	0	3	1	5	1	0	0	1	11
RG	0	4	9	1	0	2	0	2	0	0	0	0	18
UG	0	0	9	0	0	0	0	0	0	0	0	0	9
Lõpp -G	0	44	510	3	2	13	25	18	2	2	4	12	635
AH	9	62	3	7	0	0	0	0	0	0	0	0	81
EH	1	3	0	0	0	0	0	0	0	0	0	0	4
HH	0	0	1	0	0	0	0	0	0	0	0	0	1
OH	16	13	0	1	0	0	0	0	0	0	0	0	30
AH	0	2	0	4	0	0	0	0	0	0	0	0	6
Lõpp -H	26	80	4	12	0	0	0	0	0	0	0	0	122
AI	1	85	1	0	5	1	1	2	0	0	0	0	96
BI	0	10	158	14	2	2	1	4	5	2	0	1	199
CI	0	0	0	0	1	0	0	1	0	0	0	0	2
DI	0	0	66	155	37	24	68	36	13	5	4	6	414
EI	1306	2	0	0	2	9	2	1	1	0	0	0	1323
FI	0	0	0	0	0	0	0	0	5	1	0	0	6
GI	0	1	133	581	559	289	71	56	35	15	7	18	1765
HI	0	0	49	2	1	0	8	3	1	0	1	0	65
II	0	393	1	1	0	13	0	4	0	3	0	1	416
KI	0	0	61	147	127	42	87	63	40	6	12	16	601

Sõna- lõpp	Sõnavormi pikkus (tähtedes)												Kokku
	2	3	4	5	6	7	8	9	10	11	12	>12	
LI	0	1103	198	155	119	27	21	31	9	8	8	10	1689
MI	0	1	31	76	11	9	7	12	12	8	4	13	184
NI	0	8	161	158	91	78	60	50	45	13	14	34	712
OI	4	1	0	0	0	0	1	0	0	0	0	0	6
PI	0	2	23	21	8	2	8	4	4	0	1	1	74
RI	0	6	94	106	67	59	23	19	36	16	8	12	446
SI	2	45	113	183	213	82	117	71	41	13	21	36	937
ZI	0	0	0	1	1	1	0	0	0	0	0	0	3
TI	0	8	31	538	291	218	149	115	52	32	10	12	1456
UI	0	994	0	0	0	66	0	0	0	0	0	0	1060
VI	0	0	40	38	11	5	5	9	4	0	1	3	116
OI	0	396	0	0	2	0	1	0	0	0	0	0	399
ÄI	0	94	0	0	0	0	2	1	0	0	0	0	97
Lõpp -I	1313	3149	1160	2176	1548	927	632	482	303	122	91	163	12 066
AK	0	0	4	6	2	2	0	0	1	0	1	1	17
CK	0	0	3	0	0	2	1	0	0	0	0	0	6
EK	0	0	2	3	0	1	1	0	4	2	0	1	14
HK	0	60	8	0	0	1	4	1	0	0	1	0	75
IK	0	0	245	18	98	100	60	21	12	23	10	17	604
KK	0	0	57	0	0	0	2	2	0	0	0	1	62
LK	0	0	48	0	0	0	1	4	0	1	0	0	54
NK	0	0	5	1	0	2	2	1	0	0	0	0	11
OK	0	0	5	0	0	0	0	0	0	0	0	0	5
RK	0	1	18	0	0	5	4	1	2	1	2	0	34
SK	0	3	10	15	0	3	2	1	2	1	0	1	38
TK	1	0	8	0	0	0	0	0	0	1	0	0	10
ÜK	0	1	8	0	28	1	3	0	0	1	1	0	43
ÖK	0	0	1	0	0	0	0	0	0	0	0	0	1
ÄK	0	0	1	0	0	0	0	1	0	0	0	0	2
ÖK	0	0	3	0	0	0	1	0	0	0	0	0	4
ÜK	0	0	0	0	0	0	2	0	0	0	0	0	2
Lõpp -K	1	65	426	43	128	117	83	32	21	30	15	21	982
AL	0	175	363	225	227	116	58	49	40	19	20	20	1312
EL	1	74	382	236	299	120	175	116	42	54	19	52	1570

