

ТАРТУСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

На правах рукописи

ТУЛДАВА Юхан Артурович

УДК 519.766.4:809.454.5-3

ПРОБЛЕМЫ И МЕТОДЫ КВАНТИТАТИВНО-СИСТЕМНОГО
ИССЛЕДОВАНИЯ ЛЕКСИКИ

(на материале эстонского языка)

П Р И Л О Ж Е Н И Е

10.02.07 - Финно-угорские языки

10.02.21 - Структурная, прикладная и
математическая лингвистика

Диссертация на соискание ученой степени
доктора филологических наук

TARTU ÜLIKOOLI
EESTI KEELE ÕPPETOOL

ТАРТУ - 1984

П Р И Л О Ж Е Н И Е

А. Списки слов и других лингвистических единиц с частотными данными	
(А-I - А-81)	455
Б. Статистические таблицы	
(Б-I - Б-73)	580

Базовый словарь: 1200 наиболее частотных лексем (частота $F \geq 10$ в тексте объемом $N = 100\,000$) по данным частотного словаря авторской речи современной художественной прозы. Крестиком отмечены древние слова (возраст более 800 лет). В скобках слова, встретившиеся только в I или 2 подвыборках из 20 (распростр. $R \leq 2$).

	4237	+olema	218	+tundma
	3493	+ja	209	kas
	2598	+ta	207	+vastu
	1981	+see	206	+ära
	1395	+ei	200	+hea
	1300	+et	198	+välja
	1047	+kui	195	+tahtma
	879	+mis	194	+mitte
	845	+ma	192	+küll
10.	827	+tema	60.	190 pärast
	724	aga	189	+mõni
	634	+oma	185	+istuma
	613	+mina	185	+naine
	581	+ise	181	+andma
	568	+nagu	180	rääkima
	499	+minema	179	sest
	496	+tulema	176	+vana
	493	+siis	175	+panema
	465	+saama	174	+kus
20.	448	+kes	70.	174 +mõtlemä
	436	+nii	171	+käima
	434	ka	167	+ega
	428	+ning	166	+enam
	382	+kõik	165	+pääv
	373	+mees	164	+seisma
	366	+või (konj.)	164	+siin
	350	+üks	163	+asi
	339	+tegema	163	+pea
	327	+veel	159	+nägu
30.	309	kuid	80.	153 +seal
	304	+hakkama	151	+aasta
	304	+jääma	150	+sa
	304	+teine	149	+kuidas
	297	+vaatama	148	+peale
	285	+aeg	146	+tagasi
	285	+teadma	144	+läbi
	285	+võima	139	+kord
	272	+miski	138	+kaks
	272	+suur	136	+all
40.	268	+nägema	90.	135 +keegi
	264	+pidama (pidi)	133	+jõudma
	261	+võtma	133	+kõige
	260	+silm	131	+iga (pron.)
	254	+inimene	131	+vaid
	248	+juba	130	isegi
	238	+nüüd	130	+maa
	237	+käsi	126	ikka
	234	+ainult	125	+jalg
	230	+üle	125	+ju
50.	223	+ütlemä	100.	123 +sõna

Таблица А-I
(продолжение)

	123	+mingi		87	+kodu
	122	+elu		86	+kuulma
	122	+poiss		86	vist
	120	+poole		85	+juures
	116	+jälle		85	pilk
	113	+mööda		85	tuba
	112	+meri		84	+praegu
	112	+vahel		84	+sisse
	110	+kogu (pron.)		83	+pidama
110.	110	+ootama	160.	82	+noor
	109	väga		82	võib-olla
	108	mõte		81	+taga
	107	+elama		81	+tõusma
	107	+esimene		79	alles
	107	+palju		79	+edasi
	106	+enne		79	+linn
	105	+astuma		79	siiski
	105	+tõõ		79	+tooma
	105	+uus		78	+kaduma
120.	104	+leidma	170.	78	+lõõma
	101	+alla		78	+meel
	101	+isa		76	+alati
	101	+üles		76	+hoidma
	101	+ümber		76	+kokku
	100	+ajama		76	+koos
	100	+selg		76	kuigi
	99	+juurde		76	+oskama
	98	+ees		76	paar
	97	+laud		75	+kinni
130.	97	muutuma	180.	75	kohe
	97	+pikk		75	märkama
	96	sõitma		75	+pääke
	94	+maha		74	+aken
	94	väike		74	+koht
	94	+õhtu		74	+küla
	93	+ema		74	+laps
	93	+jutt		73	kunagi
	93	niisugune		73	+otsima
	93	+viimane		73	+paistma
140.	92	+arvama	190.	73	+püüdma
	92	just		72	+eest
	91	hoopis		72	ometi
	91	+laskma		72	+selline
	91	+rohken		71	+hall (adj.)
	90	+ette		71	hetk
	90	+tee (liikl.)		71	juhtuma
	90	+täis		71	+lõpp
	89	+hobune		71	+mets
	89	muidugi		71	+muu
150.	88	+maja	200.	71	peaaegu

Таблица А-I
(продолжение)

	71	+viima		53	+uskuma
	70	pisut		53	+vend
	70	+tõdruk		52	+aru (mõistus)
	69	ent		52	+sama
	69	+hääli		52	+sinna
	68	+vesi		52	+suu
	67	+uks		52	tekkima
	67	uesti		52	+õõ
	66	+jätma		52	ükski
210.	66	+tuul	260.	51	+hilja
	64	+ilus		51	+kohal
	64	+kõsima		51	+miks
	64	+mitu		51	+must
	64	+nimi		51	+näima
	64	+too (pron.)		50	+kuulama
	63	+jooksma		50	täiesti
	63	+kolm		49	+kuhu
	63	+poeg		49	+puu
	62	+mõistma		49	+suutma
220.	61	otse	270.	49	+terve
	61	+võõras		48	+algama
	60	+lugema		48	+keel
	60	päris		48	+liiga
	60	+sina		48	+saatma
	60	+taevas		48	+õde
	60	+tõmbama		48	üksteise
	60	tänav		47	kõrval
	59	hommik		47	poolt
	59	järele		47	+punane
230.	59	lahti	280.	47	+selge
	59	+ots		47	+õlg (gen. õla)
	59	+tõstma		46	tõesti
	59	üldse		46	+täna
	58	+raske		46	+tühi
	58	+süda		46	üsna
	58	+valge		45	jaoks
	57	+kasvama		45	+jõud
	57	tunne		45	+kivi
	56	jah		45	+kohta
240.	56	+küllap	290.	45	+külm
	56	+maailm		45	lihtsalt
	55	+ehk		45	+liikuma
	55	+õige		45	tund
	54	kirjutama		45	vahtima
	54	+kord		44	kõrvale
	54	+pool (num.)		44	mõnikord
	54	tunduma		44	peremees
	54	+vanker		44	+puhas
	53	kuidagi		44	+süa
250.	53	+laev	300.	44	+äkki

Таблица А-I
(продолжение)

310.

43 +hästi
43 imelik
43 +kaasa
43 +lumi
43 +sõuma
43 +viskama
42 kostma
42 +lubama
42 +naerma
42 +peal
42 +soe
42 õu
41 ikkagi
41 +kartma
41 +kuni
41 meeldima
41 otsekui
41 ringi
41 seesama
320. 41 +taha
41 +vajuma
41 +vastama
41 õieti
40 järgi
40 +kael
40 +kandma
40 +kust
40 +kutsuma
40 +kõrge
330. 40 +sein
40 +suruma
40 +vaja
40 +vari
39 eriti
39 +kaua
39 +kell
39 kirik
39 +paluma
39 +samm
340. 38 +asuma
38 +rahvas
38 +sealt
37 +heitma
37 +ilma
37 ilmuma
37 +lugu
37 niisama
37 perenaine
37 +rõõm
350. 37 talu

360.

37 teineteise
37 +unustama
37 vaikima
36 +aitama
36 +avama
36 huvitama
36 +juus (juuksed)
36 +püsti
36 sellepärast
36 +sõda
36 tähendama
36 veidi
35 haarama
35 +kerge
35 kindel
35 +langema
35 +lendama
35 +raha
35 +sattuma
370. 35 +tugev
35 tõepoolest
35 +valgus
35 varsti
35 +värav
34 +hing
34 +kitsas
34 +lõppema
34 naeratama
34 +osa
380. 34 pilt
34 +sinine
33 +abi
33 +hüüdma
33 +kiri
33 +lukkama
33 +mure
33 +mängima
33 niiviisi
33 +nina
390. 33 otsustama
33 paat
33 +palju
33 +valmis
33 vanamees
32 +hall
32 kilomeeter
32 +kuu
32 +rind
32 tuttav
400. 32 +õigus

Таблица А-І
(продолжение)

410.	31	+alt	460.	28	+kollane
	31	+ammu		28	+mõlema(d)
	31	+ilm		28	+neli
	31	koguni		28	pakkuma
	31	kukkuma		28	+põõrama
	31	+küsimus		28	+sõber
	31	+madal		28	+täht
	31	+magama		28	+vahele
	31	natuke(ne)		28	+üksik
	31	nõnda		27	+eit
	31	+nõudma		27	huvi
	31	+pihk		27	+jumal
	31	vähemalt		27	järel
	31	õhk		27	järsku
	31	+ääres		27	+kauge
420.	31	+üksi		27	+kevad
	30	+kaugele		27	kinnitama
	30	kippuma		27	kuskil
	30	+lai		27	+kõndima
	30	laiali	470.	27	+lõpuks
	30	+lehm		27	missugune
	30	+nimetama		27	mäletama
	30	+nurk		27	+paks
	30	+ostma		27	pull (sõnn)
	30	pisike		27	+rahu
	30	ruttama		27	+riie
	30	+tuhat		27	+rohi
	30	vaevalt		27	ronima
	30	vaikne		27	sakslane
	29	+armastama	480.	27	+seletama
	29	doktor		27	surema
	29	+jooma		27	säärane
	29	jälgima		27	tagant
	29	loom		27	+talv
430.	29	+lühike(ne)		27	+tohtima
	29	meenutama		27	+vihm
	29	mõõduma		27	üksnes
	29	+nõdal		26	härja
	29	+nõitama		26	igaüks
	29	+paik (koht)	490.	26	+jõgi
	29	raamat		26	+kala
	29	+siduma		26	+kallas
	29	+siit		26	+katus
	29	+surm		26	+keha
	29	täpselt		26	+lootma
	29	tükk		26	meenuma
	29	+vaba		26	+muidu
	29	+vedama		26	+pime
	28	+hele		26	pool (pp)
440.	28	igatahes		26	+põhi
450.			500.		

Таблица А-І
(продолжение)

	21	+sees		19	proovima
	21	silmitsema		19	+põlema
	21	+sobima		19	+ratas
	21	soov		19	saabuma
	21	+sundima		19	seejärel
	21	toimuma		19	+serv
	21	tõmbuma		19	+tuli
	21	+tõsi		19	+täitma
	21	+tõsine		19	vaene
610.	21	vaikus	660.	19	vanaema
	21	viis (subst.)		19	vanaisa
	21	+(õng)		19	veider
	20	algul		19	+õnn
	20	armastus		19	+õppima
	20	igasugune		19	üleni
	20	justkui		18	ainuke(ne)
	20	+kihutama		18	arutama
	20	+kiiresti		18	eemal
	20	kujutama		18	+hein
620.	20	+lahkuma	670.	18	+hoolimata
	20	+leht		18	+hull
	20	+leib		18	ial(gi)
	20	lõhnama		18	+kakskümmend
	20	minut		18	kartul
	20	+muna		18	(katel)
	20	+riippuma		18	kerkima
	20	ruum		18	+koer
	20	+segama		18	korraga
	20	+sündima		18	+lamama
630.	20	+taas	680.	18	+laulma
	20	tilluke(ne)		18	müür
	20	tõsiselt		18	+nahk
	20	tütarlaps		18	+paras
	20	uurima		18	+pilv
	20	+vastus		18	põhjus
	20	+õlu		18	rahulik
	20	+ära (verb)		18	rida
	19	ehitama		18	+seitse
	19	hirmus		18	sirutama
640.	19	+häda	690.	18	sugugi
	19	kapten		18	+sulama
	19	keset		18	+sõrm
	19	klaas		18	särama
	19	+kolmas		18	tasku
	19	(kurat)		18	+tegemine
	19	küün		18	tõttu
	19	+lähedale		18	tõõtama
	19	milleks		18	+ulatuma
	19	nähtavasti		18	+vahetama
650.	19	+pesa	700.	18	vahetevahel

Таблица А-І
(продолжение)

710.	18	+veerema	16	+kadakas
	18	+vili	16	+kallis
	17	+aed	16	+kaotama
	17	ehkki	16	+kask
	17	esialgu	16	kena
	17	+julgema	16	kuhugi
	17	+järv	16	kujutlema
	17	+kass	16	kusagil
	17	+kestma	16	+kuusk
	17	+kiitma	760.	16 kÜlaline
	17	kindlasti	16	+liiv
	17	klass	16	+lõpetama
	17	korraaks	16	+lähedal
	17	korter	16	+mänd
	17	laskuma	16	+nõu (nõuanne)
720.	17	lause	16	oh (interj.)
	17	mood	16	olukord
	17	mälestus	16	ootamatult
	17	+märg	16	otsas (pp)
	17	määrama	770.	16 +paremini
	17	müts (peakate)	16	peatuma
	17	nimelt	16	poisike(ne)
	17	paber	16	poolest (pp)
	17	parajasti	16	pruun
	17	(parun)	16	+puhuma
	17	perekond	16	+püsima
	17	+pimedus	16	rong
	17	(puhk)	16	(saag)
	17	+pääsema	16	saal
730.	17	saatus	780.	16 +saksa (adj.)
	17	sageli	16	selguma
	17	+soovima	16	+(siil) (gen. siil
	17	+sügav	16	sõit
	17	+tahtmine	16	sündmus
	17	+tarvis	16	teenima
	17	+tume	16	+tegu
	17	+tungima	16	+tihe
	17	+(vanne)	16	tõeline
	17	+vasak	16	+ujuma
740.	17	võimalik	790.	16 unenägu
	17	+vähem	16	+vahe
	17	ühine	16	vahepeal
	16	ajalugu	16	+vajama
	16	asetama	16	valu (aisting)
	16	eemale	16	+voodi
	16	eriline	16	voolama
	16	+hiir	16	värske
	16	hābenema	16	õnnelik
	16	+hūppama	16	+äärde (pp)
750.	16	jahe	800.	16 üksinda

Таблица А-I
(продолжение)

810.	16	Ulesanne	860.	14	lapsepõlv
	15	+algus		14	lõbus
	15	eks		14	+muutma
	15	erinema		14	+müüma
	15	+hingama		14	peen
	15	järjest		14	peitma
	15	+kaits(e)ma		14	+peksma
	15	kasutama		14	+pistma
	15	kavatsema		14	puhkama
	15	kujutlus		14	+puutuma
	15	+kuluma		14	+põrand
	15	+kuna		14	+põsk
	15	+kuum		14	+pärima
	15	kõõk		14	+püha
	15	(metsavaht)		14	+sadama
820.	15	+muld	870.	14	+sugulane
	15	mäng		14	surnu
	15	nukker		14	sõõgilaud
	15	nõrk		14	tajuma
	15	+näpp		14	+viha
	15	+oks		14	viimaks
	15	pood		14	+väin
	15	praegune		14	värv
	15	+sadam		14	+väsima
	15	süüdi (adv.)		14	+õhtuti
	15	+tapma		14	+õige
	15	tegelikult	880.	13	alustama
	15	tookord		13	amet
	15	vagun		13	arvamus
	15	vaikselt		13	eesti (adj.)
	15	õõsel		13	+eile
	14	ajaleht		13	hämärus
	14	arvatavasti		13	(hävituspataljon)
	14	+habe		13	+jagama
	14	+hind		13	+jagu
	14	+homme		13	juht (gen. juhu)
	14	igatsema		13	jutustama
	14	juhtima	890.	13	+kajakas
	14	kaaslane		13	kannul
840.	14	katsuma		13	kasv
	14	+kehv		13	+kaudu
	14	kergitama		13	+keelama
	14	+king		13	kleit
	14	kodune		13	konks
	14	kogunema		13	korralik
	14	kolhoosnik		13	+kott (gen. koti)
	14	+kukal		13	kraav
	14	kõlama		13	+kuivama
	14	+lagi		13	kummaline
	14	(lant)	900.	13	+kõht

Таблица А-І
(продолжение)

	13	+laht		12	+auk
	13	+lausuna		12	enam-vähem
	13	+libisema		12	esimees
	13	lill		12	+hool
	13	loodus		12	hotell
	13	+looma		12	juurest (pp)
	13	lõhn		12	kaabu
	13	+milline		12	+kange
	13	muigama		12	kasu
910.	13	muusika	960.	12	kiire
	13	noh		12	kino
	13	näide		12	kohvik
	13	omavahel		12	kurb
	13	pagan		12	kõigepealt
	13	+petma		12	kõigest
	13	+pigem		12	kõrb (gen. kõrbe)
	13	pigistama		12	+kõva
	13	puudutama		12	+kübar
	13	+pühkima		12	+labidas
920.	13	püss	970.	12	+laduma
	13	raputama		12	+lage
	13	rumal		12	+liigutama
	13	rõõmustama		12	lähedus
	13	sammuma		12	+magus
	13	seejuures		12	meeleolu
	13	silmapilk		12	naaber
	13	+sirge		12	niimoodi
	13	suhe		12	number
	13	+suits		12	(ookean)
930.	13	sulgema	980.	12	+org
	13	suudlema		12	paremal
	13	+sügis		12	piinama
	13	teade		12	+pikali
	13	tekk (vaip)		12	+puri
	13	tihti		12	rabelema
	13	+tiib		12	+rada
	13	torkama		12	rajama
	13	+tunnistama		12	ring
	13	+tütar		12	salapärase
940.	13	+ulatama	990.	12	+saun
	13	(vader)		12	teadus
	13	vaidlus		12	+tervis
	13	+viga		12	tuju
	13	viis (gen. viisi)		12	(vaeras) (gen. vaera)
	13	väljak		12	+vagu
	13	väljuma		12	vaidlema
	13	+värisema		12	+veeretama
	13	ükskord		12	+venima
	13	ükskõik		12	võimalus
950.	12	+armas	1000.	12	(võllas)

Таблица А-І
(продолжение)

	12	+vägev		11	pressima
	12	+väljas		11	probleem
	12	üpris		11	pugema
	11	film		11	+puuduma
	11	foto		11	põgenik
	11	hindama		11	põrmugi
	11	hulgas		11	+põõsas
	11	+hõmar		11	päriselt
	11	iseenest		11	püksid
1010.	11	joonistama	1060.	11	rahulikult
	11	järgnema		11	+saar (geogr.)
	11	+kaar		11	sealsamas
	11	+kahvatu		11	+sekka
	11	kalduma		11	+säär
	11	kallal (pp)		11	sõandama
	11	+kaunis		11	+taguma
	11	kergelt		11	+talvine
	11	+keskmine		11	tekitama
	11	+kevadine		11	toon (gen. tooni)
1020.	11	kiirustama	1070.	11	tormama
	11	kirg		11	tugitool
	11	korralikult		11	tähelepanu
	11	korv		11	tänane
	11	+kumbki		11	tänapäev
	11	+kõrvuti		11	tühjus
	11	+kõsk		11	uhke
	11	+kükitama		11	(usuühing)
	11	küngas		11	+vaenlane
	11	legendik		11	+vaim
1030.	11	lahtine	1080.	11	+valama
	11	(loss)		11	+võim
	11	luik		11	+võimas
	11	+lääkima		11	(õemees)
	11	maantee		11	õpetaja
	11	madrus		11	+ärkama
	11	+millal		11	üldiselt
	11	millegipärast		11	üllatama
	11	muide		11	ümmargune
	11	muretsema		10	aeglaselt
1040.	11	muusikaline	1090.	10	+ait
	11	mõis		10	armuma
	11	mõnu		10	asjalik
	11	naeratus		10	daam
	11	+noogutama		10	ebamäärane
	11	olgu(gi)		10	eelmine
	11	oluline		10	enamasti
	11	peegel		10	enamik
	11	+(piht) (gen. piha)		10	ettepoole
	11	piinlik		10	halvasti
1050.	11	plaan	1100.	10	hoog

Таблица А-І
(продолжение)

10	hunnik	10	+pind (gen. pinna)
10	hõre	10	punetama
10	+iga (subst.)	10	raadio
10	igavene	10	(raehärra)
10	ilme	10	+rand
10	juhuslik	10	rikkuma
10	+jälg	10	+saba
10	+jää	10	sajand
10	(kajut)	10	sarnanema
1110.	10 kauaks	1160.	10 +sarv
10	+kaup	10	+seas (pp)
10	kaupa (pp)	10	selgesti
10	kenasti	10	selleks
10	kobama	10	sigaret
10	kohaselt	10	silmapiir
10	+kohkuma	10	sisemine
10	kohtuma	10	+soolane
10	kohver	10	sulane
10	kombel (pp)	10	sõbralik
1120.	10 korsten	1170.	10 +sõdur
10	+kuldne	10	sõjavägi
10	+kuri	10	säng
10	+kuus (num.)	10	sööstma
10	kõikuma	10	+süli
10	kõrgus	10	taganema
10	+kätte (pp)	10	taluma
10	küljes	10	talumees
10	lahing	10	tavaline
10	+lammast	10	+turi
1130.	10 +leppima	1180.	10 tõde
10	+liikumine	10	+tõukama
10	lisama	10	+uinuma
10	+lõhkuma	10	(uss)
10	lähedalt	10	uudis
10	+lähem	10	vahest
10	+maitse(s)ma	10	valitsema
10	miljon	10	vanasti
10	mingisugune	10	varandus
10	mitmesugune	10	+vastast
1140.	10 +murdma	1190.	10 +veri
10	mõnus	10	+vihane
10	niipalju	10	+viiskümmend
10	omaette	10	võrdlema
10	+onu	10	Siendama
10	ootamatu	10	+õnnetu
10	otsast	10	õpetama
10	peamine	10	õrn
10	(perepoeg)	10	äkitselt
10	(pidalitõbine)	10	ämbre
1150.	10 pikkamööda	1200.	10 +äär
		10	ülikond
		10	ümar

Таблица А-2
(гл. III, § I)

Фрагмент распределительного алфавитно-частотного словаря словоформ эстонского языка (F - суммарная абсолютная частота, R - коэф. распространения, D - коэф. стабильности, U - коэф. употребительности)

Слово- форма	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	F	R	D	U
	AB	VG	AN	HK	JK	PK	LP	VS	HS	RS	MT	EV	AK	TK	JP	JT	AV	MU	K/N	JS				
aasta (nom.)	0	0	2	1	0	0	0	0	0	0	1	0	0	1	3	0	1	1	0	1	11	8	0,677	7
aastaajad	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
aastaajal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	1	0	0
aastad	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	1	0	0	0	1	8	4	0,369	3
aastaga	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	0	0
aastaid	0	0	2	2	0	0	0	2	2	0	2	0	1	2	2	0	1	1	1	0	18	11	0,780	14
aastaks	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0,319	1
aastakümneid	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	2	0,319	1
aastal	0	0	2	0	0	0	1	0	2	1	2	1	1	0	0	0	2	1	5	1	19	11	0,721	13
aastast	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	4	2	0,227	1
aasta (gen.)	1	1	0	0	0	0	2	0	7	0	0	0	1	3	2	0	0	0	5	1	23	9	0,644	14
aastat	1	3	3	1	0	1	1	3	4	2	3	0	3	4	3	5	1	2	2	1	43	18	0,844	35
aastate	0	0	0	0	0	0	0	0	0	1	0	0	1	0	2	0	0	0	0	2	9	4	0,572	5
aastatega	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	2	2	0,319	1
aastatel	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	3	6	3	0,464	3
aastates	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	3	3	0,464	1

Таблица А-3
(гл. III, § I)

Фрагмент распределительного рангового ЧС словоформ эстонского языка
(i - ранг слова, F - суммарная абсолютная частота, D - коэффициент ста-
бильности, U - коэффициент употребительности)

Слово- 1 форма	1 AB	2 VG	3 AH	4 HK	5 JK	6 PK	7 LP	8 VS	9 HS	10 RS	11 MT	12 EV	13 AK	14 TK	15 JP	16 JT	17 AV	18 MU	19 K/N	20 JS	F	D	U
1 ja	137	129	141	174	283	118	171	171	155	80	168	131	185	169	121	147	177	186	181	197	3221	0,974	3138
2 ta	98	78	112	45	75	95	76	123	25	68	123	78	71	127	37	87	138	69	27	50	1602	0,931	1491
3 on	23	54	21	134	22	86	29	76	15	28	163	118	76	41	119	50	79	71	144	90	1439	0,825	1187
4 ei	36	80	56	71	24	118	97	108	74	80	78	80	65	49	71	64	76	66	32	50	1375	0,943	1297
5 et	53	47	54	67	44	114	100	95	51	56	45	71	84	64	47	60	78	70	31	33	1264	0,938	1186
6 oli	43	66	122	7	84	28	69	27	91	29	21	52	56	107	32	74	112	51	11	34	1116	0,830	926
7 kui	51	54	52	32	56	36	62	64	35	64	55	41	40	30	77	56	44	37	57	52	995	0,956	951
8 see	11	39	32	33	23	11	42	48	22	32	42	58	40	58	55	33	46	37	27	24	713	0,938	669
9 oma	24	29	56	35	36	33	21	14	25	24	25	20	37	19	29	23	38	37	24	43	592	0,956	566
10 nagu	23	16	29	35	25	17	20	28	25	37	43	37	26	33	28	45	16	18	23	18	542	0,950	515
11 mis	21	30	30	30	15	21	23	33	24	29	19	26	22	24	27	26	24	25	36	32	517	0,985	509
12 ka	9	23	48	23	15	29	11	20	22	35	23	24	49	16	17	10	16	17	19	8	434	0,892	387
13 siis	13	18	27	21	16	14	18	30	20	15	17	16	27	25	25	30	31	29	17	13	422	0,964	407
14 ma	1	-	-	-	-	61	48	31	-	15	-	69	19	16	22	47	-	61	-	26	416	0,733	305
15 ning	26	18	28	1	30	3	8	6	21	17	3	22	16	12	67	26	46	9	18	16	393	0,802	315

Таблица А-4

(гл. III, § I)

Фрагмент ЧС: 226 наиболее частотных словоформ с частотой $F \geq 44$ в тексте объемом $N = 99\,898$ (i - ранг, R - коэф. распростр., F - абсол. частота, F^* - накопленная абсол. частота, P^* - накопленная относит. частота в %)

i	Словоформа	R	F	F^*	P^*
1	ja (k)	20	3221	3221	3,22
2	ta (nom)	20	1602	4823	4,83
3	on	20	1439	6262	6,27
4	ei	20	1375	7637	7,64
5	et (k)	20	1264	8901	8,91
6	oli	20	1116	10017	10,03
7	kui (k)	20	995	11012	11,02
8	see	20	713	11725	11,74
9	oma	20	592	12317	12,33
10	nagu	20	542	12859	12,87
11	mis	20	517	13376	13,39
12	ka	20	434	13810	13,82
13	siis (d_1)	20	422	14232	14,25
14	ma	12	416	14648	14,66
15	ning (k)	20	393	15041	15,06
16	aga (d_2)	20	384	15425	15,44
17	seda	20	378	15803	15,82
18	aga (k)	20	340	16143	16,16
19-	nii	20	329	16472	16,49
20	või (k)	20	329	16801	16,82
21	kes	20	297	17098	17,12
22	veel (d_1)	20	292	17390	17,41
23	ja (d_2)	20	272	17662	17,68
24	nad	20	257	17919	17,94
25	pole	20	249	18168	18,19
26	kõik	20	246	18414	18,43
27	polnud	20	238	18652	18,67
28	selle	20	237	18889	18,91

i	Словоформа	R	P	P*	P*
29	juba	20	231	19120	19,14
30	midagi	20	224	19344	19,36
31	teda	19	223	19567	19,59
32	nüüd	20	219	19786	19,81
33	ainult (d ₂)	20	218	20004	20,02
34	oleks	20	213	20217	20,24
35	ära (d ₃)	20	206	20423	20,44
36	tema (gen)	20	198	20621	20,64
37	olid (pl)	19	193	20814	20,84
38	kuid (k)	19	192	21006	21,03
39-	kas (d ₂)	20	189	21195	21,22
40	ta (gen)	18	189	21384	21,41
41	tal	20	186	21570	21,59
42	küll (d ₂)	20	185	21755	21,78
43	üle (pp)	20	180	21935	21,96
44	välja (d ₃)	20	178	22113	22,14
45	mitte	20	177	22290	22,31
46	mida	20	174	22464	22,49
47	ise	20	169	22633	22,66
48-	kus	20	166	22799	22,82
49	pärast (pp)	20	166	22965	22,99
50	enam	20	165	23130	23,15
51	talle	20	162	23292	23,32
52-	mees	20	157	23449	23,47
53	siin	20	157	23606	23,63
54	sest (k)	20	155	23761	23,79
55	seal	20	151	23912	23,94
56	kuidas	19	148	24060	24,08
57-	ega (k)	20	145	24205	24,23
58	neid	20	145	24350	24,37
59	tuli (v)	20	144	24494	24,52
60	nende	20	141	24635	24,66
61	need	20	140	24775	24,80
62-	end	17	133	24908	24,93
63	kõige (d ₁)	20	133	25041	25,07

i	Словоформа	R	F	F*	P*
64-	isegi (d ₂)	20	130	25171	25,20
65	tema (nom)	20	130	25301	25,33
66	me (nom)	13	127	25428	25,45
67	vastu (pp)	20	126	25554	25,58
68	ju	18	125	25679	25,71
69	all (pp)	20	121	25800	25,83
70	poole (pp)	20	120	25920	25,95
71	kuid (d ₂)	17	117	26037	26,06
72	tuleb	20	116	26153	26,18
73	sellest	20	115	26268	26,29
74-	läks	20	114	26382	26,41
75	meie (gen)	13	114	26496	26,52
76-	mu	13	111	26607	26,63
77	üks (n)	20	111	26718	26,75
78	mulle	11	110	26828	26,86
79	tagasi (d ₃)	19	109	26937	26,96
80	kogu (p)	20	108	27045	27,07
81	jälle	19	106	27151	27,18
82	palju (d ₁)	20	105	27256	27,28
83	väga	18	104	27360	27,39
84-	hakkas	18	99	27459	27,49
85	mind	11	99	27558	27,59
86	peale (pp)	20	96	27654	27,68
87	just	20	92	27746	27,77
88-	jäi	20	91	27837	27,87
89	vaid (k)	20	91	27928	27,96
90-	muidugi	19	89	28017	28,05
91	võib	19	89	28106	28,13
92-	keegi	19	86	28192	28,22
95	minu	10	86	28278	28,31
	rohkem	20	86	28364	28,39
	vist	18	86	28450	28,48
96-	olla	20	85	28535	28,56
97	sa	16	85	28620	28,65

Таблица А-4
(продолжение)

i	Словоформа	R	F	F*	P*
98-	kaks	18	84	28704	28,73
99	peab	19	84	28788	28,82
100-	juures (pp)	20	83	28871	28,90
101	maha (d ₃)	19	83	28954	28,98
102-	juurde (pp)	17	82	29036	29,07
103	võib-olla	18	82	29118	29,15
104	vastu (d ₃)	19	81	29199	29,23
105-	ajal (s)	15	80	29279	29,31
106	silmad (s)	19	80	29359	29,39
107-	mille	20	79	29438	29,47
110	praegu	18	79	29517	29,55
	siiski	19	79	29596	29,63
	teha	18	79	29675	29,71
111-	vaatas	17	78	29753	29,78
112	üles (d ₃)	19	78	29831	29,86
113-	alati	18	76	29907	29,94
115	hea (nom)	18	76	29983	30,01
	sai (v)	17	76	30059	30,09
116-	ikka (d ₁)	20	74	30133	30,16
117	mul	12	74	30207	30,24
118-	läbi (pp)	19	73	30280	30,31
119	ütles	16	73	30353	
120	selles	19	72	30425	30,46
121-	kinni (d ₃)	18	71	30496	30,53
123	ometi	20	71	30567	30,60
	peaaegu	18	71	30638	30,67
124-	olnud	19	70	30708	30,74
127	siis (d ₂)	18	70	30778	30,81
	vahel (pp)	19	70	30848	30,88
	üंबर (pp)	20	70	30918	30,95
128	mina	11	69	30987	31,02
129-	alles (d ₁)	19	68	31055	31,09
138	kohe (d ₁)	18	68	31123	31,15
	kokku (d ₃)	20	68	31191	31,22

