

SÕNALIIKIDE SAGEDUSEST ILUKIRJANDUSPROOSA AUTORIKÖNES

J. Tuldava, A. Villup

1. SISSEJUHATUS

Sõnavara liigendamine leksikaal-grammatilisteks klassideks ja eri klasside funktsionaalse koormuse määramine on pidevalt püsinud keeleteadlaste huvi keskpunktis. Jättes kõrvale küsimuse sõnade klassifitseerimise eri võimalustest⁺ ja võttes aluseks sõnaliikide traditsioonilise käsitlemise (eesti keele kohta vt. Valgma, Rimmel, 1968), vaatleme käesolevas uurimuses eesti keele sõnaliikide esinemust ühe allkeele - ilukirjandusproosa autorikõne piires. Meetodina kasutame statistilist vaatlust ja analüüsi.

Sõnaliikide statistilise uurimise vajadust ja tähtsust on rõhutanud mitmed tuntud keeleteadlased. Juba 1938. a. kirjutas V.V. Vinogradov: "Sõnatüüpide esinemissagedused on ilmselt erinevad kirja- ja kõnekeele eri stiilides. Täpsed uurimused selles valdkonnas võimaldaksid kindlaks teha grammatilis-struktuurseid ja osalt ka semantilisi erinevusi stiilide vahel. - - - Grammatiliste kategooriate analüüs peab selgitama nende funktsionaalse kaalu eri stiilides." (Виноградов, 1938, 155-156). Hilisemad uurimused on veenvalt näidanud, et sõnaliikide sagedused ("funktsionaalne

⁺ Sõnade liigitamise probleemi kohta lähemalt vt. Жирмунский, 1968; Стеблин-Каменский, 1974. Uusi liigitamise põhimõtteid esitavad näit. O. Sunik (Суник, 1966), I. Revzin (Ревзин, 1967), L. Sumarokova (Сумарокова, 1967, L. Andrejeva (Андреева, 1969). Struktuuraal-lingvistikas on tuntud Ch. Fries'i (1952) distributiivsete sõnaklasside süsteem. Generatiivse semantika seisukohast on sõnaliike vaadelnud J. Lyons (1966), H. Öim (Ойм, 1969) jt. Eesti keele sõnaklasside matemaatilis-lingvistilise käsitlemise (hulgateooria alusel) esitab T. Tobias (Тобиас, 1962), statistilis-kombinatoorselt vaatab eesti keele sõnaklasse H. Pak (Пак, 1965).

kaal") kujutavad endast tähtsaid kvantitatiivseid tunnuseid, mis võimaldavad eristada ja diagnoosida nii funktsionaalseid stiile ja žanre kui ka individuaalseid autoristiile (vt. näit. Ключкова, 1968; Тищенко, 1970; Кожина, 1972; Головин, 1974). Võrreldes näiteks tähtsamate käändsõnaliikide esinemissagedusi tekstis vene keele eri funktsionaalsetes stiilides (Ключкова, 1968):

	Kõnekeel	Ilukirjandus (autorikõne)	Teaduslik kirjandus
Nimisõna	13,0 %	28,7 %	36,9 %
Omadussõna	3,5 %	7,9 %	16,4 %
Asesõna	17,5 %	10,2 %	6,2 %

Ka tegusõna osakaal varieerub tunduvalt vastavalt funktsionaalsele stiilile, näit. ukraina keele kohta tehtud uurimuse andmeil (Тищенко, 1970) on draamatekstis tegusõnu keskmiselt 22,0 %, ilukirjanduslikus proosas 19,7 %, teaduslik-tehnilises kirjanduses 13,9 % ja ühiskondlik-politiilise sisuga tekstides 11,1 %. Analoogilisi andmeid sõnaliigisageduste varieeruvusest eri stiilides on saadud ka paljude teiste keelte materjali põhjal (vt. näit. Kelemen, 1964; Mistrík, 1969; Zsilka, 1973).

Eri funktsionaalsete stiilide (allkeelte, žanride) piires on omakorda võimalik täheldada suuremat või väiksemat sõnaliigisageduste hajuvust keskvärtuse ümber, mis peegeldab individuaalseid erinevusi sõnaliikide kasutamisel. Üldreeglina on individuaalsed sagedusnäitajad määratud žanrilise kuuluvusega ja jäävad kindlatesse variatsioonipiiridesse. Vastavate meetoditega on aga võimalik kindlaks teha statistiliselt olulisi hälbeid "normist" või olulisi erinevusi autorite vahel. Individuaalsed erinevused on osaliselt tingitud välistest asjaoludest, näit. teema valikustesisulaadist (jutustus, kirjeldus jne.), kuid suurel määral olenevad need ka autori isiksusest ja stiilimaneerist. Seetõttu on võimalik sõnaliikide sagedusi arvesse võtta autorsuse kindlakstegemisel (vt. näit. Таршинская, 1969). Tuleb muidugi silmas pidada, et autori stiil võib aja jooksul muutuda, nii nagu võib muutuda terve žanri üldpilt: näitena nimetatagu olulisi nihkeid vene ja inglise ajalehekeele sõnaliigisagedustes (Сиротинина, 1968; Турыгина, 1965).

Puhtstilistiliste probleemide lahendamise kõrval saab sõnaliikide statistilist analüüsi edukalt rakendada keelte tüpoloogilisel uurimisel (Андреев, 1967, 76 jj.), tekstide automaattõõtluse huvides (Белоголов, 1964; Раскина, Чепиго, 1970) ja mitmesuguste teoreetiliste lingvostatistiliste probleemide lahendamisel (Фрумкина, 1962; Клявнина, 1969; Якубайтис, 1969; Бектаев, Лукьяненок, 1971).

Sõnaliikide statistilist analüüsi on varem teostatud ka eesti keele kohta. Võib nimetada J. Valge (1970, 1972) uurimusi sõnaliikide sagedusest eri funktsionaalsetes stiilides, M. Tiitsi, L. Veileri, K. Enniko ja S. Meimani töid sõnaliikide sagedusest A. H. Tammsaare romaani "Tõde ja õigus" I köite autorikõnes (Tiits, Veiler, 1974; Enniko, Meiman, 1975), A. Villupi (1972 ja 1974), E. Keskküla (1972) ning S. Kõiva ja E. Raadiku (1974) uurimusi adverbi esinemissagedusest eesti ilukirjandusproosas. Sõnaliikide kvantitatiivset esinemust rahvalaulus on vaadelnud E. Laugaste (1969).

Käesoleva töö eesmärgiks on lähemalt analüüsida sõnaliikide sagedusi kindlalt piiritletud allkeeles - tänapäeva ilukirjandusproosa autorikõnes, kusjuures uuritakse allkeele üldisi omadusi sõnaliikide kasutamise seisukohast ning individuaalseid erinevusi autorite vahel. Artiklis käsitletakse küsimust mitmest eri aspektist: sõnaliikide sagedusi vaadeldakse nii tekstis kui ka vastava teksti juurde kuuluvast sõnastikus, määratakse kindlaks sõnaliikide informatiivne koormus, võrreldakse tekste sõnaliigisageduste omavaheliste suhete (nn. sõnaliigi-indeksite) alusel ja arvutatakse sõnaliikidevahelised korrelatsioonid. Selline kompleksne analüüs teenib esmajoones stiliiuurimisel ja praktilisi eesmärgi, kusjuures tähelepanu on pööratud ka sõnaliikide statistilise uurimise meetodika küsimustele (materjali valik ja representatiivsus, interpretatsioon jne.).

Vaatlusmaterjalina kasutatakse 10 valimit (väljavõtet) 5000 sõnet⁺ kümne autori teosest, mis on ilmunud pärast 1960. a. Valimi üldmaht on seega 50 000 sõnet. Vaadeldavad teosed ja nende lühendid on järgmised:

⁺ Sõnavarastatistika terminite ja mõistete kohta vt. Tuldava, 1971.

1. A. Beekman, Kartulikujused. Tallinn, 1968. - A.B.
2. V. Gross, Pinginaabrid. Tallinn, 1965. - V.G.
3. A. Hint, Tuuline rand IV. Tallinn, 1966. - A.H.
4. H. Kiik, Tondiõõmaja. Tallinn, 1970. - H.K.
5. J. Kross, Kolme katku vahel I. Tallinn, 1970. - J.K.
6. L. Promet, Primavera. Tallinn, 1971. - L.P.
7. V. Saar, Ukuaru. Tallinn, 1969. - V.S.
8. H. Sergo, Põgenike laev. Tallinn, 1966. - H.S.
9. M. Traat, Tants aurukatla ümber. Tallinn, 1971. - M.T.
10. E. Vetemaa, Väike romaaniraamat. Tallinn, 1968. - E.V.

Sõnaliikide sagedused teksti ja sõnastiku tasandil on arvutatud tekstidele koostatud sagedussõnastike põhjal, ülejäänud arvutustööd teostati TRÜ arvutuskeskuses.

2. MATERJALI REPRESENTATIIVSUS

Keelenähtuste statistilisel uurimisel kerkib alati probleem materjali representatiivsusest ja tulemuste usaldatavusest. Antud juhul seisneb küsimus esiteks selles, kas valimid iga autori teosest on küllalt suured, et pidada neid statistiliselt usaldatavaks, ja teiseks, kas vaadeldavad valimid kokku representeerivad küllaldase usaldatavusega vastavat üldkogumit, s. t. tänapäeva eesti ilukirjandusproosa autorikõnet.

Valimite representatiivsuse üle otsustamine on teatavasti seotud nii kvalitatiivsete kui ka kvantitatiivsete hinnangutega (vrd. Tiit, 1971, 74 jj.). Materjali sisuline valik kuulub esmajoones selle konkreetse teadusharu kompetentsi, mille valdkonnas uurimine toimub. Sellest seisukohast lähtudes peame eesti kaasaegse ilukirjandusproosa tunnustatud autorite teoseid täiesti vastavaks uurimuse eesmärkidele. Väljavallitud kümme teost on võetud kompetentsete kirjandusteadlaste abiga koostatud nimestikust juhusliku valiku alusel. Seega on valiku printsiip õige ka statistiliselt, nimelt igale üldkogumi indiviidile on tagatud võrdne tõenäosus sattuda valimisse (Tiit, 1971, 76). Üldkogumi määratlemisel oleme lähtunud keelelise materjali homogeensuse nõudest ja seepärast on vaatluse all antud etapil ainult ilu-

kirjandusproosa a u t o r i k õ n e (välja on jäetud tegelaskõne, mille leksikaalsed, grammatilised jm. omadused erinevad tunduvalt autorikõnest; tegelaskõnet on mõtet uurida eraldi).

Järgnevalt peame kontrollima statistilise vaatluse teel saadud andmete usaldatavust. Selleks rakendame keelestatistikas kasutatavaid matemaatilise statistika meetodeid. Kõigepealt vaatleme üksikvalimite ja seejärel valimite seeria (üldvalimi) statistilist usaldatavust.