Sõna- lõpp	Sõnavormi pikkus (tähtedes)												Kokku
	2	3	4	5	6	7	8	9	10	11	12	>12	
IL	0	17	84	20	110	49	31	17	24	19	6	12	389
LL	0	128	293	3	4	3	7	3	0	3	2	0	446
OL	0	14	61	4	2	11	14	3	8	1	0	0	118
UL	0	85	47	123	37	48	57	15	18	8	5	19	462
AL	0	0	21	0	0	0	0	1	0	0	0	1	23
OL	0	14	3	0	0	2	0	0	1	0	0	0	20
Lõpp - L	1	507	1254	611	679	349	342	204	133	104	52	104	4340
AM	0	0	159	8	12	29	6	4	3	3	0	1	225
EM	0	1	2	121	198	13	4	18	6	4	1	1	369
HM	1	0	14	0	0	0	0	1	0	0	0	0	16
IM	0	0	15	5	6	2	0	0	0	0	0	0	28
LM	0	10	75	1	7	1	14	2	0	0	1	0	111
MM	0	3	10	2	0	6	3	3	1	0	0	0	28
OM	0	0	9	0	1	2	1	3	0	1	0	0	17
RM	0	0	18	0	0	0	0	2	0	2	0	1	23
SM	0	0	0	0	1	2	2	2	0	0	1	0	8
TM	0	0	1	0	0	0	0	0	0	0	0	0	1
UM	0	3	4	1	1	1	4	2	7	0	2	1	26
OM	0	0	7	0	0	1	0	0	0	0	0	1	9
Lõpp - M	1	17	314	138	226	57	34	37	17	10	5	5	861
AN	0	1	46	50	34	6	14	4	1	0	0	2	158
EN	0	0	53	16	54	3	5	2	1	0	1	0	135
HN	0	0	6	0	0	0	0	0	1	0	1	0	8
IN	0	19	277	88	107	258	204	71	52	12	5	2	1095
MN	0	0	1	0	0	0	0	0	0	0	0	1	2
NN	1	25	17	0	2	16	3	5	2	0	0	0	71
ON	1272	5	6	1	4	4	24	6	3	2	2	7	1336
RN	0	3	4	0	0	0	1	0	0	0	1	0	9
UN	0	0	14	21	6	2	2	0	1	0	1	0	47
AN	0	0	2	0	0	0	0	0	0	0	0	0	2
ÜN	0	0	2	0	0	1	0	0	0	0	0	0	3
Lõpp - N	1273	53	428	176	207	290	253	88	61	14	11	12	2866
AO	0	11	0	0	0	0	1	0	0	0	0	1	13

Sõna- lõpp	Sõnavormi pikkus (tähtedes)												Kokku
	2	3	4	5	6	7	8	9	10	11	12	≥12	
UR	0	1	61	3	3	1	4	2	2	6	0	2	85
OR	0	0	2	0	0	0	0	0	0	0	0	0	2
AR	0	0	1	0	0	0	1	1	0	0	0	0	3
OR	0	0	2	0	0	1	2	1	1	0	0	1	8
ÜR	0	1	1	0	0	1	1	1	1	0	0	0	6
Lõpp - R	2	2	158	157	107	37	28	17	19	16	2	18	563
AS	0	201	173	383	1003	333	348	132	72	40	17	22	2724
ES	0	372	349	291	515	296	343	185	116	61	31	68	2627
IS	0	490	737	208	1046	159	46	39	39	34	21	37	2856
KS	0	158	267	404	329	441	137	141	118	73	57	94	2219
NS	0	2	2	0	0	0	0	0	0	0	0	0	4
OS	0	1	93	5	6	10	8	2	0	1	0	1	127
PS	0	0	17	0	0	0	0	13	2	0	0	1	33
RS	0	0	1	0	0	0	0	1	0	0	0	0	2
SS	0	4	35	59	1	1	4	4	3	4	0	2	117
TS	0	3	20	10	5	10	7	6	2	9	1	2	75
US	0	189	116	354	600	186	96	62	34	30	24	61	1752
AS	0	0	1	0	0	0	0	0	0	0	0	0	1
OS	0	1	0	0	0	0	0	0	1	0	0	0	2
OS	0	0	2	0	0	1	0	0	0	0	0	1	4
Lõpp - S	0	1421	1813	1714	3505	1437	989	585	387	252	151	289	12 543
AT	0	0	17	30	148	151	88	32	36	20	14	27	563
ET	1218	2	21	87	76	21	4	8	16	15	8	8	1484
HT	0	16	32	0	1	1	3	12	1	1	1	1	69
IT	0	16	46	10	53	62	31	13	6	19	7	14	277
KT	0	0	3	0	0	2	1	1	0	0	0	2	9
LT	0	31	14	171	358	305	381	175	209	152	72	141	2009
NT	0	68	12	5	41	2	4	8	3	6	0	1	150
OT	0	0	2	8	6	1	0	0	0	0	0	0	17
RT	0	0	50	33	18	7	2	5	1	0	0	0	116
ST	0	5	487	307	561	803	374	423	274	223	127	302	3886
TT	0	9	70	0	4	10	7	8	6	1	0	4	119
UT	0	15	8	82	7	19	35	11	8	4	7	9	205
ÖT	0	0	1	0	0	0	1	0	0	0	0	0	2