№	i	Словоформа	R	F	F*	P*
		kunagi	16	68	31259	31,29
		läbi (d ₃)	19	68	31327	31,36
		naine	17	68	31395	31,43
		nii (k ^x)	18	68	31463	31,50
		pisut	13	68	31531	31,56
		poleks	18	68	31599	31,63
		saab	18	68	31667	31,70
139-		eest (pp)	17	67	31734	31,77
140		mööda (pp)	17	67	31801	31,83
141-		iga (p)	16	66	31867	31,90
142		uuesti	19	66	31933	31,97
143		edasi (d ₃)	18	65	31998	32,03
144-		kuigi (k)	17	64	32062	32,09
145		mingi	19	64	32126	32,16
146-		aega (part)	20	63	32189	32,22
150		ole (neg)	18	63	32252	32,28
		olin	10	63	32315	32,35
		taga (pp)	20	63	32378	32,41
		võis (v)	15	63	32441	32,47
151		sel	18	62	32503	32,54
152-		ees (pp)	20	61	32564	32,60
153		ennast	15	61	32625	32,66
154		suur	19	60	32685	32,72
155		tegi	18	59	32744	32,78
156-		läheb	17	58	32802	32,84
158		võttis	16	58	32860	32,89
		õhtul	16	58	32918	32,95
159-		jah	17	56	32974	33,01
161		küllap	17	56	33030	33,06
		minna	19	56	33086	33,12
162-		isa (nom)	14	55	33141	33,17
163		mõni	15	55	33196	33,23
164-		lahti (d ₃)	19	54	33250	33,28
165		suure (a)	17	54	33304	33,34
166-		alla (pp)	17	53	33357	33,39
171		istus	17	53	33410	33,44

i	Словоформа	R	F	F *	P *
	minema	18	53	33463	33,50
	täis (a)	17	53	33516	33,55
	vana (a, nom)	16	53	33569	33,60
	ühe	18	53	33622	33,66
172-	ikka (d ₂)	19	52	33674	33,71
175	kelle	17	52	33726	33,76
	näha	19	52	33778	33,81
	sinna	18	52	33830	33,86
176-	asi	17	51	33881	33,92
181	hakkab	16	51	33932	33,97
	miks	15	51	33983	34,02
	saa (neg)	17	51	34034	34,07
	sisse (d ₃)	18	51	34085	34,12
	üldse	17	51	34136	34,17
182-	endale	18	50	34186	34,22
188	ette (d ₃)	17	50	34236	34,27
	koos (d ₁)	19	50	34286	34,32
	kord (d ₁)	14	50	34336	34,37
	käis (v)	16	50	34386	34,42
	poiss	17	50	34436	34,47
	tundis	14	50	34486	34,52
189-	kohal (pp)	16	49	34535	34,57
192	kuhu	17	49	34584	34,62
	mehe	14	49	34633	34,67
	peale (d ₃)	19	49	34682	34,72
193-	enda	19	48	34730	34,77
202	jääb	17	48	34778	34,81
	kolm	18	48	34826	34,86
	kui (d ₁)	15	48	34874	34,91
	liiga (d ₁)	17	48	34922	34,96
	pidi (v)	19	48	34970	35,01
	rääkis	18	48	35018	35,05
	saanud (neg)	19	48	35066	35,10
	täiesti (d ₁)	15	48	35114	35,15
	üle (d ₃)	18	48	35162	35,20

i	Словоформа	R	F	F*	P*
203-	pani	12	47	35209	35,24
205	poolt (pp)	16	47	35256	35,29
	seisis	13	47	35303	35,34
206-	ehk (d ₂)	18	46	35349	35,39
212	enne (pp)	16	46	35395	35,43
	korda (part)	16	46	35441	35,48
	mehed	15	46	35487	35,52
	päris (d ₁)	19	46	35533	35,57
	tea (neg)	19	46	35579	35,62
	tõesti	18	46	35625	35,66
213-	aeg	15	45	35670	35,71
219	hoopis (d ₁)	18	45	35715	35,75
	jaaks	15	45	35760	35,80
	kohta (pp)	18	45	35805	35,84
	meid	9	45	35850	35,89
	meie (nom)	12	45	35895	35,93
	neile	17	45	35940	35,98
220-	ema (nom)	12	44	35984	36,02
226	inimene	16	44	36028	36,06
	mõnikord	14	44	36072	36,11
	mööda (d ₃)	16	44	36116	36,15
	nägu	17	44	36160	36,20
	saanud	17	44	36204	36,24
	teada	15	44	36248	36,29

Список наиболее частотных существительных (с частотой $F > 10$). Звездочкой отмечены слова, встретившиеся только в I или 2 подвыборках ($R \leq 2$)

1. mees	373	koht	74
aeg	285	küla	74
silm	260	laps	74
inimene	254	40. hetk	71
käsi	237	lõpp	71
naine	185	mets	71
päev	165	tüdruk	70
asi	163	häääl	69
pea	163	vesi	68
10. nägu	159	uks	67
aasta	151	tuul	66
kord	139	nimi	64
maa	130	poeg	63
jalg	125	50. taevas	60
sõna	124	tänav	60
elu	122	hommik	59
poiss	122	ots	59
meri	112	süda	58
mõte	108	tunne	57
20. töö	105	maailm	56
isa	101	vanker	54
selg	100	laev	53
laud	97	vend	53
õhtu	94	60. aru (mõistus)	52
ema	93	suu	52
jutt (jutu)	93	öö	52
tee (liikl.)	90	puu	49
hobune	89	keel	48
maja	88	õde	48
30. kodu	87	õlg (õla)	47
pilk	85	jõud	45
tuba	85	kivi	45
linn	79	tund	45
meel	78	70. peremees	44
päike	75	lumi	43
aken	74	õu	42

kael	40	paik (koht)	29
sein	40	raamat	29
vari	40	surm	29
kell	39	tükk	29
kirik	39	sõber	28
samm	39	täht	28
rahvas	38	eit	27
80. lugu	37	huvi	27
perenaine	37	jumal	27
rõõm	37	120. kevad	27
talu	37	pull (sõnn)	27
juus (juuksed)	36	rahu	27
sõda	36	riie	27
raha	35	rohi	27
valgus	35	sakslane	27
värav	35	talv (gen talve)	27
hing	34	vihm	27
90. osa	34	härä	27
pilt	34	jõgi	26
abi	33	130. kala	26
kiri	33	kallas	26
mure	33	katus	26
nina	33	keha	26
paat	33	põhi	26
vanamees	33	sild	26
kilomeeter	32	kool	25
kuu	32	laine	25
100. rind	32	meeter	25
õigus	32	trepp	25
ilm	31	140. auto	24
küsimus	31	hulk	24
pihk	31	kõrv	24
õhk	31	külg	24
lehm	30	lõuna	24
nurk	30	põlv	24
doktor	29	uni	24
loom	29	hammas	23
110. nädal	29	huul	23

komme	23	kapten	19
150. koorem	23	190. klaas	19
pall	23	☞ kurat	19
tool	23	küün (küüni)	19
buss	22	pesa	19
laul	22	ratas	19
laut	22	serv	19
lind	22	tuli	19
masin	22	vanaema	19
mägi	22	vanaisa	19
naer	22	õnn	19
160. suvi	22	200. hein	18
hirm	21	kartul	18
häbi	21	☞ katel	18
kahju	21	koer	18
kolhoos	21	müür	18
lootus	21	nahk	18
lõug	21	pilv	18
mulje	21	põhjus	18
nali	21	rida	18
noormees	21	sõrm	18
170. pere	21	210. tasku	18
proua	21	tegemine	18
pudel	21	vili	18
põld	21	aed	17
soov	21	järv	17
tõsi	21	kass	17
vaikus	21	klass	17
vanem	21	korter	17
☞ õng	21	lause	17
armastus	20	mood	17
180. leht	20	220. mälestus	17
leib	20	müts (peakate)	17
minut	20	paber	17
muna	20	☞ parun	17
ruum	20	perekond	17
tütarlaps	20	pimedus	17
vastus	20	☞ puhk	17
õlu	20	saatus	17
häda	19	tahtmine	17

230.	vanne	17		kolhoosnik	14
	ajalugu	16	270.	kukal	14
	hiir	16		lagi	14
	kadakas	16	☞	lant	14
	kask	16		lapsepõlv	14
	kuusk	16		pool (poole)	14
	külaline	16		põrsand	14
	liiv (gen liiva)	16		põsk	14
	mänd	16		sugulane	14
	nõu	16		surnu	14
	olukord	16		söögilaud	14
240.	poisike(ne)	16	280.	viha	14
	rong	16		väin	14
☞	saag	16		värv	14
	saal	16		amet	13
☞	siil (gen siili)	16		arvamus	13
	sõit	16		hämarus	13
	sündmus	16	☞	hävituspataljon	13
	tegu	16		jagu	13
	unenägu	16		juht (gen juhu)	13
	vahe	16		kajakas	13
250.	valu (aisting)	16	290.	kasv	13
	vooli	16		kleit	13
	ülesanne	16		konks	13
	algus	15		kott (gen koti)	13
	kujutlus	15		kraav	13
	köök	15		kõht	13
☞	metsavaht	15		laht	13
	muld	15		lill	13
	mäng	15		loodus	13
	näpp	15		lõhn	13
260.	oks	15	300.	muusika	13
	pood (gen poe)	15		näide	13
	sadam	15		paar	13
	vagun	15		pagan	13
	ajaleht	14		püss	13
	habe	14		silmapilk	13
	hind	14		suhe	13
	kaaslane	14		suits	13
	king	14		sügis	13

teade	13	350. king	11
310. tekk (vaip)	13	korv	11
tiib	13	käsk	11
tütar	13	küngas	11
✠ vader	13	legendik	11
vaidlus	13	✠ loss	11
viga	13	luik	11
viis (gen viisi)	13	maantee	11
väljak	13	madrus	11
auk	12	mõis	11
esimees	12	360. mčnu	11
320. hool	12	naeratus	11
hotell	12	peegel	11
kaabu	12	✠ piht (gen piha)	11
kasu	12	plaan	11
kino	12	probleem	11
kohvik	12	põgenik	11
kõrb (gen kõrbe)	12	põõsas	11
kübar	12	püksid	11
labidas	12	saar (geogr)	11
lähedus	12	370. säär	11
330. meeleolu	12	toon (gen tooni)	11
naaber	12	tugitool	11
number	12	tähelepanu	11
✠ ookean	12	tänapäev	11
org	12	tühjus	11
puri	12	✠ usuühing	11
ring	12	vaenlane	11
saun	12	vaim	11
teadus	12	võim	11
340. tervis	12	380. võõras	11
tuju	12	✠ õemees	11
✠ vaeras (gen vaera)	12	õpetaja	11
vagu	12	ait	10
võimalus	12	daam	10
✠ võllas	12	hoog	10
film	11	hunnik	10
foto	11	iga (gen ea)	10
halb	11	ilme	10
kaar	11	jälg	10

390. jää	10	sarv	10
* kajut	10	410. sigaret	10
kaup	10	silmapiir	10
kohver	10	sulane	10
korsten	10	sõdur	10
kõrgus	10	sõjavägi	10
külm	10	säng	10
lahing	10	süli	10
lammas	10	talumees	10
liikumine	10	turi	10
400. onu	10	tuttav	10
* perepoeg	10	420. tõde	10
* pidalitõbine	10	* uss	10
pind (gen pinna)	10	uudis	10
raadio	10	varandus	10
* raehärja	10	veri	10
rand	10	ämber	10
saba	10	äär	10
sajand	10	427. ülikond	10

Список наиболее частотных глаголов (F > 10)

1. olema	4237	pidama (pidas)	83
minema	499	tõusma	81
tulema	496	tooma	79
saama	465	kaduma	78
tegema	339	lööma	78
hakkama	304	40. hoidma	76
jääma	304	oskama	76
vaatama	297	märkama	75
teadma	285	otsima	73
10. võima	285	paistma	73
nägema	268	püüdma	73
pidama (pidi)	264	juhtuma	71
võtma	261	viima	71
ütleva	223	jätma	66
tundma	218	küsimä	64
tahtma	195	50. jooksmä	63
istuma	185	mõistma	62
andma	181	lugema	60
rääkima	180	tõmbama	60
20. panema	175	tõstma	59
mõtlema	174	kasvama	57
käima	171	kirjutama	54
seisma	164	tunduma	54
jõudma	133	uskuma	53
ootama	110	tekkima	52
elama	107	60. näima	51
astuma	105	kuulama	50
leidma	104	suutma	49
ajama	100	algama	48
30. muutuma	97	saatma	48
sõitma	96	liikuma	45
arvama	92	vahtima	45
laskma	91	sööma	43
kuulma	86	viskama	43
		kostma	42

70. lubama	42	110. mõõduma	29
naerma	42	näitama	29
kartma	41	siduma	29
meeldima	41	vedama	29
vajuma	41	pakkuma	28
kandma	40	pöörama	28
kutsuma	40	kinnitama	27
suruma	40	kõndima	27
paluma	39	mäletama	27
asuma	38	ronima	27
80. heitma	37	120. seletama	27
ilmuma	37	surema	27
unustama	37	tohtima	27
vaikima	37	lootma	26
aitama	36	meenuma	26
avama	36	imestama	25
huvitama	36	seadma	25
tähendama	36	kuuluma	24
haarama	35	kõnelema	24
langema	35	lõikama	24
90. lendama	35	130. harjuma	23
sattuma	35	katma	23
lõppema	34	leiduma	23
naeratama	34	läheneda	23
vastama	34	nutma	23
hüüdma	33	tabama	23
lõkkama	33	keerama	22
mängima	33	kiskuma	22
otsustama	33	taipama	22
kukkuma	31	valima	22
100. magama	31	140. varjama	22
nõudma	31	aimama	21
kippuma	30	jätkuma	21
nimetama	30	maksma	21
ostma	30	mõjuma	21
ruttama	30	pöörduma	21
armastama	29	silmitsema	21
jooma	29	sobima	21
jälgima	29	sundima	21
meenutama	29	toimuma	21

150. tšubuma	21	190. lõpetama	16
kihutama	20	peatuma	16
kujutama	20	puhuma	16
lahkuma	20	püsuma	16
lõhnema	20	selguma	16
rippuma	20	teenima	16
segama	20	ujuma	16
sündima	20	vajama	16
uurima	20	voolama	16
ära jne. (käs- kiv kõnev.)	20	erinema	15
160. ehitama	19	200. hingama	15
proovima	19	kaits(e)ma	15
põlema	19	kasutama	15
saabuma	19	kavatsema	15
täitma	19	kuluma	15
õppima	19	tapma	15
arutama	18	igatsema	14
kerkima	18	juhtima	14
lamama	18	katsuma	14
laulma	18	kergitama	14
170. sirutama	18	210. kogunema	14
sulama	18	kõlama	14
särama	18	muutma	14
töötama	18	müüma	14
ulatuma	18	peitma	14
vahetama	18	peksma	14
veerema	18	pistma	14
julgema	17	puhkama	14
kestma	17	puutuma	14
kiitma	17	pärima	14
180. laskuma	17	220. sadama	14
määrama	17	tajuma	14
pääsema	17	väsima	14
soovima	17	alustama	13
tungima	17	jagama	13
asetama	16	jutustama	13
häbenema	16	keelama	13
hüppama	16	kuiutama	13
kaotama	16	lausuma	13
kujutlema	16	libisema	13
		230. looma	13

mulgama	13	pressima	11
petma	13	pugema	11
pigistama	13	puuduma	11
puudutama	13	sõandama	11
pühkima	13	taguma	11
raputama	13	tekitama	11
rõõmustama	13	tormama	11
sammuma	13	270. valama	11
sulgema	13	ärkama	11
240. suudlema	13	üllatama	11
torkama	13	armuma	10
tunnistama	13	kobama	10
ulatama	13	kohkuma	10
väljuma	13	kohtuma	10
värisema	13	kõikuma	10
laduma	12	leppima	10
liigutama	12	lisama	10
piinama	12	280. lõhkuma	10
rabelema	12	maits(e)ma	10
250. rajama	12	murdma	10
vaidlema	12	punetama	10
veeretama	12	rikkuma	10
venima	12	sarnanema	10
hindama	11	sööstma	10
joonistama	11	taganema	10
järgnema	11	taluma	10
kalduma	11	tõukama	10
kiirustama	11	290. uinuma	10
kükitama	11	valitsema	10
260. läikima	11	võrdlema	10
muretsema	11	õiendama	10
noogutama	11	294. õpetama	10

Список наиболее частотных прилагательных

(Табл. 10)

1. suur	270	paks	26
hea	180	40. rõõmus	26
vana	155	pisike	25
uus	103	tore	25
pikk	94	üksik	25
väike	94	lihtne	24
täis	87	pehme	24
viimane	85	pime	24
noor	75	ainus	23
10. hall	70	järgmine	23
ilus	63	tark	23
raske	58	50. terav	23
valge	57	tähtis	23
õige	54	kuiv	22
must	51	paljas	22
võõras	50	tuttav	22
selge	47	halb	21
punane	44	endine	20
tühi	44	tilluke	20
20. puhas	43	tõsine	20
imelik	41	hirmus	19
kõrge	40	60. roheline	19
soe	40	veider	19
kindel	35	paras	18
külm	35	rahulik	18
kerge	34	märg	17
sihine	34	sügav	17
tugev	34	tume	17
kitsas	33	vasak	17
30. valmis	33	võimalik	17
lai	30	kena	16
madal	30	70. pruun	16
vaikne	30	saksa	16
lühike	29	tihe	16
vaba	29	tõeline	16
hele	28	värske	16
kauge	27	õnnelik	16
kollane	27	ainuke(ne)	15

ball	15	kange	11
jahe	15	keskmine	11
kuum	15	kevadine	11
80. nukker	15	lahtine	11
praegune	15	magus	11
vaene	15	muusikaline	11
kehv	14	piinlik	11
peen	14	110. talvine	11
ühine	14	uhke	11
eesti	13	võimas	11
eriline	13	ümmargune	11
kallis	13	asjalik	10
kodune	13	eelmine	10
90. korralik	13	hõre	10
kummaline	13	hämar	10
lõbus	13	igavene	10
sirge	13	juhuslik	10
kurb	12	120. kaunis	10
kõva	12	kuldne	10
lage	12	mõnus	10
nõrk	12	sisemine	10
rumal	12	sõbralik	10
salapärane	12	vihane	10
100. terve	12	õnnetu	10
vägev	12	127. ümar	10
kahvatu	11		

Список наиболее частотных местоимений ($F \geq 10$)

1. ta		2598	üks		94
ta	2144		20. niisugune		93
nad	454		selline		72
see		1981	muu		71
see	1728		mitu		64
need	253		too		64
mis		873	too	59	
ma		845	nood	5	
ma	716		sina		60
me	129		sina	48	
tema		827	teie	12	
tema	660		sama		52
nemad	166		ükski		52
tema'p	1		üksteise		48
oma		634	seesama		41
mina		613	seesama	37	
mina	318		needsamad	4	
meie	295		30. teineteise		37
ise		580	terve		37
kes		448	palju		33
10. kõik		382	mõlema(d)		28
miski		272	missugune		27
mõni		189	säärane		27
teine		185	igaüks		26
sa		150	iseenese v. -enda		23
sa	134		samasugune		22
te	16		igasugune		20
keegi		135	40. milline		13
iga		131	kumbki		11
mingi		123	mingisugune		10
kogu		108	43. mitmesugune		10

Список наиболее частотных наречий

(d_1 - самостоятельное, d_2 - модальное, d_3 - префиксальное наречие)
($F \geq 10$, $R \geq 3$)

F		d_1	d_2	d_3	F		d_1	d_2	d_3
1330	ei	-	1330	-	94	maha	11	-	83
492	siis	422	70	-	92	just	-	92	-
434	ka	-	434	-	91	hoopis	47	44	-
384	aga	-	384	-	89	muidugi	-	89	-
368	nii	329	39	-	89	rohkem	89	-	-
308	veel	292	16	-	86	vist	-	86	-
272	ja	-	272	-	84	praegu	84	-	-
248	juba	231	17	-	82	võib-olla	-	82	-
238	nüüd	230	8	-	81	vastu	-	-	81
218	ainult	218	-	-	79	alles	68	1	10
206	ära	-	-	206	79	edasi	14	-	65
198	välja	19	-	179	79	siiski	-	79	-
189	kas	-	189	-	76	alati	76	-	-
189	küll	4	185	-	76	kokku	8	-	68
177	mitte	-	177	-	75	kinni	4	-	71
174	kus	166	8	-	75	kohe	68	7	-
166	enam	166	-	-	73	kunagi	68	5	-
163	siin	160	3	-	72	ometi	-	72	-
153	seal	153	-	-	71	läbi	3	-	68
149	kuidas	148	1	-	71	peaaegu	71	-	-
133	kõige	133	-	-	70	pisut	70	-	-
130	isegi	-	130	-	67	uesti	67	-	-
126	ikka	74	52	-	64	sisse	13	-	51
125	ju	-	125	-	61	otse	20	41	-
117	kuid	-	117	-	60	päris	46	14	-
117	tagasi	8	-	109	59	lahti	5	-	54
116	jälle	112	4	-	59	üldse	-	59	-
109	väga	109	-	-	56	ette	5	-	51
107	palju	107	-	-	56	küllap	-	56	-
101	üles	23	-	78	54	kord	50	4	-

F		d ₁	d ₂	d ₃	F		d ₁	d ₂	d ₃
53	kuidagi	20	33	-	37	niisama	32	5	-
52	kui	48	4	-	37	või	-	37	-
52	peale	3	-	49	36	et	-	36	-
52	sinna	52	-	-	36	püsti	9	-	27
51	hilja	51	-	-	36	veidi	36	-	-
51	miks	51	-	-	35	ees	35	-	-
50	koos	46	-	4	35	ning	-	35	-
50	täiesti	48	2	-	35	sellepärast	35	-	-
50	üle	1	-	49	35	tõepoolest	-	35	-
49	kuhu	49	-	-	35	varsti	35	-	-
48	alla	15	-	33	33	niiviisi	33	-	-
48	liiga	48	-	-	31	ammu	25	6	-
46	ehk	-	46	-	31	koguni	-	31	-
46	mööda	2	-	44	31	natuke(ne)	31	-	-
46	tõesti	-	46	-	31	nõnda	29	2	-
46	täna	46	-	-	31	vähemalt	-	31	-
46	üsna	43	3	-	31	üksi	26	5	-
45	lihtsalt	10	35	-	31	ümber	1	-	30
44	mõnikord	44	-	-	30	kaugele	30	-	-
44	siia	44	-	-	30	laiali	30	-	-
44	äkki	39	5	-	30	vaevalt	6	24	-
43	hästi	43	-	-	29	järele	-	-	29
43	kaasa	-	-	43	29	siit	29	-	-
42	vahel	42	-	-	29	täpselt	21	8	-
41	ikkagi	-	41	-	28	igatahes	-	28	-
41	ringi	-	-	41	27	järsku	27	-	-
41	õieti	3	38	-	27	kuskil	27	-	-
40	kust	40	-	-	27	kõrvale	3	-	24
40	vaid	-	40	-	27	lõpuks	22	5	-
40	vaja	-	-	40	27	üksnes	-	27	-
39	enne	38	1	-	26	muidu	16	10	-
39	eriti	33	6	-	26	nagu	-	26	-
39	kaua	39	-	-	26	samuti	18	8	-
38	ent	-	38	-	26	teisiti	26	-	-
38	sealt	38	-	-	26	umbes	26	-	-

F		d ₁	d ₂	d ₃	F		d ₁	d ₂	d ₃
25	kaugel	25	-	-	17	nimelt	1	16	-
24	küllalt	24	-	-	17	parajasti	4	13	-
24	pealegi	-	24	-	17	sageli	17	-	-
24	pärast	24	-	-	17	vähem	17	-	-
24	sest	-	24	-	16	eemale	7	-	9
24	vara	24	-	-	16	kuhugi	16	-	-
23	aeg-ajalt	23	-	-	16	kusagil	16	-	-
23	samas	23	-	-	16	ootamatult	16	-	-
22	aina	-	22	-	16	paremini	16	-	-
22	ega	-	22	-	16	vahepeal	16	-	-
22	harva	22	-	-	16	üksinda	14	2	-
22	vähe	22	-	-	15	all	15	-	-
21	ilmselt	-	21	-	15	eks	-	15	-
21	lausa	-	21	-	15	järjest	9	6	-
21	seepärast	21	-	-	15	süüdi	15	-	-
20	algul	20	-	-	15	tarvis	-	-	15
20	kiiresti	20	-	-	15	tegelikult	-	15	-
20	taas	20	-	-	15	tookord	15	-	-
20	tõsiselt	20	-	-	15	vaikselt	15	-	-
19	lähedale	19	-	-	15	öösel	15	-	-
19	milleks	19	-	-	14	arvatavasti	-	14	-
19	nähtavasti	-	19	-	14	homme	14	-	-
19	seejärel	19	-	-	14	sees	14	-	-
19	veelgi	19	-	-	14	viimaks	13	1	-
19	üleni	19	-	-	14	õhtuti	14	-	-
18	eemal	18	-	-	14	õige	11	3	-
18	ial(gi)	18	-	-	13	eile	13	-	-
18	korraga	18	-	-	13	omavahel	13	-	-
18	sugugi	7	11	-	13	pigem	-	13	-
18	taga	5	-	13	13	seejuures	-	13	-
18	vahete- vahel	18	-	-	13	tihti	13	-	-
17	esialgu	17	-	-	13	ükskord	13	-	-
17	juurde	1	-	16	13	ükskõik	-	13	-
17	kindlasti	2	15	-	12	enam-vähem	12	-	-
17	korraaks	17	-	-	12	kuigi	12	-	-

F		d ₁	d ₂	d ₃	F		d ₁	d ₂	d ₃
12	kõigepealt	12	-	-	11	rahulikult	11	-	-
12	kõigest	-	12	-	11	sealsamas	11	-	-
12	niimoodi	12	-	-	11	üldiselt	-	11	-
12	paremal	12	-	-	10	aeglaselt	10	-	-
12	pikali	12	-	-	10	enamasti	10	-	-
12	väljas	12	-	-	10	ettepoole	10	-	-
12	üpris	12	-	-	10	halvasti	10	-	-
11	iseenesest	7	4	-	10	kauaks	10	-	-
11	kergelt	11	-	-	10	kenasti	10	-	-
11	korralikult	11	-	-	10	omaette	10	-	-
11	kõrvuti	11	-	-	10	pikkamööda	10	-	-
11	millal	11	-	-	10	selgesti	10	-	-
11	millegi- pärast	11	-	-	10	vahest	-	10	-
11	muide	-	11	-	10	vanasti	10	-	-
11	põrmugi	6	5	-	10	äkitselt	10	-	-
11	päriselt	10	1	-					

Список наиболее частотных союзов ($F \geq 10$)

1. ja	3221	kuigi	64
et	1264	otsekui	41
kui	995	kuni	32
nagu	542	ent	31
ning	393	enne (k^x)	21
aga	340	ei (k^x)	20
või	329	justkui	20
kuid	192	20. kas (k^x)	20
sest	155	ehkki	17
10. ega	145	mitte (k^x)	17
vaid	91	ainult (k^x)	16
nii (k^x)	68	24. kuna	15

Фрагмент распределительного ЧС лексем (F - суммарная и $\bar{F}$ - средняя частота, σ - стандартное отклонение)

[illegible]

Лексема	I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	F	$\bar{F}$	δ
meev 'муж(чина)'	52	25	12	31	10	18	9	16	36	18	27	16	12	9	20	17	6	14	11	14	373	18,7	11,0
naine 'жен(щин)а'	39	17	13	5	1	8	8	6	14	10	1	9	2	1	4	7	13	14	6	7	185	9,3	8,4
inimene 'человек'	11	20	15	15	5	13	10	9	15	22	8	13	8	10	8	15	9	14	1	33	254	12,7	6,8
silm 'глаз'	15	18	7	18	24	7	18	13	13	17	5	8	10	12	7	26	10	19	8	5	260	13,0	6,2
käsi 'рука'	23	7	12	14	6	3	23	10	18	18	12	7	4	10	9	9	17	14	11	10	237	11,9	5,7 ⁰
pea 'голова'	14	10	4	5	11	4	9	6	12	6	15	4	7	8	17	8	5	11	3	4	163	8,2	4,1
nägu 'лицо'	9	9	1	7	15	2	13	8	11	8	9	14	3	11	4	13	1	7	3	11	159	8,0	4,4
jalg 'нога'	26	3	6	12	8	4	4	4	7	14	5	4	3	8	2	6	3	2	2	2	125	6,3	5,6

Таблица А-12
(гл. III, § I)

50 наиболее частотных словоформ, ранжированных по абсолютной частоте F и коэффициенту стабильности D

(i_D - ранг по D , i_F - ранг по F)

i_D	Словоформа	D	F	i_F	Словоформа	D	F	i_F
1	mis	0,985	517	11	26 üle (pp)	0,917	180	43
2	ja (k)	974	3221	1	27 ära (d_3)	908	206	35
3	juba (d_1)	974	231	29	28 ainult (d_2)	899	218	33
4	siis (d_1)	964	422	13	29 ka	892	434	12
5	selle	958	237	28	30 kus	892	166	48
6	kui (k)	956	995	7	31 kas (d_2)	889	189	39
7	oma	956	592	9	32 või (k)	888	329	20
8	kõik	951	246	26	33 nad	887	257	24
9	nagu	950	542	10	34 midagi	835	224	30
10	veel (d_1)	945	292	22	35 teda	830	223	31
11	välja (d_3)	944	178	44	36 oli	830	1116	6
12	ei	943	1375	4	37 ise	828	169	47
13	kes	939	297	21	38 tal	828	186	41
14	et (k)	938	1264	5	39 mida	828	174	46
15	see	938	713	8	40 ja (d_2)	827	272	23
16	seda	937	378	17	41 küll (d_2)	827	185	42
17	nii (d_1)	936	329	19	42 oleks	826	213	34
18	nüüd	936	219	32	43 on	825	1439	3
19	aga (d_2)	935	384	16	44 pole	823	249	25
20	ta (nom)	931	1602	2	45 polnud	815	238	27
21	mitte	927	177	45	46 ta (gen)	815	189	40
22	pärast (pp)	922	166	49	47 kuid (k)	813	192	38
23	aga (k)	921	340	18	48 olid (pl)	803	193	37
24	enam	921	165	50	49 ning	802	393	15
25	tema (gen)	919	198	36	50 ma	733	416	14

Таблица А-13
(гл. III, § I)

Стабильность частоты словоформ

(по значению коэффициента стабильности D)

D > 0,95

0,985	mis	0,967	ees (pp)	0,956	kui (k)
0,974	ja (k)	0,964	siis (d ₁)	0,956	oma
0,974	juba (d ₁)	0,958	selle	0,951	kõik

0,95 ≥ D > 0,90

0,950	nagu	0,936	tagasi (d ₃)	0,917	kogu (p)
0,945	veel (d ₁)	0,935	aga (d ₂)	0,917	neid
0,944	välja (d ₃)	0,931	ta (nom)	0,917	üle (pp)
0,943	ei	0,929	need	0,914	panna
0,939	kes	0,927	mitte	0,908	ära (d ₃)
0,938	et (k)	0,925	kõige	0,905	palju (d ₁)
0,938	see	0,922	pärast (pp)	0,902	just
0,937	seda	0,921	aga (k)	0,901	nende
0,936	nii (d ₁)	0,921	enam		
0,936	nüüd	0,919	tema (gen)		

0,90 ≥ D > 0,85

0,899	ainult (d ₂)	0,889	kas (d ₂)	0,883	talle
0,898	peale (d ₃)	0,889	vastu (pp)	0,882	alles (d ₁)
0,898	tegi	0,888	või (k)	0,882	sisse (d ₃)
0,892	ka	0,887	nad	0,869	isegi (d ₂)
0,892	kus	0,885	kuidas		

Таблица А-13
(продолжение)

85 > D > 0.80

aega (part)	lahti (d ₃)	sai (v)
alati	läbi (d ₃)	seal
all (pp)	läbi (pp)	sel
alla (pp)	läks	selles
asi	maha (d ₃)	sellest
eest (pp)	mida	sest
enda	midagi	siiski
endale	mingi	silmad (s)
enne (pp)	minna	sinna
ette (d ₃)	muidugi	suur
hakkas	mõni	ta (gen)
ikka (d ₁)	nii (k ^x)	taga (pp)
ikka (d ₂)	ning	tal
ise	näha	tea (neg)
istus	oleks	teda
ja (d ₂)	oli	teha
ju	olid (pl)	tema (nom)
juures (pp)	olla	tuleb
jäi	olnud	tuli (v)
jälle	ometi	uuesti
kaks	on	vahel (pp)
keegi	peab	vaid (k)
kelle	peale (pp)	vastu (d ₃)
kokku (d ₂)	pidi (v)	võib
kolm	pole	võib-olla
koos (d ₁)	polnud	ühe
korda (part)	poole (pp)	üks (n)
kuhu	poolt (pp)	üle (d ₃)
kuid (k)	praegu	üles (d ₃)
kuigi (k)	päris (d ₁)	ümber (pp)
küll (d ₂)	saanud (neg)	

Таблица А-14
(гл. III, § I)

Десять наиболее частотных словоформ ранжированных по абсолютной частоте F и коэффициенту употребительности U (i_F - ранг по F ; i_U - ранг по U)

i_F	Словоформа	F	D	U	i_U
1	ja (k)	3221	0,974	3138	1
2	ta (nom)	1602	0,931	1491	2
3	on	1439	0,825	1187	4
4	ei (d_2)	1375	0,943	1297	3
5	et	1264	0,938	1186	5
6	oli	1116	0,830	926	7
7	kui (k)	995	0,956	951	6
8	see	713	0,938	669	8
9	oma	592	0,956	566	9
10	nagu	542	0,950	515	10

Таблица А-15
(гл. III, § 1)

Динамика изменения состава ЧС лексем
(частоты в группах по 100 лексем)

Ранги Часть речи	I-	10I-	20I-	30I-	40I-	50I-	60I-	70I-	80I-	90I-	100I-	110I-	Всего слов	
	100	200	300	400	500	600	700	800	900	1000	1100	1200	число	%
Существительное	15	28	28	31	33	39	38	40	43	49	39	39	422	35,2
Глагол	24	22	20	32	26	21	30	23	29	26	20	21	294	24,5
Прилагательное	3	7	11	11	12	13	9	16	13	10	18	16	139	11,6
Наречие	22	22	24	17	17	15	16	15	12	12	18	11	201	16,8
Местоимение	15	5	7	3	4	2	1	0	0	1	1	2	41	3,4
Числительное	3	1	2	0	2	2	3	0	0	0	0	3	16	1,3
Релятивные слова ⁺	7	14	7	4	6	8	3	4	2	1	3	8	67	5,6
Союз	11	1	0	2	0	0	0	1	1	0	1	0	17	1,4
Междометие	0	0	1	0	0	0	0	1	0	1	0	0	3	0,2
Все слова	100	100	100	100	100	100	100	100	100	100	100	100	1200	100,0

⁺ Релятивные слова включают послелог и предлог, из которых большинство может выступать также в роли наречия.