Valimid igast vaadeldavast teosest koosnevad meie katse puhul viiest osavalimist à 1000 sõnet. Osavalimid on võetud juhuväljavõtu printsiibil teose eri osadest terviklike lõikudena (välja jättes tegelaskõne). Selliseid 1000-sõnelisi tekstilõike nimetatakse sõnavarastatistikas "standardvalimiteks" ja need võetakse tavaliselt aluseks igasuguste sõnavarastatistiliste uurimuste läbiviimisel (vrd. Андреев, 1967, 86; Лукьяненко, 1974, 326; Якубайтис, Стурите, 1974, 45). Statistilise vaatluse teel saadud andmeid - sõnaliikide suhtelisi sagedusi - võib iga teose puhul vaadelda kui keskmisi sagedusi 1000-sõneliste tekstilõikude kohta. Statistilise usaldatavuse kontrollimisel tuleb seepärast kindlaks teha sageduste hajuvus viie tekstilõigu (osavalimi) ulatuses. Et statistilised hajuvushinnangud osutusid kõigil autoritel väga lähedasteks, toome näitena andmed ainult ühe autori - M. Traadi kohta (vt. tabel 1). Tabelis esitatakse sõnaliikide absoluutsagedused eri osavalimites (m_1), sageduste summa ($\sum m_1$), keskmine suhteline sagedus (\bar{p}) protsentides ning hajuvushinnangud 95%-lisel usaldusnivool: keskvärtuse absoluutne viga ehk nn. piirviga ($\epsilon_{\bar{p}}$), keskvärtuse usalduspiirid ($\bar{p} \pm \epsilon_{\bar{p}}$) ja suhteline viga ($\epsilon_{\bar{p}}^{\%}$) protsentides. Peale selle hinnati iga valimirea homogeensust hii-ruudu (χ^2) abil. (Arvutuste kohta lähemalt vt. Tuldava, 1969 ja 1970)

Nagu tabelist nähtub, on vaadeldavas teoses nimisõna keskmine suhteline sagedus 30,7 % piirveaga $\pm 1,7$. Usalduspiirid on seega 29,0 ... 32,4 %, kusjuures keskvärtuse suhteline viga moodustab 5,5 %. Hii-ruudu väärtus (2,38) näitab, et sageduste kõikumine eri osavalimeis on juhuslikku laadi ja et kogu valimit võib nimisõnade esinemuse suhtes pidada statistiliselt homogeenseks (hii-ruudu kriitiline väärtus 5%-lisel olulisusnivool on antud katse puhul 9,49).

T a b e l 1

Sõnaliikide sagedused tekstis ja nende hajuvushinnangud M. Traadi teose
"Tants aurukatla ümber" põhjal (5 osavalimit à 1000 sõnet)

Sõnaliik	Absoluutsagedused osavalimites					Kokku ($\sum m_i$)	Kesk- mine \bar{p} (%)	Piir- viga $\xi_{\bar{p}}$ (%)	Usalduspiirid (%)	Suhte- viga $\delta_{\bar{p}}$ (%)	χ^2
	m_1	m_2	m_3	m_4	m_5						
Nimisõna	297	305	324	319	293	1538	30,7	$\pm 1,7$	29,0 ... 32,4	5,5	2,38
Omadussõna	72	57	64	60	55	308	6,2	$\pm 0,8$	5,4 ... 7,0	12,9	2,94
Arvsõna	9	3	13	12	12	49	1,0	$\pm 0,5$	0,5 ... 1,5	50,0	6,84
Aesesõna	120	125	121	134	137	637	12,7	$\pm 1,0$	11,7 ... 13,7	7,9	1,86
Tegusõna	221	228	223	228	233	1133	22,6	$\pm 0,6$	22,0 ... 23,2	2,7	0,39
Määrsõna	166	169	149	154	155	793	15,9	$\pm 1,6$	6,8 ... 10,0	6,9	7,83
Kaassõna	29	28	32	16	23	128	2,6	$\pm 0,8$	1,8 ... 3,4	30,8	6,13
Sidesõna	86	79	73	76	89	403	8,1	$\pm 0,8$	7,3 ... 8,9	9,9	2,25
Hüüdsõna	-	6	1	1	3	11	0,2	$\pm 0,1$	0,1 ... 0,3	50,0	-
Kokku	1000	1000	1000	1000	1000	5000	100,0		-		

Seega on nimisõna esinemissagedus vaadeldavas teoses küllaltki stabiilne ja selle keskvärtus usaldusväärne, mida kinnitab ka suhtelise vea madal värtus (5,5 %).⁺ Vaadeldes teiste sõnaliikide sageduste hajuvushinnanguid tabeli 1 põhjal, võime samuti konstateerida tulemuste suhteliselt head usaldatavust ja sageduste stabiilsust antud valimi ulatuses. Ainult väikese sagedusega arvsõnad, kaassõnad ja hüüdsõnad osutavad suuremat hajuvust, kuid see mõjustab vähe sõnaliikide sagedusjaotust tervikuna. Võib teha järelduse, et ilukirjandusproosa autorikõne puhul on 5000-sõnelised valimid (jaotatuna osavalimitesse 5 x 1000) sõnaliikide statistilise vaatluse jaoks küllalt representatiivsed.

Vaadeldes sõnaliigisageduste jaotumust üldvalimis, s.o. kümne autori teose ulatuses (tabel 2), võime nentida, et suuremal osal juhtudest on sõnaliikide keskmised suhtelised sagedused (\bar{p}) statistiliselt usaldatavad. Suhteline viga (95%-lisel usaldusnivool) ei ületa põhiliste sõnaliikide puhul 10 %, asesõna ja kaassõna sageduste viga on vastavalt 15,8 ja 12,9 ning ainult arvsõna ja hüüdsõna puhul tõuseb viga üle 20 %. Viimati nimetatud kaks sõnaliiki moodustavad aga kokku ainult ca 1 % ja nende hajuvus ei saa oluliselt mõjustada teiste sõnaliikide sagedusi. Hii-ruudu väärtused näitavad, et mõningate sõnaliikide puhul (nimisõna, asesõna ja tegusõna) on sageduste jaotumus eri valimite kaupa ebaühtlasem, kui võiks oodata homogeense valimirea korral (hii-ruudu kriitiline väärtus 5%-lisel olulisusnivool on antud juhul 16,9). See tähendab, et nimetatud sõnaliikide sagedused on eriti arvestatavad individuaalseid stile eristavate faktoritena.

Kokkuvõttes võib väita, et meie katses osalevad kümme valimit üldmahuga 50 000 sõnet representeerivad küllaldase statistilise usaldatavusega vaadeldavat allkeelt - tänapäeva ilukirjandusproosa autorikõnet.

⁺ Teatavasti peetakse keelestatistilistes töödes küllaldaseks, kui suhteline viga on 10-20 %, kusjuures sõnavara-statistilistes uurimustes on lubatud suhtelise vea suurus kuni 30 % (vt. Бектаев, Пиотровский, 1974, 189).

Tabel 2

Sõnaliikide suhtelised sagedused tekstis ja nende hajuvushinnangud ilukirjandus-
proosa autorikõne põhjal (10 valimit à 5000 sõnet)

Sõnaliik	Suhteline sagedus (%)	A u t o r i d										Keskmine \bar{p} (%)	Piirviga $\epsilon_{\bar{p}}$ (%)	Usalduspiirid (%)	Suhteline viga $\delta_{\bar{p}}$ (%)	χ^2^*
		A.B.	V.G.	A.H.	H.K.	J.K.	L.P.	V.S.	H.S.	M.T.	A.V.					
Nimisõna	35,9	30,4	32,5	34,8	32,8	30,4	26,8	36,8	30,7	26,8	31,7	$\pm 2,6$	29,1 ... 34,3	8,2	36,6 [■]	
Omadussõna	5,4	7,4	5,6	5,2	7,4	5,5	4,9	5,8	6,2	6,3	6,0	$\pm 0,6$	5,4 ... 6,6	10,0	11,2	
Arvsõna	0,6	1,4	1,3	1,3	1,7	0,9	0,9	1,4	1,0	0,5	1,1	$\pm 0,3$	0,8 ... 1,4	27,3	12,0	
Assesõna	8,3	11,7	12,9	9,3	9,1	13,8	13,8	7,9	12,7	14,4	11,4	$\pm 1,8$	9,6 ... 13,2	15,8	49,3 [■]	
Tegusõna	25,0	22,2	19,1	24,2	18,3	24,0	23,5	22,0	22,6	24,0	22,5	$\pm 1,6$	20,9 ... 24,1	7,1	19,7 [■]	
Määrsõna	13,9	17,4	16,6	15,0	16,5	13,9	17,3	14,8	15,9	16,3	15,8	$\pm 0,9$	14,9 ... 16,7	5,7	9,5	
Kaassõna	3,7	2,5	3,5	2,4	3,9	2,9	3,7	3,6	2,6	2,6	3,1	$\pm 0,4$	2,7 ... 3,5	12,9	10,0	
Sidesõna	7,2	6,8	8,3	7,7	10,1	8,5	9,5	7,6	8,1	9,0	8,2	$\pm 0,7$	7,5 ... 8,9	8,4	11,5	
Hüüdsõna	-	0,2	0,2	0,1	0,2	0,1	0,4	0,1	0,2	0,1	0,2	$\pm 0,1$	0,1 ... 0,3	50,0	-	
Kokku (%)	100	100	100	100	100	100	100	100	100	100	100	-	-	-	-	

* Märkus: χ^2 väärtused on arvatatud sõnaliikide absoluutsageduste alusel.

3. SÖNALIIKIDE SAGEDUSED TEKSTIS

3.1. Üldised märkused

Tabelis 2 esitatakse sõnaliikide sagedused t e k s - t i s . Statistilise vaatluse tulemused näitavad, et tänapäeva eesti ilukirjandusproosa autorikõnes esineb kõige rohkem nimisõnu (keskmiselt 31,7 %), seejärel tegusõnu (22,5 %), määrsõnu (5,8 %) ning asesõnu (11,4 %). Selline on sagedusjärjestus ka kõigi vaadeldavate autorite teostes, välja arvatud J. Kross, kellel on sidesõnu rohkem kui asesõnu. Erinevus nimetatud sõnaliikide keskmiste sageduste vahel on statistiliselt kindlustatud (võrreldagu keskmiste suhteliste sageduste usalduspiire, näit. määrsõnal 14,9 16,7 % ja asesõnal 9,6 13,2, s. t. usalduspiirid ei kattu, vt. tabel 2). Võib aga täheldada seda, et üksikutel autoritel on erinevused mõnede sõnaliikide sageduste vahel väiksemad kui teistel autoritel. Nii näiteks on V. Saarel ja A. Valtonil tegusõnade osakaal tekstis lähedane nimisõnade osakaalule ja vastavaid sagedusi lahutab teineteisest vaid mõni protsent. Määrsõnu esineb üldvalimis oluliselt rohkem kui asesõnu, kuid erandina kasutab L. Promet mõlemaid sõnaliike ligikaudu ühepalju (määrsõnu 13,9 ja asesõnu 13,8 %).

Nimetatud neljale kõige kasutatavamale sõnaliigile järgnevad sagedusjärjestuses sidesõna (8,3 %), omadussõna (6,0 %) ja kaassõna (3,1 %, sellest eessõnu 0,4 ja tagasõnu 2,7 %). Järjestus on samasugune kõigil autoritel, välja arvatud V. Gross, kes kasutab valimi andmeil omadussõnu rohkem kui sidesõnu (vastavalt 7,4 ja 6,8 %). Sõnaliikide sagedusjärjestuse lõpus paiknevad arvsõna (1,1 %) ja hüüdsõna (0,1 %).

Ilukirjandusproosa autorikõne sõnaliigisagedusi on mõtet võrrelda teiste eesti keele kohta teostatud statistiliste uurimustega. Tabelis 3 esitamise lisaks meie katse andmetele sõnaliikide tekstisagedused A.H. Tammsaare romaani "Tõde ja õigus" I köite autorikõnes (valimi maht 60 000 sõnet; uurimuse teostasid A. Villup, K. Enniko ja S. Meiman), ajalehekeele sporditeadetes ja TASS-i sõnumites (valimi üldmaht 10 000 sõnet, vt. Valge, 1972) ja rahvalaulus (Lau-gaste, 1959).