Eõna- lõpp	Sõnavormi pikkus (tähtedes)												Kokku
	2	3	4	5	6	7	8	9	10	11	12	>12	
Lõpp -T	1218	162	763	733	1273	1384	931	696	560	441	236	509	8906
AU	4	1	0	1	0	0	0	0	1	0	0	0	7
BU	0	0	5	14	0	1	0	2	2	0	0	0	24
CU	0	0	0	0	0	0	1	0	0	0	0	0	1
DU	0	5	29	96	2	2	10	9	8	4	2	1	168
EU	0	0	0	1	0	0	0	0	0	0	0	0	1
GU	0	1	775	76	85	113	29	32	17	5	5	4	1142
HU	0	3	77	0	1	2	1	0	0	0	0	0	86
IU	0	0	35	1	0	0	0	0	3	0	0	0	39
JU	123	3	76	143	0	2	10	7	7	5	0	1	377
KU	0	0	30	156	74	61	48	49	34	23	17	31	523
LU	0	83	52	37	3	8	13	18	6	6	5	2	233
MU	110	0	30	67	0	2	10	5	5	0	0	4	233
NU	0	8	150	26	3	3	2	6	17	3	0	2	220
OU	0	1	0	0	0	0	0	0	0	0	0	0	1
PU	0	0	7	18	0	1	0	0	0	1	0	1	28
RU	0	49	26	20	14	6	8	3	6	2	1	1	136
SU	15	14	23	27	36	2	3	6	6	5	2	1	140
TU	0	0	52	317	42	55	42	22	13	16	9	11	579
UU	0	51	0	2	0	12	6	6	3	4	0	0	84
VU	0	0	0	25	0	0	1	1	0	0	0	1	28
OU	3	9	0	0	2	2	1	0	0	0	0	0	17
ÜU	0	0	0	0	0	0	1	0	0	0	0	0	1
Lõpp - U	255	228	1367	1027	262	272	186	168	128	74	41	60	4068
AV	0	0	13	28	35	28	31	10	3	2	3	5	158
EV	0	0	30	21	15	19	6	12	3	3	2	1	111
HV	0	0	6	0	0	0	1	0	0	0	0	0	7
IV	0	0	11	4	7	6	0	0	5	1	0	1	35
LV	0	0	8	0	0	0	0	2	0	0	0	0	10
OV	0	0	4	2	0	1	2	0	0	0	0	0	9
RV	0	1	11	0	0	0	1	1	1	2	0	0	17
SV	0	1	3	0	0	0	0	0	0	0	0	0	4
UV	0	0	0	6	13	1	0	0	0	3	0	0	23

Sõna- lõpp	Sõnavormi pikkus (tähtedes)												Kokku
	2	3	4	5	6	7	8	9	10	11	12	>12	
Lõpp - V	0	2	86	61	70	55	41	25	11	11	5	7	374
AA	0	10	0	0	1	0	0	0	0	0	0	0	11
Lõpp - A	0	10	0	0	1	0	0	0	0	0	0	0	11
OO	7	25	0	1	4	2	3	0	1	1	0	0	44
Lõpp - O	7	25	0	1	4	2	3	0	1	1	0	0	44
UU	0	2	0	2	0	0	0	0	0	0	0	0	4
Lõpp - U	0	2	0	2	0	0	0	0	0	0	0	0	4
KOKKU	10180	9839	15715	13899	13155	9552	7236	4925	3238	2057	1257	2367	93420
%	10,90	10,53	16,82	14,88	14,08	10,22	7,75	5,27	3,47	2,20	1,35	2,53	100,00

Tabel 2

Sõnavormide lõpptähed eestikeelses tekstis ja sõnastikus (tänapäeva ilukirjandusproosa autorikõne põhjal). Teksti maht N = 93420 sõnet, sõnastiku maht V = 26025 sõnavormi.