Таблица А-16
(гл. III, § I)

Динамика изменения состава ЧС лексем
(накопленные частоты в %)

Ранг Часть речи	100	200	300	400	500	600	700	800	900	1000	1100	1200	Весь ЧС
Существительное	15,0	21,5	23,7	25,5	27,0	29,0	30,3	31,5	32,8	34,4	34,8	35,2	61,9
Глагол	24,0	23,0	22,0	24,5	24,8	24,2	25,0	24,8	25,2	25,3	24,8	24,5	14,0
Прилагательное	3,0	5,0	7,0	8,0	8,8	9,5	9,4	10,3	10,6	10,5	11,2	11,6	11,7
Наречие	22,0	22,0	22,7	21,3	20,4	19,5	19,0	18,5	17,8	17,2	17,3	16,8	9,6
Местоимение	15,0	10,0	9,0	7,5	6,8	6,0	5,3	4,6	4,1	3,8	3,5	3,4	0,4
Числительное	3,0	2,0	2,0	1,5	1,6	1,7	1,9	1,6	1,4	1,3	1,2	1,3	1,0
Релятивные слова	7,0	10,5	9,3	8,0	7,6	7,7	7,0	6,6	6,1	5,6	5,4	5,6	1,0
Союз	11,0	6,0	4,0	3,5	2,8	2,2	2,0	1,9	1,8	1,6	1,5	1,4	0,2
Междометие	0,0	0,0	0,3	0,2	0,2	0,2	0,1	0,2	0,2	0,3	0,3	0,2	0,5
Все слова (%)	100	100	100	100	100	100	100	100	100	100	100	100	100

Таблица А-17
(гл. III, § I)

Распределение частей речи среди 100 наиболее частотных лексем эстонского языка в словаре (С) и тексте (Т); Т/С - средняя повторяемость части речи в тексте

Часть речи	С л о в а р ь		Т е к с т		Т/С
	число (С)	%	число (Т)	%	
Существительное	15	15,0	2913	7,0	194
Глагол	24	24,0	10303	24,9	429
Прилагательное	3	3,0	648	1,6	216
Наречие	22	22,0	6064	14,7	276
Местоимение	15	15,0	10665	25,8	711
Числительное	3	3,0	792	1,9	264
Релятивные слова	7	7,0	1261	3,0	180
Союз	11	11,0	8712	21,1	792
Междометие	-	-	-	-	-
Всего	100	100,0	41358	100,0	414

(41,4 % всего
текста)

Таблица А-13
(гл. III, § I)

Распределение частей речи среди 100 наиболее частотных лексем по данным 40 разных языков (Частотный словарь русского языка, 1977; Saukkonen P. et al., 1979; Engwall G., 1974)

Язык Часть речи	Эстонский (худ.проза)	Финский (смешанный)	Французский (худ.проза)	Русский (смешанный)
Существительное	15	15	12	12
Глагол	24	25	19	12
Прилагательное	3	7	3	3
Местоимение	15	17	28	25
Числительное	3	2	2	3
Наречие	29	26	14	14
Пред-/послелог			14	15
Союз	11	8	7	8
Частицы	-	-	-	8
Другие	-	-	1	-
Всего	100	100	100	100

Таблица A-19
(гл. III, § I)

20 наиболее частотных лексем в эстонском языке

Авторская речь Роман А.Х. Таммсааре Газета: внеш. Учебники
худож. прозы авт.речь речь персон. политика IV класса

olema	olema	olema	olema	plema
ja	ja	mina	ja	ja
ta	ta	see	see	see
see	see	ei	aasta	t(em)a
ei	tema	ja	et	kui
et	kui	ma	ta	mis
kui	et	siis	ei	m(in)a
mis	ei	sa	valitsus	ei
ma	aga	kui	mis	et
tema	nagu	aga	riik	saama
aga	oma	sina	oma	kõik
oma	üttelema	ta	ka	ka
mina	siis	mis	kui	arv
ise	mis	et	saama	siis
nagu	sest	tema	uus	või
minema	minema	minema	kes	üks
tulema	ka	tulema	ise	teine
siis	ise	saama	suur	vesi
saama	tegema	oma	üks	s(in)a
kes	saama	tegema	kõik	maa

Таблица А-20
(гл. III, § I)

20 наиболее частотных имен существительных в эстонском
языке

Авторская речь худож.прозы	Роман А.Х. Таммсааре авт.речь речь персон.	Газета: внеш. политика	Учебники IV класса	
mees	poiss	laps	aasta	arv
aeg	peremees	isa	valitsus	vesi
silm	laps	mees	riik	maa
inimene	perenaine	peremees	rahvas	aasta
käsi	aeg	ema	partei	sõna
naine	silm	asi	poliitika	näide
päev	asi	käsi	suhe	hulk
asi	mees	perenaine	aeg	õhk
pea	sõna	õigus	osa	osa
nägu	käsi	eit	õigus	inimene
aasta	tüdruk	inimene	sõda	laul
kord	kord	vesi	liit	aeg
maa	kodu	jumal	maa	joonis
jalg	jalg	poiss	nõukogu	punkt
sõna	süda	aeg	otsus	kord
elu	isa	kodu	huvi	lause
poiss	poeg	süda	ajaleht	abi
meri	ema	jalg	kord	taim
mõte	meel	loom	tulemus	käsi
tõõ	vesi	silm	võim	päike

Таблица А-21
(гл. III, § I)

20 наиболее частотных глаголов в эстонском языке

Авторская речь худож.прозы	Роман А.Х. Таммсааре авт.речь речь персон.	Газета: внеш. политика	Учебники IV класса
olema	olema	olema	olema
minema	üttelema	saama	saama
tulema	minema	andma	kirjutama
saama	tegema	võtma	andma
tegema	saama	tulema	tulema
hakkama	tulema	toimuma	minema
jääma	vastama	tegema	leidma
vaatama	pidama (pidi)	pidama (pidi)	nimetama
teadma	võima	pidama (pidas)	hakkama
võima	vastama	jääma	nägema
nägema	hakkama	minema	võima
pidama (pidi)	tahtma	asuma	tegema
võtma	jääma	kuuluma	võtma
üttelema	võtma	tähendama	kasutama
tundma	teadma	moodustama	arvutama
tahtma	arvama	viima	panema
istuma	panema	algama	laulma
andma	nägema	märkima	üttelema
rääkima	rääkima	nägema	jääma
päiema	seisma	võima	pidama (pidi)

Таблица А-22
(гл. III, § I)

10 наиболее частотных имен прилагательных
в эстонском языке

Авторская речь худож.прозы	Роман А.Х. Таммсааре авт. речь речь персон.	Газета: внеш. политика	Учебники IV класса
suur	suur	hea	uus suur
hea	vana	suur	suur järgmine
vana	hea	kallis	poliitiline uus
uus	uus	vana	rahvuslik väike
pikk	noor	kange	viimane suurem
väike	täis	hull	kindel
täis	räske	õige	ametlik
noor	viimane	täis	sotsialistlik
hall	väike	uus	tähtis
ilus	endine	terve	peamine

Таблица А-23
(гл. III, § I)

20 наиболее частотных лексем в разных языках (по данным: 40 русского яз., 1977; 40 суч. укр. худ. прозы, 1981; Saukkonen P. et al., 1979; Latviešu val. biez., 1972; Engwall G., 1974; Kušera H., Francis W.N., 1967; Meier H., 1964; Hassler-Göransson C., 1966)

Эстонский яз. (худож. проза)		Финский яз. (худож. проза)		Латышский яз. (худож. проза)		Французский яз. (худож. проза)		Украинский яз. (худож. проза)	
olema	4237	olla	2692	un	3588	le, la, les	I333I	1 ₁ (союз)	I4276
ja	3493	ja	I874	būt	2394	de	9780	в	I279I
ta	2598	se	I052	vinš	2I96	je	6768	не	II030
see	I98I	ei	I038	tas	I859	il	5280	на _I (предлог)	I0388
5. ei	I395	hän	80I	ar	I246	être	4746	з	8338
et	I300	minä	542	kas	I225	avoir	4633	в in	8200
kui	I047	että	424	es	II65	un, une	4475	а _I (союз)	625I
mis	879	mutta	386	kā	I042	et	4II9	до _I (предлог)	482I
ma	845	niin	38I	bet	989	à	3774	я	4783
10. tema	827	joka	378	ka	909	qui	29I3	1 ₂ (частица)	4666
aga	724	kun	37I	uz	904	ne	2887	вона	402I
oma	634	kuin	29I	no	903	elle	2865	що ₂ (союз)	3333
mina	6I3	tulla	276	par	736	que (conj)	2788	що _I (местоим.)	3052
ise	58I	sinä	268	nebūt	720	se	2384	бути ₂ (связка)	2874
15. nagu	568	he	246	viss	646	des	2303	за _I (предлог)	282I
minema	499	nyt	239	arī	585	son (pron)	I977	ти ₂ (местоим.)	2764
tulema	496	mikä	229	vai	583	ils	I975	вои	2680
siis	493	ne	227	tad	548	pas	I873	як ₄ (союз)	2670
saama	465	saada	2I9	jau	536	ce	I7I8	св _{III} 2 (местоим.)	2504
20. kes	448	kaikki	208	mēs	530	en (prep.)	I660	ж	2448
Всего	24I23		I2I42		23304		82249		II47II
Объем текста	99898		469I2		I00000		200000		500357
Объем сло- варя	I4654		8578		II439		I0868		3339I

Таблица А-23
(продолжение)

20 наиболее частотных лексем
(лемматизация ЧС английского, немецкого и шведского яз. сделана автором)

	Русский яз. (смеш. тексты)	Английский яз. (смеш. тексты)	Немецкий яз. (смеш. тексты)	Шведский яз. (смеш. тексты)
5.	в (во) 42854 и 36266 не 19228 на 17262 я 13839 быть 13307 что 13185 он 13143 с (со) 12975	the 69971 be 36441 of 36411 and 28852 a (an) 26974 to 26149 in 21341 he 19159 have 11513	der, die, das 1139270 und 320072 ein(e) 251565 sein (Verb) 245902 in (im) 242750 zu 172625 er 133028 ich 131288 haben 127171 werden 125694 nicht 114518 von (vom) 113201	och 14273 den, det 12802 i 9644 vara 8503 en, ett 8450 att 8003 jag 7934 ha(va) 6309 på 6028 som 5290 han 5173 till 4482 för 4303 med 3927
10.	а 10719 как 7425 это 6940 вы 6547 ты 6475	it 10614 that 10595 for 9485 they 8026 with 7289	daß 7250 dieser 6742 es 6354 auf 5896 sie (Sg.) 5378	av 3903 vi 3877 de 3781 så 3685 om 3040
15.	к (ко) 6246 мы 6220 этот 6061 она 5836 они 5712	as 7250 on 6742 I 6354 she 5896 at 5378	so 74273	hon 3003
20.	но 5176	by 5305		
Всего 255416		359745	3794617	126410
Объем текста 1056382		1014232	10910777	500000
Объем словаря 39268		(50406 слово- форм)	(258173 словоформ)	

Таблица А-24
(гл. III, § I)

20 наиболее частотных имен существительных в
художественной прозе разных языков

Эстонский	Финский	Латышский	Французский	Русский
mees	aika	cilvēks	jour	человек
aeg	mies	laiks	ceil	рука
silm	päivä	roka	main	глаз
inimene	käsi	diena	père	дело
käsi	poika	māja	heure	жизнь
naine	ihminen	acs	homme	голова
päev	lapsi	darbs	temps	лицо
asi	asia	gads	fois	день
pea	maa	galva	femme	мать
nāgu	pää	māte	vie	время
aasta	nainen	vārds	tête	земля
kord	isä	tēvs	chose	вода
maa	äiti	vieta	mot	слово
jalg	tie	ceļš	voix	раз
sõna	silmä	vīrs	chambre	товарищ
elu	ovi	reize	enfant	место
poiss	hetki	kāja	nuit	отец
meri	sana	puse	soir	дом
mõte	vesi	meža	an	год
tõõ	puu	dzīve	monde	нога

Таблица А-25
(гл. III, § I)

20 наиболее частотных глаголов в художественной
прозе разных языков

Эстонский	Финский	Латышский	Французский	Русский
olema	olla	būt, nebūt	être	быть
minema	tulla	varēt, nevarēt	avoir	сказать
tulema	saada	redzēt	dire	говорить
saama	voida	iet	faire	мочь
tegema	sanoa	sākt	voir	знать
hakkama	mennā	sacīt	aller	стать
jääma	nähdä	zināt	pouvoir	видеть
vaatama	pitää	domāt	savoir	идти
teadma	tietää	vajadzēt	vouloir	спросить
võima	tehdä	darīt	venir	пойти
nägema	lähteä	teikt	prendre	стоять
pidama (pidi)	katsoa	klūt	parler	думать
võtma	alkaa	gribēt	falloir	хотеть
üttelema	kuulla	palikt	devoir	жить
tundma	ottaa	nākt	regarder	ответить
tahtma	antaa	likties	passer	смотреть
istuma	käydä	stāvēt	aimer	казаться
andma	istua	dzirdēt	croire	сидеть
rääkima	puhua	aiziet	penser	начать
panema	näyttää	saprast	laisser	взять

Таблица А-26
(гл. III, § I)

10 наиболее частотных имен прилагательных
в художественной прозе разных языков

<u>Эстонский</u>	<u>Финский</u>	<u>Латвийский</u>	<u>Французский</u>	<u>Русский</u>
suur	suuri	liels	petit	большой
hea	pieni	jauns	grand	новый
vana	hyvä	vecs	bon	маленький
uus	koko	labs	long	белый
pikk	vanha	mazs	beau	черный
väike	uusi	pēdējais	vieux	красный
täis	kaunis	balts	noir	старый
viimane	pitkä	mēlms	blanc	молодой
noor	korkea	skaists	vrai	хороший
hall	musta	pilns	jeune	последний

Таблица А-27

Десять наиболее частотных словоформ в разных языках (данные взяты из работ: Saukkonen P., 1977; Engwall G., 1974; Meier H., 1964; Kučera H., Francis N.W., 1967; Гурченкова В.П., 1972; Allén S., 1970)

Эстонский	Финский	Французский	Немецкий	Английский	Шведский	Русский
ja (k)	ja	de	die	the	och	в
ta (nom)	on	la	der	of	i	и
on	että	le	und	and	att	на
ei	ei	et	in	to	en	не
et	oli	il	zu	a	som	с
oli	se	à	den	in	det	что
kui	mutta	les	das	that	är	по
see	niin	je	nicht	is	av	о
oma	ovat	l'	von	was	den	к
nagu	kuin	un	sie	he	pa	из

Покрытие

текста (%)

12,7

11,7

19,7

18,1

24,3

20,2

14,5

Таблица А-28
(гл. III, § 2)

Десять наиболее частотных словоформ в текстах авторской речи современной эстонской художественной прозы (подвыборки объемом 5000 словоупотреблений)

№ пп.	Э. Бээкман		В. Гросс		А. Хинт		Х. Кийк		Я. Кросс	
1	137	ja	129	ja	141	ja	174	ja	283	ja
2	98	ta	78	ei	122	oli	129	on	84	oli
3	64	Benita	78	ta	112	ta	68	ei	75	ta
4	47	et	66	oli	56	oma	63	et	54	kui
5	43	oli	54	on	54	ei	45	ta	40	et
6	40	kui	53	kui	48	ka	35	nagu	38	Pall
7	38	mees	46	et	47	et	35	oma	36	oma
8	35	ei	39	see	45	kui	33	see	30	ning
9	26	ning	30	mis	33	polnud	30	kui	26	ta (gen)
10	24	oma	29	oma	32	see	30	mis	25	ja (d)
Σ	552		602		690		642		691	
%	11,0		12,0		13,8		12,9		13,8	

(Продолжение следует)

Таблица А-28
(продолжение)

№ пп.	П. Куусберг		Л. Промет		В. Саар		Х. Серро		Р. Сирге	
1	118	ja	171	ja	171	ja	155	ja	80	ja
2	117	ei	100	et	123	ta	91	oli	77	ei
3	112	et	97	ei	100	ei	73	ei	68	ta
4	95	ta	76	ta	87	et	50	et	61	kui
5	86	on	69	oli	76	on	33	kui	54	et
6	61	ma	59	kui	61	kui	31	Martin	37	nagu
7	34	kui	48	ma	48	see	25	nagu	35	ka
8	33	oma	42	see	33	mis	25	oma	32	see
9	29	ka	29	on	31	ma	25	ta	30	võl
10	28	oli	29	seda	30	siis	24	mis	29	mis
Σ	713		720		760		532		503	
%	14,3		14,4		15,3		10,8		10,1	

(Продолжение следует)

Таблица А-28
(продолжение)

№	М. Траат		Э. Ветемаа		А. Каал		Т. Каллас		Я. Пээгель	
пп.										
1	168	ja	131	ja	185	ja	169	ja	121	ja
2	163	on	114	on	83	et	127	ta	119	on
3	123	ta	78	ta	76	on	107	oli	74	kui
4	76	ei	75	ei	71	ta	62	et	71	ei
5	54	kui	69	ma	63	ei	57	see	67	ning
6	43	et	68	et	56	oli	43	ei	55	see
7	43	nagu	58	see	49	ka	41	on	42	et
8	42	see	52	oli	40	see	33	nagu	37	ta
9	33	nad	35	kui	38	kui	27	kui	32	oli
10	33	pole	35	nagu	37	oma	25	aga (d)	29	oma
Σ	778		715		698		691		647	
%	15,4		14,3		14,1		13,8		13,2	

(Продолжение следует)

Таблица А-28
(продолжение)

№ пп.	Ю. Туулик			А. Валтон			М. Унт		Э. Нийт, Я. Кросс		Ю. Смуул	
1	147	ja		177	ja		186	ja	181	ja	197	ja
2	87	ta		138	ta		71	on	144	on	90	on
3	74	oli		112	oli		70	et	54	kui	50	ei
4	62	ei		79	on		69	ta	36	mis	50	ta
5	55	et		76	ei		64	ei	34	aga (d)	43	oma
6	55	kui		74	et		61	ma	30	siin	42	kui
7	50	on		46	ning		51	oli	28	ei	34	oli
8	47	ma		46	see		37	oma	28	et	32	et
9	45	nagu		44	kui		37	see	27	see	32	mis
10	33	see		38	oma		36	kui	27	ta	28	kes
Σ	655			830			682		589		598	
%	13,1			16,7			13,7		11,5		11,9	

Список однокоренных слов на букву А (по данным 40 авторской речи современной эстонской художественной прозы; $N = 99898$, $L = 14654$), отсутствующих в Орфографическом словаре 1976 г. (объемом 115 000).

aasaserv	aknaauk	armuvidin
aastakümnetepikkune	aknavalge (s)	arupõld
aastaringi	akuraadio	asenditunne
aastatuhandete-vanune	alakivi	atlassdress
abielutehnika	algkapital	atrama
absoluutsus	algkoolipäevad	augustiõõ
abusoon	algühtsus	aunaalune (s)
adrypeenar	alkoholismivastane	aupalk
aed-aun	alpitaim	autodejõgi
agulipunker	alraunileht	autodevool
ah-ah-haa	alumiiniumilõikeline	autopeegel
ahjaa	alumiiniumiplaadike	autouks
ahjukõks	alusmüüriderägane	autovargus
ahjupanek	ananassitaoline	avaldusviis
ahnevõitu	andumisheldus	avameelsus-
aiaplank	antimagistraalkraav	vajadus
aidarähistas	antiurbanism	
ainitine	antvärgisell	(65 слов
ainusündinud	apostlipilt	из 313
ainusünnis	arenemispinge	однокоренных
aisaots	arengulugu	слов на
ajaloonäide	arenguseadus	букву А)
ajalootund	arithmeticus (lad.)	
ajaloopik	armuküllane	
akendepoolne	armusuhe	

Таблица А-30
(гл. IV, § I)

Распределение частот букв в начальной и терминальной позициях слова по данным словаря эстонского языка (Oigekeelsuse sõnaraamat, 1960; Rückläufiges estnisches Wörterbuch, 1979). Объем словаря: 105000 лексем.

№ пп.	В начале слова		В конце слова	
	Буква	Частота (%)	Буква	Частота (%)
1	K	16,15	E	19,85
2	P	10,33	S	16,41
3	S	9,16	A	14,49
4	T	8,17	K	7,15
5	V	6,31	T	7,04
6	M	6,31	D	5,36
7	L	6,13	N	4,45
8	A	5,45	R	4,24
9	R	5,38	L	3,85
10	H	4,15	I	3,83
11	E	3,22	U	3,34
12	N	3,09	M	3,32
13	I	2,41	V	2,01
14	J	2,17	G	1,97
15	D	2,10	P	1,29
16	F	1,73	O	0,36
17	O	1,49	Ü	0,32
18	B	1,49	B	0,20
19	G	1,24	F	0,15
20	U	1,00	H	0,11
21	Ü	0,86	Ž	0,11
22	Š	0,68	Š	0,08
23	A	0,31	U	0,04
24	Š	0,25	A	0,03
25	Ö	0,12	Ö	0,002
26	Ž	0,12	Y	0,001
27	C	0,12	C	-
28-32	Q, Z, W, X, Y	0,06	J, Q, Z, W, X	-
	Всего (%)	100,0	Всего (%)	100,0

Распределение частот п е р в ы х букв словоформ в текстах разных подъязыков эстонского языка

Начальная буква	Авторская Речь	Речь пер- сонажей	Научный текст	Газета	В с е г о	
					число	%
a	4,3	6,7	5,0	4,9	218	5,3
b	0,4	0,1	0,1	0,2	8	0,2
d	0,0	0,2	1,1	0,3	15	0,4
e	5,4	7,2	6,9	5,9	264	6,4
f	0,1	0,1	0,6	0,5	11	0,3
g	0,0	0,0	0,0	0,0	0	0,0
h	4,3	2,4	2,2	1,5	110	2,7
i	1,4	2,3	2,6	2,9	91	2,2
j	6,5	5,3	4,2	6,3	229	5,6
k	13,7	11,7	15,1	17,3	580	14,1
l	4,7	3,3	3,9	3,1	154	3,8
m	6,7	12,5	7,3	5,6	342	8,3
n	5,6	4,1	4,4	4,2	189	4,6
o	4,9	7,4	7,6	8,6	287	7,0
p	8,5	6,7	6,7	7,9	305	7,4
r	2,3	2,3	3,7	2,3	106	2,6
s	8,2	10,3	8,1	8,2	361	8,8
t	10,0	9,7	8,6	9,6	390	9,5
u	0,9	0,3	0,8	0,7	27	0,7
v	8,1	4,7	7,3	6,2	267	6,5
õ	0,8	0,6	1,2	0,9	35	0,9
ä	0,7	0,8	0,2	0,0	20	0,5
ö	0,3	0,1	0,0	0,2	6	0,1
ü	2,2	1,2	2,4	2,7	84	2,1
Всего (%)	100,0	100,0	100,0	100,0	-	100,0
Букв	1106	1230	900	864	4100	-

Таблица А-32
(гл. IV, § I)

Распределение частот последних (терминальных) букв словоформ в текстах разных подязыков эстонского языка

Конечная буква	Авторская речь	Речь пер- сонажей	Научный текст	Газета	В с е г о	
					число	%
a	18,2	23,8	14,6	18,1	782	19,1
b	1,6	1,7	2,1	1,4	70	1,7
d	12,6	9,6	12,1	12,7	475	11,6
e	14,2	15,5	23,9	18,4	724	17,7
g	0,6	0,2	0,6	0,5	18	0,4
h	0,1	0,9	0,0	0,2	14	0,3
i	12,1	12,1	10,3	11,5	474	11,6
k	1,3	0,9	1,3	0,9	45	1,1
l	5,4	4,9	5,5	6,1	220	5,4
m	1,2	1,1	0,7	0,5	37	0,9
n	4,1	5,9	3,2	2,2	166	4,0
o	0,1	0,3	0,1	0,5	10	0,2
p	0,0	0,3	0,0	0,0	4	0,1
r	0,5	0,0	0,3	0,7	15	0,4
s	14,5	12,7	11,0	13,8	534	13,0
t	8,5	6,2	9,5	8,9	333	8,1
u	4,6	3,4	4,5	3,4	164	4,0
v	0,2	0,3	0,3	0,2	11	0,3
z	0,0	0,1	0,0	0,0	1	0,02
õ	0,2	0,1	0,0	0,0	3	0,1
Всего(%)	100,0	100,0	100,0	100,0	-	100,0
Букв	1106	1230	900	864	4100	-

Таблица А-33
(гл. IV, § I)

Распределение частот первых и последних букв по данным сводного текста из 4 подязыков эстонского языка (авторская речь и речь персонажей художественной прозы, газета, научный текст) и данные о коррелятивной функции (КФ)

Буква	Общая частота (Т)	В начале слова (Н)	КФ (Н/Т)	В конце слова (К)	КФ (К/Т)
а	12,0	5,3	0,44	19,1	1,59
b	0,7	0,2	0,29	1,7	2,43
d	4,4	0,4	0,09	11,6	2,64
e	11,3	6,4	0,57	17,7	1,56
f	0,1	0,3	3,00	0,0	-
g	1,9	0,0	-	0,4	0,21
h	1,7	2,7	1,59	0,3	0,18
i	9,5	2,2	0,23	11,6	1,22
j	1,8	5,6	3,11	-	-
k	5,3	14,1	2,66	1,1	0,21
l	6,4	3,8	0,59	5,4	0,84
m	3,8	8,3	2,18	0,9	0,24
n	4,7	4,6	0,96	4,0	0,85
o	3,5	7,0	2,00	0,2	0,06
p	1,8	7,4	4,11	0,1	0,06
r	2,8	2,6	0,93	0,4	0,14
s	8,8	8,8	1,00	13,0	1,49
t	7,1	9,5	1,34	8,1	1,14
u	6,1	0,7	0,11	4,0	0,65
v	2,2	6,5	2,95	0,3	0,14
õ	1,3	0,9	0,69	0,0	-
ä	1,5	0,5	0,33	0,0	-
õ	0,3	0,1	0,33	0,1	0,33
ü	1,0	2,1	2,10	0,1	-
Всего %	100,0	100,0	-	100,0	-
Букв	24796	4100	-	4100	-

Таблица А-34
(гл. IV, § I)

Распределение наиболее частотных биграмм в начале словоформ в тексте и в соответствующем словаре словоформ по данным выборки из 4 подязыков эстонского языка. Объем выборки: 4100 словоупотреблений; объем словаря: 2840 словоформ. (Двухбуквенные словоформы типа *te* 'вы', *ta* 'он, она' и др. включены.)

Ранг i	Т е к с т		С л о в а р ь	
	Биграмма	Частота (%)	Биграмма	Частота (%)
1.	se-	3,4	ka-	4,2
2.	ka-	3,3	va-	4,0
3.	te-	3,3	ko-	3,7
4.	ku-	3,0	ku-	2,9
5.	ol-	2,9	la-	2,5
6.	mi-	2,7	se-	2,2
7.	va-	2,6	ta-	2,2
8.	ko-	2,5	si-	2,0
9.	ta-	2,3	pi-	2,0
10.	ke-	1,6	su-	1,7
11.	me-	1,6	ma-	1,7
12.	si-	1,5	pa-	1,7
13.	tu-	1,5	sa-	1,6
14.	pe-	1,4	tu-	1,6
15.	kõ-	1,4	kõ-	1,5
16.	ma-	1,3	ke-	1,5
17.	pa-	1,3	vi-	1,4
18.	võ-	1,3	pe-	1,4
19.	sa-	1,2	te-	1,4
20.	ni-	1,1	ra-	1,3
21.	ra-	1,1	kä-	1,3
22.	jā-	1,0	na-	1,2
23.	ki-	1,0	mu-	1,2
24.	mu-	1,0	pu-	1,2
25.	ne-	1,0	kü-	1,1
Всего (%)		46,3		48,5

Таблица А-35
(гл. IV, § I)

Распределение наиболее частотных биграмм в к о н ц е словоформ в тексте и в соответствующем словаре словоформ по данным выборки из 4 подязыков эстонского языка. Объем выборки: 4100 словоупотреблений; объем словаря: 2840 словоформ. (Двухбуквенные словоформы включены.)

Ранг i	Т е к с т		С л о в а р ь	
	Диграмма	Частота (%)	Диграмма	Частота (%)
1.	-st	3,7	-st	6,5
2.	-le	3,6	-as	6,3
3.	-ud	3,4	-ud	5,8
4.	-id	3,2	-le	4,9
5.	-se	3,1	-id	4,1
6.	-ks	2,7	-es	3,8
7.	-ga	2,8	-is	3,4
8.	-te	2,8	-ga	3,0
9.	-is	2,6	-se	3,0
10.	-da	2,6	-us	3,0
11.	-es	2,6	-lt	2,7
12.	-as	2,5	-ks	2,6
13.	-ma	2,5	-ma	2,6
14.	-ad	2,4	-da	2,2
15.	-us	2,1	-ad	2,0
16.	-el	2,0	-el	1,8
17.	-ne	1,8	-ta	1,7
18.	-gi	1,8	-ne	1,6
19.	-lt	1,7	-ti	1,6
20.	-de	1,6	-na	1,5
21.	-ed	1,5	-de	1,4
22.	-li	1,5	-al	1,2
23.	-al	1,4	-ed	1,2
24.	-ti	1,3	-te	1,2
25.	-na	1,1	-gi	1,2
Всего (%)		58,5	Всего (%)	70,3

Таблица А-36
(гл. IV, § I)

Частотность биграмм в тексте по данным выборки из 4 подязыков эстонского языка (в равных пропорциях: авторская речь и речь персонажей художественной прозы, газета, научный текст). Объем выборки: 4100 словоупотреблений. В скобках частота биграмм - двухбуквенных словоформ (типа te 'вы', ta 'он, она', ja 'и' и др.).