T a b e l 3

Sõnaliikide sagedused (%) tekstis eesti keele erinevate allkeelte põhjal

Sõnaliik	Ilukirjan- dusproosa, autorikõne	"Tõde ja õigus" I, autorikõne	A j a l e h t		Rahva- laul
			Spordi- teated	TASS-i sõnumid	
Nimisõna	31,7	30,1	34,2	49,1	53,2
Tegusõna	22,5	23,6	22,8	19,6	24,9
Määrsõna	15,8	16,5	14,5	5,9	6,3
Asesõna	11,4	11,8	9,1	6,3	2,5
Sidesõna	8,3	10,3	6,2	6,2	0,6
Omadussõna	6,0	4,1	7,7	7,7	9,5
Kaassõna	3,1	2,8	2,4	3,4	1,8
Arvsõna	1,1	0,8	3,1	1,8	0,9
Hüüdsõna	0,1	0,0	0,0	0,0	0,3
Kokku	100,0	100,0	100,0	100,0	100,0

Võrreldes kaasaegse ilukirjandusproosa sõnaliigisagedusi A.H. Tammsaare omadega, võime nentida, et statistilised näitajad on väga lähedased. See tähendab, et tänapäeva keskmine vastab üldiselt A.H. Tammsaare sõnaliikide kasutusele. Ainult omadussõna ja sidesõna puhul võib täheldada statistiliselt olulist vahet: A.H. Tammsaare kasutab nimelt omadussõnu vähem ja sidesõnu rohkem kui kaasaegsed autorid keskmiselt. Omadussõnade protsent (4,1 %) on A. H. Tammsaarel madalam kui ühelgi teisel vaadeldaval autoril. Sidesõnu on palju ka J. Krossil (10,1 %), V. Saarel (9,5 %) ja A. Valtonil (9,0 %), kuid need autorid erinevad A. H. Tammsaarest selle poolest, et näit. J. Krossil on omadussõnade protsent kõrge (7,4 %) ja V. Saarel ning A. Valtonil on nimisõnade osakaal tekstis kõige väiksem vaadeldavate autorite seas (vastavalt 26,0 ja 26,8 %), kuna aga A.H. Tammsaarel on nimisõnu 30,1 %.

Vaadeldes meie katse taustal ajalehetekste, näeme, et sporditeated on ilukirjandusproosa autorikõnele küllaltki lähedased, kuid siseriigi- ja välissõnumid erinevad ilukirjandusproosast oluliselt ja peamiselt selle poolest, et sõnumites on nimisõnade osakaal väga suur (49,1 %), tegusõnade ning asesõnade osakaal aga suhteliselt väike (vastavalt 19,6 ja 6,3 %).

Sõnaliigisagedused rahvalaulus peegeldavad žanri omapära, nimelt võib täheldada nimisõnade rohkust (53,2 %), kusjuures tegusõnade osatähtsus on samuti suur (24,9 %). Rahvalaulu eripäraks võib pidada ka omadussõnade suurt sa-

gedust (9,5 %). Paistab silma asesõnade, sidesõnade ja kaassõnade vähene kasutamine, võrreldes ilukirjandusproosa autorikõnega.

Huvitav on kõrvutada eesti keele andmeid vastavate uurimuste tulemustega teistest keeltest. Tabelis 4 esitame sõnaliikide sagedused viie keele ilukirjanduslikus tekstis: soome (P. Saukkoneni andmeil), ungari (Zsilka, 1973), läti (Якубайтис, 1974), vene (Ключкова, 1968) ja ukraina keel (Тищенко, 1970).

T a b e l 4

Sõnaliikide esinemissagedus (protsentides)
eri keelte ilukirjandusproosa tekstis⁺

Sõnaliik	Eesti keel	Soome keel	Ungari keel	Läti keel	Vene keel	Ukraina keel
Nimisõna	31,7	24,8	28,8	30,1	28,7	29,2
Omadussõna	6,0	9,1	6,8	5,5	7,9	6,8
Arvsõna	1,1	0,9	2,5	1,3	1,2	1,1
Asesõna	11,4	12,1	5,8	12,2	10,2	9,0
Tegusõna	22,5	29,0	21,6	23,0	18,3	19,7
Määrsõna	15,8	11,4	8,8	10,0	6,0	6,2
Kaassõna	3,1	2,3	24,6	5,9	12,2	12,5
Sidesõna	8,2	10,2		8,0	8,5	9,1
Hüüdsõna	0,2	0,2	0,1	0,3	0,0	0,0
Mitmesugused partiklid	-	-	-	3,7	7,0	6,3
Kokku	100,0	100,0	100,0	100,0	100,0	100,0

⁺ Eesti ja vene keele andmed pärinevad ilukirjandusproosa autorikõnest, soome keele näitajad on arvatud kõige suurema sagedusega sõnade põhjal (mis katavad ilukirjandusproosa tekstist umbes 80 %), teiste keelte sagedused baseeruvad ilukirjandusproosa tekstidel (autorikõne ja tegelaskõne koos).

3.2. Üksikute sõnaliikide sagedused tekstis

Meie vaatluse andmeil (tabel 2) on nimisõna keskmine esinemissagedus ilukirjandusproosa autorikõnes 31,7 %. Matemaatilise statistika abil arvatud usalduspiirid näitavad, et hulgaliste katsete korral langeks keskmine sagedus 95-l juhul sajast piiridesse 29,1...34,3%.

Neid piire võime kokkuleppeliselt vaadelda antud allkee-
le "normina", pidades kõiki individuaalseid sagedusi, mis
langevad neisse piiridesse, stilistiliselt neutraalseteks
(mittemarkeerituteks) ja käsitades usalduspiiridest välju-
vaid sagedusi oluliste individuaalsete hälvetena allkeele
keskmisest.

Nimisõnade kasutamise seisukohast ületavad "normi"
H. Sergo (36,8 %), A. Beekman (35,9 %) ja H. Kiik (34,8 %),
"alla normi" on nimisõna sagedus A. Valtonil (26,8) ja
V. Saarel (26,0).

Nimisõnade rohkust tekstis peetakse lingvostilistikas
sageli "nominaalse" stiili tundemärgiks. Kuid niisuguse ot-
sustuse tegemisel peame olema ettevaatlikud, kuna kvantita-
tiivne näitaja üksikult võttes ei võimalda alati adekvaat-
selt hinnata stiili omadusi. On vaja teostada ka kvalita-
tiivne analüüs ja vaadelda nimisõnade semantilisi jm. eri-
jooni tekstides. Kvantitatiivse analüüsi täiustamiseks on
otstarbekohane kindlaks teha tunnuste omavahelised suhted,
näiteks võib antud juhul uurida, milliste sõnaliigisagedus-
te arvel on paisutatud nimisõnade sagedus ja milline on
tekstisageduste ja vastava sõnastiku sageduste suhe, s. t.
tuleb kindlaks teha, kas tekstis esinevad nimisõnad on suu-
re korduvusega või on tegemist nimisõnade mitmekesisusega.
Neid küsimusi vaatleme artikli järgnevates osades.

Nimisõnade sagedusi võib käsitleda ka diferentseeri-
tult, näiteks jaotades nimisõnad kahte peamisse semantilise
rühma - üldnimed ja pärisnimed. Ilukirjandusproosa auto-
rikõnes jaotuvad nende sagedused järgmiselt:

üldnimed	26,9 % (85 % nimisõnade üldarvust)
pärisnimed	4,8 % (15 % " ")
kokku nimi- sõnu	31,7 % (100 %)

Üllatab pärisnimede küllaltki suur protsent tekstis.
Üksikute autorite järgi kõigub pärisnimede sagedus 3,3 %
(V. Gross) ja 6,0 % (H. Sergo) vahel. Võrdluseks toome and-
med vene keele kohta: M. Solohhovi romaanis "Ülesküntud
uudismaa" registreeriti tekstis 3,8 % pärisnimesid, mis
moodustab 16,1 % nimisõnade üldarvust (Лятина, 1968). Läti
sagedussõnastiku andmeil on ilukirjanduslikes tekstides pä-
risnimesid 2,9 %, kuid valimisse on kaasa arvatud ka draa-

ma- ja luuletekstid, kus pärisnimesid esineb suhteliselt vähe. Registreeritud on ka pärisnimede sagedused läti keele teistes allkeeltes, nimelt ajalehekeeles - 5,8 % ja teaduslik-tehnilistes tekstides - 0,4 % (Latviešu valodas biežuma vārdnīca, 1972).

Tegusõna katab keskmiselt 22,5 % tekstist. Keskmine sagedus on arvatud piirveaga \pm 1,6, seega on usalduspiirid 20,9 ... 24,1 %. Nagu nähtub tabelist 2, on tinglikust ülempiirist suuremad väärtused A. Beekmanil (25 %) ja H. Kiigel (24,2 %). Ülempiirile lähenevad ka L. Prometi ja A. Valtoni näitajad (mõlemal 24,0 %). Huvitav on märkida, et A. Beekmanil on nii tegusõnade kui ka nimisõnade protsent oluliselt suurem allkeele keskmistest sagedustest. Kui tegusõnade sagedust seostada stiili "verbaalsusega", siis peab antud juhul nentima, et kirjaniku stiil on üheaegselt nii nominaalne kui verbaalne. Nähtavasti on aga otstarbekam nimetatud mõisteid defineerida nende omavahelise suhtena või seoses teiste sõnaliikide sagedusega (vt. allpool p. 5).

Alla "normi" kasutavad tegusõnu A. Hint (19,1 %) ja J. Kross (18,3 %). Ülejäänud autorid jäävad tegusõnade esinemissageduse poolest allkeele keskmise piiridesse.

Ka tegusõnade esinemust võib vaadelda grammatilise jm. liigituse seisukohalt. Jaotades tegusõnad käändelisteks ja pöördelisteks vormideks, saame järgmised tekstisagedused:

teigusõna käändelised vormid	8,2 % (36 % tegusõnade üldarvust)
-"- pöördelised vormid	14,3 % (64 % tegusõnade üldarvust)
<hr/>	
kokku tegusõnu	22,5 % (100 %)

Autorite seas paistab silma V. Saar tegusõna käändeliste vormide rohkuse poolest (10,4 %). Oluliselt alla keskmise esineb käändelisi vorme H. Kiigel (6,5 %) ja J. Krossil (6,4 %). Käändeliste vormide rohkus võib olla tingitud erinevatest asjaoludest (liitaegade kasutamine, seos modaalverbidega, kesksõnade esinemine täiendina jm.). Edaspidine analüüs peab välja selgitama nii ühised kui ka erinevad jooned autorite keelekasutuses. Toome siinkohal näiteks mõned tüüpilised laused V. Saare teosest "Ukuaru" (pöö-

ratagu tähelepanu tegusõnade käändeliste vormide kasutamisele):

Polnud mulle antud mahti aastaid lugeda ega arvutada, pole seda vajagi olnud, sest mu aastad on seisnud elavaina mu ümber, sirgunud, kasvanud ja nõudnud ikka rohkem ja rohkem. (lk. 8)

Niisuguses meeleolus võib üksnes mööda metsa hulkuda, sihitult hulkuda, et rahuneda ja koguda jõudu, või siis lasta kohiseda mändidel ja voolata mälestustejõel ... (lk. 11)

Minnal pole midagi maailma ja inimeste vastu, tal pole midagi varjata, peita ega häbeneda, ent ometi on tal raske tühja mannerguga läbi metsa minna; ta ei lähe viima, ta läheb tahtma ja tooma; ta ei lähe andma, vaid nuruma ja saa-ma. (lk. 70)

Kes teab, mitu põlvkonda siin nabasid paigast venitada des on kangutanud ja ubinud kivimürakaid, püüdnud neid aedadeks-aunadeks kokku veeretada. (lk. 71)

... vaatamata, mis ma panen või panemata jätan, läks ta eidega jumalaga jätma. (lk. 7)

Tegusõna pöördelisi vorme esineb eriti palju H. Kiigel (17,7 % tekstist ja 73,1 % tegusõnade üldarvust). See on tingitud peamiselt lihtaegade domineerimisest H. Kiige verbikasutuses. Keskmisest oluliselt rohkem kasutavad verbi pöördelisi vorme ka A. Beekman (16,3 %), L. Promet ja A. Valton (mõlemal 16,0 %). Kui vaadelda pöördeliste vormide protsenti tegusõnade üldarvust, siis võib esile tõsta M. Traati, kes pöördeliste vormide sageduselt (68,6 %) järgneb H. Kiigele. Teksti katmuse seisukohast jääb aga M. Traadi nimetatud näitaja keskmise piiridesse (15,5 %).