Lõpp- täht	Tekstis		Sõnastikus*	
	arv	%	arv	%
a	19690	21,08	3657	14,05
b	1968	2,11	582	2,24
d	10244	10,96	3887	14,94
e	12689	13,58	4431	17,03
f	9	0,01	3	0,01
g	635	0,68	87	0,33
h	122	0,13	17	0,07
i	12066	12,92	2367	9,10
k	982	1,05	334	1,28
l	4340	4,65	1172	4,50
m	861	0,92	223	0,86
n	2866	3,07	501	1,93
o	297	0,32	68	0,26
p	138	0,15	42	0,16
r	563	0,60	174	0,67
s	12543	13,43	3780	14,52
t	8906	9,53	3383	13,00
u	4068	4,35	1064	4,09
v	374	0,40	234	0,90
ä	11	0,01	2	0,01
õ	44	0,05	14	0,05
ü	4	0,004	3	0,01
Kokku	93420	100,0	26025	100,0

* Andmed sõnastiku kohta on saadud käesolevas kogumikus avaldatud pöördtsagedussõnastiku põhjal.

О СООТНОШЕНИИ КОНЦА СЛОВА И ДЛИНЫ СЛОВА В ЭСТОНСКОМ ТЕКСТЕ

Ü. Kaasik, J. Tuldava

Резюме

В статье приводятся данные о статистике конца слова в эстонском тексте в комбинации с данными о длине слова, измеряемой количеством букв. Наиболее частотной формой в эстонском тексте является словоформа, оканчивающаяся на букву a и состоящая из 2 букв (обозначение $-a/2$). Такие формы вместе взятые покрывают 6,33 % текста (примеры: ja 'и', ta 'он, она, его, ее', ka 'тоже', ma 'я', sa 'ты'). Далее следуют формы $-a/4$ (3,83%), $-a/6$ (3,75%), $-e/5$ (3,38%), $-i/3$ (3,37%). Десять наиболее частотных форм (типов) покрывают 32,38 %, двадцать наиболее частотных форм - 48,83 % текста. Полный обзор частот комбинаций конечных букв и пар конечных букв с длиной словоформ см. табл. 1. Самыми частотными конечными буквами словоформ в тексте являются a, e, s, i, d, t, l, n, u, b; в соответствующем словаре - e, d, a, a, i, i, i, u, b, a (см. табл. 2). Средняя длина словоформ в тексте - 5,8 букв. Все данные получены на основе текстов из авторской речи современной эстонской художественной прозы (объем выборки - 93420 словоупотреблений, объем словаря - 26025 словоформ).

ON THE RELATION BETWEEN WORD ENDING AND WORD LENGTH IN ESTONIAN TEXTS

Ü. Kaasik, J. Tuldava

Summary

The article presents statistical data on word ending in combination with word length (measured in letters) in Estonian texts of fiction (non-conversational material only). The most frequent form in the text is the word form ending in the letter a and consisting of 2 letters (marked as $-a/2$). Such forms account for 6.33 per cent of all words in the text (examples: ja 'and', ta 'he, she, his, her', ka 'too', ma 'I', sa 'you' (2nd person singular)). The following most frequent forms are $-a/4$ (3.83 %), $-a/6$ (3.75 %), $-e/5$ (3.38 %), $-i/3$ (3.37 %). The ten most frequent forms make up 32.38 %, the twenty most frequent forms - 48.83 % of all word forms in the text. A complete survey about the frequencies of the combinations of one or two ending letters and the word length, see in Table 1. The most frequent ending letters in the text are: a, e, s, i, d, t, l, n, u, b; in the corresponding vocabulary: e, d, a, a, i, i, i, u, b, a (see Table 2). The mean length of the word in the text is 5.8 letters. The volume of the sample - 93,420 words (tokens), the volume of the vocabulary - 26,025 word forms (types).