№ пп.	Би- грамма	Ч а с т о т а			В с е г о	
		в начале словоф.	в конце словоф.	внутри- словно	число	%
1	2	3	4	5	6	7
1.	se	138	130	196	464	2,2
2.	is	27	111	261	399	1,9
3.	st	2	152	238	392	1,9
4.	te	127 (+9)	105 (+9)	111	352	1,7
5.	le	20	150	176	346	1,7
6.	as	22	101	194	317	1,5
7.	us	5	84	227	316	1,5
8.	li	30	64	213	307	1,5
9.	es	21	110	173	304	1,5
10.	ta	64 (+31)	44 (+31)	165	304	1,5
11.	al	31	60	204	295	1,4
12.	el	17	81	197	295	1,4
13.	va	106	19	156	281	1,4
14.	ma	37 (+18)	83 (+18)	140	278	1,3
15.	si	64	40	167	271	1,3
16.	ja	19 (+113)	30 (+113)	108	270	1,3
17.	in	22	26	216	264	1,3
18.	tu	64	24	173	261	1,3
19.	ku	124	33	104	261	1,3
20.	id	7	131	119	257	1,2
21.	mi	110	11	121	242	1,2
22.	ka	102 (+32)	19 (+32)	74	227	1,1
23.	da	-	111	113	224	1,1
24.	ee	30	46	138	214	1,0
25.	ol	117	4	83	204	1,0

(Продолжение следует)

Таблица А-36
(продолжение)

1	2	3	4	5	6	7
26.	ud	-	140	62	202	1,0
27.	ik	10	32	157	199	1,0
28.	ad	2	99	97	198	1,0
29.	at	-	23	174	197	1,0
30.	ks	1	118	75	194	0,9
31.	ne	40	74	76	190	0,9
32.	me	54 (+12)	38 (+12)	75	179	0,9
33.	nd	-	19	148	167	0,8
34.	ga	-	113	54	167	0,8
35.	de	3	66	96	165	0,8
36.	ei	5 (+55)	- (+55)	104	164	0,8
37.	ul	-	34	126	160	0,8
38.	et	11 (+49)	10 (+49)	88	158	0,8
39.	ra	43	18	93	154	0,7
40.	an	21	11	122	154	0,7
41.	oo	3	4	139	146	0,7
42.	la	35	21	87	143	0,7
43.	na	32	45	66	143	0,7
44.	ko	104	-	38	142	0,7
45.	on	2 (+88)	2 (+88)	50	142	0,7
46.	it	-	11	128	139	0,7
47.	ni	45	24	69	138	0,7
48.	ti	4	56	76	136	0,6
49.	il	9	29	97	135	0,6
50.	ak	9	1	125	135	0,6
51.	en	24	5	105	134	0,6
52.	sa	41 (+10)	18 (+10)	62	131	0,6
53.	ll	-	14	115	129	0,6
54.	li	-	17	108	125	0,6
55.	ed	5	64	51	120	0,6
56.	aa	4	8	106	118	0,6
57.	ke	67	21	30	118	0,6
58.	ri	16	10	91	117	0,6

(Продолжение следует)

Таблица 1-56
(продолжение)

1	2	3	4	5	6	7
59.	ek	9	-	108	117	0,6
60.	am	2	9	106	117	0,6
61.	gi	-	72	43	115	0,5
62.	em	2	12	101	115	0,5
63.	nu	4	9	98	111	0,5
64.	ar	31	1	77	109	0,5
65.	ag	37	-	68	105	0,5
66.	uu	12	3	89	104	0,5
67.	ea	-	16	87	103	0,5
68.	su	39	11	53	103	0,5
69.	ju	35 (+8)	13 (+8)	44	100	0,5
70.	ts	-	7	92	100	0,5
71.	im	3	-	97	100	0,5
72.	er	15	9	75	99	0,5
73.	re	10	11	77	98	0,5
74.	he	14	17	66	97	0,5
75.	ki	39	25	30	94	0,4
76.	eg	9	-	84	93	0,4
77.	mu	41 (+3)	5 (+3)	42	91	0,4
78.	ŝi	9	16	65	90	0,4
79.	ut	-	9	81	90	0,4
80.	lu	12	8	66	86	0,4
81.	ŝr	15	2	69	86	0,4
82.	lt	-	70	15	85	0,4
83.	ui	1	45	38	84	0,4
84.	gu	-	31	53	84	0,4
85.	du	-	6	76	82	0,4
86.	un	4	1	75	80	0,4
87.	pa	55	3	21	79	0,4
88.	pe	60	-	19	79	0,4
89.	ah	-	5	73	78	0,4
90.	aj	19	-	56	75	0,4
91.	am	23	1	47	71	0,4

(Продолжение следует)

Таблица А-36
(продолжение)

1	2	3	4	5	6	7
92.	di	6	18	46	70	0,4
93.	nn	-	1	67	68	0,3
94.	av	3	3	61	67	0,3
95.	vð	55	-	10	65	0,3
96.	ha	24	15	25	64	0,3
97.	ev	-	3	61	64	0,3
98.	ai	14	-	47	61	0,3
99.	vi	30	3	27	60	0,3
100.	pi	34	-	24	58	0,3

100 наиболее частых биграмм	2563	3374	10144	16081	77,8 %
Остальные	1097	286	2769	4152	22,2 %
В с е г о	3660	3660	12913	20233	100,0 %
	(4100 вместе с двухбуквенными словоформами)				

Таблица А-37
(гл. IV, § I)

Частотность биграмм в словаре словоформ по данным выборки из 4 подязыков. Объем словаря: 2840 словоформ.

№ пп.	Би- грамма	Ч а с т о т а			В с е г о	
		в нач. СФ	в конце СФ	в ост. поз.	число	%
1	2	3	4	5	6	7
1.	as	21	179	203	403	2,4
2.	st	1	185	165	350	2,1
3.	se	61	86	198	345	2,0
4.	is	23	97	216	336	2,0
5.	ta	61	48	172	281	1,7
6.	us	2	84	190	276	1,6
7.	al	16	35	218	269	1,6
8.	ma	47	73	136	256	1,5
9.	le	22	138	94	254	1,5
10.	va	114	18	120	252	1,5
11.	es	5	109	135	249	1,5
12.	ud	4	165	64	233	1,4
13.	si	58	28	138	224	1,3
14.	tu	46	19	153	218	1,3
15.	ka	118	13	83	214	1,3
16.	el	7	52	154	213	1,3
17.	ku	81	29	96	206	1,2
18.	id	-	115	68	183	1,1
19.	la	71	19	85	175	1,0
20.	li	21	17	132	170	1,0
21.	at	=	20	149	169	1,0
22.	sa	46	25	89	160	0,9
23.	na	38	13	106	157	0,9
24.	in	9	4	143	156	0,9
25.	da	1	63	90	154	0,9

(Продолжение следует)

Таблица А-37
(продолжение)

1	2	3	4	5	6	7
26.	ik	4	16	130	150	0,9
27.	nu	2	6	142	150	0,9
28.	ga	-	86	63	149	0,9
29.	te	39	34	75	148	0,9
30.	ad	-	56	90	146	0,9
31.	na	35	44	64	143	0,8
32.	ul	5	13	124	142	0,8
33.	an	5	-	134	139	0,8
34.	ks	-	74	62	136	0,8
35.	it	-	13	115	128	0,8
36.	ja	28	25	75	128	0,8
37.	ko	105	-	17	122	0,7
38.	de	-	39	82	121	0,7
39.	ri	27	16	75	118	0,7
40.	ar	5	4	108	117	0,7
41.	oo	7	3	106	116	0,7
42.	ne	5	46	61	112	0,7
43.	il	7	19	85	111	0,7
44.	su	47	10	51	108	0,6
45.	ti	58	45	5	108	0,6
46.	ak	2	3	100	105	0,6
47.	et	8	8	89	105	0,6
48.	en	20	1	84	105	0,6
49.	ke	43	17	45	105	0,6
50.	un	4	1	99	104	0,6

50 наиболее

частых биграмм 1327 2213 5479 9019 53,3 %

Остальные 1513 627 5761 7901 46,7 %

В с е г о 2840 2840 11240 16920 100,0 %

Таблица А-38
(гл. IV, § I)

Распределение фонетических типов биграмм в эстонском тексте по данным выборки из 4 подязыков объемом 4100 словоупотреблений (V - гласная, C - согласная буква)

Тип би-граммы	Число возможных биграмм	В начале словоформы				В конце словоформы			
		Реализовано		Частота в тексте		Реализовано		Частота в тексте	
		число	%	число	%	число	%	число	%
VV ⁺	(9 ² =) 81	15	18,5	156	3,8	21	25,9	252	6,1
VC	(9·14=) 126	45	35,7	872	21,3	57	45,2	1488	36,3
CV	(14·9=) 126	102	81,0	3013	73,5	57	45,2	1906	46,5
CC ⁺	(14 ² =) 196	13	6,6	59	1,4	30	11,7	454	11,1
Всего	529	175	33,1	4100	100,0	165	31,2	4100	100,0

+ Из них с одинаковыми буквами:

VV - в начале словоформы - реализовано 6, в тексте 55 (1,3 %); в конце - реализ. 7, в тексте
 CC - " " " 0, " 0 ; " " 5, 82 (2,0%)
 30 (0,7%)

Таблица А-39
(гл. IV, § 1)

Наиболее частотные CV-структуры словоформ в эстонском словаре и тексте по данным базового ЧС словоформ (V = 1168, N = 53027); C - согласный, V - гласный звук; двоеточие после символа означает долготу звука (в письме двойные буквы)

№ пп.	Тип словоформы (CV-структура)	Ч а с т о т а		Примеры (с указанием частоты)
		в словаре	в тексте	
1.	CVCV	86	4614	nagu 536, seda 378
2.	CVCVC(:)	56	1273	tuleb 116, vahel 70
3.	CVCCVC(:)	54	1357	polnud 258, rohkem 86
4.	CVVCV	43	1160	kõige 133, peale 96
5.	CVCCV	43	1454	välja 178, nende 141
6.	CVC:V	41	1699	selle 237, mitte 177
7.	CV:C(:)	38	2198	siis 422, veel 292
8.	CVVC(:)	37	1890	kõik 246, kuid 192
9.	CVCVCV	36	1000	midagi 224, tagasi 109
10.	CVVCVC(:)	30	736	kuidas 148, seisis 47
11.	CV:CVC(:)	30	720	juures 83, vaatas 78
12.	VCV	29	3775	oli 1116, oma 592
13.	CVCC	28	1518	ning 393, sest 155
14.	CVC:VC(:)	25	598	hakkas 99, selles 72
15.	VCVC	24	918	olid 180, enam 165
16.	VCCVC(:)	24	702	ütles 73, olnud 70
17.	CVV	23	1897	kui 929, või 306
18.	VCCV	21	465	enda 48, otse 41
19.	VC:V	19	690	olla 85, ekki 76
20.	CV:CV	19	640	poole 120, keegi 86
21.-				
166.	Остальные (146)	462	23723	-
	В с е г о	1168	53027	

Таблица А-40
(гл. IV, § I)

Распределение CV-структур по функциональной нагрузке Т/С
(Т - частота в тексте, С - частота в словаре) по данным
базового ЧС словоформ

№ пп.	CV-струк- тура	Т	С	Т/С	Примеры
1.	CV	6628	13	509,8	ja, ta, ka
2.	VC(:)	2753	6	458,8	on, et
3.	VV	1404	4	351,0	ei, ðe
4.	VCV	3775	29	130,2	oli, oma
5.	CVC(:)	1870	15	124,7	mis, kes
6.	CV:	1374	14	98,1	see, nii
7.	CVV	1897	23	82,5	kui, vði
8.	CV:C	2198	38	57,8	siis, veel
9.	CVCC	1518	28	54,2	ning, sest
10.	CVCV	4614	86	53,7	nagu, seda
11.	CVVC(:)	1890	37	51,1	kõik, kuid
12.	CVC:V	1699	41	41,4	selle, mitte
13.	VCVC(:)	918	24	38,3	olid, enam
14.	VC:V	690	19	36,3	olla, äkki
15.	CVCCV	1454	43	33,8	välja, nende
16.	CV:CV	640	19	33,7	poole, keegi
17.	VCCVC	702	24	29,3	ütles, olnud
18.	CVCVCV	1000	36	27,8	midagi, tagasi
19.	CVVCV	1160	43	27,0	kõige, peale
20.	CVCCVC	1357	54	25,1	polnud, rohkem

Таблица А-41
(гл. IV, § 1)

Распределение CV-структур по длине в словаре и тексте по данным
базового ЧС словоформ (СФ) эстонского языка

Длина СФ в слогах	Число разных структур	Ч а с т о т а		Наиболее частотные структуры	
		в словаре	в тексте	в словаре	в тексте
1	20	235	23305	CV:C(:),CVVC(:)CVCC	CV,VC(:),CV:C(:), CVV,CVVC(:)
2	58	682	24568	CVCV,CVCVC,CVCCVC	CVCV,VCV,CVC:V
3	56	201	4312	CVCVCV,VCVCV,CVCVCVC	CVCVCV,VCVCV,CVVCVCV
4	26	44	764	VCVCVCV,OVVCVCV	VCVCVCV,VCVCVCVC
5	5	5	65	VCVCVCCVCV	CVCVCVCVCVC
6	1	1	13	CVCVCVCCVCVCCVC	-
В с е г о	166	1168	53027	-	-

Таблица А-42
(гл. IV, § I)

Распределение CV-структур односложных словоформ (ОСФ) в словаре и тексте эстонского языка по длине словоформ в звуках. Объем словаря - 3293 ОСФ, объем текста - 3000 односложных словоупотреблений

Длина СФ	Струк- тура	Словарь число	Словарь %	Текст %	Примеры
1	V:	4	0,12	0,03	õõ, ee, aa, oo
2	CV(:)	60	1,8	30,5	ma, maa
	V:C(:)	34	1,0	0,7	kus, uut
	VC(:)	19	0,6	10,8	on, õnn, ätt
	VV	9	0,3	5,6	ei, au
3	CV:C(:)	513	15,6	10,6	laad, laat
	CVC(:)	362	11,0	9,3	tal, tall, kott
	CVV	113	3,5	9,0	kui, pea
	CCV:	6	0,2	0,03	prii, truu
	VCC	67	2,0	2,4	üks, ehk
	VVC	43	1,3	0,4	aed, ait
	V:CC(:)	4	0,1	0,4	eest, härt
4	CVC(:)C(:)	835	25,3	8,0	kurg, kurk
	CVVC(:)	470	14,3	9,8	vaid, wait
	CV:CC(:)	160	4,9	0,7	keeld, keelt
	CCV:C(:)	76	2,3	0,2	pruun, pruut
	CCVC:	55	1,7	0,03	trumm, trepp
	CCVV	3	0,1	0,03	krae, prao
	CCCV:	1	0,03	0,00	skraa
	VCCC	1	0,03	0,00	unts
	VVCC(:)	18	0,5	0,03	aust, belt

(Продолжение следует)

Таблица А-42
(продолжение)

Длина СФ	Струк- тура	Словарь число	%	Текст %	Примеры
5	CVVC(:)C(:)	236	7,2	1,2	loend, vðike
	CVCC(:)C(:)	100	3,0	0,07	kimps, punkt
	CV:CCC(:)	1	0,03	0,00	paavst
	CCVCC(:)	62	1,9	0,03	sport, skalp
	CCVVC(:)	7	0,21	0,00	kloun, kleit
	CCV:C(:)C	5	0,15	0,03	kriips, prääks
	CCCV:C(:)	4	0,12	0,03	stroof, skvääär
	CCCV:	1	0,03	0,00	sprott
6	CCVCC(:)C	15	0,45	0,03	prints, skunks
	CCVVC(:)C(:)	3	0,09	0,00	praost, šveits
	CCCVCC(:)	2	0,06	0,00	sprint, splint
	CCCV:CC(:)	3	0,09	0,00	skraalt
	CCCVVC(:)	1	0,03	0,00	streik
В с е р о		3293	100 %	100 %	-

Таблица А-43
(гл. IV, § 1)

Канонические формы односложных словоформ
эстонского языка с дифференциацией гласного ядра
(по данным словаря)

№ п/п	Канонич. форма	Я д р о			Всего	
		краткий гласный	долгий гласный	дифтонг	число	%
1.	V	-	4	9	13	0,4
2.	CV	-	60	113	173	5,3
3.	VC	19	34	43	96	2,9
4.	CCV	-	6	3	9	0,3
5.	CVC	362	513	470	1345	40,9
6.	VCC	67	4	18	89	2,7
7.	CCCV	-	1	-	1	0,03
8.	CCVC	55	76	7	138	4,2
9.	CVCC	835	160	236	1231	37,4
10.	VCCC	1	-	-	1	0,03
11.	CCCVCC	1	4	1	6	0,2
12.	CCVCC	62	5	3	70	2,1
13.	CVCCC	100	1	-	101	3,0
14.	CCCVCCC	2	3	-	5	0,1
15.	CCVCCC	15	-	-	15	0,5
Всего		1519	871	903	3293	-
%		46,1	26,5	27,4	-	100,0

Таблица А-44

(гл. IV, § 1)

Канонические формы односложных словоформ
эстонского языка с дифференциацией гласного ядра

(по данным текста)

№ пп.	Канонич. форма	Я д р о			Всего	
		краткий гласный	долгий гласный	дифтонг	число	%
1.	V	-	1	168	169	5,6
2.	CV	-	916	270	1186	39,5
3.	VC	323	20	13	356	11,9
4.	CCV	-	1	1	2	0,07
5.	CVC	278	317	296	891	29,7
6.	VCC	73	11	1	85	2,8
7.	CCCV	-	-	-	-	-
8.	CCVC	1	5	-	6	0,23
9.	CVCC	239	22	38	299	10,0
10.	VCCC	-	-	-	-	-
11.	CCCVC	-	1	-	1	0,03
12.	CCVCC	1	1	-	2	0,07
13.	CVCCC	2	-	-	2	0,07
14.	CCCVCC	-	-	-	-	-
15.	CCVCCC	1	-	-	1	0,03
Всего		918	1295	787	3000	-
%		30,6	43,2	26,2	-	100,0

Таблица А-45

(гл. IV, § 2)

Распределение частот производных слов в тексте эстонской художественной прозы (авторская речь и речь персонажей) и в соответствующем словаре (по данным исследования В. Кюауса)

Часть речи	Т е к с т			С л о в а р ь		
	Всего (число)	Из них производных (число)	(%)	Всего (число)	Из них производных (число)	(%)
Существительные	1642	373	22,7	1295	339	26,2
Прилагательное	371	251	67,7	345	235	68,1
Глагол	913	216	23,4	566	193	34,1
Наречие	826	171	20,7	369	114	30,9
Местоимение	456	22	4,8	100	14	14,0
Числительное	83	9	10,8	43	6	13,9
Остальные	566	-	-	184	-	-
Всего	4857	1042	21,5	2902	901	31,0

Таблица А-46
(гл. IV, § 2)

Распределение частот суффиксов производных существительных в словаре и тексте по данным 40 лексем авторской речи современной эстонской художественной прозы (N = 99898, L = 14654)

№ пп.	Суф- фикс	Словарь		Т е к с т		Примеры сущ-х
		число	%	число	%	
1.	-us	727	26,5	1751	25,6	valgus; küsimus
2.	-mine	483	17,6	715	10,4	tegemine, tahtmine
3.	-(i)k	256	9,3	574	8,4	rätik; tulek
4.	-ja	239	8,7	353	5,2	õpetaja, müüja
5.	-e	147	5,3	603	8,8	teade, hõpe
6.	-u	92	3,4	729	10,7	nägu, elu
7.	-lane	77	2,8	223	3,3	vaenlane, sugulane
8.	-nu, -tu	74	2,7	102	1,5	surnu; tundmatu
9.	-kas	73	2,7	129	1,9	kajakas, jõmpsikas
10.	-ke(ne)	73	2,7	122	1,8	lauake, osake
11.	-n	64	2,3	84	1,2	kumin, mürin
12.	-line	39	1,4	70	1,0	külaline, tööline
13.	-nd	37	1,3	71	1,0	loend, olend
14.	-ng	37	1,3	61	0,9	areng, nauding
15.	-kond	36	1,3	104	1,5	perekond, ülikond
16.	-is	35	1,3	58	0,8	näidis, petis
17.	-nik	32	1,2	83	1,2	kõlhoosnik, põgenik
18.	-ndus	18	0,7	48	0,7	varandus, kirjandus
19.	-stik	18	0,7	29	0,4	ilmastik, koristik
20.	-ur	18	0,7	39	0,6	sõdur, kalur
Остальные		168	6,1	898	13,1	-
Всего произ- водн. сущ-х		2743	100,0	6846	100,0	-
Всего существ- тельных		9025	-	31650	-	-

Таблица А-47
(гл. IV, § 2)

Распределение морфологических структур производных существительных в тексте эстонской художественной прозы и в соответствующем словаре: Т - основа, С - суффикс

Структура	Частота (%)		Наиб. частотные суффиксы
	в тексте	в словаре	
Т (глагол.) + С	62,9	34,1	-us, -mine, -u
Т (сущ.) + С	24,6	32,9	-(i)k, -kas, -ne
Т (прил.) + С	8,2	14,1	-us, -(i)k, -is
Т (нареч.) + С	2,4	9,4	-us, -(i)k, -ne
Т (сущ.) + С+С	1,6	4,7	
Т (числ.) + С	0,3	4,7	
Всего (%)	100,0	100,0	-
Объем выборки	1000	85	-

Отношение текст/словарь

(функц. нагрузка):

$$1000/85 = 11,8$$

Таблица 1-48
(гл. IV, § 2)

Распределение морфологических структур производных прилагательных в тексте эстонской художественной прозы и в соответствующем словаре: Т - основа, С - суффикс

Структура	Частота (%)		Наиб. частотные суффиксы
	в тексте	в словаре	
Т (сущ.) + С	66,0	34,1	-ne, -lik, -line
Т (нареч.) + С	13,2	17,0	-ne, -mine, -ik
Т (глагол.) + С	10,1	9,8	-lik, -tu, -s
Т (прилаг.) + С	8,9	27,9	-ne, -kas, -lik
Т (числ.) + С	1,6	7,3	
Т (местоим.) + С	0,2	4,9	
	100,0	100,0	
	1000	41	

Отношение текст/словарь:

$$1000/41 = 24,4$$

IV. Прилагательные (n = 1): 12

V. Числительные (n = 2): 18, 10

VI. Местоимения (n = 2): 4, 20

Сложные слова в базовом словаре эстонского языка (объем словаря - 1202 лексемы, из них 76 сложных; n - число слов, R - число подвыборок, F - частота)

I. Существительные (n = 28 + 6 суд. при R < 3)

maailm (F=56)	ajalugu	16	meeleolu	12
peremees 44	olukord	16	maantee	11
perenaine 37	unenägu	16	tugitool	11
vanamees 33	ülesanne	16	tähelepanu	11
noormees 21	ajaleht	14	tänapäev	11
tütarlaps 20	lapsepõlv	14	õemees	11
vanaema 19	sõõgilaud	14	silmapiir	10
vanaisa 19	silmapiik	13	sõjavägi	10
perekond 17	esimees	12	talumees	10
			ülikond	10

II. Наречия (n = 27)

võib-olla 82	seepärast	21	kõigepealt	12
peaaegu 71	seejärel	19	niimoodi	12
mõnikord 44	vahetevahel	18	iseenesest	11
niisama 37	esialgu	17	millegipärast	11
sellepärast 36	vahapeal	16	sealsamas	11
tõepoolest 35	seejuures	13	ettepoole	10
niiviisi 33	ükskord	13	niipalju	10
igatahes 28	ükskõik	13	omaette	10
aeg-ajalt 23	enam-vähem	12	pikkamööda	10

III. Местоимения (n = 10)

niisugune 93	teineteise	37	samasugune	22
üksteise 48	missugune	27	igasugune	20
seesama 41	iseene	23	mingisugune	10
			mitmesugune	10

IV. Прилагательные (n = 1): salapärane 12

V. Числительные (n = 2): kakskümmend 18, viiskümmend 10

VI. Союзы (n = 2): otsekui 41, justkui 20

Таблица А-50

(гл. IV, § 2)

Распределение частот сложных словоформ (ССФ) по длине
в тексте авторской речи современной художественной прозы по дан-
ным выборок из 10 произведений

Длина ССФ в слогах	Э.Бэзман	В.Гросс	Х.Кийк	Я.Кросс	П.Куусберг	Ю.Смуул	М.Траат	М.Унт	А.Валтон	Э.Ветемаа	Всего	в среднем (%)	Ошибка ср. при $\alpha=0,05$
2	3	7	7	3	7	4	1	4	2	2	40	4,3	$\pm 1,6$
3	23	26	32	22	19	13	29	15	10	21	210	22,8	$\pm 5,0$
4	33	24	37	46	36	29	42	40	11	36	334	36,2	$\pm 7,2$
5	26	20	14	25	24	20	20	19	10	29	207	22,4	$\pm 4,1$
6	6	9	10	16	11	9	6	7	5	13	92	10,0	$\pm 2,5$
7	1	6	2	2	8	-	2	4	1	4	30	3,3	-
8	-	1	-	1	1	-	-	4	-	1	8	0,8	-
9	-	-	1	1	-	-	-	-	-	-	2	0,2	-
Всего	92	93	103	116	106	75	100	93	39	106	923	100,0	-
Объем выборки	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	10.000	-	-
Доля ССФ в тексте (%)	9,2	9,3	10,3	11,6	10,6	7,5	10,0	9,3	3,9	10,6	9,2	-	$\pm 1,5$

Таблица 1-51

(гл. IV, § 2)

Распределение сложных словоформ (ССФ) по длине в
тексте эстонской художественной прозы

Длина ССФ в слогах	Авторская речь (%)		Речь персон. (%) (соврем. литер.)
	Соврем. литер.	Литерат. 1900-1940	
2	4,3	3,1	8,6
3	22,8	27,9	23,0
4	36,2	46,9	42,2
5	22,4	16,3	15,6
6	10,0	5,1	8,6
7	4,3	0,7	2,0
Всего (%)	100,0	100,0	100,0
Доля ССФ в тексте (%)	9,2	7,4	4,9
Ср. длина ССФ в слогах	4,25	3,95	4,00

Таблица 1-52
(гл. IV, § 2)

Структурные типы сложных словоформ (ССФ) в тексте эстонской художественной прозы (по числу слогов в первом и во втором компоненте сложного слова)

Структ. тип	Частота (%)		Примеры (словоформы)
	Авт. речь	Речь персон.	
1-1	4,3	8,6	seekord, niipea, noormees
1-2	12,6	12,7	peaasi, ausamba, sealsamas
1-3	3,8	6,6	niisugune, eesmärgiga
1-4	1,6	0,4	poolunustatud, õunrohelisteks
1-5	0,2	-	kunstmootorratturid
1-6	-	0,4	puhttehnoloogiliselt
2-1	10,2	10,2	igaüks, mõnikord, ülikool
2-2	31,4	35,2	päevakava, abielu, ebamäärast
2-3	15,2	12,3	tumepunase, õhtupoolikul
2-4	5,9	3,7	kooliõpetaja, vahetpidamata
2-5	0,9	0,4	sissejuhatusega, üliintelli- gentsed
2-6	0,1	-	endastmõistetavusega
3-1	1,0	0,4	sadamakai, valgusekeelt
3-2	5,4	2,9	tähelepanu, vahetevahel
3-3	2,6	2,1	räämatupidaja, edasijõudnutel
3-4	1,5	-	teravalõualine, joonestuspa- berile
3-5	0,4	-	kaheksakorruseline, kohviku- tuttavatega
4-1	0,2	-	kaardimängutoaks
4-2	1,2	2,9	sigaretihimu, otsustamiskiire
4-3	0,8	0,4	valitsejaisandat
4-4	0,2	-	vägilasevuntsidega
4-5	0,1	0,4	neljakuune-tuhandelise, tel- liskivi-vaheseinaga
5-1	0,1	-	laudatulekuteel
5-2	0,1	0,4	unenäguderohket, pesupesemismasin
5-4	0,1	-	depressioonimeeleolu
6-2	0,1	-	kaheinimesevoodilt
Всего (%)	100,0	100,0	-
Число	923	249	-
Объем текста	10000	5000	-

Таблица А-53
(гл. IV, § 2)

Словообразовательный потенциал имен существительных по базовому словарю эстонского языка (каждое десятое существительное по порядку убывания частот в ЧС): C_I - первый компонент, C_2 - второй компонент сложного слова, Д - производное слово

Лексема	Частота (число)	Число C_I	Число сложных слов C_2	Число слов C_I+C_2	Число произв. слов(Д)	$C_I+C_2+Д$
1. mees	373	101	186	287	17	304
2. nāgu	159	26	8	34	8	42
3. tõõ	105	244	248	492	15	507
4. kodu	87	95	8	103	12	115
5. hetk	71	5	4	9	3	12
6. taevas	60	36	11	47	4	51
7. aru	52	18	1	19	9	28
8. peremees	44	6	11	17	3	20
9. lugu	37	4	45	49	-	49
10. osa	34	36	87	123	13	136
11. rind	32	73	12	85	22	107
12. nādal	29	14	8	22	2	24
13. kevad	27	42	4	46	3	49
14. kala	26	138	69	207	9	216
15. auto	24	64	27	91	1	92
16. koorem	23	10	28	38	5	43
17. suvi	22	74	8	82	11	93
18. pere	21	17	21	38	5	43
19. leib	20	84	28	112	7	119
20. pesa	19	24	45	69	11	80

(Продолжение следует)

Лексема	Частота (число)	Число сложных слов			Число произв. слов(II)	C_1+C_2+D
		C_1	C_2	C_1+C_2		
21. koer	18	45	39	84	5	89
22. aed	17	82	64	146	5	151
23. perekond	17	26	6	32	4	36
24. külaline	16	10	6	16	-	16
25. tegu	16	27	20	47	18	65
26. määng	15	71	109	180	8	188
27. kukal	14	18	-	18	4	22
28. viha	14	13	3	16	10	26
29. kleit	13	5	39	44	1	45
30. näide	13	4	3	7	2	9
31. tiib	13	25	14	39	16	55
32. kaabu	12	1	4	5	-	5
33. number	12	17	32	49	1	50
34. vagu	12	5	23	28	12	40
35. lagendik	11	1	6	7	-	7
36. põgenik	11	-	1	1	-	1
37. vaim	11	45	24	69	10	79
38. jälg	10	2	17	19	9	28
39. onu	10	3	2	5	1	6
40. sõdur	10	8	13	21	1	22
В с е р о	1530	1519	1284	2803	267	3070
В среднем	38,25	37,98	32,1	70,08	6,68	76,75

Таблица А-54

(гл. IV, § 2)

Распределение частот частей речи (в %) в словаре эстонского языка по данным выборок разного объема (лексемы).
 I: подвыборка, II: 4С лексем, III: ОС-76 (Õigekeelsussõnaraamat, 1976)

Часть речи	I	II	III
Существительное	44,2	61,6	75,0
Глагол	26,6	14,0	8,5
Прилагательное	9,5	11,7	11,4
Наречие	10,7	9,6	4,2
Местоимение	4,4	0,4	0,07
Числительное	1,3	1,0	0,2
Релят. слова (пред- и послелоги)	2,2	1,0	0,1
Союз	0,9	0,2	0,03
Междометие	0,2	0,5	0,5
Всего (%)	100,0	100,0	100,0
Объем словаря	2200	14650	115000
Объем текста	5000	100000	-

Таблица А-55
(гл. IY, § 2)

Динамика изменения долей частей речи (в %) по частотным зонам в словаре по данным 40 лексем авторской речи современной эстонской художественной прозы, исключая собственные имена ($n = 1180$) и числа ($n = 65$). Общий объем словаря $L = 14654$ лексем; объем текста $N = 99898$ словоупотреблений.