Süntaktilise funktsiooni järgi liigitatakse tegusõnad tavaliselt iseseisvateks, abi- ja modaalverbideks (Valgma, Rimmel, 1968, 114). Pakub huvi tegusõna olema, mis võib esineda abiverbina, köitmena ja iseseisvas tähenduses. Meie vaatluse andmeil on tegusõna olema keskmine sagedus tekstis 5,4 %, iseseisvas tähenduses aga 2,1 % (mis moodustab ligikaudu 40 % tegusõna olema üldsagedusest). Köitme ja abiverbi funktsioonis esineb olema ligikaudu võrdse sagedusega -

vastavalt 1,7 ja 1,6 % (seega mõlemal juhul umbes 30 % üldsagedusest). Tegusõna olema osatähtsust tekstis tuleb hinnata ka selle järgi, et tekstis esinevate verbide üldsagedusest moodustab olema keskmiselt 25 %, s. t. iga neljas verb tekstis on olema (kõige sagedamini vormides on, oli, olnud).

Käändeliste ja pöördeliste vormide jaotumuse seisukohast on tegusõna olema vastavad näitajad tekstis 0,7 ja 4,7 %, s. o. 13 ja 87 % verbi olema sagedusest. Selles suhtes erineb tegusõna olema jaotumus tunduvalt verbide üldjaotumusest, nimelt on käändeliste ja pöördeliste vormide sagedused verbi kogu esinemuse taustal vastavalt 36 ja 64 % (vt. eespool). Tegusõna olema eripäraks on seega pöördeliste vormide suur ülekaal tekstis.

O m a d u s s õ n a sageduste kõikumus ilukirjandusproosa tekstis ei ole kuigi suur: keskmiselt 5 ja 7,5% vahel. Meie katse andmeil on kõige suuremad sagedused V. Grossil ja J. Krossil - mõlemal 7,4 %. Suhteliselt vähe kasutavad omadussõnu V. Saar (4,9 %) ja H. Kiik (5,2 %). Valimi põhjal arvutatud keskmine suhteline sagedus on 6,0% piirveaga $\pm 0,6$. Omadussõnade sageduse järgi tekstis on sagedeli hinnatud stiili "kvalitatiivsust", s. t. autori kalduvust kasutada jutustuses ja kirjelduses kvalitatiivset hinnangut. Peab aga silmas pidama, et just omadussõnade puhul osutub eriti vajalikuks kvantitatiivset uurimist seostada sisulise analüüsiga, sest omadussõnade kasutamine (mis on suhteliselt vähe seotud lause süntaktilise struktuuriga) peegeldab eriti ilmekalt autorite individuaalseid erinevusi (vrd. Шайкевич, 1968; Kraus, 1972). Omadussõnade statistilis-lingvistiline analüüs väärrib kahtlemata eri uurimust.

Kolme vaadeldud sõnaliiki - nimisõnu, tegusõnu ja omadussõnu - võib pidada ilukirjandusliku teose "võtmesõnadeks", kuna need sõnaliigid kannavad peamist i n f o r m a t i i v s e t koormust ilukirjanduses (Гальперин, 1974, 128; vt. ka allpool, p. 4.2.). Nende kogusagedus ilukirjandusproosa autorikõnes ületab 70 %. Stiiliuurimise seisukohast pakuvad huvi ka ülejäänud sõnaliigid, eriti m ä ä r s õ n a d, mis katavad tekstist meie katse andmeil keskmiselt 15,8 % (usalduspiirid 14,9 ... 16,7 %). Kaks kirjanikku - V. Gross ja V. Saar - paistavad silma eriti

kõrge adverbiprotsendiga (vastavalt 17,4 ja 17,3 %), kuna aga A. Beekman ja L. Promet kasutavad määrsõnu oluliselt vähem allkeele keskmisest (mõlemad 13,9 %). Määrsõnad liigitatakse eesti keeles teatavasti kolme liiki: iseseisvad, rõhu- ja abimäärsõnad. Nende sagedusjaotumus mele valimis on järgmine:

iseseisev määrsõna	8,2 % (52 % määrsõnade üldarvust)
rõhumäärsõna	5,6 % (35 % " ")
abimäärsõna	2,0 % (13 % " ")
määrsõnu kokku	15,8 % (100 %)

Iseseisvaid määrsõnu esineb kõige rohkem V. Saarel (9,4 %, kusjuures iseseisvad määrsõnad moodustavad tema poolt kasutatud määrsõnade üldarvust 54,4 %). V. Grossil on iseseisvaid määrsõnu 9,1 % (52,4 % määrsõnade üldarvust), tema rõhumäärsõnade näitaja aga (6,8 % ehk 39 % määrsõnade üldarvust) on kõrgem kui ühelgi teisel autoril. Abimäärsõnu kasutab kõige rohkem A. Beekman (3,5 % tekstist, 25 % kõigist teoses kasutatud määrsõnadest).

Huvitav on vaadelda ka a s e s õ n a d e sageduste jaotumust eri alaliikide järgi. Kokku katavad asesõnad tekstist 11,4 %, millest 6,7 % moodustavad isikulised ja näitavad asesõnad. Järgnevas tabelis esitame andmed asesõna alaliikide sageduste kohta tekstis ja asesõnade üldarvu põhjal:

isikuline asesõna	4,4 % (39 % asesõnade üldarvust)
näitav	" 2,3 % (20 % " ")
umbmäärane	" 1,9 % (17 % " ")
küsiv-siduv	" 1,4 % (12 % " ")
omastav	" 0,7 % (6 % " ")
enesekohane	" 0,6 % (5 % " ")
vastastikune	" 0,1 % (1 % " ")
asesõnu kokku	11,4 % (100 %)

Autoritest kasutavad palju asesõnu A. Valton (14,4 %), L. Promet ja V. Saar (mõlemal 13,8 %). Väga vähe leidub asesõnu H. Sergio tekstis (7,9 %).

A. Valtoni kõrge asesõnade protsent sõltub peamiselt näitajate ja umbmääraste asesõnade sagedusest tekstis (vastavalt 3,7 % ja 2,5 %). Isikulisi asesõnu on A. Valtonil

4,9 % (suure sagedusega esineb sõna ta). Isikuliste asesõnade kõige kõrgema näitajaga paistavad silma L. Promet (7,1 %) ja V. Saar (6,8 %). Nende tekstis prevaleerib asesõna mina oma vormidega. Mõlemad autorid kasutavad teoses minavormi (V. Saar küll ainult osaliselt). Peab aga lisama, et L. Prometil ja V. Saarel on ka näitavate asesõnade sagedus tekstis küllaltki suur: vastavalt 2,7 ja 2,8 %. Võrdluseks võib nimetada, et A. Beekmanil on näitavaid asesõnu ainult 1,1 %.

K a a s s õ n u esineb autorikõne tekstis keskmiselt 3,1 %, neist eessõnu 0,4 % (13 % kaassõnade üldarvust) ja tagasõnu 2,7 % (87 %). Oluliselt üle "normi" esineb kaassõnu J. Krossil - 3,9 %, neist tagasõnu 3,5 %, s.o. 90 % kõigist tema poolt kasutatud kaassõnadest. Kõrge kaassõnade protsent on ka V. Saarel ja A. Beekmanil (mõlemal 3,7 %), kusjuures ees- ja tagasõnade sageduste suhe on mõlemal autoril ligikaudu võrdne (1:6). Võib arvata, et kaassõnade rohke kasutamine iseloomustab autori individuaalset stiili, näiteks kalduvust konkretiseerida ruumilisi, põhjuslikke jm. suhteid. Illustreerime seda mõne näitega J. Krossi ja A. Beekmani teostest.

Ja kõik kiikasid muudkui üles ja üles ja keerasid kaelu vasaku õla poolé kõveraks. (J. K., 8).

... punakuldne kaabu tegi vahva kaare üle sinise tae-va. (J. K., 7).

Ise vahtisid nad õhku ja kõigutasid tühjuse kohal koi-bi. (J. K., 8).

Vanker seisis parajasti kahe pärna vahel. (A.B., 5).

Küüni otsas lahtise kuuri all seisis turbahunniku kõr-val looreha. (A. B., 5).

Ringi vaadanud, nägi mees kuuseheki vahelt korralikku põhukuhja. (A. B., 6).

S i d e s õ n a kvantitatiivset esinemust peavad paljud uurijad oluliseks stiilitunnuseks (Kelemen, 1964; Мистрик, 1967). Meie katse põhjal on sidesõnade keskmine sagedus autorikõnes 8,2 % usalduspiiridega 7,5 ... 8,9 %. Sidesõna alaliikide jaotumus tekstis on järgmine:

rinnastav sidesõna	5,6 %	(70 % sidesõnade üldarvust)		
alistav	"	2,6 %	(30 %	" ")
kokku sidesõnu	8,2 %	(100 %)		

Eriti palju sidesõnu esineb J. Krossil (10,1 %), V. Saarel (9,5 %) ja A. Valtonil (9,0 %). Kõigil neil autoritel domineerivad rinnastavad sidesõnad. Paistab silma rinnastava sidesõna ja kasutamine J. Krossi tekstis: valimis (5000 sõnet) esineb ja 274 korda, s. t. katab tekstist 5,5%. Sidesõnade rohke esinemine peegeldab tavaliselt teksti süntaktilise struktuuri keerukust, näit. liitlausete või koondlausete eelistamist. Toome näiteks lause J. Krossi teosest, juhtides tähelepanu sidesõna ja kasutamisele:

Nii et ta teadis hästi, kuidas seal välja näeb, kui sada viiskümmend vahaküünalt sinab ja kõik need imelised kujud, Saalomon ja kes nad seal olid, seinavaipadelt alla vaatavad, ja kui hulk raeisandaid muskuse- ja lavendli- ja muskaadilõhna pilves kõrgel toolidel koos on ja vahivad sulle otsa ja sa pead seal midagi ütleva, ja sul on tunne, et häält pole sul ollagi, sest kõri on kuiva saepuru täis ... (J. K., 59)

J. Krossile on omane ka nn. teksti kompaktsus (Мукт-пик, 1967, 47), mis ilmneb selles, et lauseid seostatakse määrsõnaliselt kasutatud sidesõnaga järgneva lause alguses. Näiteks:

... täpipealt nagu isa talliuksest Kalamaja rannas. Ja väheke kõrgemalt lõhnas väravavahe vastjähvatatud niisust ... (J. K., lk. 10).