[illegible]

Таблица А-56
(гл. IV, § 2)

Распределение частот частей речи (в %) в тексте авторской речи современной эстонской художественной прозы по данным выборки из 10 произведений (10 × 5000 словоупотреблений)

Автор \ Часть речи	Э. Бээкман	Э. Гросс	А. Хинт	Х. Кийк	Э. Кресс	Л. Промет	В. Саар	Х. Серго	М. Траат	А. Балтон	В среднем (± ошибка ср. при $\alpha = 0,05$)	
Существит.	35,9	30,4	32,5	34,8	32,8	30,4	26,8	36,8	30,7	26,8	31,7	±2,6
Глагол	25,0	22,2	19,1	24,2	18,3	24,0	23,5	22,0	22,6	24,0	22,5	±1,6
Прилагат.	5,4	7,4	5,6	5,2	7,4	5,5	4,9	5,8	6,2	6,3	6,0	±0,6
Наречие	13,9	17,4	16,6	15,0	16,5	13,9	17,3	14,8	15,9	16,3	15,8	±0,9
Местоим.	8,3	11,7	12,9	9,3	9,1	13,8	13,8	7,9	12,7	14,4	11,4	±1,8
Числит.	0,6	1,4	1,3	1,3	1,7	0,9	0,9	1,4	1,0	0,5	1,1	±0,3
Релят.	3,7	2,5	3,5	2,4	3,9	2,9	3,7	3,6	2,6	2,6	3,1	±0,4
Союз	7,2	6,8	8,3	7,7	10,1	8,5	9,5	7,6	8,1	9,0	8,2	±0,7
Междом.	-	0,2	0,2	0,1	0,2	0,1	0,4	0,1	0,2	0,1	0,2	±0,1
Всего (%)	100	100	100	100	100	100	100	100	100	100	100	-

Таблица А-57
(гл. IV, § 2)

Распределение частот частей речи (в %) в тексте разных подъязыков
(стилей) эстонского языка

Часть речи	Авт. речь	Авт. речь	Газета		Народ-	Речь персон.
	совр. худ. прозы	А.Х.Тамм- сааре	Спорт. новости	Сообщ. ТАСС	ная песня	совр. худ. прозы
Существит.	31,7	30,1	34,2	49,1	53,2	17,4
Глагол	22,5	23,6	22,8	19,6	24,9	26,6
Прилагат.	6,0	4,1	7,7	7,7	9,5	5,1
Наречие	15,8	16,5	14,5	5,9	6,3	21,8
Местоим.	11,4	11,8	9,1	6,3	2,5	20,0
Числит.	1,1	0,8	3,1	1,8	0,9	0,6
Релят.	3,1	2,8	2,4	3,4	1,8	1,7
Союз	8,2	10,3	6,2	6,2	0,6	5,0
Междом.	0,2	0,0	0,0	0,0	0,3	1,8
Всего (%)	100,0	100,0	100,0	100,0	100,0	100,0

Распределение частот частей речи (в %) в тексте художественной прозы разных языков (данные взяты из работ: Saukkonen P. et al., 1979; Zsilka T., 1973; Якубайтис Т.А., 1974; Ключкова Э.А., 1968; Тищенко В., 1970)

Язык Часть речи	Эстонский	Финский	Венгер- ский	Латыш- ский	Русский	Украин- ский
Существит.	31,7	29,7	30,0	30,1	28,7	29,2
Глагол	22,5	27,6	22,4	23,0	18,3	19,7
Прилагат.	6,0	7,8	10,0	5,5	7,9	6,8
Наречие	15,8	10,9	8,0	10,0	6,0	6,2
Местоим.	11,4	11,2	5,7	12,2	10,2	9,0
Числит.	1,1	1,2	1,5	1,3	1,2	1,1
Пред-, послелог	3,1	3,6	21,9	5,9	12,2	12,5
Союз	8,2	7,3		8,0	8,5	9,1
Междом.	0,2	0,7	0,5	0,3	0,0	0,0
Частицы	-	-	-	3,7	7,0	6,4
Всего (%)	100,0	100,0	100,0	100,0	100,0	100,0
Энтропия (H)	2,62	2,60	2,47	2,76	2,85	2,82

Таблица А-59

(гл. IV, § 2)

Частотность лексико-морфологических классов слов (в %) в тексте по данным разных подъязмов (стилей) эстонского языка

Морфол. группы частей р.	Авт. речь совр. худ. прозы	Авт. речь А.Х.Тамм- сааре	Газета Спортивн. новости	Сообщ. ТАСС	Народ- ная песня	Речь персон. совр. худ. прозы
Склоняемые слова	50,2	46,8	54,1	64,9	66,1	43,1
Глагол:						
склон. формы	8,2					6,6
сприг. формы	14,3	22,5	22,8	19,6	24,9	20,0
Все изменяемые слова	72,7	70,4	76,9	84,5	91,0	69,7
Неизмен. слова	27,3	29,6	23,1	15,5	9,0	30,3
Все слова (%)	100,0	100,0	100,0	100,0	100,0	100,0

Таблица А-60
(гл. IV, § 2)

Частотность лексико-морфологических классов слов (в %) в текстах разных языков

Морфол. группа \ Язык	Финский	Венгер-ский	Латыш-ский	Русский	Украин-ский
Склоняемые слова	49,2	47,2	49,1	48,0	46,1
Глагол	27,6	22,4	23,0	18,3	19,7
(в т.ч. склоняемые формы)	(?)	(4,6)	(?)	(3,5)	(?)
Все изменяемые слова	76,8	69,6	72,1	66,3	65,8
Неизменяемые слова	23,8	30,4	27,0	33,7	34,2
Все слова (%)	100,0	100,0	100,0	100,0	100,0

Таблица А-61
(гл. IV, § 3)

Распределение частот лексико-семантических групп имен существительных в авторской речи современной эстонской художественной прозы (по данным сводной выборки из 20 произведений)

Лексико-семантическая группа	Словарь		Текст		Отнош.	
	число (С)	%	число (Т)	%	Т/С	
<u>Собственные имена:</u>						
Имена людей	476	5,4	2699	8,6	5,7	
Прочие	704	8,1	1367	4,3	1,9	
Всего собственных имен	1180	13,5	4066	12,9	3,5	
<u>Нарицательные сущ. (по ДСП):</u>						
НIII: лицо	} одуш.	1078	12,4	3961	12,6	3,7
НIIIО: нелицо		162	1,9	568	1,8	3,5
НIOI: природное	} неодуш.	1139	13,1	5634	17,9	4,9
НIOO: неприродное		2207	25,3	7185	22,8	3,3
Всего конкретных сущ.		4586	52,7	17348	55,1	3,8
НОII: действие	} акт.	1084	12,5	2256	7,2	2,1
НОIO: недействие		309	3,5	908	2,9	2,9
НОOI: свойство	} неакт.	145	1,7	468	1,5	3,2
НОOO: несвойство		1403	16,1	6461	20,5	4,6
Всего абстрактных сущ.		2941	33,8	10093	32,0	3,4
Всего нарицательных сущ.		7527	86,5	27441	87,1	3,6
Всего существительных		8707	100,0	31507	100,0	3,6
Всего слов в авторской речи		14654	-	99898	-	6,8

Таблица А-62

(гл. IV, § 3)

Распределение частот лексико-семантических групп имен существительных в авторской речи I-го тома романа А.Х. Таммсааре "Правда и справедливость" (эст.)

Лексико-семантическая группа	Словарь число (С) %		Текст число (Т) %		Отнош. Т/С	
<u>Собственные имена:</u>						
Имена людей	93	2,3	6443	18,8	69,3	
Прочие	58	1,5	1438	4,2	24,8	
Всего собственных имен	151	3,8	7881	23,0	52,2	
<u>Нарицательные сущ. (по ДСП):</u>						
Н111: лицо	} одуш.	451	11,2	5060	14,7	11,2
Н110: нелицо		136	3,4	992	2,9	7,3
Н101: природное	} неодуш.	543	13,5	5679	16,6	10,5
Н100: неприродное		1002	25,0	5580	16,3	5,6
Всего конкретных сущ.		2132	53,1	17311	50,5	8,1
Н011: действие	} акт.	736	18,3	2283	6,6	3,1
Н010: недействие		175	4,4	884	2,6	5,1
Н001: свойство	} неакт.	145	3,6	500	1,5	3,4
Н000: несвойство		675	16,8	5424	15,8	8,0
Всего абстрактных сущ.		1731	43,1	9091	26,5	5,3
Всего нарицательных сущ.		3863	96,2	26402	77,0	6,8
Всего существительных		4014	100,0	34283	100,0	8,5
Всего слов в авторской речи		7348	-	114124	-	15,5

Таблица А-63
(гл. IV, § 3)

Распределение частот лексико-семантических групп имен существительных в речи персонажей I-го тома романа А.И. Тами-сааре "Правда и справедливость" (зст.)

Лексико-семантическая группа	Словарь		Текст		Отнош.	
	число (С)	%	число (Т)	%	Т/С	
<u>Собственные имена:</u>						
Имена людей	77	4,7	961	10,2	12,5	
Прочие	46	2,8	486	5,1	10,6	
Всего собственных имен	123	7,5	1447	15,3	11,8	
<u>Нарицательные сущ. (по ДСП):</u>						
НИИ: лицо	} одуш.	238	14,5	1984	21,0	8,3
НИО: нелицо		94	5,7	410	4,4	4,4
НИОІ: природное	} неодуш.	250	15,2	1589	16,8	6,4
НИОО: неприродное		415	25,3	1680	17,8	4,0
Всего конкретных сущ.		997	60,7	5663	60,0	5,7
НОІІ: действие	} акт.	143	8,7	416	4,4	2,9
НОІО: недействие		36	2,2	150	1,6	4,2
НООІ: свойство	} неакт.	35	2,1	191	2,0	5,5
НООО: несвойство		309	18,8	1576	16,7	5,1
Всего абстрактных сущ.		523	31,8	2333	24,7	4,5
Всего нарицательных сущ.		1520	92,5	7996	84,7	5,3
Всего существительных		1643	100,0	9443	100,0	5,7
Всего слов в речи персонажей		3135	-	46232	-	14,7

Таблица А-64
(гл. IV, § 3)

Список наиболее частотных собственных имен по данным 40 авторской речи современной эстонской художественной прозы; числа обозначают частоту и количество подвыборок (макс. 20).

А. Имена людей

39/6	Jaan	20/3	Heino
32/4	Juhan	17/3	Jüri
23/4	Mihkel	10/3	Paul
17/4	Villem	10/3	Robert
35/3	Enn	9/3	Lembit
28/3	Anna	6/3	Aadu
28/3	Ester		

В двух подвыборках: Aadam, Aleksander, Alfred, Andrus, Ants, Bernhard, Eduard, Elmar, Irene, Jaak, Jaanus, Jakob, Johannes, Joosep, Joss, Kaarel, Katharina, Laas, Liiv, Maarja, Maria, Mati, Minna, Ott, Riina, Roosi, Sass, Tõnis, Vambo.

Б. Географические названия

33/12	Tallinn	9/3	Eestimaa
24/9	Tartu	7/3	Siber
16/9	Eesti	6/3	Kadrioriorg
9/5	Venemaa	6/3	Saksa
8/5	Moskva	4/3	Narva
9/4	Saaremaa	4/3	Pärnu
4/4	Emajõgi	3/3	Toompea
4/4	Saksamaa	3/3	Volga

В двух подвыборках: Aafrika, Euroopa, Haapsalu, Inglismaa, Kuivastu, Kuressaare, Leningrad, Liivimaa, Murmansk, Nõmme, Paide, Pariis, Riia, Rooma, Toonemägi, Viljandi, Virtsu, Ulemiste.

Список наиболее частотных существительных — названий людей (группа NIII) — по данным 40 авторской речи современной эстонской художественной прозы; числа обозначают частоту в тексте объемом 100 тыс. словоупотреблений.

373	mees	'муж(чина)'	16	külaline	'гость'
254	inimene	человек	16	poisike(ne)	мальчишка
185	naine	жен(щин)а	14	kaaslane	сотоварищ
122	poiss	мальчик	14	kolhoosnik	колхозник
101	isa	отец	14	augulane	родственник
93	ema	мать	14	surru	покойник
74	laps	ребенок	13	tütar	дочь
70	tüdruk	девочка	12	naaber	сосед
63	pöeg	сын	12	esimees	председатель
53	vend	брат	11	madrus	матрос
48	õde	сестра	11	õpetaja	учитель
44	peremees	хозяин	11	vaenlane	враг
37	perenaine	хозяйка	11	võõras	чужой, гость
33	vanamees	старик	11	põgenik	беженец
28	sõber	друг	10	daam	дама
27	eit	баба, старуха	10	zulane	батрак
27	sakslane	немец	10	sõdur	солдат
26	härja	господин	10	talumees	крестьянин
21	proua	госпожа	10	onu	дядя
21	vanem(ad)	родители	10	tuttav	знакомец
21	noormees	юноша	Коллективы:		
20	tütarlaps	девочка, девушка	38	rahvas	народ
19	vanaema	бабушка	21	pere	семья
19	vanaisa	дедушка	17	perikond	семейство
19	kapten	капитан	10	sõjavägi	армия

Таблица А-66

(гл. IV, § 3)

Список наиболее частотных существительных - названий
птиц и животных (группа IIIO).

89	hobune	'лошадь'	6	hani	'гусь'
30	lehm	корова	6	kõrb (kõrvi)	гнездо
29	loom	животное	6	lõvi	лев
26	kala	рыба	6	põrsas	поросенок
22	lind	птица	5	haug	щука
18	koer	собака	5	mullikas	телка
17	kass	кошка	5	part	утка
16	hiir	мышь	5	sipelgas	муравей
13	kajakas	чайка	5	vares	ворона
11	luik	лебедь	4	jänes	заяц
10	lammas	овца	4	karu	медведь
10	uss	змея	4	kukk	петух

Таблица А-67

(гл. IV, § 3)

Список наиболее частотных существительных — названий органов и частей тела.

260	silm	'глаз'	24	kõrv	'ухо'
237	käsi	рука	24	põlv	колени
163	pea	голова	23	hammas	зуб
159	nägu	лицо	23	huul	губа
125	jalg	нога	21	lõug	подбородок
100	selg	спина	18	nahk	кожа
58	süda	сердце	18	sõrm	палец; перст
52	suu	рот	15	näpp	палец
48	keel	язык	14	habe	борода
47	õlg	плечо	14	kukal	затылок
40	kael	шея	14	põsk	щека
36	juus	волос	13	kõht	живот
33	nina	нос	11	säär	голень
32	rind	грудь	10	süli	лоно (руки)
31	pihk	ладонь	10	turi	заплечье, затылок
26	keha	тело	10	veri	кровь

Относящиеся только к животным:

20	muna	'яйцо'
19	pesa	гнездо
13	tiib	крыло
10	saba	хвост
10	sarv	рог

Список наиболее частотных существительных группы НГОІ
("природное").

А. Земля, воды, погода и т.п.

I30 maa	'земля'	27 vihm	'дождь'	I4 väin	'пролив'
II2 meri	море	26 jõgi	река	I3 laht	бухта
68 vesi	вода	26 kallas	берег	I3 loodus	природа
66 tuul	ветер	25 laine	волна	I2 kõrre	пустыня
56 maailm	мир, свет	22 mägi	гора	I2 org	долина
45 kivi	камень	I8 pilv	облако	II künkas	холм
43 lumi	снег	I7 järv	озеро	II lagendik	равнина
3I ilm	погода	I6 liiv	песок	II saar	остров
3I õhk	воздух	I5 muld	земля, почва	I0 jää	лед

Б. Растения и деревья

7I mets	'лес'	I6 kask	'береза'
49 puu	дерево	I6 kuusk	ель
27 rohi	трава	I6 männ	сосна
20 leht	лист	I5 oks	ветвь
I8 hein	сено	I3 lill	цветок
I6 kadakas	можже- вельник	II põõsas	куст

В. Небесные тела и направления света

75 päike	'солнце'	24 lõuna	'юг'
60 taevae	небо	8 lään	запад
32 kuu	месяц, луна	5 ida	восток
28 täht	звезда	5 planeet	планета
26 põhj	север	2 kosmos	космос

Таблица А-69
(гл. IV, § 3)

Список наиболее частотных существительных группы Н100
("артефакты").

А. Предметы быта, орудия труда

97 laud	'стол'	I3 kott	'мешок'	10 kohver	'чемодан'
39 kell	часы	I3 püss	ружье	10 säng	постель
35 pall	мяч	I3 konka	крючок	10 ämber	ведро
35 tool	стул	I3 tekk	одеяло	9 kirves	топор
22 masin	машина	I2 labidas	лопата	9 käru	тележка
21 pudel	бутылка	II korv	корзинка	9 küünl	свеча
I9 ratas	колесо	II peegel	зеркало	9 pink	скамейка
I6 voodi	кровать	II tugitool	кресло	8 kast	ящик
I4 sõõgilaud	обеденный стол	IO kaup	товар	8 pakk	пакет(ик)

Б. Названия одежды

27 riie	'одежда; ткань'	I2 kaabu	'шляпа'	9 rätik	'платок'
I8 tasku	карман	I2 kübar	шляп(к)а	9 seelik	юбка
I7 müts	шапка	II püksid	брюки	8 mantel	пальто
I4 king	ботинок	IO ülikond	костюм	8 saabas	сапог
I3 kleit	платье	9 kasukas	шуба	6 kuub	пиджак

В. Названия пищи и т.п.

20 leib	'хлеб'	8 piim	'молоко'	6 sool	'соль'
20 õlu	пиво	8 toit	пища	6 vein	вино
I8 kartul	картофель	7 kohv	кофе	3 jook	напиток
IO sigaret	сигарета	7 liha	мясо	3 rasv	жир
9 viin	водка	7 sõõk	еда	2 jahu	мука

(Продолжение следует)

Таблица А-69
(продолжение)Г. Жилища и строения

88 maja	'дом'	40 sein	'стена'	20 ruum	'помещение'
87 kodu	дом	39 kirik	церковь	18 müür	ограда, стена
85 tuba	комната	35 värav	ворота	17 korter	квартира
74 aken	окно	26 katus	крыша	16 saal	зал
67 uks	дверь	25 trepp	лестница	14 põrand	пол

Д. Средства и пути передвижения

90 tee	'дорога'	26 sild	'мост'	15 vagun	'вагон'
60 tänav	улица	24 auto	авто- машина	12 rada	тропа
55 laev	корабль	22 buss	автобус	11 maantee	шоссе
54 vanker	телега	16 rong	поезд	9 jaam	станция
33 paat	лодка	15 sadam	гавань	7 jalg- ratas	велосипед

Е. Поселения и т.п.

79 linn	'город'	22 mõis	'мыза'
74 küla	деревня	21 kolhoos	колхоз
37 talu	хутор	7 vald	волость

Ж. Культура и искусство

34 pilt	'картина'	12 kino	'кино(театр)'
33 kiri	письмо	10 raadio	радио
29 raamat	книга	7 ülikool	университет
25 kool	школа	6 ajakiri	журнал
14 ajaleht	газета	6 teater	театр

Таблица А-70

(гл. IV, § 3)

Список наиболее частотных абстрактных существительных по признаку "активность" (процесс, действие и результат действия).

108	mõte	'мысль'	17	tahtmine	'хотение'
105	töö	работа	16	nõu (nõuanne)	совет
93	jutt	рассказ	16	sõit	поездка
71	lõpp	конец	16	sündmus	происшествие
57	tunne	чувство	16	tegu	дело
33	abi	помощь	15	algus	начало
31	küsimus	вопрос	15	kujutlus	представление
27	huvi	интерес	13	arvamus	мнение
22	naer	смех	13	juht (juhu)	случай
21	lootus	надежда	13	kasv	рост
21	soov	желание	13	teade	извещение
20	armastus	любовь	13	vaidlus	спор
20	vastus	ответ	11	käsk	приказ
18	tegemine	дело	11	naeratus	улыбка
17	mälestus	воспоминание	10	liikumine	движение

Таблица А-71
(гл. IV, § 3)

Список наиболее частотных абстрактных существительных по признаку "неактивность":

А. Свойство, качество

44 jõud	'сила'	13 hämarus	'сумрак'
35 valgus	свет	12 kasu	польза
32 õigus	право	12 lähedus	близость
21 kahju	вред	11 tühjus	пустота
21 tõsi	истина	10 halb (subst.)	худо
21 vaikus	тишина	10 kõrgus	высота
20 hea (subst.)	добро	10 külm (subst.)	холод
17 pimedus	темнота	10 tõde	правда

Б. Состояние, чувства

122 elu	'жизнь'	19 hädä	'беда'
78 meel	чувство	19 õnn	счастье
57 tunne	чувство, ощущение	16 olukord	положение
37 rõõm	радость	16 valu	боль
33 mure	скорбь	14 viha	гнев
29 surm	смерть	12 meeleolu	настроение
27 rahu	мир	12 tervis	здоровье
24 uni	сон	12 tuju	настроение
21 hirm	страх	11 king	страсть
21 häbi	стыд	11 mõnu	удовольствие

(Продолжение следует)

Таблица А-71
(продолжение)В. Время

285 aeg	'время'	27 kevad	'весна'
165 päev	день	27 talv	зима
151 aasta	год	22 suvi	лето
94 õhtu	вечер	13 silmapilk	миг
71 hetk	мгновение	13 sügis	осень
59 hommik	утро	11 tänapäev	сегодняшний день
52 öö	ночь	10 iga (subst.)	возраст
29 nädal	неделя	10 sajand	век

Г. Количество и мера

139 kord	'раз;порядок'	20 minut	'минута'
45 tund	час	14 pool	половина
34 osa	часть	13 jagu	часть
32 kilomeeter	километр	12 number	номер
29 tükk	кусок	12 paar	пара
25 meeter	метр	10 enamik	большинство
24 hulk	множество	10 hunnik	куча
23 koorem	воз	6 sekund	секунда

Д. Место и др.

74 koht	'место'	18 rida	'ряд'
59 ots	конец, кончик	12 auk	яма
30 nurk	угол	12 ring	круг
24 külg	бок, сторона	11 kaar	дуга
19 serv	край, грань	10 pind	почва

Таблица А-72
(гл. 17, § 3)

"Плюсовая" и "минусовая" полисемия. Имена существительные эстонского языка (цифры обозначают: число значений по данным толкового словаря и частоту по данным ЧС авторской речи художественной прозы).

А. Плюсовые слова (повышение количества значений по сравнению с нормой):

11/139	kord	'порядок;раз'	8/19	pesa	'гнездо'
10/74	koht	место;местность	8/26	põhi	дно
10/33	kiri	письмо	8/45	kivi	камень
9/10	kõrgus	высота	7/74	laps	ребенок
9/56	maailm	мир, свет	6/10	pind	почва
9/78	meel	чувство	6/11	kaar	дуга
9/122	elu	жизнь	6/29	paik	место

Б. Минусовые слова (понижение количества значений по сравнению с нормой):

1/151	aasta	'год'	1/21	soov	'желание'
1/27	kevad	весна	1/17	järv	озеро
1/33	paat	лодка	1/17	müts	шапка
1/32	kilomeeter	километр	1/14	king	ботинок
1/27	sakslane	немец	1/14	sugulane	родственник
1/25	meeter	метр	2/254	inimene	человек
1/24	auto	автомашина	2/185	naine	жен(щин)а
1/22	buss	автобус	2/89	hobune	лошадь
1/22	naer	смех	2/71	hetk	мгновение
1/21	kolhoos	колхоз	2/71	mets	лес
1/21	mulje	впечатление	3/165	päev	день
1/21	pudel	бутылка	3/122	poiss	мальчик

Таблица А-73

(гл. V, § 2)

Распределение слов в объединенном тексте при частоте
 $F = 5$ (А - Э. Бэкиман, В - Л. Промет)

A - B	A - B	A - B	A - B	A - B	A - B
5 - 0	4 - 1	3 - 2	2 - 3	1 - 4	0 - 5
iseenese	jälgima	alt	algus	huvi	aru
kahmama	käugele	jagama	juus	jätma	ehitama
kannul	lahtine	jutt	jõud	pöeg	kiri
kobama	liiga	kuhu	kord (d_1)	sulgema	kohta(pp)
lehm	lähedus	kõrvale	külm	trepp	lamama
muld	mägas	loopima	lëndama	voodi	linn
piits	too (p)	lõikama	lõug	üksnes	nutma
pitsitama	viskama	mingi	mõjuma		nõudma
rakendama	värav	otsekui	poolt (pp)		peegel
vagu	äär	peremees	rahvas		
vedama		raske	soe		
		riputama	uurima		
		siiski	valmis		
		taevas	väga		
		taga			
(Э) 11	10	15	14	7	9
(Т) 2	10	21	21	10	2
(Э-Т)+ 9	0	- 6	- 7	- 3	+ 7

Таблица А-74

(гл. У, § 2)

Частоты 40 полновзначных слов в 20 текстах (выборках из авторской речи современной эстонской художественной прозы): F - суммарная частота.

Автор	AB	VG	АН	HK	JK	PK	LP	VS	HS	RS	MT	EV	AK	TK	JP	JT	AV	MU	K/N	JS	F
Лексема	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
suur	7	15	18	8	11	6	11	5	24	5	7	12	16	10	28	12	16	20	21	18	270
vana	3	5	18	3	4	9	6	2	11	10	10	7	8	2	20	5	5	7	10	11	156
hea	2	7	3	17	4	10	6	6	5	9	5	13	11	16	15	8	21	12	5	6	181
ilus	1	1	1	1	2	-	5	2	1	1	5	3	5	8	4	5	10	5	3	1	64
see	2	2	1	1	1	-	2	-	-	4	2	-	1	4	5	5	-	7	-	2	39
rõõmus	1	1	-	2	3	2	-	-	-	-	3	-	2	2	-	2	4	-	1	2	25
imelik	1	2	-	-	3	5	3	4	-	1	3	7	-	4	-	7	-	1	-	-	41
hall	2	6	8	1	10	-	1	4	-	-	2	-	2	7	3	10	1	6	5	2	70
nägema	14	10	10	21	11	15	16	14	14	14	9	10	10	14	17	25	8	19	6	11	268
vaatama	21	7	2	26	18	5	22	11	13	13	16	17	9	16	16	35	19	19	3	8	296
kuulma	5	9	3	4	1	8	4	7	5	5	3	3	5	2	2	9	-	1	3	7	86
kuulama	4	-	1	1	2	2	2	1	1	2	4	6	-	1	1	4	4	4	-	10	50
rääkima	5	3	6	6	7	9	9	6	8	10	7	9	15	17	4	27	9	17	3	4	181
küsima	-	3	2	2	4	5	8	2	7	3	-	1	3	10	3	2	-	7	-	2	64
kõima	6	6	8	11	3	11	5	10	13	11	15	1	11	2	11	15	10	11	3	8	171
sõitma	2	5	2	6	4	7	4	6	6	4	8	1	6	6	6	3	-	7	7	6	96
teadma	5	15	18	11	12	14	17	19	20	12	14	11	9	7	21	13	22	17	10	12	279
mõtleva	6	7	6	10	4	13	17	6	4	2	15	8	5	17	3	10	21	11	-	8	173

(Продолжение следует)

Таблица А-74
(продолжение)

	AB	VG	AH	HK	JK	PK	LP	VS	HS	RS	MT	EV	AK	TK	JP	JT	AV	MU	K/N	JS	F
tundma	12	10	10	12	13	16	20	7	9	17	4	19	3	13	6	18	10	9	3	10	221
tahtma	12	7	14	9	3	14	17	15	9	10	5	11	12	5	8	12	10	15	2	5	195
kartma	1	1	-	1	1	5	2	2	1	1	3	2	4	1	3	4	2	6	-	1	41
lootma	3	2	-	3	-	-	2	3	1	1	2	2	3	-	-	1	3	-	-	-	26
naerma	4	1	4	4	3	3	3	2	1	1	-	2	2	2	1	3	2	2	1	1	42
meeldima	-	3	-	-	-	3	1	1	-	-	2	6	2	6	2	2	4	5	3	-	40
kevad	-	3	5	1	-	-	3	6	1	1	1	-	2	-	1	-	-	3	-	-	27
õhtu	2	6	3	4	1	4	5	5	13	4	4	2	1	7	3	8	7	12	-	3	94
mõte	4	3	5	11	5	5	4	7	3	4	6	4	5	6	7	3	10	6	2	5	105
tunne	4	3	-	1	1	1	3	-	-	7	1	5	1	5	-	3	13	3	1	2	54
elu	2	15	12	10	1	2	4	3	7	1	8	1	2	7	13	6	7	5	4	12	122
surn	1	-	6	-	-	-	2	2	4	1	1	2	-	-	-	3	2	-	2	3	29
rõõm	-	2	3	1	-	2	8	1	-	2	-	2	1	2	4	3	3	2	-	1	37
mure	3	-	8	2	-	2	1	2	-	-	1	-	3	1	7	-	-	-	-	1	33
töö	-	13	6	12	-	3	1	6	8	4	7	4	5	-	12	3	6	1	2	7	104
õna	3	7	5	7	1	12	3	6	11	5	13	6	2	2	4	7	5	3	1	15	122
mees	52	25	12	31	10	18	9	15	36	18	27	16	12	7	20	17	6	14	11	14	372
naine	39	17	13	5	1	8	8	6	14	10	1	9	2	1	4	7	13	14	6	7	185
süda	3	-	8	1	1	2	3	2	3	7	4	5	2	1	2	7	2	2	1	2	58
silma	15	18	7	18	24	7	18	13	13	17	4	8	10	12	6	26	10	19	8	5	258
päike	3	1	-	3	-	-	7	5	7	3	6	9	2	-	8	8	-	7	5	1	75
ilm	-	2	1	-	-	-	1	1	6	-	1	-	1	1	4	2	2	1	-	8	31

Корреляционная матрица 20 текстов на основе
40 однозначных слов

[illegible]

Таблица А-76

(гл. У, § 2)

Факторная матрица 20 текстов
на основе 40 полнозначных слов
(варимакс-решение)

Автор		Факторные нагрузки		Общность (h^2)
		F_1	F_2	
Э. Бэзкман	(AB)	0,40	0,51	0,43
В. Гросс	(VG)	0,36	0,77*	0,71
А. Хинт	(AH)	0,17	0,75*	0,58
Х. Кийк	(HK)	0,68	0,53	0,74
Я. Кросс	(JK)	0,76*	0,29	0,66
П. Куусберг	(PK)	0,56	0,55	0,61
Л. Промет	(LP)	0,85*	0,26	0,79
В. Саар	(VS)	0,62	0,53	0,67
Х. Серго	(HS)	0,37	0,86*	0,87
Р. Сирге	(RS)	0,71*	0,49	0,75
М. Траат	(MT)	0,39	0,62	0,54
Э. Ветемаа	(EV)	0,70*	0,42	0,66
А. Каал	(AK)	0,56	0,54	0,61
Т. Каллас	(TK)	0,80*	0,08	0,65
Ю. Пээгель	(JP)	0,29	0,79*	0,71
Ю. Туулик	(JT)	0,87*	0,23	0,81
А. Валтон	(AV)	0,59	0,33	0,46
М. Унт	(MU)	0,77*	0,46	0,79
Э. Нийт,				
Я. Кросс	(K/N)	0,22	0,75*	0,61
Ю. Смуул	(JS)	0,22	0,79*	0,67
Вклад фактора		6,89	6,45	-
%		34,5	32,3	66,8

Таблица А-77

(гл. У, § 2)

Лексико-стилистические признаки по данным 10 текстов современной эстонской художественной прозы (авторская речь)

№ пп.	Лексико-стилистический признак	Э. Безман	В. Гросс	А. Хант	Х. Кийк	Я. Кросс	Л. Пронет	В. Саар	Х. Серго	М. Траит	А. Балтон	Среднее
1.	Частота существительных (%)	35,9	30,4	32,5	34,8	32,8	30,4	26,0	36,8	30,7	26,8	31,7
2.	" прилагательных (%)	5,4	7,4	5,6	5,2	7,4	5,5	4,9	5,8	6,2	6,3	6,0
3.	" глаголов (%)	25,0	22,2	19,1	24,2	18,3	24,0	23,5	22,0	22,6	24,0	22,5
4.	" союзов (%)	7,2	6,8	8,3	7,7	10,1	8,5	9,5	7,6	8,1	9,0	8,3
5.	R (сложность текста)	2,40	2,57	2,92	2,13	3,00	2,09	2,46	2,41	2,62	2,33	2,49
6.	C (концентрация - %)	10,0	8,1	8,2	10,0	7,6	7,5	9,1	8,9	7,7	9,5	8,7
7.	$V_1/N \cdot 100$ (%)	45,4	43,8	36,2	43,2	46,8	42,1	36,9	46,6	41,7	36,1	41,9
8.	$F_{\max}/N \cdot 100$ (%)	2,74	2,57	2,82	3,48	5,67	3,41	3,44	3,13	3,33	3,55	3,41
9.	St (стереотипность)	3,55	3,74	3,85	3,88	4,06	3,95	4,20	3,64	4,14	4,53	3,95
10.	$a \cdot 10^{-2}$ (потенц. богатство)	I23	III	I22	83	I34	I34	III	I37	III	I65	I23

Таблица А-78
(гл. V, § 2)

Средняя длина слова ($\bar{l}$) и предложения ($\bar{j}$) и
измерение сложности текста по формуле $R = \bar{l} \lg \bar{j}$.
Данные 10 текстов современной эстонской художе-
ственной прозы (авторская речь).