Ning selles tähelepanekus oli midagi kehutavat, teadagi, ja midagi kurvastavat kuidagi ka. Aga riiuleil ja riiulite all poepõrandal oli muidugi imeväärt kraami. (J.K., 60).

A r v s õ n a d e sagedus tekstis on keskmiselt 1,1 %. Sellest moodustavad põhjarv sõnad 0,8 % ja järgarv sõnad 0,3 %. Suhteliselt palju esineb arvsõnu J. Krossil (1,7 %), V. Grossil ja H. Sergol (mõlemal 1,4 %), näit.:

Maret kaalub viiskümmend kolm kilo. (V. G., 9).

Vanavanemad elasid veel, isa ja ema - see teeb neli.

Siis kaks säluas isaõde ja kolm last, Juhanist nooremad. Kokku ühiksa. Juhan ise ka - ongi kümme. (H.S., 10).

Paistavad silma veel ligikandseid arve väljendavad sõnad, nagu: tosinkond (J. K.), paarkümmend, kümmekond, versta viis, tuhat korda armsam (H. S.).

H ü ü d sõ n a d e osatähtsus autorikõne tekstis on suhteliselt väike, esinevad peamiselt sellised tavalised sõnad nagu jah, ei, oh ei, noh, no, näe, ah. Kõige rohkem leidub hüüdsõnu V. Saare tekstis (22 juhtu 5000-sõnelise valimi kohta). Teistel autoritel esineb hüüdsõnu niisama suure valimi kohta keskmiselt 5 - 10 korda, A. Beekmanil ainult üks kord.

3.3. Sõnaliikide rühmitused

Sõnaliike on võimalik ühendada rühmadeks, lähtudes mitmesugustest formaalsetest või sisulistest kriteeriumidest. Sõnaliigirühmade moodustamine ja nende statistiline analüüs võib pakkuda huvi individuaalsete või funktsionaalsete stiilide uurimisel, eri keelte tüpoloogilisel kõrvutamisel, rakenduslingvistiliste probleemide lahendamisel jne. (Букович, 1972; Зореф, 1972; Кривоносов, 1973). Järgnevalt vaatleme eesti keele sõnaliike sellise rühmitamise alusel, mis võimaldab võrrelda stiile ja osaliselt kõrvutada eesti keelt mõningate teiste keeltega.

M o r f o l o o g i l i s e s t seisukohast võime eesti keele sõnad jaotada kahte suurde rühma: muutuvad sõnad ja muutumatud sõnad. Muutuvate sõnade rühma kuuluvad nimi-sõnad, omadussõnad, asesõnad, arvsõnad ja tegusõnad. Neid võib omakorda jaotada käändsõnadeks (neli esimesena nimetatud sõnaliiki), tegusõna käändelisteks vormideks ja pöörde-listeks vormideks (pöördsõnadeks). Muutumatud sõnad oleksid seega kõigi ülejäänud sõnaliikide sõnad, kusjuures siinkohal ei võeta arvesse teatavate määrsõnade ja kaassõnade osalist muutumist kohakäänetes (näit. peal:peale:pealt). (Selliselt vaadeldakse muutuvaid ja muutumatuid sõnu ka eesti keele traditsioonilises grammatikas, vrd. Valgma, Rimmel, 1968, 52).

Tabelis 3 esitati andmed üksikute sõnaliikide sageduste kohta erinevate allkeelte tekstis. Koondades sõnalii-

gid morfoloogilisteks rühmadeks, saame järgmised tulemused (vt. tabel 5):

T a b e l 5

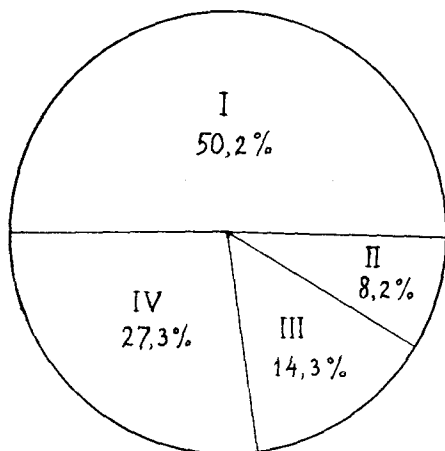
Morfoloogiliste sõnaliigirühmade sagedused (%)
eesti keele eri allkeelte tekstis

Sõnaliigirühm	Tänapäeva ilukirjan- dusproosa autorikõne	"Tõde ja õigus" I autori- kõne	A j a l e h t		Rahva- laul
			Sport	Sõnumid	
Käändsõnad	50,2	46,8	54,1	64,9	66,1
Tegusõna käände- lised vormid	8,2	22,5 23,6	22,8	19,6	24,9
Tegusõna pöörde- lised vormid	14,3				
Muutuvad sõnad kokku	72,7	70,4	76,9	84,5	91,0
Muutumatud sõ- nad kokku	27,3	29,6	23,1	15,5	9,0
Kokku	100,0	100,0	100,0	100,0	100,0

Tänapäeva ilukirjandusproosa autorikõnes domineerivad käändsõnad, moodustades tekstis 50,2 % (piirveaga $\pm 1,4$, usalduspiirid on 48,8 ... 51,6 %). Meie vaatluse andmetel on käändsõnade protsent kõige suurem A. Hindil (52,3 %) ja H. Sergol (51,9 %), kõige väiksem aga V. Saarel (45,6 %) ja A. Valtonil (48,0 %). Muutuvaid sõnu on kaasaegsete kirjani-
ke tekstides kokku keskmiselt 72,7 % ($\pm 1,6$). Kõige rohkem esineb neid A. Beekmanil (75,2 %), H. Kiigel (74,8 %) ja L. Prometil (74,6 %). Võrreldes tänapäeva autoreid A.H. Tammsaarega, võime nentida, et tänapäeva autorid kasutavad keskmiselt rohkem käändsõnu ja ka muutuvaid sõnu kokku, kuid mitmel üksikautoril on A.H. Tammsaarega lähedased sõnaliigirühmade sagedused. Võib teha järelduse, et ilukirjandusproosa autorikõne kvantitatiivsed näitajad on antud juhul küllaltki stabiilsed ja püsivad teatud kindlates variatsioonipiirides.

Morfoloogiliste sõnaliigirühmade jaotumust tänapäeva

ilukirjandusproosa autorikõne tekstis aitab näitlikustada
joonis 1.



Joon. 1. Morfoloogiliste sõnaliigirühmade sagedus-
jaotumus ilukirjandusproosa autorikõnes
(tekstis): I - käändsõnad, II - tegusõna
käänelised vormid, III - tegusõna pöör-
delised vormid, IV - muutumatud sõnad

Pöördudes uuesti tabeli 5 poole, võrdleme ilukirjan-
dusproosa statistilisi näitajaid teiste allkeelte omadega.
Sõnaliigirühmade sagedused toovad eriti selgelt esile eri-
nevused allkeelte vahel. Isegi ajalehe sporditeadetes on
käändsõnu tunduvalt rohkem kui ilukirjandusproosas, ja eri-
ti tõuseb käändsõnade osatähtsus ajalehe sõnumites (64,9 %).
Kuigi sõnumites on tegusõnade protsent suhteliselt väike,
saame kokkuvõttes muutuvate sõnade sageduseks tekstis 84,5%.
Muutuvate sõnade protsent on eriti suur rahvalaulus (91,0%),
selle tingivad aga teistsugused põhjused kui need, millest
sõltub muutuvate sõnade osakaal ajalehekeeles. Rahvalaulus
võib täheldada nii käändsõnade kui ka pöördõnade ülikül-
lust, ning muutumatuid sõnu esineb vaid 9 %.

Kui tahame võrrelda eesti keele statistilisi andmeid
teiste keelte vastavate näitajatega, siis on nõutav, et võr-
reldaks samu allkeeli või žanre. Olemasolevad andmed mõnin-
gate teiste keelte kohta pärinevad ilukirjandusproosa teks-
te käsitlevatest uurimustest, milles autorikõnet ja tege-

laskõnet pole diferentseeritud. Tabelis 6 esitame morfoloogiliste sõnaliigirühmade sagedused ungari, läti, vene, ukraina ja rumeenia keele (Kelemen, 1964) kohta.

T a b e l 6
Morfoloogiliste sõnaliigirühmade teksti sagedused eri keeltes (%)

Sõnaliigirühm	Ungari keel	Läti keel	Vene keel	Ukraina keel	Rumeenia keel
Käändsõnad	43,9	49,1	48,0	46,1	40,6
Tegusõnad	21,6	23,0	18,3	19,7	23,0
(käändel. vormid	3,2)		(3,5)		
(pöördel. vormid	18,4)		(14,8)		
Muutuvad sõnad	65,5	72,1	66,3	65,8	63,6
Muutumatud sõnad	34,5	27,9	33,7	34,2	36,4
K o k k u	100,0	100,0	100,0	100,0	100,0

Siinjuures võib märkida, et sõnade liigitamine muutuvateks ja muutumatuteks sõnadeks on täiesti kindlapiiriline slaavi, romaani ja germaani keeltes (vrd. Кривоносов, 1973) ning võimalik ka ungari ja läti keeles (vt. näit. ungari sõnalikide rühmitamist T. Zsilka uurimuses, 1973, lk. 101 jj.). Toodud andmeid võime võrrelda eesti keelega, silmas pidades teatavat erinevust allkeeltes, s. t. tegelaskõne mõju võõrkeelse ilukirjandusproosa kvantitatiivsetele karakteristikutele. Võib aga arvata, et antud juhul autorikõne ja tegelaskõne tasakaalustuvad: ühelt poolt esineb tegelaskõnes vähem käändsõnu, teiselt poolt rohkem tegusõnu, seega peaks muutuvaid sõnu olema mõlemas ligikaudu võrdselt. Muutuvate sõnade osatähtsusest tekstis on eesti keel lähedane läti keelele, ületades teised vaadeldavad keeled tunduvalt.

Väärrib tähelepanu sõnade rühmitamine s e m a n t i i l i s - s ü n t a k t i l i s e printsiibi järgi, mille kohaselt moodustuvad kaks suurt rühma: "autosemantilised" ehk täistähenduslikud sõnad (vene k. ПОЛНОЗНАЧНЫЕ, ЗНАМЕ-

НАТЕЛЬНЫЕ СЛОВА; saksa k. Vollwörter) ja "sünsemantilised" sõnad, mida võib tinglikult nimetada "abisõnadeks" või "struktuurisõnadeks" (vene k. строение, структурные, служебные слова; saksa k. Leerwörter, Strukturwörter). Niisugune sõnade liigitamine kaheks suureks rühmaks on tuntud juba Aristotelese ajast ja selle alusel on korduvalt uuritud stiile ja võrreldud keeli tüpoloogilisest seisukohast. Täpse piiri tõmbamine täistähenduslike ja mittetäistähenduslike sõnade vahele on esile kutsunud vaidlusi. Keelestatistikas on välja kujunenud kaks skeemi, millest lähtuvad kvantitatiivsed uurimused. P. Guiraud' (1959) järgi peetakse täistähenduslikeks sõnadeks ainult nimisõnu, omadussõnu, tegusõnu ja määrsõnu (eesti keele seisukohast iseseisvaid määrsõnu). On tuntud teinegi liigitus (näit. Tšitilová, 1972), mis arvab täistähenduslike sõnade hulka ka arvsõnad ja asesõnad. Ka eesti keele grammatikas eristatakse täistähenduslike sõnu, (Mihkla jt., 1974, 12), lähtudes nii semantilisest kui ka süntaktilisest printsibist.

Käesolevas töös võtame aluseks mõlemad skeemid: I - täistähenduslikud sõnad kitsamas mõttes (nimisõnad, omadussõnad, tegusõnad, iseseisvad määrsõnad) ja ülejäänud sõnad; II - täistähenduslikud sõnad laiemas mõttes (kaasa arvatud arvsõnad ja asesõnad) ning abisõnad.⁺ Sellise jaotuse alusel saame eri autorite kohta järgmised tekstisagedused (protsentides):

A.B. V.G. A.H. H.K. J.K. L.P. V.S. H.S. M.T. A.V.

I - 73,9 69,0 65,5 71,3 67,3 66,3 63,8 72,6 67,9 65,8

II - 82,8 82,1 79,7 81,9 78,1 81,0 78,5 81,9 81,6 80,7

Esimese skeemi järgi esineb autorikõne tekstis keskmiselt 68,3 % täistähenduslike sõnu (usalduspiirid 65,9 70,7 %). Variatsioonilatus on suur, nimelt kõiguvad sagedused 63,8 % (V. Saar) ja 73,9 % (A. Beekman) vahel. See tähendab, et täistähenduslike sõnade protsent on esimese skeemi järgi individuaalseid stiile eristav faktor. Teise skeemi järgi, s. t. arvestades täistähenduslike sõnade hulka

⁺ Rangelts võttes tuleks täistähenduslike sõnade hulgast välja arvata tegusõna olema abiverbi funktsioonid (mis moodustab keskmiselt 1,6 % tekstist). Käesoleval juhul arvestame verbi olema sagedust tervikuna, et andmed oleksid paremini võrreldavad teiste keeltega.

ka arv- ja asesõnad, on autorikõne keskmine 80,8 %, kusjuures kõikumus on väike (usalduspiirid: 79,7 ... 81,9 %, suhteline viga ainult 1,4 %!). Juba varem on juhitud tähelepanu sellele, et täistähenduslike sõnade protsent tekstis on teise skeemi järgi ühe keele ulatuses väga stabiilne (näit. tšehhi keeles kõigub täistähenduslike sõnade sagedus 80% ümber, vt. Těšitelová, 1972). Nimetatud tunnus võimaldab keeli uurida tüpologia seisukohast. Esitame alljärgnevalt kõrvutavad andmed täistähenduslike sõnade sageduse kohta kahe skeemi järgi eesti keeles ja mõnedes teistes keeltes.

	I	II
Eesti keel (ilukirjandusproosa autorikõne)	68,3 %	80,8 %
Ungari keel (ilukirjandusproosa)	69,5 %	78,9 %
Läti keel (ilukirjandusproosa)	68,6 %	82,1 %
Vene keel (ilukirjandusproosa autorikõne)	60,9 %	72,3 %
Ukraina keel (ilukirjandusproosa)	61,9 %	72,0 %
Saksa keel (ilukirjandusproosa autorikõne) ⁺	55,2 %	71,9 %

Paistab silma täistähenduslike sõnade suhteliselt kõrge sagedus eesti, läti ja ungari keeles (nii I kui ka II skeemi järgi).

4. SÕNASTIKU- JA TEKSTISAGEDUSED; FUNKTSIONAALNE JA INFORMATIIVNE KOORMUS

4.1. Sõnaliikide sagedused sõnastikus

Sõnaliikide statistilisel uurimisel pakub huvi nende esinemissagedus vaadeldavate tekstide juurde kuuluvais sõnastikes. Käesolevas töös on sõnastikud koostatud 7 teksti kohta sõnavormide tasandil. Sõnaliikide suhtelised sagedused ja hajuvushinnangud esitame tabelis 7. Nagu näha, on ka sõnastikus nimisõnad sageduselt esikohal (keskmiselt 44,2 %), teisel kohal on tegusõnad (26,6 %) ja kolmandal kohal määrsõnad (10,7 %). Tuleb aga silmas pida, et meie näites on sagedused arvestatud sõnavormide sõnastiku põhjal. Võrdluseks võib tuua andmed uurimusest

⁺ Saksa keele andmed on võetud L. Pilleri artiklist (Piller, 1971). Teiste keelte kohta kasutatakse varem nimetatud uurimuste andmeid.

Tabel 7

Sõnalikide suhteline sagedus sõnastikus (sõnavormide tasandil) eesti ilukirjandus-
proosa autorikõne põhjal (7 valimit à 5000 sõnet)

Suhte- line Sõna- sage- dus liik (%)	A u t o r i d							Keskmine \bar{p} (%)	Piirviga $\epsilon_{\bar{p}}$ (%)	Usalduspiirid (%)	Suhte- line viga $\delta_{\bar{p}}$ (%)
	A.B.	V.G.	H.K.	J.K.	L.P.	V.S.	H.S.				
Nimisõna	40,3	42,7	45,9	45,9	44,9	39,3	50,2	44,2	$\pm 3,5$	40,7 ... 47,7	7,9
Omadussõna	8,4	12,0	8,3	12,2	8,5	8,7	8,3	9,5	$\pm 1,7$	7,8 ... 11,2	17,9
Arvsõna	0,6	1,6	1,6	1,5	1,0	1,1	1,7	1,3	$\pm 0,4$	0,9 ... 1,7	30,8
Asesõna	3,5	5,0	4,9	3,5	4,9	5,6	3,8	4,4	$\pm 0,8$	3,6 ... 5,2	17,8
Tegusõna	32,6	24,5	27,1	20,1	27,6	30,9	23,7	26,6	$\pm 4,0$	22,6 ... 30,6	15,0
Määrsõna	11,3	11,1	9,1	13,1	10,0	10,6	9,3	10,7	$\pm 1,3$	9,4 ... 12,0	12,1
Kaassõna	2,2	2,0	2,1	2,5	2,1	2,5	2,0	2,2	$\pm 0,2$	2,0 ... 2,4	9,1
Sidesõna	1,0	0,8	0,7	0,9	0,9	1,0	0,8	0,9	$\pm 0,1$	0,8 ... 1,0	11,1
Hüüdsõna	0,1	0,3	0,3	0,3	0,1	0,3	0,2	0,2	$\pm 0,1$	0,1 ... 0,3	50,0
Kokku (%)	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	-	-	-
Sõnastiku maht (V)	2870	2783	2643	2821	2721	2447	2892	keskm. 2740	-	-	-

A.H. Tammsaare romaani "Tõde ja õigus" I köite autorikõne kohta (Enniko, Meiman, 1975), kus sõnaliikide sõnastikusa-
gedused on antud sõnade (lekseemide) tasandil: nimisõnu -
43,4 %, tegusõnu - 22,8 %, omadussõnu - 8,6 %, arvsõnu -
0,9 %, asesõnu - 2,1 %, määrsõnu - 17,2 %, kaassõnu - 3,5 %,
sidesõnu - 1,4 %, hüüdsõnu - 0,1 %. Lekseemide tasandil vä-
heneb mõnevõrra muutuvate sõnade osakaal, kuna aga muutuma-
tute sõnade osatähtsus suureneb (näiteks on määrsõnu sõna-
vormide sõnastikus keskmiselt 10,7 %, A.H. Tammsaare leksee-
mide sõnastikus aga 17,3 %).

Morfoloogilise rühmituse järgi jagunevad sagedused ilu-
kirjandusproosa autorikõne sõnastikus sõnavormide ja lek-
seemide tasandil järgmiselt:

	Sõnavormid (7 autorit)	Lekseemid (A.H. Tammsaare)
Käändsõnad	59,4 %	54,0 %
Tegusõna käändel. vormid	12,0 %	26,5 %
Tegusõna pöördel. vormid	14,6 %	
Muutuvad sõnad (kokku)	86,0 %	77,5 %
Muutumatud sõnad (kokku)	14,0 %	22,5 %

Täistähenduslikke sõnu (sõnavorme) on kaasaegse ilu-
kirjandusproosa autorikõne sõnastikus I skeemi järgi (s. o.
nimisõnad, omadussõnad, tegusõnad ja iseseisvad määrsõnad)
87,6 % ning II skeemi järgi (kaasa arvatud ase- ja arvsõ-
nad) - 93,3 %. Lekseemide tasandil on vastavad sagedused
A.H. Tammsaare sõnastikus: 83,0 % ja 86,4 %.

4.2. Sõnaliikide funktsionaalne ja informatiivne koormus

Sõnaliikide funktsionaalset koormust võib määritleda
mitmel viisil. Kõige lihtsam on arvestada funktsionaalse
koormusena vastava sõnaliigi suhtelist sagedust, s. o. suh-
telist osatähtsust tekstis või sõnastikus. Selline vaatlus-
viis ei too aga iseenesest esile mõningaid olulisi sisemisi
seaduspärasusi, mis on omased sõnaliikide funktsioneerimi-
sele tekstis. On otstarbekohane vaadelda sõnaliikide esine-
must teksti ja sõnastiku sageduste suhtena, mis annab meile
vastavasse sõnaliiki kuuluvate sõnade keskmise

k o r d u v u s e t e k s t i s :

$$KK = t/s,$$

kus KK tähistab korduvus- ehk iteratsiooni-indeksit, t - tekstiasagedust ja s - sõnastikusagedust (antud juhul sõnavormide tasandil). Tabelis 8 on toodud andmed eri sõnaliikide keskmise korduvuse kohta kaasaegse ilukirjandusproosa autorikõne tekstis. Kõige suurema keskmise korduvusega on sidesõnad: igauks neist kordub keskmiselt 17 korda 5000-sõnelises tekstis. Selle põhjal võib väita, et sidesõnad on tekstis kõige suurema koormusega. Kõige väiksema korduvusega on omadussõnad (KK = 1,1). Kui sõnade korduvus tekstis on väike, siis tähendab see ühtlasi seda, et nende mitmekesisus on suur.

Pakub huvi vaadelda sõnaliikide korduvust ka morfoloogiliste rühmituste järgi. Käändsõnad korduvad 5000-sõnelises valimist keskmiselt 1,5 korda, tegusõna käändelised vormid 1,3 ja pöördelised vormid 1,8 korda. Muutuvad sõnad (sõnavormid) korduvad keskmiselt 1,6 korda, muutumatud sõnad aga 3,6 korda.