Текст (автор)	Объем выборки	Число предлож.	$\bar{l}$	$\bar{j}$	R
Э. Бэзкман	4972	476	2,36	10,4	2,40
В. Гросс	5018	450	2,45	11,2	2,57
А. Хинт	4997	266	2,29	18,8	2,92
Х. Кийк	5015	550	2,22	9,1	2,13
Я. Кросс	5026	263	2,34	19,1	3,00
Л. Промет	5009	629	2,31	8,0	2,09
В. Саар	4971	357	2,15	13,9	2,46
Х. Серго	4963	455	2,32	10,9	2,41
М. Траат	5082	358	2,27	14,2	2,62
А. Валтон	5001	421	2,17	11,9	2,35
В среднем	-	-	2,30	12,8	2,49

Таблица А-79

(гл. V, § 2)

Корреляционная матрица ($r \cdot 100$) на основе сравнения лексико-стилистических признаков 10 текстов современной эстонской художественной прозы (авторская речь)

2	3	4	5	6	7	8	9	10	П р и з н а к и
-05	-12	-49	03	19	73	-06	-81	-24	1. Частота существительных
	-53	06	53	-51	35	39	09	27	2. " прилагательных
		-39	-89	60	-08	-49	02	-11	3. " глаголов
			33	-27	-31	79	70	36	4. " союзов
				-47	00	37	01	09	5. Сложность текста (R)
					-09	-31	-10	-14	6. Концентрация (C)
						26	-63	-16	7. V_1/N
							40	21	8. F_{max}/N
								38	9. Стереотипность (St)
								-	10. Потенц. богатство (a)

(Критическое

значение $r_{0,10;8} = 0,54$)

Факторная матрица (варимакс-решение) лексико-стилистических признаков по данным
10 текстов современной эстонской художественной прозы (авторская речь)

Лексико-стилистический признак	Факторные нагрузки				Общность (h^2)
	F_1	F_2	F_3	F_4	
1. Частота существительных	0,91*	-0,06	-0,09	-0,14	0,86
2. " прилагательных	0,05	-0,41	0,09	0,78*	0,79
3. " глаголов	-0,08	0,95*	-0,23	-0,15	0,98
4. " союзов	-0,46	-0,26	0,84*	-0,07	0,99
5. Сложность текста (R)	-0,01	-0,89	0,12	0,15	0,83
6. Концентрация (C)	0,10	0,50	-0,11	-0,40	0,43
7. Разнообразие (V_1/N)	0,88*	0,07	0,14	0,40	0,96
8. $F_{\max}/N$	0,04	-0,25	0,91*	0,25	0,95
9. Стереотипность (St)	-0,83	0,09	0,44	0,13	0,91
10. Потенц. богатство лексики (a)	-0,28	-0,05	0,22	0,24	0,19
Вклад фактора	2,60	2,26	1,89	1,14	-
%	26,0	22,6	18,9	11,4	78,9

Таблица А-81

(гл. V, § 2)

Индивидуальные значения (оценки) факторов для
10 текстов

Т е к с т (автор)	Значения факторов				Значимость ($f \geq 0,7 $)			
	f_1	f_2	f_3	f_4	f_1	f_2	f_3	f_4
Э. Бээкман	1,2	1,1	-0,4	-0,7	+	+	0	-
В. Гросс	-0,1	-0,2	-1,7	1,9	0	0	-	+
А. Хинт	-0,5	-2,1	-1,0	-1,6	0	-	-	-
Х. Кийк	0,7	0,8	0,1	-0,6	+	+	0	0
Я. Кросс	0,7	-1,3	2,2	1,1	+	-	+	+
Л. Промет	-0,1	0,8	0,4	-0,1	0	+	0	0
В. Саар	-1,2	0,3	0,7	-1,2	-	0	+	-
Х. Серго	1,3	-0,2	-0,2	-0,3	+	0	0	0
М. Траат	-0,4	0,1	-0,3	0,8	0	0	0	+
А. Валтон	-1,8	0,7	0,2	0,7	-	+	0	+

Таблица Б-І
(к гл. II, § I)

Распределение слов древнего происхождения по частотным зонам эстонского словаря: $F(i)$ - абсолютная частота и $p(i)$ - относительная частота древних слов в данной частотной зоне (i); $F^*(i)$ и $p^*(i)$ - соответствующие накопленные частоты; n - объем зоны ($n = 100$), n^* - накопленный объем зон

i	$F(i)$	$F^*(i)$	$p(i) = \frac{F(i)}{n}$	$p^*(i) = \frac{F^*(i)}{n^*}$
I	91	91	0,91	0,91
2	77	168	0,77	0,84
3	70	238	0,70	0,79
4	66	304	0,66	0,76
5	60	364	0,60	0,73
6	54	418	0,54	0,69
7	46	464	0,46	0,66
8	44	508	0,44	0,64
9	43	551	0,43	0,61
10	37	588	0,37	0,59
11	30	618	0,30	0,56
12	34	652	0,34	0,54
Всего	652	-	-	-

Таблица Б-2
(к гл. II, § I)

Древняя лексика в эстонском тексте: распределение слов соответственно частотным зонам (i) словаря.

$N(i)$ - покрываемость текста всеми словами данной частотной зоны, $F_t(i)$ - покрываемость текста древними словами, $N^*(i)$ и $F_t^*(i)$ - соответствующие накопленные частоты

i	$N(i)$	$N^*(i)$	$F_t(i)$	$F_t^*(i)$	$F^*(i)/N^*(i)$
I	41358	41358	38877	38877	0,940
2	8911	50269	7417	46294	0,921
3	5429	55698	3820	50114	0,900
4	3719	59417	2458	52572	0,885
5	2839	62256	1703	54275	0,872
6	2291	64547	1237	55512	0,860
7	1914	66461	883	56395	0,849
8	1644	68105	726	57121	0,839
9	1408	69513	604	57725	0,830
10	1249	70762	461	58186	0,822
11	1091	71853	331	58517	0,814
12	1000	72853	340	58857	0,808
Всего	72853	-	58857	-	-

Таблица Б-3
(к гл. II, § I)

Эмпирическое и теоретическое распределение частот древних слов в словах трех языков; параметр α и коэффициент убывания $e^{-\alpha}$ по экспоненциальному закону

$$F(i) = 100 e^{-\alpha i}$$

Ранг i (зона)	Эстонский		Французский		Немецкий	
	Эмп.	Теор.	Эмп.	Теор.	Эмп.	Теор.
I (1-100)	91	91	91	94	94	93
2 (101-200)	77	82	84	89	88	87
3 (201-300)	70	74	84	84	83	81
4 (301-400)	66	67	71	79	73	76
5 (401-500)	60	61	73	74	63	70
6 (501-600)	54	55	52	70	55	66
7 (601-700)	46	50	57	66	55	61
8 (701-800)	44	45	55	62	59	57
9 (801-900)	43	41	52	58	52	53
10 (901-1000)	37	37	61	55	53	50
Параметры:						
α	0,1		0,06		0,07	
$e^{-\alpha}$	0,905		0,942		0,932	

Таблица Б-4
(к гл. II, § I)

Интегральное распределение частот древних слов в шести языках:
эмпирические и теоретические данные; параметры распределения Вейбулла

Ранг (зона)	Эстонский		Французский		Немецкий		Английский		Русский		Чешский	
	Эмп.	Теор.	Эмп.	Теор.	Эмп.	Теор.	Эмп.	Теор.	Эмп.	Теор.	Эмп.	Теор.
1 (1-100)	91	91	91	93	94	96	92	94	84	84	75	76
2 (101-200)	168	170	175	177	182	180	162	158	141	142	138	133
3 (201-300)	238	241	259	255	265	257	215	212	193	193	188	184
4 (301-400)	304	305	330	327	338	329	255	258	244	242	231	230
5 (401-500)	364	364	403	395	401	395	302	300	286	282	268	273
6 (501-600)	418	417	455	459	456	458	334	337	318	321	304	313
7 (601-700)	464	466	512	519	511	516	363	371	360	359	349	351
8 (701-800)	508	510	567	576	570	570	399	408	395	394	391	387
9 (801-900)	551	551	619	630	622	622	430	431	428	428	423	421
10 (901-1000)	588	588	680	680	675	670	461	458	463	460	455	453
Параметры:												
F_n	1000		1600		1500		900		1900		1500	
α	0,095		0,060		0,066		0,110		0,045		0,052	
β	0,971		0,965		0,953		0,807		0,792		0,836	
i_e	11,3		18,5		17,3		15,3		50,4		34,5	

Таблица Б-5
(к гл. II, § I)

Эмпирическое и теоретическое интегральное распределение
древних слов в эстонском тексте по функции

$$F_t^*(i) = 61000 (1 - e^{-\sqrt{i}}).$$

Зона ЧС (i)	Эмпир. $F_t^*(i)$	Теоретич. $F_t^*(i)'$
I (I-100)	38877	38559
2 (I-200)	46294	46170
3 (I-300)	50114	50208
4 (I-400)	52572	52745
5 (I-500)	54275	54480
6 (I-600)	55512	55733
7 (I-700)	56395	56672
8 (I-800)	57121	57395
9 (I-900)	57725	57963
10 (I-1000)	58186	58418
II (I-II00)	58517	58787
12 (I-I200)	58857	59091

Таблица Б-6

(к гл. II, § 2)

Эмпирические и теоретические данные роста
лексики эстонского литературного языка

(L'_T - экспоненциальное, L''_T - логистическое распределение)

№ пп.	Год	T	L_T (эмпир.)	L'_T (теорет.)	L''_T (теорет.)
1.	1660	0,6	10 000	2 000	2 000
2.	1780	1,8	14 000	14 000	13 000
3.	1818	2,2	21 000	24 000	24 000
4.	1869	2,7	50 000	50 000	50 000
5.	1893	2,9	60 000	67 000	60 000
6.	1930	3,3	120 000	120 000	86 000
7.	1960	3,6	105 000	185 000	105 000
8.	1976	3,8	115 000	247 000	115 000
-	2000	4,0	прогноз:	330 000	124 000

Таблица Б-7
(гл. III, § I)

Основные статистические характеристики ЧС разных языков (Алексеев П.М., 1968; Бектаев К.Б., 1978; Григорьева А.С., 1981; Част. слов. совр. укр. худ. прозы (пробн. тетрадь), 1969; Kučera H., Francis W.N., 1967; Latviešu val. biež. vārdn., 1972)

Язык	N	V	L	V/N	N/V	L/N	N/L	L/V	V/L
Эстонский (худ. проза)	99 898	30 733	14 654	0,31	3,3	0,15	6,8	0,48	2,1
Украинский (худ. проза)	100 000	27 570	13 954	0,28	3,6	0,14	7,2	0,51	2,0
Латышский (худ. проза)	100 000	-	11 439	-	-	0,11	8,7	-	-
Казахский (детск. лит.)	98 040	23 350	10 076	0,24	4,2	0,10	9,7	0,43	2,3
Русский (эпист. речь)	96 800	15 842	8 064	0,16	6,1	0,08	12,0	0,57	1,7
Немецкий (научно-техн.)	100 000	14 434	-	0,14	6,9	-	-	-	-
Английский (общелит. тексты)	101 566	13 706	-	0,13	7,4	-	-	-	-
Английский (электроника)	100 000	7 853	5 197	0,079	12,7	0,052	19,2	0,66	1,5
Французский (электроника)	100 000	8 108	4 572	0,081	12,3	0,046	21,9	0,56	1,8

Таблица Б-8
(гл. III, § I)

Статистическая зависимость между количеством лексем (L) и количеством словоформ (V), а также между мерой анализа (L/V) и объемом текста (N) по данным I-го тома романа А.Х. Таммсааре "Правда и справедливость" (эст.). Теоретические значения вычислены по формулам:

$$L \text{ (теор.)} = 2,75 V^{0,81};$$

$$L/V \text{ (теор.)} = 1,74 N^{-0,12}.$$

N	V	L	L	L/V	L/V
		(эмпир.)	(теор.)	(эмпир.)	(теор.)
10 000	3 636	2 114	2 116	0,58	0,58
20 000	5 944	3 124	3 136	0,53	0,53
30 000	7 503	3 781	3 787	0,50	0,51
114 124	16 750	7 348	7 257	0,44	0,43
(весь роман)					

Таблица Б-9
(гл. III, § I)

Стабилизация средней частоты ($\bar{F}$) лекси́мы шлеме 'идти' по мере увеличения числа подвыборок (i) по 5000 словоупотреблений (F - абсолютная частота, F* - накопленная абсол. частота)

i	F	F*	$\bar{F}$
1	21	21	21,0
2	20	41	20,5
3	28	69	23,0
4	26	95	23,8
5	15	110	22,0
6	24	134	22,3
7	25	159	22,7
8	39	198	24,8
9	34	232	25,8
10	24	256	25,6
11	27	283	25,7
12	24	307	25,6
13	27	334	25,7
14	12	346	24,7
15	22	368	24,5
16	43	411	25,7
17	20	431	25,4
18	41	472	26,2
19	12	484	25,5
20	15	499	25,0

Проверка на нормальность распределения слова (лексемы) меев 'муж(чина)' с помощью "чисел Вестергарда" (Пиотровский Р.Г. и др., 1977, с. 344).

Вариационный ряд абсолютных частот F_i
(20 подвыборок по 5000 словоупотреблений):

6-9-9-10-11-12-12-14-14-16-16-17-18-18-20-27-25-31-36-52.

Гипотеза о нормальности распределения не отвергается, если:

в пределах $\bar{F} \pm 0,36$ находится не менее 25 % частот;

" $\bar{F} \pm 0,76$ " " 50 % " ;

" $\bar{F} \pm 1,16$ " " 75 % " ;

" $\bar{F} \pm 3,06$ " " 99,8 % " .

В данном случае средняя частота $\bar{F} = 18,7$ (при суммарной частоте $F = 373$), стандартное отклонение $\sigma = 11$. Распределение частот:

$\bar{F} \pm 0,36$ (15...22) - 6 подвыборок - 30 %;

$\bar{F} \pm 0,76$ (11...26) - 12 " - 60 %;

$\bar{F} \pm 1,16$ (6,5...31) - 17 " - 85 %;

$\bar{F} \pm 3,06$ (0 ...52) - 20 " - 100 %.

Эта группировка показывает, что распределение частот слова меев подчиняется нормальному закону (гипотеза о нормальности распределения не отвергается).

Таблица Б-II
(гл. III, § 2)

Частотная структура текста по данным 4С лексем авторской речи эстонской художественной прозы (i - ранг, F_i - частота, $m(F_i)$ - количество слов с данной частотой). Объем текста $N = 99\ 898$; объем словаря $L = 14\ 654$.

i	F_i	$m(F_i)$	i	F_i	$m(F_i)$
1	4237	I	31-33	304	3
2	3493	I	34	297	I
3	2598	I	35-37	285	3
4	1981	I	38-39	272	2
5	1395	I	40	268	I
6	1300	I	41	264	I
7	1047	I	42	261	I
8	879	I	43	260	I
9	845	I	44	254	I
10	827	I	45	248	I
11	724	I	46	238	I
12	634	I	47	237	I
13	613	I	48	234	I
14	581	I	49	230	I
15	568	I	50	223	I
16	499	I	51	218	I
17	496	I	52	209	I
18	493	I	53	207	I
19	465	I	54	206	I
20	448	I	55	200	I
21	436	I	56	198	I
22	434	I	57	195	I
23	428	I	58	194	I
24	382	I	59	192	I
25	373	I	60	190	I
26	366	I	61	189	I
27	350	I	62-63	185	2
28	339	I	64	181	I
29	327	I	65	180	I
30	309	I	66	179	I

Таблица Б-II
(продолжение)

i	F_i	$m(F_i)$
67	I76	I
68	I75	I
69-70	I74	2
71	I71	I
72	I67	I
73	I66	I
74	I65	I
75-76	I64	2
77-78	I63	2
79	I59	I
80	I53	I
81	I51	I
82	I50	I
83	I49	I
84	I48	I
85	I46	I
86	I44	I
87	I39	I
88	I38	I
89	I36	I
90	I35	I
91-92	I33	2
93-94	I31	2
95-96	I30	2
97	I26	I
98-99	I25	2
I00	I24	I
I01	I23	I
I02-I03	I22	2
I04	I20	I
I05	II6	I
I06	II3	I
I07-I08	II2	2
I09-II0	II0	2
III	I09	I

i	F_i	$m(F_i)$
II2	I08	I
II3-II5	I07	3
II6	I06	I
II7-II9	I05	3
I20	I04	I
I21-I24	I01	4
I25-I26	I00	2
I27	99	I
I28	98	I
I29-I31	97	3
I32	96	I
I33-I35	94	3
I36-I39	93	4
I40-I41	92	2
I42-I44	91	3
I45-I47	90	3
I48-I49	89	2
I50	88	I
I51	87	I
I52-I54	86	3
I55-I57	85	3
I58-I59	84	2
I60	83	I
I61-I62	82	2
I63-I64	81	2
I65-I69	79	5
I70-I72	78	3
I73-I79	76	7
I80-I83	75	4
I84-I87	74	4
I88-I91	73	4
I92-I94	72	3
I95-202	71	8
203-204	70	2
205-206	69	2

Таблица Б-II

(продолжение)

i	P_i	$m(P_i)$
207	68	1
208-209	67	2
210-211	66	2
212-217	64	6
218-220	63	3
221	62	1
222-223	61	2
224-229	60	6
230-235	59	6
236-238	58	3
239-241	57	3
242-244	56	3
245-247	55	3
248-252	54	5
253-256	53	4
257-263	52	7
264-268	51	5
269-270	50	2
271-274	49	4
275-280	48	6
281-285	47	5
286-289	46	4
290-299	45	10
300-305	44	6
306-312	43	7
313-318	42	6
319-329	41	11
330-339	40	10
340-346	39	7
347-349	38	3
350-360	37	11
361-369	36	9
370-382	35	13
383-389	34	7
390-404	33	15

i	P_i	$m(P_i)$
405-412	32	8
413-428	31	16
429-445	30	17
446-466	29	21
467-479	28	13
480-509	27	30
510-529	26	20
530-543	25	14
544-564	24	21
565-586	23	22
587-612	22	26
613-650	21	38
651-676	20	26
677-707	19	31
708-747	18	40
748-793	17	46
794-856	16	63
857-893	15	37
894-940	14	47
941-1017	13	77
1018-1080	12	63
1081-1169	11	89
1170-1293	10	124
1294-1414	9	121
1415-1567	8	153
1568-1779	7	212
1780-2044	6	265
2045-2389	5	345
2390-2980	4	591
2981-3918	3	938
3919-5972	2	2054
5973-14654	1	8682

Таблица В-12
(гл. III, § 2)

Частотно-ранговое распределение лексики по данным ЧС
л е к с е м авторской речи эстонской художественной про-
зы (i - ранг, F_i - частота, F_i^* - накопленная частота,
 P_i^* (%) - покрываемость текста)

i	F_i	F_i^*	$P_i^*(\%)$	i	F_i	F_i^*	$P_i^*(\%)$
1	4237	4237	4,24	400	33	59583	59,64
2	3493	7730	7,74	500	27	62517	62,58
3	2598	10328	10,34	600	22	64948	65,01
4	1981	12309	12,32	700	19	66986	67,05
5	1395	13704	13,72	800	16	68733	68,80
6	1300	15004	15,02	900	14	70282	70,35
7	1047	16051	16,07	1000	13	71622	71,70
8	879	16930	16,95	1001 - 1017	13	71843	71,92
9	845	17775	17,79	1018 - 1080	12	72599	72,67
10	827	18602	18,62	1081 - 1169	11	73578	73,65
20	448	24123	24,15	1170 - 1293	10	74818	74,89
30	309	27867	27,90	1294 - 1414	9	75907	75,98
40	268	30743	30,77	1415 - 1567	8	77131	77,21
50	223	33192	33,23	1568 - 1779	7	78615	78,70
60	190	35201	35,24	1780 - 2044	6	80205	80,29
70	174	36999	37,04	2045 - 2389	5	81930	82,01
80	153	38634	38,67	2390 - 2980	4	84294	84,38
90	135	40070	40,11	2981 - 3918	3	87108	87,20
100	124	41358	41,40	3919 - 5972	2	91216	91,31
200	71	50284	50,34	5973 - 14654	1	99898	100,00
300	44	55785	55,84				

Таблица Б-13
(гл. III, § 2)

Частотно-ранговое распределение лексики по данным ЧС
с л о в о ф о р м авторской речи эстонской художествен-
ной прозы (i - ранг, F - частота, F^* - накопленная
частота, P^* (%) - покрываемость текста)

i	F_i	F_i^*	$P_i^*(\%)$	i	F_i	F_i^*	$P_i^*(\%)$
1	3221	3221	3,22	200	48	35066	35,10
2	1602	4823	4,83	300	34	38592	38,63
3	1439	6262	6,27	400	25	41484	41,53
4	1375	7637	7,64	500	21	43755	43,80
5	1264	8901	8,91	600	17	45634	45,68
6	1116	10017	10,03	700	15	47264	47,31
7	995	11012	11,02	800	13	48703	48,75
8	713	11725	11,74	900	12	49966	50,02
9	592	12317	12,33	1000	11	51109	51,16
10	542	12859	12,87	1001- 1034	11	51483	51,54
20	329	16801	16,82	1035- 1168	10	52823	52,88
30	224	19344	19,36	1169- 1299	9	54002	54,06
40	189	21384	21,41	1300- 1500	8	55610	55,67
50	165	23130	23,15	1501- 1753	7	57381	57,44
60	141	24635	24,66	1754- 2131	6	59649	59,71
70	120	25920	25,95	2132- 2650	5	62244	62,31
80	108	27045	27,07	2651- 3460	4	65484	65,55
90	89	28017	28,05	3461- 5088	3	70368	70,44
100	83	28871	28,90	5089- 8973	2	78138	78,22
				8974- 30733	1	99898	100,00

Таблица В-14
(гл. III, § 2)

Частотно-ранговое распределение лексики по данным под-
выборки - 40 л е к с е м авторской речи Э. Бээкман
(N = 4998; L = 1953)

i	F_i	F_i^*	$P_i^*(\%)$	i	F_i	F_i^*	$P_i^*(\%)$
1	148	148	2,96	29	20	1217	24,35
2	137	285	5,70	30	19	1236	24,73
3	113	398	7,96	31-32	18	1272	25,45
4	86	484	9,68	33	17	1289	25,79
5	53	537	10,74	34	15	1304	26,09
6	52	589	11,78	35-37	14	1346	26,93
7	45	634	12,69	38-45	13	1450	29,01
8	41	675	13,51	46-50	12	1510	30,21
9	40	715	14,31	51-58	11	1598	31,97
10	39	754	15,09	59-70	10	1718	34,37
11	37	791	15,83	71-86	9	1862	37,25
12	33	824	16,49	87-96	8	1942	38,86
13	31	855	17,11	97-124	7	2138	42,78
14	27	882	17,65	125-158	6	2342	46,86
15	26	908	18,17	159-195	5	2527	50,56
16	25	933	18,67	196-258	4	2779	55,60
17	24	957	19,15	259-371	3	3118	62,38
18	23	980	19,61	372-669	2	3714	74,31
19-25	22	1134	22,69	670-1953	1	4998	100,00
26-28	21	1197	23,95				

Таблица Б-15
(гл. III, § 2)

Частотно-ранговое распределение лексики по данным под-
выборки - ЧС с л о в о ф о р м авторской речи Э. Бээкман
(N = 4998, V = 2869)

i	F_i	F_i^*	$P_i^*(\%)$	i	F_i	F_i^*	$P_i^*(\%)$
1	137	137	2,74	17	17	694	13,89
2	98	235	4,70	18-19	16	726	14,53
3	64	299	5,98	20-23	15	771	15,43
4	47	346	6,92	23	14	785	15,71
5	43	389	7,78	24-28	13	850	17,01
6	40	429	8,58	29-30	12	874	17,49
7	38	467	9,34	31-36	11	940	18,81
8	35	502	10,04	37-42	10	1000	20,01
9	26	528	10,56	42-52	9	1099	21,99
10	24	552	11,04	53-58	8	1147	22,95
11	23	575	11,50	59-78	7	1280	25,61
12	22	597	11,94	79-101	6	1418	28,37
13-14	21	639	12,79	102-129	5	1558	31,17
15	20	659	13,19	130-185	4	1782	35,65
16	18	677	13,55	186-303	3	2136	42,74
				304-599	2	2728	54,58
				600-2869	1	4998	100,00

Таблица Б-16
(гл. III, § 2)

Значения параметров распределения Ципфа C , k и γ по данным ЧС эстонского языка

Тип словаря и параметры	Сводный словарь (авт. речь худ. пр.)	Индивид. словарь (Э. Бээкман)
<u>ЧС лесем:</u>		
Объем словаря (L)	14654	1953
Параметры: C	6823	196
k	0,068	0,039
γ	0,92	0,72
<u>ЧС словоформ:</u>		
Объем словаря (V)	30733	2869
Параметры: C	4122	129
k	0,041	0,026
γ	0,86	0,68
Объем текста (N)	99898	4998

Значения параметров распределения Ципфа-Мандельброта
к, γ, В по данным ЧС разных языков (наш материал; Бектаев К.Б., 1978;
Калинина Е.А., 1968; Sigurd B., 1970)

Язык и подъязык	Тип единиц	Объем текста	Объем словаря	П а р а м е т р ы		
				к	γ	В
Эстонский (худ. проза)	Лексемы	99898	14654	0,068	0,92	-
				0,148	1,04	2,3
	Словоформы	99898	30733	0,041	0,86	-
Казахский (худ. проза)				0,046	0,87	0,5
	Словоформы	105494	22642	0,059	0,87	-
				0,079	0,92	4,96
Русский (электроника)	Словоформы	100000	14062	0,052	0,84	1,9
Шведский (смеш.)	Словоформы	100000	16217	0,058	1,31	5,72

Таблица Б-18
(гл. III, § 2)

Значения параметра γ в разных частотных зонах по данным ЧС эстонского и русского языков (I - Засорина Л.Н., 1966; II - ЧС русского языка, 1977)

Частотная зона (i)	Эстонский язык		Русский язык (лексемы)	
	лексемы	словоформы	I	II
I ÷ 30	0,83	0,77	0,70	0,71
30 ÷ 2500	0,98	0,89	0,94	0,95
> 2500	1,03	0,75	1,24	1,51
Весь словарь	0,92	0,86	0,93	1,0
Объем словаря	14654	30733	10830	39268
Объем текста	99898	99898	120474	1056382

Таблица Б-19
(гл. III, § 2)

Значения параметра γ в разных частотных зонах по данным ЧС лексем авторской речи и речи персонажей I-го тома романа "Правда и справедливость" ("Tõde ja õigus") А.Х.Таммсааре

Частотная зона (i)	Авторская речь	Речь персонажей
I ÷ 30	0,68	0,59
30 ÷ 1500	1,11	1,21
> 1500	1,43	1,47
Весь словарь	1,0	1,01
Объем словаря	7348	3135
Объем текста	114124	46232

Таблица Б-20
(гл. III, § 2)

Частотно-ранговое распределение лексики по данным 40 лексем авторской речи I-го тома романа "Правда и справедливость" ("Tõde ja õigus") А.Х. Таммсааре

i	F_i	F_i^*	$P_i^*(\%)$	i	F_i	F_i^*	$P_i^*(\%)$
1	4755	4755	4,17	100	172	59207	51,88
2	4142	8897	7,80	200	94	72079	63,16
3	3594	12491	10,95	300	59	79418	69,59
4	2537	15028	13,17	400	42	84349	73,91
5	1909	16937	14,84	500	34	88126	77,22
6	1708	18645	16,34	600	26	91095	79,82
7	1672	20317	17,80	700	21	93454	81,89
8	1665	21982	19,26	800	17	95371	83,57
9	1602	23584	21,01	900	14	96955	84,96
10	1525	25109	22,00	1000	12	98296	86,13
20	703	33440	29,30	1001-1055	12	98956	86,71
30	510	39217	34,36	1056-1107	11	99528	87,21
40	405	43777	38,36	1108-1175	10	100208	87,81
50	334	47328	41,47	1176-1257	9	100946	88,45
60	285	50455	44,21	1258-1388	8	101994	89,37
70	261	53155	46,58	1389-1534	7	103016	90,27
80	207	55427	48,57	1535-1721	6	104138	91,25
90	192	57422	50,32	1722-1968	5	105373	92,33
				1969-2333	4	106833	93,61
				2334-2914	3	108576	95,14
				2915-4028	2	110804	97,09
				4029-7348	1	114124	100,00

Таблица Б-21
(гл. III, § 2)

Частотно-ранговое распределение лексики по данным ЧС
лексем р е ч и п е р с о н а ж е й I-го тома романа
"Правда и справедливость" ("Tõde ja õigus") А.Х.Таммсааре

1	F_i	F_i^*	$P_i^*(\%)$	i	F_i	F_i^*	$P_i^*(\%)$
1	2413	2413	5,22	100	73	28791	62,28
2	1203	3616	7,82	200	36	33876	73,27
3	1178	4794	10,37	300	21	36687	79,35
4	1142	5936	12,84	400	15	38459	83,19
5	1068	7004	15,15	500	11	39729	85,93
6	970	7974	17,25	501-522	11	39971	86,46
7	932	8906	19,26	523-553	10	40281	87,13
8	810	9716	21,02	554-594	9	40650	87,93
9	784	10500	22,71	595-644	8	41050	88,79
10	742	11242	24,32	645-712	7	41526	89,82
20	416	16493	35,67	713-801	6	42060	90,98
30	234	19487	42,16	802-913	5	42620	92,19
40	202	21652	46,83	914-1067	4	43236	93,52
50	158	23404	50,62	1068-1296	3	43923	95,01
60	129	24841	53,73	1297-1766	2	44863	97,04
70	113	26068	56,39	1767-3135	1	46232	100,00
80	96	27148	58,72				
90	82	28020	60,61				

Таблица Б-22
(гл. III, § 2)

Закон Ципфа. Наблюдаемые и ожидаемые частоты (F_i) по данным 40 лексем авторской речи эстонской художественной прозы. Вычисления по формулам:

$$\begin{array}{lll}
 \text{I:} & F_i = C i^{-\gamma} & (\text{Ципф}) \\
 \text{II:} & F_i = C (i+B)^{-\gamma} & (\text{Мандельброт}) \\
 \text{III:} & F_i = C i^{-(\gamma + \varphi \ln i)} & (\text{Алексеев}) \\
 \text{IV:} & F_i = C (i+B_1)^{-1} - B_2 & (\text{Крылов})
 \end{array}$$

Ранг i	F_i (набл.)	F_i (ожд.)			
		I	II	III	IV
I	4237	6823	4267	4000	3583
2	3493	3606	3240	2616	2749
3	2598	2483	2607	2005	2231
4	1981	1906	2178	1651	1876
5	1395	1552	1868	1413	1619
10	827	820	1086	854	961
50	223	187	241	233	226
100	124	99	120	126	115
500	27	23	23	27	23
1000	13	12	11	13	11
1500	8	8	7	8	7
3000	4	4	4	4	4
5000	2	3	2	2	2
10000	1	1	1	1	1

Параметры:

C	6823	14770	4000	11824
γ	0,92	1,04	0,59	(1,0)
$B_{(1)}$	-	2,3	-	2,3
B_2	-	-	-	0,4
φ	-	-	0,035	-

($N = 99898$; $L = 14654$; $F_{max} = 4237$)

Таблица Б-23
(гл. III, § 2)

Модель Орлова. Наблюдаемые и ожидаемые частоты (F_i) по данным ЧС лексем авторской речи эстонской художественной прозы. Вычисления по формуле:

$$F_i = C(i+B)^{-\gamma}$$

- I: $Z = N$; $\gamma = 1$
 II: $Z \neq N$; $\gamma = 1$
 III: $Z \neq N$; $\gamma \neq 1$

Ранг i	F_i (набл.)	F_i (ожд.)		
		I	II	III
I	4237	4242	4226	6174
2	3493	3131	3068	3364
3	2598	2482	2409	2341
4	1981	2055	1982	1806
5	1395	1754	1684	1476
10	827	1012	961	785
50	223	231	217	180
100	124	118	110	95
500	27	24	22	22
1000	13	12	11	11
1500	8	8	8	8
3000	4	4	4	4
5000	2	2	2	3
10000	1	1	1	1
Параметры:				
C		11960	11200	6570
γ		(1,0)	(1,0)	0,92
B		1,82	1,65	0,07
Z		(99898)	175000	175000

($N = 99898$; $L = 14654$; $F_{\max} = 4237$)

Таблица В-24
(гл. III, § 2)

Дробно-линейная модель. Наблюдаемые и ожидаемые частоты (F_i) по данным ЧС лексем авторской речи эстонской художественной прозы (на основе линейной связи между $1/F_i$ и i ; ср. рис. 3.9). Вычисления по формуле:

$$F_i = (a + bi)^{-1}$$

(или $F_i = C(i+B)^{-1}$, где $C = b^{-1}$ и $B = aC$)

i	F_i (набл.)	$1/F_i$	F_i (ожд.)
1	4237	$0,024 \cdot 10^{-2}$	4444
2	3493	0,029	3333
3	2598	0,038	2667
4	1981	0,050	2222
5	1395	0,072	1904
10	827	0,12	1111
50	223	0,45	256
100	124	0,81	130
500	27	3,70	27
1000	13	7,69	13
1100	11	9,09	12
1200	10	10,00	11
1300	9	11,11	10
1400	9	11,11	9
1500	8	12,50	9
2000	6	16,67	7
3000	4	25,00	4
5000	2	33,33	3
10000	1	100,00	1
Параметры:			
a		0,00015	-
b		0,000075	-
C		-	13333
B		-	2

Таблица Б-25
(гл. III, § 2)

Модель Арапова-Ефимовой-Шрейдера ("дискретный аналог" закона Ципфа). Наблюдаемые и ожидаемые частоты (F_i) по данным ЧС авторской речи эстонской художественной прозы. Вычисления по формуле

$$F_i = \frac{\beta (L+1)^{\gamma}}{1-\gamma} \left[(i+1)^{1-\gamma} - i^{1-\gamma} \right]$$

Ранг i	ЧС лексем		ЧС словоформ	
	F_i (набл.)	F_i (ожд.)	F_i (набл.)	F_i (ожд.)
1	4237	3182	3221	3479
2	3493	1945	1602	2197
3	2598	1418	1439	1636
4	1981	1123	1375	1316
5	1395	932	1264	1106
10	827	514	542	633
50	223	121	165	164
100	124	64	83	91
500	27	15	21	23
1000	13	8	11	13
1500	8	6	8	9
3000	4	3	4	5
5000	2	2	3	3
10000	1	1	1	2
20000	-	-	1	1
Параметры:				
L	14654		30733	
γ	0,92		0,86	
β	0,6562		0,6608	

($N = 99898$)

Таблица В-26
(гл. III, § 2)

Значения параметров функции Вейбулла c и k по данным
40 словоформ разных языков
(Пиотровский Р.Г. и др., 1977; Борисевич А.Д., 1972; Гонча-
ренко В.В., 1972)

Язык (и подязык)	N	c	k
Эстонский (худ. проза)	99 898	0,053	0,379
Русский (офиц. тексты)	100 000	0,054	0,446
Английский (смеш. тексты)	100 000	0,114	0,348
Английский (строит. материалы)	200 000	0,126	0,394
Немецкий (публицистика)	200 000	0,155	0,289
Французский (электроника)	200 000	0,143	0,357

Таблица Б-27
(гл. III, § 2)

Наблюдаемые и ожидаемые накопленные частоты (F_i^*) по данным ЧС лексем авторской речи эстонской художественной прозы. Вычисления по формулам:

$$\text{I: } F_i^* = \frac{C}{\gamma - 1} [1 - (i + 1)^{1-\gamma}] \quad (\gamma \neq 1)$$

$$\text{II: } F_i^* = \frac{C}{\gamma - 1} [(1 + B)^{1-\gamma} - (i + 1 + B)^{1-\gamma}] \quad (\gamma \neq 1)$$

$$\text{III: } F_i^* = N(1 - e^{-ci^k}) \quad (\text{Вейбулл})$$

Ранг i	F_i^* (набл.)	F_i^* (ожд.)		
		I	II	III
I	4237	4863	3708	5540
10	18602	18036	19090	14841
100	41358	38089	45300	39413
1000	71622	62937	71962	72087
5972	91216	85707	91255	94226
14654	99898	98436	100449	98604
Параметры:				
C		6823	14770	-
γ		0,92	1,04	-
B		-	2,3	-
c		-	-	0,057
k		-	-	0,45

($N = 99898$)

Таблица Б-28
(гл. III, § 2)

Частотно-спектральное распределение л е к с е м в
с л о в а р е по данным ЧС авторской речи эстонской худо-
жественной прозы (m_F - число слов с частотой F ; m_F^* - на-
копленное число слов; p_F и p_F^* - соотв. относительные ча-
стоты в %)

F	m_F	m_F^*	$p_F(\%)$	$p_F^*(\%)$
1	8682	8682	59,3	59,3
2	2054	10736	14,0	73,3
3	938	11674	6,4	79,7
4	591	12265	4,0	83,7
5	345	12610	2,4	86,1
6	265	12875	1,8	87,9
7	212	13087	1,4	89,3
8	153	13240	1,0	90,3
9	121	13361	0,8	91,1
10	124	13485	0,9	92,0
11 - 20	519	14004	3,5	95,5
21 - 30	222	14226	1,5	97,0
31 - 40	99	14325	0,7	97,7
41 - 50	61	14386	0,4	98,1
51 - 100	144	14530	1,0	99,1
101 - 200	70	14600	0,5	99,6
201 - 300	21	14621	0,2	99,8
301 - 400	10	14631	0,0	99,8
> 400	23	14654	0,2	100,0
Σ	14654	-	100,0	-

Таблица Б-29
(гл. III, § 2)

Частотно-спектральное распределение л е к с е м в
т е к с т е по данным ЧС авторской речи эстонской худо-
жественной прозы

F	m_F	m_F^*	$p_F(\%)$	$p_F^*(\%)$
1	8682	8682	8,7	8,7
2	4108	12790	4,1	12,8
3	2814	15604	2,8	15,6
4	2364	17968	2,4	18,0
5	1725	19693	1,7	19,7
6	1590	21283	1,6	21,3
7	1484	22767	1,5	22,8
8	1224	23991	1,2	24,0
9	1089	25080	1,1	25,1
10	1240	26320	1,2	26,3
11 - 20	7568	33888	7,6	33,9
21 - 30	5543	39431	5,5	39,4
31 - 40	3458	42889	3,5	42,9
41 - 50	2721	45610	2,7	45,6
51 - 100	10303	55913	10,3	55,9
101 - 200	9953	65866	10,0	65,9
201 - 300	5253	71119	5,2	71,1
301 - 400	3358	74477	3,4	74,5
> 400	25421	99898	25,5	100,0
Σ	99898	-	100,0	-

Таблица В-30
(гл. III, § 2)

Частотно-спектральное распределение словоформ в
словаре по данным ЧС авторской речи эстонской худо-
жественной прозы

F	m_F	m_F^*	p_F (%)	p_F^* (%)
1	21760	21760	70,8	70,8
2	3885	25645	12,6	83,4
3	1628	27273	5,3	88,7
4	810	28083	2,7	91,4
5	519	28602	1,7	93,1
6	378	28980	1,2	94,3
7	253	29233	0,8	95,1
8	201	29434	0,7	95,8
9	131	29565	0,4	96,2
10	134	29699	0,4	96,6
11 - 20	529	30228	1,7	98,3
21 - 30	175	30403	0,6	98,9
31 - 40	90	30493	0,3	99,2
41 - 50	62	30555	0,2	99,4
51 - 100	96	30651	0,3	99,7
101 - 200	47	30698	0,2	99,9
201 - 300	15	30713	0,1	100,0
301 - 400	6	30719		
> 400	14	30733		
Σ	30733	-	100,0	-

Таблица Б-3I
(гл. III, § 2)

Частотно-спектральное распределение словоформ в
тексте по данным ЧС авторской речи эстонской художественной прозы

F	m_F	m_F^*	P_F (%)	P_F^* (%)
1	21760	21760	21,8	21,8
2	7770	29530	7,8	29,6
3	4884	34414	4,9	34,5
4	3240	37654	3,2	37,7
5	2595	40249	2,6	40,3
6	2268	42517	2,3	42,6
7	1771	44288	1,8	44,4
8	1608	45896	1,6	46,0
9	1179	47075	1,2	47,2
10	1340	48415	1,3	48,5
11 - 20	7623	56038	7,6	56,1
21 - 30	4300	60338	4,3	60,4
31 - 40	3171	63509	3,2	63,6
41 - 50	2808	66317	2,8	66,4
51 - 100	6637	72954	6,6	73,0
101 - 200	6752	79706	6,8	79,8
201 - 300	3608	83314	3,6	83,4
301 - 400	2127	85441	2,1	85,5
> 400	14457	99898	14,5	100,0
Σ	99898	-	100,0	-

Таблица Б-32
(гл. III, § 2)

Частотно-спектральное распределение лексем в словаре и тексте
по данным ЧС авторской речи I-го тома романа "Правда
и справедливость" ("Tõde ja õigus") А.Х. Таммсааре

F	С л о в а р ь				Т е к с т			
	m_F	m_F^*	$P_F(\%)$	$P_F^*(\%)$	m_F	m_F^*	$P_F(\%)$	$P_F^*(\%)$
I	3320	3320	45,18	45,18	3320	3320	2,91	2,91
2	III4	4434	15,16	60,34	2228	5548	1,95	4,86
3	581	5015	7,91	68,25	I743	7291	1,53	6,39
4	365	5380	4,97	73,22	I460	8751	1,28	7,67
5	247	5627	3,36	76,58	I235	9986	1,08	8,75
6	I87	5814	2,54	79,12	II22	III08	0,98	9,73
7	I46	5960	1,99	81,11	I022	I2130	0,90	10,63
8	I31	6091	1,78	82,89	I048	I3178	0,92	11,55
9	82	6173	1,12	84,01	738	I3916	0,65	12,20
10	68	6241	0,93	84,94	680	I4596	0,60	12,80
II-20	387	6628	5,27	90,21	5654	20250	4,95	17,75
2I-30	I86	6814	2,53	92,74	4653	24903	4,07	21,82
3I-40	II9	6933	1,62	94,36	4252	29155	3,73	25,55
4I-50	78	7011	1,06	95,42	3520	32675	3,08	28,63
5I-100	I48	7159	2,01	97,43	I0351	43026	9,07	37,70
10I-400	I49	7308	2,03	99,46	27321	70347	23,94	61,64
> 400	40	7348	0,54	100,00	43777	II4124	38,36	100,00
Σ	7348	-	100,0		II4124	-	100,0	

Таблица Б-33
(гл. III, § 2)

Частотно-спектральное распределение лексем в словаре и тексте
по данным ЧС речи персонажей I-го тома романа "Правда
и справедливость" ("Tõde ja õigus") А.Х. Таммсааре

F	С л о в а р ь				Т е к с т			
	m_F	m_F^*	ρ_F (%)	ρ_F^* (%)	m_F	m_F^*	ρ_F (%)	ρ_F^* (%)
I	1369	1369	43,67	43,67	1369	1369	2,96	2,96
2	470	1839	14,99	58,66	940	2309	2,03	4,99
3	229	2068	7,30	65,96	687	2996	1,49	6,48
4	154	2222	4,91	70,88	616	3612	1,33	7,81
5	112	2334	3,57	74,45	560	4172	1,21	9,02
6	89	2423	2,84	77,29	534	4706	1,16	10,18
7	68	2491	2,17	79,46	476	5182	1,03	11,21
8	50	2541	1,59	81,05	400	5582	0,87	12,07
9	41	2582	1,31	82,36	369	5951	0,80	12,87
10	31	2613	0,99	83,35	310	6261	0,67	13,54
11-20	218	2831	6,95	90,30	3200	9461	6,92	20,46
21-30	66	2897	2,11	92,41	1654	11115	3,58	24,04
31-40	56	2953	1,79	94,19	1929	13044	4,17	28,21
41-50	38	2991	1,21	95,41	1732	14776	3,75	31,96
51-100	65	3056	2,07	97,48	4404	19180	9,53	41,49
101-400	59	3115	1,88	99,36	10559	29739	22,84	64,33
> 400	20	3135	0,64	100,00	16493	46232	35,67	100,00
Σ	3135	-	100,0	-	46232	-	100,0	-

Таблица Б-34
(гл. III, § 2)

Частотно-спектральное распределение словоформ в словаре по
данным выборок из авторской речи 20 произведений эстонской худо-
жественной прозы: F - частота словоформы, m_F - количество сло-
воформ с частотой F

$m_F \backslash F$	A. Beekman	V. Gross	A. Hint	M. Kiik	J. Kross	P. Kuusberg	L. Promet	V. Saar
1	2270	2202	1805	2156	2336	1906	2112	1834
2	296	298	335	286	269	305	296	304
3	118	118	104	108	81	110	116	107
4	56	54	79	55	62	46	47	50
5	28	27	34	34	27	26	29	30
6	23	18	22	25	19	20	17	23
7	19	12	20	10	8	22	8	13
8	6	11	9	9	7	11	16	13
9	11	7	15	8	8	8	11	8
10	6	6	1	5	3	9	6	8
11-20	22	29	16	27	26	26	26	32
> 20	14	16	23	15	15	19	14	17
$\sum m_F = V$	2869	2798	2463	2738	2861	2508	2698	2439
N	4998	5026	4993	4996	4991	4976	5014	4975

(продолжение следует)

Таблица Б-34
(продолжение)

$F_i \backslash m_{F_i}$	H.Sergo	R.Sirge	M.Traat	E.Vetemaa	A.Kaal	T.Kallas	J.Peegel	J.Taulik
1	2306	2507	2102	1983	2012	2097	2099	1845
2	289	309	290	287	292	264	280	319
3	107	103	106	101	119	112	104	129
4	38	45	51	39	32	59	58	52
5	32	20	22	28	28	35	34	31
6	23	16	19	19	19	25	24	24
7	21	12	20	12	13	13	10	20
8	15	7	11	15	6	9	6	13
9	7	4	6	15	7	8	5	8
10	5	9	5	5	11	6	5	3
11-20	18	22	26	32	29	25	25	27
> 20	15	13	16	16	18	16	19	20
$\sum m_{F_i} = V$	2876	3067	2674	2552	2586	2669	2669	2491
N	4949	4975	5043	5001	4967	5001	4921	4986

610

(продолжение следует)

Таблица Б-34
(продолжение)

F_i	m_{F_i}	A.Valton	M.Unt	J.Kross, E.Niit	J.Smuul	(с р е д н е е)		
						m_F	$p\%$	$p^{\Sigma}\%$
1		1797	1890	2490	2364	2106	78,3	78,3
2		270	284	275	303	292	10,9	89,2
3		121	110	108	92	109	4,1	93,3
4		47	60	55	52	52	1,9	95,2
5		27	35	33	35	30	1,1	96,3
6		20	16	19	20	21	0,8	97,1
7		14	14	18	10	14	0,5	97,6
8		10	8	14	11	10	0,4	98,0
9		11	12	5	4	8	0,3	98,3
10		8	6	2	7	6	0,2	98,5
11-20		29	28	22	20	25	0,9	99,4
> 20		19	20	12	15	17	0,6	100,0
$\Sigma m_{F_i}=V$		2373	2483	3053	2933	2690	100,0	-
N		4984	4971	5123	5008	4995	-	-

Таблица Б-35
(гл. III, § 2)

Вычисление энтропии ($H = -\sum p_i \log_2 p_i$) лексико-частотного спектра словаря на основе сводного ЧС словоформ ($V=99898$, $V = 30733$) и ЧС подвыборки (по средним данным: $V = 4995$, $V = 2690$)

F_i	Сводный словарь		Подвыборка	
	p_i	$-p_i \log_2 p_i$	p_i	$-p_i \log_2 p_i$
1	0,708	0,3527	0,783	0,2763
2	0,126	0,3766	0,109	0,3485
3	0,053	0,2246	0,041	0,1889
4	0,026	0,1369	0,019	0,1086
5	0,017	0,0999	0,011	0,0716
6	0,012	0,0766	0,008	0,0612
7	0,008	0,0557	0,005	0,0382
8	0,007	0,0501	0,004	0,0319
9	0,004	0,0319	0,003	0,0251
10	0,004	0,0319	0,002	0,0179
11-20	0,017	0,0999	0,009	0,0612
> 20	0,018	0,1043	0,006	0,0443
Σ	1,000	$H = 1,6411$	1,000	$H = 1,2737$

Таблица В-36
(гл. III, § 2)

Энтропия (H) лексико-частотного спектра словаря по
данным ЧС словоформ 20 текстов (при $N = 5000$)

Текст (автор)	H
1. Э. Бээкман	1,2410
2. В. Гросс	1,2387
3. А. Хинт	1,4616
4. Х. Кийк	1,2588
5. Я. Кросс	1,1132
6. М. Траат	1,2503
7. Э. Ветемаа	1,3054
8. А. Каал	1,2819
9. Т. Каллас	1,2779
10. Ю. Пээгель	1,2577
11. П. Куусберг	1,3681
12. Л. Промет	1,2625
13. В. Саар	1,4007
14. Х. Серго	1,1880
15. Р. Сирге	1,0843
16. Ю. Туулик	1,4381
17. А. Валтон	1,3971
18. М. Унт	1,3807
19. Э. Нийт, Я. Кросс	1,1351
20. Ю. Смуул	1,1550
В среднем	1,2748

Таблица Б-37
(гл. III, § 2)

Наблюдаемые и ожидаемые численности слов $m(F)$ с частотой F по данным 40 лексем авторской речи эстонской художественной прозы ($N = 99898$, $L = 14654$). Вычисления по формулам:

$$\begin{aligned}
 \text{I: } m(F) &= c F^{-(1+\alpha)} && (\text{при } \gamma \neq 1) \text{ (по Ципфу)} \\
 \text{II: } m(F) &= c F^{-2} && (\text{при } \gamma = 1) \text{ (по Ципфу)} \\
 \text{III: } m(F) &= \alpha F^{-\beta} && (\text{аппроксим.}) \\
 \text{IV: } m(F) &= \alpha (F - R)^{-\beta} && (\text{аппроксим.})
 \end{aligned}$$

F	$m(F)$ (набл.)	$m(F)$ - (о ж и д.)			
		I	II	III	IV
1	8682	15928	(8682)	7350	8646
2	2054	3749	2171	2111	2000
3	938	1608	965	1017	932
4	591	882	543	606	554
5	345	534	347	406	373
10	124	130	87	117	113
20	26	31	22	33	35
30	17	13	10	16	18
40	10	7	5	10	11
60	6	3	3	5	6
100	2	1	1	2	2
Параметры:		$c = \alpha L = 15928$		$\alpha = 7350$	$\alpha = 4800$
		$\alpha = 1/\gamma = 1,087$		$\beta = 1,8$	$\beta = 1,65$
		$c = m(1) = 8682$			$R = 0,3$

Наблюдаемые и ожидаемые численности слов $m(F)$ с частотой F по данным 40 лексем авторской речи эстонской художественной прозы ($N = 99898$, $L = 14654$). Вычисления по формулам:

$$\text{I: } m(F) = C \ln \frac{F + \beta_2}{(F + \beta_2)^2 - 1} \quad (\text{Крылов})$$

$$\text{II: } i_{F+1} = \frac{L(Z)}{(1-x)^F} \left(\frac{\ln x}{x-1} - \sum_{k=1}^F \frac{(1-x)^{k-1}}{k} \right) \quad (\text{Орлов})$$

$$\text{III: } m(F) = \beta^\alpha (L+1) [(\beta+F-1)^\alpha - (\beta+F)^\alpha] \quad (\text{Арапов})$$

$$\text{IV: } L_{i+1} = L_i \frac{\alpha + (i-1)}{x + (i-1) + 1} \quad (\text{Уэринг-Хердан})$$

F_i	$m(F)$ (набл.)	$m(F)$ - (о ж и д.)			
		I	II	III	IV
1	8682	8440	8010	9298	(8682)
2	2054	2255	2411	2151	2308
3	938	1070	1149	941	1049
4	591	627	650	524	593
5	345	413	442	332	379
10	124	110		80	91
20	26	28		19	28
30	17	13		8	18
40	10	7		5	11
60	6	3		2	6
100	2	1		1	2
Параметры:		$C = 11824$	$Z = 175000$	$\alpha = 1/\gamma = 1,087$	$\alpha = 0,7651$
		$\beta_2 = 0,4$	$L(Z) = 19637$	$\beta = 0,656$	$x = 1,8773$
			$\chi = Z/N$		

Таблица Б-39
(гл. III, § 2)

Интегральный частотный спектр по данным 40 лексем авторской речи эстонской художественной прозы: наблюдаемые и ожидаемые накопленные численности слов $m^*(F)$. Вычисления по формулам:

$$\text{I: } m^*(F) = L [1 - (F+c)^{-\alpha}] \quad (\text{по Ципфу})$$

$$\text{II: } m^*(F) = L (1 - e^{-\alpha F^{\beta}}) \quad (\text{Вейбулл})$$

F	$m^*(F)$	I	II
I	8682	8435	8983
5	12610	12637	12529
10	13485	13594	13584
50	14386	14451	14581
100	14530	14557	14639
400	14631	14652	14640
4237	14654	14653	14654
Параметры:		$\alpha = \frac{1}{\delta} = 1,087$ $c = 1,2$	$\alpha = 0,95$ $\beta = 0,44$

Таблица Б-40
(гл. III, § 2)

Распределение частот лексем в тексте (интегральное частотно-спектральное распределение) по данным ЧС авторской речи эстонской художественной прозы: F - частота слова, $p^*(F)$ - покрытие текста на основе накопленных частот, z - квантили нормального распределения

F	$p^*(F)$	z
1	0,087	-1,36
2	0,128	-1,14
3	0,156	-1,01
4	0,180	-0,92
5	0,197	-0,85
6	0,213	-0,80
7	0,228	-0,75
8	0,240	-0,71
9	0,251	-0,67
10	0,263	-0,63
20	0,335	-0,43
30	0,405	-0,24
40	0,434	-0,17
50	0,456	-0,11
100	0,559	0,15
200	0,659	0,41
300	0,712	0,58
400	0,745	0,66
500	0,783	0,78
600	0,794	0,82
1000	0,850	1,04
1300	0,863	1,09
2000	0,900	1,28
3500	0,960	1,75
4237	1,000	-

($N = 99\ 898$; $L = 14\ 654$)

Таблица В-41
(гл. III, § 3)

Зависимость между объемом словаря (V — число словоформ) и объемом текста (N — число словоупотреблений) по данным подвыборки ЧС авторской речи эстонской художественной прозы (средние данные). Вычисления ожидаемого роста словаря по формулам:

$$\text{I:} \quad V = N \alpha (N + \ell)^{-1}$$

$$\text{II:} \quad V = N \alpha (N + \ell)^{-1} + c$$

N	V (набл.)	V (I)		V (II)	
1000	686	654	(-32)	686	(0)
2000	1241	1240	(-1)	1237	(-4)
3000	1762	1767	(+5)	1756	(-6)
4000	2238	2244	(+6)	2245	(+7)
4995	2690	2676	(-14)	2704	(+14)
Параметры:		α		11837	
		ℓ		18776	
		c		31016	
				100	

Таблица Б-42
(гл. III, § 3)

Зависимость между объемом словаря (V - число словоформ, L - число лексем) и объемом текста (N - число словоупотреблений) по данным разных языков (Алексеев П.М., 1968; Sambor J., 1970; Billmeier G., 1968; *Latviešu valodas...*, 1969). Вычисления ожидаемого роста словаря по формуле $V' \text{ (или } L') = N \alpha (N + \beta)^{-1}$

Английский язык (электроника)			Польский язык (худож. лит.)			Немецкий язык (газета)			Латышский язык (газета)		
N	V	V'	N	L	L'	N	L	L'	N	L	L'
20000	2948	2964	17141	4388	4366	15000	4280	4264	50000	7065	6966
30000	4040	3982	29787	6146	6246	30000	7472	7637	100000	9834	10314
40000	4836	4807	48255	8026	8042	45000	10460	10373	200000	13389	13577
50000	5399	5490	64510	9250	9105	60000	12832	12635	300000	16103	15178
			Параметры:								
$\alpha = 12710$			$\alpha = 14995$			$\alpha = 36560$			$\alpha = 19860$		
$\beta = 65770$			$\beta = 41720$			$\beta = 113610$			$\beta = 92550$		

Таблица Б-43

(гл. III, § 3)

Наблюдаемый и ожидаемый рост словаря по данным
авторской речи современной эстонской художественной прозы (словоформы)

N	V (набл.)	V (ожд.) по формулам:									
		(2a)		(3a)		(3б)		(3в)		(4a)	
1000	686	688	(+2)	684	(-2)	686	(0)	685	(-1)	685	(-1)
2000	1241	1240	(-1)	1249	(+8)	1245	(+4)	1244	(+3)	1246	(+5)
3000	1762	1752	(-10)	1760	(-2)	1756	(-6)	1756	(-6)	1757	(-5)
4000	2238	2237	(-1)	2234	(-4)	2237	(-1)	2237	(-1)	2237	(-1)
5000	2692	2705	(+13)	2680	(-12)	2697	(+5)	2693	(+1)	2693	(+1)
Прогноз:											
10000	-	4880		4623		4791		4715		4753	
50000	-	19196		14115		17690		14229		16867	
100000	-	34625		20912		30746		19398		28432	
Параметры:											
a		1,9248		-0,0456		-6,4732		$-6,728 \cdot 10^{-5}$		-0,0040	
B		0,8510		0,3071		0,2492		0,8780		2,3535	
V _п		-		-				24000		-	

Таблица В-44
(гл. III, § 3)

Наблюдаемый и ожидаемый рост словаря по данным романа "Правда и право" (т. I) А.Х. Таммсааре (эстонский язык; авторская речь, лексемы); звездочкой отмечены точки наблюдения (по данным: Villup A., 1978)

N	V (набл.)	V (ожд.) по формулам:		
		(3a)	(3б)	(4a)
10000	2114*	*	*	*
20000	3124*	*	*	*
Полный объем:		Прогноз:		
114124	7348	6259	8052	7207
Параметры:				
	a	-0,1461	-20,4413	-0,006714
	B	0,2567	-0,4837	2,4521

Таблица В-45
(гл. III, § 3)

Наблюдаемый и ожидаемый рост словаря по данным "Пиковой дамы" А.С. Пушкина (по данным: Орлов Ю.К., 1978) - лексеми

N	V (набл.)	V (ожд.) по формулам:			
		(3а)	(3б)	(3в)	(4а)
1000	462*	*	*	*	*
2000	787*	*	*	*	*
3000	1067*	*	*	*	*
Прогноз:					
4000	1348	1308	1325	1326	1321
5000	1541	1527	1564	1567	1556
6000	1752	1727	1790	1794	1776
Полный объем:					
6861	1928	1887	1976	1981	1957
Параметры:					
	a	-0,1128	-5,376	-0,0001094	-0,01694
	B	0,2663	-0,3126	0,7766	1,976
	V _n	-	-	20000	-

Таблица Б-46
(гл. III, § 3)

Наблюдаемый и ожидаемый рост словаря по данным ЧС
английского языка Х. Кучеры и В. Фрэнсиса (Ки́сера, Н.,
Francis W.N., 1967) - словоформы

N	V (набл.)	V (ожд.) по формулам:			
		(3а)	(3б)	(3в)	(4а)
101566	13706*	*	*	*	*
253538	23655*	*	*	*	*
Прогноз:					
1014232	50406	47500	53105	49076	50617
10 ⁷	-	93290	199430	95332	148000
50721	8749	8737	8961	8891	8853
10051	3009	2729	3241	3138	2968
2000	700-1000	760	1082	1095	902
1000	-	427	649	694	525
500	-	437	378	438	300
100	-	57	94	151	77
10	-	7	10	33	9
Параметры:					
a		-0,237	-20,287	-7,21 · 10 ⁻⁵	-0,00891
B		0,185	-0,429	0,6612	2,215
V _n		-	-	100000	-

Таблица Б-47
(гл. III, § 3)

Наблюдаемый объем (V) и ожидаемый объем (V') словаря в зависимости от длины текста (N) по данным разных языков по формуле: $V' = Ne^{-a(\ln N)^B}$

а) Латвийский язык - газеты, лексемы (Latviešu val., 1969)			б) Чешский язык - технич. тексты, словоформы (Вежка, 1972)			в) Казахский язык - газеты, словоформы (Ахабаев, 1971)		
N	V	V'	N	V	V'	N	V	V'
50000	7065	7025	25000	4829	4827	25000	9088	9161
100000	9834	9919	75000	9603	9626	50000	15047	14875
200000	13389	13510	125000	13056	13050	100000	23895	23523
300000	16103	15912	175000	15858	15853	150000	29785	30378
500000	-	19200	500000	-	28200	500000	-	61000
10^8	-	24000	10^8	-	40000	10^8	-	87000
10^7	-	37000	10^7	-	114000	10^7	-	230000
Параметры:								
a	0,003736		0,01123			0,001372		
B	2,6304		2,1539			2,8488		

Продолжение следует.

Таблица Б-47
(продолжение)

г) Польский язык - А.Мицкевич, словоформы, (Sambor, 1970)			д) Украинский язык - О.Довженко, словоформы (Дарчук, 1975)			е) Английский язык - суд.механизмы, словоформы (Лукьяненко, Нешитой, 1975)		
N	V	V'	N	V	V'	N	V	V'
12172	3434	3458	5000	1629	1629	50495	4871	4849
29787	6146	6044	10000	2637	2646	100970	6858	6882
48255	8026	7998	15000	3504	3482	201966	9470	9520
64510	9250	9398	20000	4195	4214	302156	11314	11360
100000	-	11800	100000	-	11500	403966	12975	12832
500000	-	25000	500000	-	28000	500000	-	14000
10^6	-	33000	10^6	-	40000	10^6	-	18000
10^7	-	60000	10^7	-	110000	10^7	-	36000
Параметры:								
a	0,00364			0,01055			0,01235	
B	2,6081			2,1783			2,2019	

Продолжение следует.

Таблица Б-47
(продолжение)

ж) Английский язык - электроника, словоформы (Алексеев, 1968)			з) Румынский язык - электроника, словоформы (Ешан, 1966)			и) Русский язык - электроника, словоформы (Калинина, 1968)		
N	V	V'	N	V	V'	N	V	V'
50000	5399	5437	50000	6785	6841	50000	9464	9388
100000	7853	7728	100000	10281	10070	100000	14062	14168
150000	9361	9371	150000	12477	12479	150000	17263	17803
200000	10582	10682	200000	14292	14454	200000	21468	20818
500000	-	15600	500000	-	22400	500000	-	33000
10^6	-	20000	10^6	-	30000	10^6	-	45000
10^7	-	38000	10^7	-	68000	10^7	-	94000
Параметры:								
a	0,009152			0,008148			0,004284	
b	2,3057			2,3086			2,5058	

Таблица Б-48

(гл. IV, § I)

Распределение словоформ (СФ) по длине в тексте по данным
выборки из четырех подязыков эстонского языка

Длина СФ в буквах	Авт. речь		Речь перс.		Г а з е т а		Научн. текст		В с е г о	
	число	%	число	%	число	%	число	%	число	%
2	113	10,2	178	14,5	70	8,1	76	8,4	437	10,7
3	101	9,1	207	16,8	81	9,4	59	6,6	448	10,9
4	183	16,6	235	19,1	78	9,1	77	8,5	573	14,0
5	154	13,9	194	15,8	122	14,2	106	11,8	576	14,1
6	181	16,4	168	13,6	80	9,3	100	11,1	529	12,9
7	100	9,0	114	9,3	90	10,5	100	11,1	404	9,9
8	110	10,0	55	4,5	76	8,8	105	11,7	346	8,4
9	60	5,4	43	3,5	65	7,5	86	9,6	254	6,2
10	43	3,9	18	1,5	40	4,6	60	6,7	161	3,9
11	29	2,6	9	0,7	46	5,3	41	4,5	125	3,1
12	12	1,1	6	0,5	37	4,3	33	3,7	88	2,1
13	8	0,7	2	0,2	23	2,7	22	2,4	55	1,3
>14	12	1,1	1		53	6,2	35	3,9	101	2,5
Всего СФ	1106	100,0	1230	100,0	861	100,0	900	100,0	4097	100,0
Букв	6414	-	5911	-	6079	-	6370	-	24774	-
Ср. длина СФ	5,80		4,81		7,06		7,08		6,05 (букв)	

Таблица Б-49
(гл. IV, § I)

Распределение словоформ (СФ) по длине в эстонском тексте: I - смешанный текст (см. табл. Б-48); II - художественная проза (авт. речь и речь перс.); III - газета "Rahva Hääl" 1968 (Kaasik U. и др., 1975); IV - художественная проза и газета (Kaasik U. и др., 1975). Всего - около 57 % худож. прозы, 40 % публицистики, 3 % научного текста.

Длина СФ в буквах	Ч а с т о т а (%)				В с е г о	
	I	II	III	IV	число	%
2	10,7	11,0	10,0	10,3	3950	10,5
3	10,9	10,2	8,9	10,5	3874	10,3
4	14,0	16,6	14,2	16,6	6056	16,1
5	14,1	14,7	13,2	14,3	5379	14,3
6	12,9	14,0	11,7	13,0	4928	13,1
7	9,9	9,6	9,5	9,4	3573	9,5
8	8,4	8,2	8,5	7,9	3084	8,2
9	6,2	5,4	6,9	5,9	2182	5,8
10	3,9	3,4	5,3	4,1	1542	4,1
11	3,1	2,5	3,2	2,6	1016	2,7
≥ 12	5,9	4,4	7,6	5,2	2031	5,4
Всего (%)	100,0	100,0	100,0	100,0	-	100,0
Объем текста	4097	9857	4907	18754	37615	-
Средняя длина СФ	6,05	5,7	6,3	6,0	6,0	-

Распределение словоформ (СФ) по длине в эстонском тексте и в соответствующем словаре по данным выборки из авторской речи романа Я. Кресса "Три чуми" I ("Kolme katku vahel" I)

Длина словоформы в буквах	Ч а с т о т а				Отно- шение T / C
	в тексте		в словаре		
	число	% (T)	число	% (C)	
2	518	10,7	17	0,6	17,8
3	472	9,7	86	3,1	3,1
4	727	15,0	290	10,3	1,5
5	716	14,7	413	14,7	1,0
6	693	14,3	463	16,5	0,9
7	444	9,1	377	13,4	0,7
8	432	8,9	364	13,0	0,7
9	263	5,4	238	8,5	0,6
10	198	4,1	182	6,5	0,6
11	142	2,9	134	4,8	0,6
12	71	1,5	66	2,4	0,6
13	71	1,5	63	2,3	0,6
14	55	1,1	54	1,9	0,6
15	24	0,5	24	0,9	0,6
16	16	0,3	16	0,6	0,5
17	5	0,3	5	0,5	0,6
18	5		5		
19	1		1		
20	1		1		
21	1		1		
23	2		2		
Всего	4857	100,0	2802	100,0	-
Средняя длина СФ	6,0		7,3		

Таблица Б-51
(гл. IV, § 1)

Распределение длин словоформ (СФ) в тексте авторской речи
современной эстонской художественной прозы (10 авторов)

Длина СФ в слогах	Э. Бээкман	В. Гросс	Х. Кийк	Я. Кросс	П. Куусберг	Ю. Смуул	М. Траат	М. Унт	А. Валтон	Э. Ветемаа	В с е г о	Средняя (%)	Ошибка ср. (ур. знач. 5%)
1	206	209	257	253	230	261	269	238	312	247	2482	24,8	+2,2
2	406	390	436	395	427	402	393	417	374	375	4015	40,2	+1,5
3	256	220	191	192	183	196	186	199	190	210	2023	20,2	+1,6
4	97	130	83	107	101	101	114	100	91	101	1025	10,3	+1,0
5	28	31	20	33	36	28	29	29	24	45	303	3,0	+0,5
6	6	13	10	16	12	12	7	8	7	17	108	1,1	+0,3
7	1	6	2	2	10		2	4	2	4	33	0,3	
8		1		1	1			5		1	9	0,1	
9			1	1							2	0,0	
Всего	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	10.000	100,0	
Средняя длина СФ в слогах	2,36	2,45	2,22	2,34	2,37	2,27	2,27	2,33	2,17	2,39	Итого- вая ср.	2,32	+0,05 слогов

Таблица Б-52
(гл. IY, § I)

Распределение длин словоформ (СФ) в тексте авторской речи
художественной прозы периода до 1940 г. (4 автора - классика)

Длина СФ в слогах	О. Лутс	А.Х. Тамм- сааре	Ф. Туглас	Э. Вильде	Всего	Сред- няя	Ошибка (ур. знач. 5%)
1	254	267	296	288	1105	27,6	+3,1
2	455	464	420	437	1776	44,4	+3,1
3	179	179	178	167	703	17,6	+1,0
4	79	69	90	89	327	8,2	+1,6
5	26	18	11	15	70	1,7	+1,0
6	7	3	4	3	17	0,5	
7	-	-	1	-	1		
8	-	-	-	1	1		
Всего	1000	1000	1000	1000	4000	100,0	-
Средняя длина СФ	2,19	2,12	2,12	2,12	Итоговая средняя	2,14	+0,06 слогов

Таблица Б-53
(гл. IV, § 1)

Распределение длин словоформ (СФ) в тексте по данным выборок из художественной прозы (авторская речь + речь персонажей) разных языков. Данные взяты из работ: Fucks W., 1956; Herdan G., 1966; Zvilka T., 1974; Якубайтис Т., 1963.