Sõnaliikide funktsionaalse koormuse mõõtmine korduvusindeksi alusel võimaldab kindlaks teha olulisi erinevusi autorite individuaalsetes stiilides. Vaatleme näiteks nimisõnade esinemust tekstis ja sõnastikus ning iga nimi-sõna keskmist korduvust seitsme autori tekstis.

	t	s	KK = t/s
A. Beekman	1796	1157	1,55
V. Gross	1519	1189	1,28
H. Kiik	1739	1214	1,43
J. Kross	1642	1295	1,27
L. Promet	1519	1222	1,24
V. Saar	1298	961	1,35
H. Sergio	1842	1452	1,27

A. Beekmani ja H. Sergio tekstis esineb palju nimisõnu, kuid korduvusindeks näitab olulist erinevust nende kasutuses. A. Beekmanil on nimisõnade korduvus suurem (keskmiselt 1,55) ja seega nende mitmekesisus väiksem, H. Sergio aga korduvad nimisõnad suhteliselt vähe (1,27), s. t. nende valik on suurem. Teiste sõnaliikide osas võib aga olukord olla vastupidine (nimetatud kirjanike puhul täheldamegi

Tabel 8

Sõnaliikide keskmine esinemissagedus tekstis ja sõnastikus, keskmine korduvus ja informatiivne koormus

Sõnaliik	Keskml. absol. sagedus		Keskmine korduvus (t/s)	Keskml. suhtel. sagedus (%)		Informatiivne koormus (p_s/p_t)
	tekstis (t)	sõnastikus (s)		tekstis (p_t)	sõnastikus (p_s)	
Nimisõna	1585	1213	1,3	31,7	44,2	1,4
Omadussõna	300	260	1,1	6,0	9,5	1,6
Arvsõna	55	36	1,5	1,1	1,3	1,1
Asesõna	570	121	4,4	11,4	4,4	0,4
Tegusõna	1125	727	1,5	22,5	26,6	1,2
Määrsõna	790	292	2,7	15,8	10,7	0,7
Kaassõna	155	61	2,5	3,1	2,2	0,7
Sidesõna	410	24	17,0	8,2	0,9	0,1
Hüüdsõna	10	6	1,7	0,2	0,2	1,0
Kokku	5000	2740	1,8	100,0	100,0	-

seada tegu-, määr- ja omadussõnade esinemuses). Esitame alljärgnevalt omadussõnade, tegusõnade ja määrõnade korduvusindeksi väärtused vaadeldava seitsme valimi lõikes:

	A.B.	V.G.	H.K.	J.K.	L.P.	V.S.	H.S.
Omadussõnu	1,12	1,11	1,18	1,08	1,18	1,17	1,21
Tegusõnu	1,34	1,63	1,69	1,61	1,60	1,55	1,61
Määrõna	2,14	2,82	3,12	2,24	2,57	3,34	2,76

Omadussõnade korduvus on kõige väiksem J. Krossil (KK = 1,08), tegu- ja määrõnade korduvusindeksid on aga kõige väiksemad A. Beekmanil (vastavalt 1,34 ja 2,14). Suured indeksite väärtused (ja väiksem mitmekesisus) iseloomustavad H. Sergo omadussõnu (1,21), H. Kiige tegusõnu (1,69) ja V. Saare määrõnu (3,34).

Korduvus- ehk iteratsiooni-indeksi kasutamine funktsionaalse koormuse mõõduna on ülevaatlik ja sisuliselt arusaadav. Indeksi puuduseks on asjaolu, et eri andmete võrdlemisel peab lähtuma ühesuurustest valimitest. Indeksi väärtus sõltub täielikult valimi mahust. Seepärast on mõtet otsida teisi võimalusi sõnaliikide tekstikoormuse mõõtmiseks, et saaks võrrelda eri uurimuste andmeid. Normeerides absoluutsagedused sel lihtsal viisil, et arvutame vastavad suhtelised sagedused, võime funktsionaalset koormust mõõta teksti ja sõnastiku suhteliste sageduste suhtena. Uurimused on näidanud, et sõnaliikide suhtelised sagedused jäävad valimi mahu (teksti pikkuse) suurenedes enamvähem konstantseteks nii sõnastiku kui ka teksti tasandil (sõnastiku tasandil vähenevad küll selliste sõnaliikide nagu ase-, kaas- ja sidesõnade suhtelised sagedused, kuid nende osatähtsus sõnastikus on väike ja nad ei avalda olulist mõju arvutusetele). Sõnaliikide funktsionaalset koormust võib seega määrata teksti ja sõnastiku suhteliste sageduste jagatise abil:

$$FK = p_t / p_s ,$$

kus FK on normeeritud funktsionaalse koormuse mõõt, p_t - sõnaliigi suhteline sagedus tekstis ja p_s - sõnaliigi suhteline sagedus sõnastikus.

Kirjeldataud suhet võib vaadelda ka ümberpöörduvalt, s. o. jagades sõnastikusageduse tekstisagedusega. Tuginedes informatsiooniteooria põhimõtetele, võib sel viisil saadavat suhet sisuliselt interpreteerida kui sõnaliigi i n f o r -

mativset koormust (IK):

$$IK = p_g/p_t .$$

Mida suurem on IK väärtus, seda suurem on sõnaliigi suhteline osakaal sõnastikus ja järelikult on iga esinemine tekstis suhteliselt haruldasem ning ühtlasi informatiivsem. Informatiivne koormus kirjeldatud mõttes on lähedane nn. mitmekesisuse indeksile TTR ("type-token-ratio", vt. Tulda-va, 1971, 220) ja väljendab samuti sõnavara mitmekesisust või "rikkust", antud juhul sõnaliigi kohta terviklikult. Võrreldes TTR-indeksiga sõltub aga IK väga vähesel määral valimi suurusest ja seepärast on võimalik IK abil võrrelda erinevatest katsetest ja erinevate keelte põhjal saadud tulemusi (võrreldavate valimite mahud ei tohi siiski olla liiga erinevad).

Tabelis 8 on toodud informatiivse koormuse (IK) väärtused eri sõnaliikide kohta meie katse materjali põhjal. Kõrvutatakse keskmisi suhtelisi sagedusi sõnastikus ja tekstis. Informatiivse koormuse alusel võib sõnaliigid ja-gada kahte rühma: sõnaliigid, mille IK väärtus on üle 1, ja sõnaliigid, mille IK väärtus on 1 või alla selle. Esimesse rühma kuuluvad eelkõige omadussõnad (IK = 1,6), nimisõnad (IK = 1,4) ja tegusõnad (IK = 1,2). Suhteliselt kõrget in-formatiivset koormust tekstis omavad meie katse andmeil ka arvsõnad (IK = 1,1). Määrsõnade informatiivne koormus teks-tis on kokkuvõttes suhteliselt madal. Vaadeldes määrsõna eri liike omaette, saame järgmised IK väärtused: iseseisvad määrsõnad - 0,9, rõhumäärsõnad - 0,4 ja abimäärsõnad - 0,5.

Sõnaliikide informatiivset koormust võib vaadelda di-ferentseeritult ka morfoloogiliste rühmade kaupa. Ilukir-jandusproosa autorikõne kohta saame järgmised tulemused: käändsõnad - 1,2, tegusõna käändelised vormid - 1,5, tegu-sõna pöördelised vormid - 1,0; muutuvad sõnad kokku - 1,2, muutumatud sõnad - 0,5. Seega osutuvad kõige informatiivse-maks tegusõna käändelised vormid.

Informatiivsuse indeks IK peegeldab täiesti ootuspära-selt ka seda seika, et täistähenduslikel sõnadel on tekstis märgatavalt suurem informatiivne koormus kui abisõnadel. Nii saame näiteks I skeemi alusel (vt. 3.3.) täistähenduslike sõnade IK väärtuseks 1,3: ülejäänud sõnade IK väärtuseks (kokku) aga ainult 0,4. II skeemi järgi (arvestades täis-

tähenduslike sõnade hulka ka arv- ja asesõnad) on vastavad väärtused 1,1 ja 0,3.

5. SÕNALIIGI-INDEKSID

Funktsionaalsete ja individuaalsete stiilide kvantitatiivsel uurimisel on paljudes töodes lähtunud sõnaliigisageduste omavahelistest suhetest. Eriti huvipakkuvad on niisugused sõnaliikide suhted, mida saab sisuliselt interpreteerida. Nii näiteks on juba ammu tuntud ja stilostatistikas ning psühholingvistilistes uurimustes kasutatud nn. Busemanni koefitsient, mis arvutatakse omadussõna ja tegusõna tekstisageduste suhtena (vt. Busemann, 1925 ja 1948; Boder, 1940; Antosch, 1969; Fischer, 1965; Zsilka, 1974). Omadussõna ja tegusõna vastandamine põhineb tähelepanekul, et mõlemad sõnaliigid on oma semantilise funktsiooni poolest predikatiivse toimega, s. t. nad ütlevad midagi uut kõnesoleva objekti kohta (vrd. Гальперин, 1974, 164-165). Olenevalt individuaalsest stiilist kasutavad ühed autorid esemete ja nähtuste kirjeldamisel rohkem adjektiivset ("kvalitatiivset") hinnangut, teised aga eelistavad verbaalset väljendusviisi. Seepärast saab omadussõnade ja tegusõnade sageduste suhet vaadelda kui stiili "kvalitatiivsuse" või selle vastandi - "verbaalsuse" või "tegevuslikkuse" hinnangut (vrd. vene keele stilistikas kasutusel olevaid termineid качественность ja действенность vt. näit. Папина, Иванова-Маркова, 1974).

Stilostatistikas on laialdast kasutamist leidnud veel mitmesugused muud sõnaliigisageduste suhted, mida saab sisuliselt tõlgendada ja mis on uurimuste tulemusena osutunud individuaalseid või funktsionaalseid stiile eristavateks ja diagnoosivateks faktoriteks (vt. näit. Мистрик, 1967; Головин, 1971; Muller, 1967; Rensky, 1965; Тěшителová, 1972).

Käesolevas töös vaatleme mõningaid tuntuimaid sõnaliigi-indekseid eesti ilukirjandusproosa autorikõne põhjal. Tulemused esitame kokkuvõtlikul kujul tabelis 9. Alljärgnevalt kommenteerime lühidalt tabeli andmeid.

1. Esimese sõnaliigi-indeksina on tabelis toodud omadussõnade ja verbi pöördeliste vormide sageduste suhe: A/V_p .

See suhe ("Busemanni koefitsient") määrab stiili k v a - l i t a t i i v s u s e astme, s.t. adjektiivide abil väljendatud kvalitatiivse hinnangu osatähtsuse, võrreldes verbaalse väljendusviisiga. Funktsionaalsete stiilide võrdlemisel on kvalitatiivsuse indeks alati suurem publitsistlikus ja teaduslikus tekstis ja väiksem ilukirjanduses, eriti draamas. Allkeelte piirides võib omakorda täheldada individuaalseid erinevusi kvalitatiivsuse määras. Meie katse andmeil on kümne valimi keskmine indeksi väärtus 0,43. Arvestades usalduspiire 95%-lisel usaldusnivool ($0,43 \pm 0,08$), võib nentida, et kahel autoril (J. Krossil ja V. Grossil) on indeksi väärtus märgatavalt üle keskmise. Antud juhul korreleeruvad kvalitatiivsuse hinnang ja omadussõnade sagedus tekstis (vt. tabel 2). Kuid alati ei tarvitse see nii olla. Näiteks A. Hindil ja L. Prometil on omadussõnade protsent tekstis ligikaudselt võrdne: vastavalt 5,6 ja 5,5 %. Tegusõna pöördelisi vorme on aga A. Hindil 11,1 % ja L. Prometil 16,0 %. Seega erinevad kvalitatiivsuse hinnangud neil autoritel tublisti: A. Hindi stiil on "kvalitatiivsem" ülal kirjeldatud mõttes (indeksi väärtus - 0,50), L. Prometi stiil on aga "verbaalsem" (indeksi väärtus - 0,34).

2. Omadussõnade keskmist sagedust iga nimisõna kohta väljendab nn. modifikatsioon- ehk a d j e k t i i v s u s e indeks A/S. Ilukirjandusproosa autorikõne keskmine on 0,19. Seega esineb tekstis keskmiselt üks omadussõna viie nimisõna kohta. Võrreldes teiste keelte analoogiliste andmetega on see näitaja eesti keeles suhteliselt madal. Näiteks ungari ilukirjandusproosas on vastav indeksi väärtus 0,33 (Zsilka, 1973, 118), vene keeles E. Šteinfeldti sagedussõnastiku andmeil (Штейнфельдт, 1963, 31) - 0,31. Eesti keele ajalehetekstide sporditeadetes on omadussõnade ja nimisõnade suhe 0,22 ja välisteadetes 0,16 (Valge, 1972, 32-33). A.H. Tammsaare "Tõde ja õiguse" autorikõnes on adjektiivsuse indeksi väärtuseks 0,13 (Enniko, Meiman, 1975). Käesolevas töös vaadeldavate autorite seas on suhteliselt kõrged indeksi väärtused V. Grossil ja A. Valtonil (kummalgi 0,24) ning J. Krossil (0,23).