Кол-во слогов	Английский	Французский	Немецкий	Румынский	Венгерский	Русский яз.	Латышский	Эстонский	Латинский	Турецкий
1	66,8	55,8	51,7	45,5	34,4	33,9	33,8	28,8	24,2	18,8
2	21,1	27,9	31,6	28,4	30,4	30,3	38,4	40,2	32,1	37,8
3	8,2	12,9	11,1	18,6	20,7	21,4	19,4	18,5	28,7	37,0
4	3,3	2,9	4,3	6,7	9,9	9,7	7,0	8,8	11,6	12,1
5	0,5	0,5	1,0	0,7	3,4	3,5	1,2	2,5	2,8	3,6
более 5	0,1	0,0	0,3	0,1	1,2	1,2	0,2	1,2	0,6	0,7
Ср. длина СФ в словах	1,50	1,64	1,82	1,89	2,22	2,23	2,04	2,20	2,39	2,46

Таблица Б-54
(гл. IV, § I)

Распределение длин словоформ (СФ) в эстонском тексте (по данным выборки из авторской речи художественной прозы). Ожидаемые частоты (%) вычислены с помощью формулы В. Фукса (1957):

$$P_i = \frac{e^{-(\bar{x}-1)} (\bar{x}-1)^{i-1}}{(i-1)!}$$

Длина СФ в слогах	Ч а с т о т а (%)		Разница x - y
	наблюд. (x)	ожд. (y)	
1	24,8	26,7	-1,9
2	40,2	35,3	+4,9
3	20,2	23,3	-3,1
4	10,3	10,2	+0,1
5	3,0	3,4	-0,4
6	1,1	0,9	+0,2
7	0,3	0,2	+0,1
8	0,1	0,0	+0,1
Всего (%)	100,0	100,0	-
Средняя длина СФ в слогах	2,32	2,32	-

Таблица Б-55

(гл. IV, § I)

Комплексное распределение словоформ (СФ) по длине в звуках и слогах и средние данные по выборке из авторской речи книги Ю. Смуула ("Mahulaste imelikud juhtumised...", Tallinn, 1963). Объем выборки - 1000 словоупотреблений.

Звуков Слогов	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Число СФ	Число слогов	Число звуков	Ср. длина в звуках		
																		СФ	слога	
1	118	86	53	4											261	261	717	2,75	2,75	
2		46	108	150	77	19	2								402	804	1931	4,80	2,40	
3				10	42	81	44	18	1						196	588	1393	7,11	2,37	
4						4	24	41	20	9	3				101	404	924	9,15	2,29	
5							1		1	4	5	10	2	4	1	28	140	328	11,71	2,34
6											2	2	4	2	2	12	72	156	13,00	2,17
Всего	118	132	161	164	119	105	70	60	25	16	15	6	6	3	1000	2269	5449	5,45	2,40	

Таблица В-56
(гл. IV, § I)

Вычисление константы П. Гиро (C), указывающей на устойчивую связь длины слова (x) с его рангом (i) в частотном списке по данным ЧС словоформ эстонского языка.

Средний ранг (i)	x	C = (ln2i)/x
10 (5-15)	2,30	0,57
100 (90-110)	4,30	0,54
200 (190-210)	4,50	0,57
300 (290-310)	4,96	0,56
400 (390-410)	5,30	0,55
600 (590-610)	5,20	0,59
900 (890-910)	5,84	0,56
1000 (990-1010)	5,94	0,57

600 (1-600)	В среднем:	0,56
700 (1-700)		
800 (1-800)		
900 (1-900)		
1000 (1-1000)		
10000 (1-10000)		
30000 (1-30000)		
Параметры:		

Связь между длиной словоформ (СФ) и ее частотой по данным ЧС эстонского языка ($N = 99898$, $V = 30733$). Ожидаемая средняя длина СФ в буквах вычисляется по формулам

$$\text{I: } \bar{X}_i = \alpha + \beta \ln i \quad (\text{логарифмический закон})$$

$$\text{II: } \bar{X}_i = \bar{X}_n (1 - e^{-ci^k}) \quad (\text{закон Вейбулла})$$

Здесь $\bar{X}_i$ - средняя длина СФ: $\bar{X}_i = \frac{1}{i} \sum x_i$;

$\bar{X}_n$ - предел средней длины СФ;

i - ранг СФ в ЧС;

α, β, c, k - параметры.

(Значения параметров вычислены методом наименьших квадратов на основе эмпирических данных в пределах $i = 50 \div 1000$.)

Ранг i	$\bar{X}_i$ (набл.)	$\bar{X}_i$ (ожд.)	
		I	II
50 (I-50)	3,48	3,47	3,48
100 (I-100)	3,89	3,85	3,84
200 (I-200)	4,20	4,23	4,22
300 (I-300)	4,42	4,45	4,44
400 (I-400)	4,64	4,60	4,61
500 (I-500)	4,71	4,72	4,73
600 (I-600)	4,82	4,82	4,83
700 (I-700)	4,94	4,91	4,92
800 (I-800)	4,97	4,98	4,99
900 (I-900)	5,06	5,04	5,06
1000 (I-1000)	5,12	5,10	5,12
10000 (I-10000)	-	6,35	6,30
30000 (I-30000)	-	6,95	6,74
Параметры:	-	$\alpha = 1,35$ $\beta = 0,543$	$c = 0,282$ $k = 0,203$ $\bar{X}_n = 7,5$

Таблица Б-58

(гл. IV, § 3)

Распределение групп полнзначных слов эстонского языка в зависимости от количества словарных значений (существительные и глаголы с частотой $F \geq 10$ в тексте объемом $N = 100\,000$ словоупотреблений)

Количество значений	Существительные		Глаголы		Всего	
	число	%	число	%	число	%
I	64	24,2	21	11,0	85	18,7
2	62	23,5	35	18,4	97	21,4
3	53	20,1	33	17,4	86	19,0
4	35	13,3	34	17,9	69	15,2
5	19	7,2	18	9,5	37	8,2
6	11	4,2	14	7,4	25	5,6
7	3	1,1	7	3,7	10	2,2
8	10	3,8	5	2,6	15	3,3
9	4	1,5	10	5,2	14	3,1
10	2	0,7	3	1,6	5	1,1
11	1	0,4	1	0,5	2	0,4
12	-	-	2	1,0	2	0,4
15	-	-	2	1,0	2	0,4
16	-	-	2	1,0	2	0,4
18	-	-	1	0,5	1	0,2
19	-	-	1	0,5	1	0,2
24	-	-	1	0,5	1	0,2
Всего	264	100,0	190	100,0	454	100,0

Таблица Б-59

(гл. IV, § 3)

Сдвиги в распределении вероятностей
эстонских существительных (в процентах) в за-
висимости от количества значений в разрезе от-
дельных частотных зон

Колич. значений	Процент слов	В частотных зонах:				
		F > 50	F > 40	F > 30	F > 20	F > 10
I		2,4	4,2	6,2	17,7	24,2
2		13,9	12,8	15,4	15,0	23,5
3		26,2	25,5	26,0	23,0	20,1
4		10,0	10,6	12,3	14,2	13,3
5		16,1	14,9	10,8	11,5	7,2
6		7,6	6,4	7,7	5,3	4,2
7		2,4	4,2	3,1	1,8	1,1
8		9,5	10,6	9,3	6,1	3,8
9		7,1	6,4	4,6	2,7	1,5
10		2,4	2,2	3,1	1,8	0,7
10		2,4	2,2	1,5	0,9	0,4
Всего (%)		100,0	100,0	100,0	100,0	100,0
Число слов		43	47	67	117	264

Таблица Б-60
(гл. IV, § 3)

Соотношение между количеством значений (m) и вероятностью слов с данным количеством значений (p) по данным английского, венгерского и русского языков; ожидаемая вероятность $p' = \alpha e^{-\ell \sqrt{m}}$

m	Английский яз.		Венгерский яз.		Русский яз. (глаголы)	
	p	p'	p	p'	p	p'
I	0,427	0,426	0,504	0,558	0,615	0,627
2	0,203	0,205	0,265	0,200	0,254	0,189
3	0,117	0,117	0,118	0,090	0,071	0,075
4	0,072	0,073	0,052	0,046	0,030	0,035
5	0,048	0,048	0,024	0,025	0,013	0,017
6	0,035	0,033	0,013	0,015	0,007	0,009
7	0,023	0,023	0,008	0,009	0,003	0,005
8	0,016	0,017	0,005	0,006	0,002	0,003
9	0,013	0,012	0,003	0,004	0,002	0,002
10	0,009	0,009	0,002	0,003	0,002	0,001
11	0,0073	0,0071	0,0014	0,0017		
12	0,0060	0,0055	0,0012	0,0012		
13	0,0053	0,0042	0,0009	0,0008	0,001	(0,03)
14	0,0034	0,0033	0,0007	0,0006		
15	0,0032	0,0026	0,0007	0,0004		
15	0,014	(0,015)	0,002	(0,04)		
Всего	1,0	(1,0)	1,0	(1,0)	1,0	(1,0)
Кол-во слов	29593	-	58317	-	9502	-
Параметры:						
α	2,5		6,8		11,4	
ℓ	1,77		2,5		2,9	

Таблица Б-6I

(гл. IV, § 3)

Соотношение между средней частотой употребления (F) и средним количеством значений (m) основных полнозначных слов эстонского языка (существительные и глаголы с частотой $F \geq 10$ в тексте объема $N = 100\ 000$ словоупотреблений);

n - число слов с данной частотой, $\sum m_i$ - общее число значений

Интервал частот	Ср. частота F	Существительное			Глагол			Существительное + глагол		
		n	$\sum m_i$	m	n	$\sum m_i$	m	n	$\sum m_i$	m
10-20	15	146	368	2,5	86	304	3,5	232	672	2,9
21-30	25	51	145	2,8	34	127	3,7	85	272	3,2
31-40	35	19	72	3,8	22	110	5,0	41	182	4,4
41-50	45	5	23	4,6	9	29	3,2	14	52	3,7
51-60	55	6	29	4,8	4	19	4,8	10	48	4,8
61-70	65	3	12	4,0	4	24	6,0	7	36	5,1
71-80	75	10	45	4,5	8	51	6,4	18	96	5,3
81-90	85	4	14	3,5	2	16	8,0	6	30	5,0
91-100	95	2	9	4,5	4	30	7,5	6	39	6,5
101-150	125	7	50	7,1	4	19	4,8	11	69	6,3
151-200	175	6	25	4,2	6	73	12,2	12	98	8,2
>200	300	4	24	4,8	6	66	11,0	11	90	8,2
-	-	264	816	(3,1)	189	868	(4,6)	453	1684	(3,7)

Таблица Б-62
(гл. IV, § 3)

Соотношение между частотой употребления и количеством значений слова в разных частотных зонах по данным эстонского языка; ожидаемые величины вычислены по формулам: $m_I = 8,5 - 7e^{-0,012 F}$; $m_{II} = F^{0,38}$

Частотная зона	Ср. частота (F)	Среднее количество значений		
		наблюд. (m)	ожд. (m_I)	ожд. (m_{II})
10-20	15	2,9	2,7	2,9
21-30	25	3,2	3,3	3,4
31-40	35	4,4	3,9	3,9
41-50	45	3,7	4,4	4,2
51-60	55	4,8	4,9	4,6
61-70	65	5,1	5,3	4,9
71-80	75	5,3	5,7	5,2
81-90	85	5,0	6,0	5,4
91-100	95	6,5	6,3	5,6
101-150	125	6,3	6,9	6,3
151-200	175	8,2	7,7	7,1
>200	300	8,2	8,3	8,7
Прогноз:	10	-	2,5	2,5
	5	-	2,0	1,9
	1	-	1,6	1,05

Таблица Б-63
(гл. IV, § 3)

Соотношение между частотой употребления и количеством значений слова по данным русского языка (по материалам исследования А.А. Поликарпова, 1976);

$$m_I = 8,3 - 5,2 e^{-0,0009} ; \quad m_{II} = 1,2 F^{0,23}$$

Интервал	Ср. частота (F)	Среднее количество значений		
		наблюд. (n)	ожд. (m_I)	ожд. (m_{II})
I-100	4369	8,2	8,2	8,2
101-200	758	5,4	5,7	5,5
201-300	468	4,9	4,9	4,9
301-400	374	4,4	4,6	4,7
401-500	250	4,6	4,2	4,3
501-600	234	4,3	4,1	4,2
601-700	200	3,9	4,0	4,0
701-800	176	4,4	3,9	3,9
801-900	155	3,3	3,8	3,8
901-1000	140	3,7	3,7	3,7
Прогноз:	100	-	3,6	3,5
	50	-	3,3	2,9
	10	-	3,2	2,0
	1	-	3,1	1,2

Таблица Б-64
(гл. V, § I)

Количественные показатели богатства и разнообразия лексики по данным 20 выборок из авторской речи современной эстонской художественной прозы (N - объем текста; V - объем словаря в словоформах; V_1 - количество одноразовых словоформ; $\gamma_1, \gamma_2, \gamma_3$ - константы распределения Ципфа для разных частотных зон; $S_{\bar{x}}$ - стандартное отклонение средней частоты)

А в т о р	N	V	V/N	V_1	V_1/N	V_1/V	γ_1	γ_2	γ_3
1. Э. Бээкман	4998	2869	0,574	2270	0,454	0,791	0,83	0,65	0,50
2. В. Гросс	5026	2798	0,557	2202	0,438	0,787	0,60	0,69	0,57
3. А. Хинт	4993	2463	0,493	1805	0,362	0,733	0,75	0,71	0,60
4. Х. Кийк	4996	2738	0,548	2156	0,432	0,787	0,86	0,70	0,55
5. Я. Кросс	4991	2861	0,573	2336	0,468	0,817	0,94	0,72	0,48
6. П. Куусберг	4976	2508	0,504	1906	0,383	0,760	0,87	0,72	0,60
7. Л. Промет	5014	2698	0,538	2112	0,421	0,783	0,76	0,71	0,53
8. В. Саар	4975	2439	0,490	1834	0,369	0,752	0,85	0,71	0,57
9. Х. Серго	4949	2876	0,581	2306	0,466	0,802	0,95	0,67	0,50
10. Р. Сирге	4975	3067	0,617	2507	0,504	0,817	0,57	0,73	0,47
11. М. Траат	5043	2674	0,530	2102	0,417	0,786	0,91	0,75	0,52
12. Э. Ветемаа	5001	2552	0,510	1983	0,397	0,777	0,54	0,77	0,54
13. А. Каал	4967	2586	0,521	2012	0,405	0,778	0,62	0,79	0,53
14. Т. Каллас	5001	2669	0,534	2097	0,419	0,786	0,93	0,68	0,53
15. Ю. Пээгелъ	4921	2669	0,542	2099	0,427	0,786	0,69	0,72	0,52
16. Ю. Туулик	4986	2491	0,500	1845	0,370	0,741	0,49	0,73	0,60
17. А. Валтон	4984	2373	0,476	1797	0,361	0,757	0,76	0,81	0,57
18. М. Унт	4971	2483	0,500	1890	0,380	0,761	0,58	0,78	0,56
19. Э. Нийт, Я. Кросс	5123	3053	0,596	2490	0,486	0,816	0,95	0,67	0,48
20. Ю. Смуун	5008	2933	0,586	2364	0,472	0,806	0,73	0,69	0,49
Среднее ($\bar{x}$)	4995	2690	0,539	2106	0,422	0,783	0,75	0,72	0,54
П е д о л ж е н н а я о ш и б к а ($3 \cdot S_{\bar{x}}$)	-	132	0,027	150	0,030	0,018	0,10	0,03	0,03

Таблица Б-65
(гл. V, § I)

Соотношение объема словаря (V) и объема текста (N) у разных авторов; параметры функции $V = \frac{aN}{N+b}$, вычисленные методом наименьших квадратов в интервале $N = 2000 \div 5000$

А в т о р	Объем текста (N)					Параметры		
	1000	2000	3000	4000	~ 5000	a	b	$a/b = \varphi$
1. Э. Бээкмаа	731	1333	1865	2404	2869	I2338	16600	0,743
2. В. Гресс	677	1315	1859	2358	2798	II104	14890	0,746
3. А. Хинт	649	1116	1597	2034	2463	I2190	19864	0,614
4. Х. Нийк	710	1351	1828	2315	2738	8300	10363	0,801
5. Я. Кросс	723	1315	1914	2382	2861	I3395	18299	0,732
6. П. Куусберг	645	1166	1674	2075	2508	10544	16061	0,656
7. Н. Промет	674	1212	1720	2207	2698	I3408	20192	0,664
8. В. Саар	633	1128		2045	2439	II064	17617	0,628
9. Х. Серго	734	1326	1885	2416	2876	I3708	18706	0,733
10. Р. Сарге	764	1397	2017	2572	3067	I5893	20724	0,767
11. М. Траат	689	1235	1722	2226	2674	II112	16072	0,691
12. Э. Ветемаа		1208	1734	2162	2552	10168	14780	0,688
13. А. Каал	651	1204	1690	2119	2586	10219	15029	0,680
14. Т. Каллас	663	1226	1733	2179	2669	II207	16325	0,687
15. Ю. Пээгель	690	1224	1700	2223	2669	12645	18806	0,672
16. Ю. Туулик	624	1135	1560	2005	2491	I0269	16259	0,632
17. А. Валтои	588	1036	1468	1955	2373	I6543	30118	0,549
18. М. Унт	658	1176	1678	2133	2483	I0477	15738	0,666
19. Э. Нийт, Я. Кросс	740	1357	1923	2473	3053	I4556	19515	0,746
20. Ю. Смуул	732	1361	1917	2473	2933	I2773	16811	0,760

Таблица Б-66

(гл. V, § 1)

Корреляционная матрица лексикостатистических
характеристик по данным 20 текстов

—	V/N	V_1/N	V_1/V	γ_1	γ_2	γ_3	α	φ
V/N	—	0,99	0,92	0,19	-0,59	-0,85	0,38	0,85
V_1/N	0,99	—	0,96	0,23	-0,54	-0,89	0,40	0,85
V_1/V	0,92	0,96	—	0,31	-0,39	-0,93	0,38	0,80
γ_1	0,19	0,23	0,31	—	-0,43	-0,26	0,15	0,14
γ_2	-0,59	-0,54	-0,39	-0,43	—	0,31	-0,02	-0,56
γ_3	-0,85	-0,89	-0,93	-0,26	0,31	—	-0,48	-0,65
α	0,38	0,40	0,38	0,15	-0,02	-0,48	—	-0,11
φ	0,85	0,85	0,80	0,14	-0,56	-0,65	-0,11	—
$\sum \alpha $	4,77	4,86	4,69	1,71	2,84	4,37	1,92	3,96

Таблица Б-67
(гл. V, § 2)

Распределение частот в объединенном тексте по формуле $(p+q)^F$; p и q — вероятности того, что слово относится к тексту А или к тексту В соответственно; F — частота слова в объединенном тексте, f — подчастота.

$F \backslash f$	0	1	2	3	4	...	$(p + q)^F$
1	p	q					$(p + q)^1$
2	p^2	$2pq$	q^2				$(p + q)^2$
3	p^3	$3p^2q$	$3pq^2$	q^3			$(p + q)^3$
4	p^4	$4p^3q$	$6p^2q^2$	$4pq^3$	q^4		$(p + q)^4$
...							...

В случае, если $p = q = 1/2$:

$F \backslash f$	0	1	2	3	4	...
1	$1/2$	$1/2$				
2	$1/4$	$2/4$	$1/4$			
3	$1/8$	$3/8$	$3/8$	$1/8$		
4	$1/16$	$4/16$	$6/16$	$4/16$	$1/16$	
...						...

Таблица Б-68

(гл. V, § 2)

Распределение частот слов в двух сравниваемых текстах А (Э. Бээкман) и В (Л. Промет) и в "объединенном" тексте А + В. N - объем текста, L - объем словаря, F - частота слова, m_F - число слов с данной частотой.

F	Текст А $N_A = 4998$	Текст В $N_B = 5014$	Текст А + В $N_{A,B} = 10012$
	m_F	m_F	m_F
1	947	872	1819
2	299	228	407
3	120	109	163
4	72	64	89
5	57	55	66
6	38	31	39
7	40	35	42
8	27	28	30
9	21	16	21
10	20	18	20
> 10	125	122	125
Σ	1766 L_A	1578 L_B	2821 $L_{A,B}$

Таблица Б-69
(гл. У, § 2)

Распределение частот слов в объединенном тексте
(на основе произведений Э. Бээкман и Л. Промет)

F	m _F	Под частоты (теоретич. / эмпирич.)										
		0	1	2	3	4	5	6	7	8	9	10
1	1891	$\frac{910}{947}$	$\frac{910}{872}$									
2	407	$\frac{102}{179}$	$\frac{204}{120}$	$\frac{102}{108}$								
3	163	$\frac{20}{54}$	$\frac{61}{35}$	$\frac{61}{31}$	$\frac{20}{43}$							
4	89	$\frac{6}{25}$	$\frac{22}{14}$	$\frac{33}{14}$	$\frac{22}{19}$	$\frac{6}{17}$						
5	66	$\frac{2}{11}$	$\frac{10}{10}$	$\frac{21}{15}$	$\frac{21}{14}$	$\frac{10}{7}$	$\frac{2}{9}$					
6	39	$\frac{1}{8}$	$\frac{4}{6}$	$\frac{9}{5}$	$\frac{12}{7}$	$\frac{9}{3}$	$\frac{4}{9}$	$\frac{1}{1}$				
7	42	$\frac{0}{7}$	$\frac{2}{2}$	$\frac{7}{5}$	$\frac{12}{8}$	$\frac{12}{6}$	$\frac{7}{7}$	$\frac{2}{5}$	$\frac{0}{2}$			
8	30	$\frac{0}{2}$	$\frac{1}{4}$	$\frac{3}{3}$	$\frac{7}{2}$	$\frac{8}{7}$	$\frac{7}{1}$	$\frac{3}{6}$	$\frac{1}{1}$	$\frac{0}{4}$		
9	21	$\frac{0}{5}$	$\frac{0}{0}$	$\frac{2}{2}$	$\frac{3}{3}$	$\frac{5}{5}$	$\frac{5}{5}$	$\frac{3}{5}$	$\frac{2}{1}$	$\frac{0}{1}$	$\frac{0}{0}$	
10	20	$\frac{0}{2}$	$\frac{0}{0}$	$\frac{1}{1}$	$\frac{2}{2}$	$\frac{4}{4}$	$\frac{5}{2}$	$\frac{4}{2}$	$\frac{2}{4}$	$\frac{1}{3}$	$\frac{0}{0}$	$\frac{0}{0}$

Таблица В-70
(гл. 7, § 2)

Вычисление критерия согласия хи-квадрат по формуле $\chi^2 = \sum \frac{(З-Т)^2}{Т}$ при сравнении эмпирических (З) и теоретических (Т) подчастот (на основе произведений Э. Базкиан и Л. Промет); г - число групп, k - число классов, n - число наблюдений.

Группа	Класс	З	Т	З-Т	$(З-Т)^2 : Т$
I (m ₁)	1.	947	910	+37	1369:910 = 1,50
	2.	872	910	-38	1444:910 = 1,58
II (m ₂)	3.	179	102	+77	5929:102 = 58,12
	4.	120	204	-84	7056:204 = 34,58
	5.	108	102	+6	36:102 = 0,35
III (m ₃)	6.	54	20	+34	1156:20 = 57,80
	7.	35	61	-26	676:61 = 11,08
	8.	31	61	-30	900:61 = 14,75
	9.	43	20	+23	529:20 = 26,45
IV (m ₄)	10.	25	6	+19	361:6 = 60,16
	11.	14	22	-8	64:22 = 2,90
	12.	14	33	-19	361:33 = 10,93
	13.	19	22	-3	9:22 = 0,40
	14.	17	6	+11	121:6 = 20,16
V (m ₅)	15.	11	2	+9	81:2 = 40,50
	16.	10	10	0	0
	17.	15	21	-6	36:21 = 1,71
	18.	14	21	-7	49:21 = 2,33
	19.	7	10	-3	9:10 = 0,90
	20.	9	2	+7	49:2 = 24,50
VI (m ₆)	21.	8 } 14 6 }	1 } 5 4 }	+9	81:5 = 16,20
	22.	5	9	-4	16:9 = 1,77
	23.	7	12	-5	25:12 = 2,08
	24.	3	9		
	25.	9 } 10 1 }	4 } 5 1 }	+5	25:5 = 5,00
г = 6		k = 19	n = 2583	2585	$\chi^2 = 395,75$

Опыт № 1. Исходные данные: покрываемость текста словоформами (%)

№ текста	Автор	Р а н г и с л о в о ф о р м								
		i = I	10	50	100	500	1000	1500	2000	2500
1.	Э.Безикман	2,7	11,0	21,4	28,3	50,6	62,6	72,6	82,6	92,6
2.	В.Гросс	2,6	12,0	23,3	30,4	52,4	64,2	74,2	84,1	94,1
3.	А.Хинт	2,8	13,8	26,7	33,7	57,5	70,7	80,7	90,7	100,0
4.	Х.Кийк	3,5	12,9	24,4	31,2	53,6	65,2	75,2	85,2	95,2
5.	Я.Кресс	5,7	13,8	25,0	31,3	52,2	62,7	72,7	82,7	92,8
6.	Н.Куусберг	2,4	14,3	27,0	34,7	57,6	69,7	79,7	89,8	99,8
7.	Н.Промет	3,4	14,4	25,5	32,7	54,4	66,1	76,1	86,1	96,1
8.	В.Саар	3,4	15,3	27,9	35,7	58,9	71,1	81,1	91,2	100,0
9.	Х.Серго	3,1	10,7	20,3	27,2	50,6	62,1	72,2	82,3	92,4
10.	Р.Сирге	1,6	10,1	20,2	27,1	47,2	58,5	68,5	78,6	88,6
11.	М.Траат	3,3	15,4	27,0	33,9	55,5	66,8	76,7	86,6	96,5
12.	Э.Ветемая	2,6	14,3	27,2	35,2	57,6	69,0	79,0	89,0	99,0
13.	А.Каал	3,7	14,1	27,2	34,8	56,5	68,1	78,1	88,2	98,3
14.	Т.Каллас	3,4	13,8	25,4	32,3	55,2	66,6	76,6	86,6	96,6
15.	Ю.Пээгель	2,5	13,1	25,4	32,1	54,5	66,1	76,2	86,4	96,6
16.	Ю.Туулик	2,9	13,1	25,9	33,4	57,1	70,1	80,1	90,2	100,0
17.	А.Валтон	3,6	16,7	30,2	38,0	61,1	72,5	82,5	92,5	100,0
18.	М.Унт	3,7	13,7	27,4	35,1	58,2	70,2	80,2	90,3	100,0
19.	Э.Нийт/Я.Кресс	3,5	11,5	21,4	27,6	48,9	59,9	69,7	79,4	89,2
20.	Ю.Смуун	3,9	11,9	22,6	28,8	50,0	61,4	71,4	81,4	91,4

Опыт № 2. Исходные данные: лексический спектр (в словаре) -
доля словоформ (%) с данной частотой F

№ текста	Ч а с т о т а с л о в о ф о р м											
	$F=1$	2	3	4	5	6	7	8	9	10	11-20	>20
1.	79,12	10,32	4,11	1,95	0,98	0,80	0,66	0,21	0,38	0,21	0,77	0,49
2.	78,70	10,65	4,22	1,93	0,96	0,64	0,43	0,39	0,25	0,22	1,04	0,57
3.	73,29	13,60	4,22	3,21	1,38	0,89	0,81	0,37	0,61	0,04	0,65	0,93
4.	78,74	10,45	3,94	2,01	1,24	0,91	0,37	0,33	0,29	0,18	0,99	0,55
5.	81,65	9,40	2,83	2,17	0,94	0,66	0,28	0,25	0,28	0,11	0,91	0,52
6.	76,00	12,16	4,38	1,83	1,04	0,80	0,88	0,44	0,32	0,36	1,03	0,76
7.	78,28	10,97	4,30	1,74	1,08	0,63	0,30	0,59	0,41	0,22	0,96	0,52
8.	75,20	12,46	4,39	2,05	1,23	0,94	0,53	0,53	0,33	0,33	1,31	0,70
9.	80,18	10,05	3,72	1,32	1,11	0,80	0,73	0,52	0,25	0,17	0,63	0,52
10.	81,74	10,08	3,36	1,47	0,65	0,52	0,39	0,23	0,13	0,29	0,72	0,42
11.	78,61	10,85	3,96	1,91	0,82	0,71	0,75	0,41	0,22	0,19	0,97	0,60
12.	77,70	11,25	3,96	1,53	1,10	0,74	0,47	0,59	0,59	0,19	1,25	0,63
13.	77,80	11,29	4,60	1,24	1,08	0,74	0,50	0,23	0,27	0,43	1,12	0,70
14.	78,57	9,89	4,19	2,21	1,31	0,94	0,49	0,34	0,30	0,22	0,94	0,60
15.	78,64	10,49	3,90	2,17	1,27	0,90	0,38	0,22	0,19	0,19	0,94	0,71
16.	74,07	12,81	5,18	2,09	1,25	0,96	0,80	0,52	0,32	0,12	1,08	0,80
17.	75,73	11,38	5,10	1,98	1,14	0,84	0,59	0,42	0,46	0,34	1,22	0,80
18.	76,12	11,44	4,43	2,42	1,41	0,64	0,56	0,32	0,48	0,24	1,13	0,81
19.	81,56	9,01	3,54	1,80	1,08	0,62	0,59	0,46	0,16	0,07	0,72	0,39
20.	80,60	10,33	3,14	1,77	1,19	0,68	0,34	0,38	0,14	0,24	0,68	0,51

Опыт № 3. Исходные данные: динамика роста словаря
(число разных словоформ при различных объемах текста N)

№ текста	Автор	Объем текста				
		N = 1000	2000	3000	4000	5000
1.	Э.Бэекман	731	1383	1865	2404	2869
2.	В.Гросс	677	1315	1859	2358	2791
3.	А.Хинт	649	1116	1597	2034	2463
4.	Х.Кийк	710	1351	1828	2315	2738
5.	Я.Кросс	723	1315	1914	2382	2861
6.	П.Куусберг	645	1166	1674	2075	2508
7.	Л.Промет	674	1212	1720	2207	2694
8.	В.Саар	633	1128	1700	2045	2439
9.	Х.Серго	734	1326	1885	2416	2876
10.	Р.Сирге	764	1397	2017	2572	3067
11.	М.Траат	689	1235	1722	2226	2656
12.	Э.Ветемаа	680	1208	1734	2162	2552
13.	А.Каал	651	1204	1690	2119	2586
14.	Т.Каллас	663	1226	1733	2179	2668
15.	Ю.Пээгель	690	1224	1700	2223	2669
16.	Ю.Туулик	624	1135	1560	2005	2491
17.	А.Валтон	588	1036	1468	1955	2373
18.	М.Уит	658	1176	1678	2133	2483
19.	Э.Нийт/Я.Кросс	740	1357	1923	2473	2983
20.	Ю.Смуул	732	1361	1917	2473	2929