T a b e l 9

Sõnaliigi-indeksid ilukirjandusproosa autorikõne põhjal
(10 valimit à 5000 sõnet)

Autor Indeks	A.B.	V.G.	A.H.	H.K.	J.K.	L.P.	V.S.	H.S.	M.T.	A.V.	Kesk- m \bar{x}	Pliir- viga $\epsilon_{\bar{x}}$	Suh- tel- viga $\delta_{\bar{x}}(\%)$
	A/V _p ⁺ kvalita- tiivsus	0,33	<u>0,58</u>	0,50	0,29	<u>0,63</u>	0,34	0,38	0,44	0,40	0,40	0,43	±0,08
A/S adjektiiv- sus	0,15	<u>0,24</u>	0,17	0,15	<u>0,23</u>	0,18	0,19	0,16	0,20	<u>0,24</u>	0,19	±0,02	10,5
S/V substan- tiivsus	1,44	1,37	<u>1,70</u>	1,44	<u>1,80</u>	1,27	1,11	<u>1,68</u>	1,36	1,12	1,43	±0,17	11,9
V _p /(S+P+A) verbaalsus	<u>0,33</u>	0,26	0,22	<u>0,36</u>	0,24	<u>0,32</u>	0,29	0,26	0,31	<u>0,34</u>	0,29	±0,03	10,3
P/S pronomi- naalsus	0,23	0,38	0,40	0,27	0,28	<u>0,45</u>	<u>0,53</u>	0,22	0,41	<u>0,54</u>	0,37	±0,08	21,6
D _i /V _p adverbi- aalsus	0,47	<u>0,71</u>	<u>0,75</u>	0,40	<u>0,74</u>	0,41	<u>0,72</u>	0,61	0,54	0,55	0,59	±0,10	16,9

+ Lühendite tähendused: A - adjektiiv, V - verb, V_p - verbi pöördeline vorm,
S - substantiiv, P - pronomen, D_i - iseseisev adverb

3. Nominaalsust ja verbaalsust on stilostatistilistes töödes sageli hinnatud nimisõnade ja tegusõnade sageduste suhtena. Sellist suhet peetakse tähtsaks kvantitatiivseks karakteristikuks funktsionaalsete stiilide kõrvutaval uurimisel (Кожина, 1972, 140). Mõnede uurijate arvates toob nimisõnade ja tegusõnade sageduste suhe esile erinevused stiilide "staatilisuse" ja "dünaamilisuse" vahel (Zsilka, 1967, 156). Suhtet võib esitada mõlemapoolselt, s. o. nii nimisõnade suhtena tegusõnadesse kui ka ümberpöörduvalt. Käesoleva materjali põhjal arvutame nimisõnade sageduse jaगतise tegusõnade sagedusega ja nimetame selle s u b s t a n t i i v s u s e indeksiks. Indeksi keskmine on autorikõne valimite põhjal 1,43. Ka teistes keeltes on ilukirjandusproosa substantiivsuse hinnang lähedane eesti keelele: ungari keeles - 1,45, tšehhi keeles - 1,44 (T. Zsilka andmeil), ukraina keeles - 1,49 (Тященко, 1970, 217 jj.). Eesti ajalehekeele sporditeated on lähedased ilukirjandusproosa autorikõnele (indeksi väärtus - 1,51), kuid välis- teadetes on substantiivsuse aste märksa kõrgem - 2,51 (Valge, 1972, 32-33). A.H. Tammsaare autorikõnes on verbide ja nimisõnade sageduste suhe 1:1,22.

Kaasaegse ilukirjandusproosa autorikõne keskmise ülevatavad oluliselt J. Kross, A. Hint ja H. Sergo. Väga madalat substantiivsust võib nentida V. Saare ja A. Valtoni (vaadeldavates) teostes - indeksi väärtused on vastavalt 1,11 ja 1,12.

4. Tuntud nõukogude keelestatistik B. Golovin on teinud ettepaneku stiili verbaalsust ja ühtlasi "tegevuslikkust" ning "dünaamilisust" määratleda verbisageduste suhtena noomenite (nimisõnade, omadussõnade ja asesõnade) sagedustesse tekstis (ГОЛОВИН, 1971, 147). Kasutades seda v e r b a a l s u s e hinnangut, saame autorikõne keskmiseks 0,29, kusjuures keskmisest oluliselt kõrgemat verbaalsust võib täheldada eriti H. Kiigel (0,36) ja A. Valtonil (0,34), samuti A. Beekmanil (0,33) ja L. Prometil (0,32).

5. Asesõnade ja nimisõnade suhet (P/S) võib vaadelda p r o n o m i n a a l s u s e hinnanguna. See näitaja on tundlik individuaalsete erinevuste suhtes ja võimaldab avastada ka sellise iseärasuse nagu minavormi kasutamine,

juhul kui asesõnadest arvestada ainult isikulisi asesõnu indeksi komponendina. Meie katse andmeil on kõrge pronomi-naalsuse aste L. Prometi ja V. Saare (kes kasutavad mina-vormi) ning A. Valtoni stilil.

6. Iseseisvate määrsõnade ja tegusõna pöördeliste vormide sageduste suhe näitab, kui palju keskmiselt kasutatakse täistähenduslikke määrsõnu iga verbi (pöördelise vormi) kohta tekstis. Adverbiaalsuse indeksi väärtused kõiguvad vaadeldavates valimites 0,40 (H. Kiik) ja 0,75 (A. Hint) vahel. Indeksi keskmise väärtuse ületavad oluliselt ka J. Krossi, V. Saare ja V. Grossi näitajad (vastavalt 0,74; 0,72; 0,71). Adverbiaalsuse suhteliselt kõrget astet võib täheldada ka A.H. Tammsaare autorikõnes (0,70).

Kokkuvõttes tuleb nentida, et sõnaliikide sageduste suhteid väljendavad sõnaliigi-indeksid aitavad täpsustada sõnaliikide funktsioone tekstis ja võimaldavad sügavamalt läheneda sõnaliikide stilostatistilisele analüüsile.

6. SÕNALIIKIDEVAHELISED KORRELATSIOONID

Sõnaliikide kasutamist tekstis on mõtet vaadelda kui süsteemset nähtust, s. o. elementide omavaheliste seoste ja sõltuvuste alusel. Eespool kirjeldatud sagedussuhete uurimise kõrval võib sõnaliikidevahelisi seoseid kindlaks teha ja mõõta statistilise korrelatsioonanalüüsi abil. Ilukirjandusproosa autorikõne kümne valimi põhjal arvutatud korrelatsioonid (Pearsoni-Bravais' meetodi järgi, vt. lähemalt: Tuldava, 1972) esitatakse tabelis 10.

Korrelatsioonikordajate (r) kriitilised väärtused on 10%-lisel olulisusnivool 0,54 ja 5%-lisel olulisusnivool 0,63. Peab märkima, et statistiliselt olulisi korrelatsioonide ei ole meie katse puhul kuigi palju. See tähendab, et sõnaliike kasutatakse tekstis suhteliselt sõltumatult üksteisest. Siiski võib juhtida tähelepanu mõningatele olulistele seostele, näiteks tugevale negatiivsele korrelatsioonile nimisõna ja asesõna sageduste vahel (-0,89), mõõdukale negatiivsele korrelatsioonile nimisõna ja määrsõna sageduste vahel (-0,60) ning negatiivsele seosele tegusõna ja oma-

dussõna vahel (-0,54). Nimisõna ja asesõna vaheline negatiivne seos on sisuliselt hästi seletatav: need sõnaliigid konkureerivad esemete ja nähtuste nimetamisel kõnes. Tegusõna ja omadussõna seos tuleneb "kvalitatiivsuse" ja "tegevuslikkuse" (verbaalsuse) konkurentsist esemete ja nähtuste kirjeldamisel (vt. eespool, p. 5). Raskem on leida otsesest seletust nimisõna ja määrsõna vahelisele seosele, samuti tegusõna ja arvsõna sageduste negatiivsele korrelatsioonile ($r = -0,75$). Ilmselt on siin tegemist vastastikuste korrelatsioonide läbipõimumisega, mida aitaks lahti mõtestada osa- ja mitmese korrelatsiooni arvutamine (lähemalt vt. Tiit, 1972, 93 jj.).

T a b e l 10

Korrelatsioonid sõnaliikide tekstisageduste vahel

Sõnaliik	2.	3.	4.	5.	6.	7.	8.
1. Nimisõna	0,05	-0,89	-0,12	-0,60	0,24	-0,49	0,40
2. Omadussõna		-0,10	-0,54	0,43	-0,13	0,06	0,46
3. Asesõna			0,11	0,41	-0,36	0,33	-0,44
4. Tegusõna				-0,48	-0,42	-0,39	-0,75
5. Määrsõna					-0,02	0,32	0,30
6. Kaassõna						0,40	0,18
7. Sidesõna							0,05
8. Arvsõna							-

Korrelatsioonimaatriksist lähtudes (vt. tabel 10) võib teha mõningaid järeldusi ka korrelatsiooni puudumise alusel. Nii näiteks on huvitav täheldada, et nimisõnade ja tegusõnade aktiivsus tekstis ei ole omavahel seotud ($r = -0,12$), samuti puudub oluline seos nimisõnade ja kaassõnade sageduste vahel ($r = 0,24$). Esimesel juhul tuleb ilmsiks juba varem mainitud tõsiasi, et nominaalsus ja verbaalsus võivad esineda kõrvuti ühes ja samas stiilis, kui neid defineerida sõnaliikide sageduste ja mitte nende suhte alusel. Teisel juhul võib tõdeda, et kaassõnade kasutamine, mis kaasneb nimisõnade kasutamisega, allub individuaalsele valikule ja võib seetõttu olla autoristile eristavaks teguriks.

Korrelatsioonide põhjalikumaks iseloomustamiseks on võimalik pöörduda faktoranalüüsi poole, mis aitab sõnaliikidevahelisi seoseid esitada kompaktsel kujul. Meie materjali põhjal arvatatud kolm peamist faktorit on järgmised: nominaalsuse faktor (positiivse faktorkoormusega nimisõnal), verbaalsuse faktor (bipolaarne faktor, mille teiseks pooluseks on kvalitatiivsus põhiliste koormustega omadussõnal, määrsõnal ja arvsõnal) ning analüütilisuse faktor (peamiselt kaas- ja sidesõnade aktiivsuse arvel).⁺

V i i d a t u d k i r j a n d u s

- A n t o s c h , F. The Diagnosis of Literary Style with the Verb-Adjective Ratio. - Statistics and Style. Edited by L. Doležel and R.W. Bailey. New York, 1969, 57-68.
- B o d e r , D.P. The Adjective-Verb Quotient: A Contribution to the Psychology of Language. - Psychological Record, III, 1940, 309-343.
- B u s e m a n n , A. Die Sprache der Jugend als Ausdruck der Entwicklungsrhythmik. Jena, 1925.
- B u s e m a n n , A. Stil und Charakter. Untersuchungen zur Psychologie der individuellen Redeform. Meisenheim/Glan, 1948.
- E n n i k o , K., M e i m a n , S. Sõnaliikide kvantitatiivne esinemus A.H. Tammsaare romaani "Tõde ja õigus" I köite autorikõnes. TPedI kursusetöö, juh. A. Villup. Tln., 1975.
- F i s c h e r , H. Entwicklung und Beurteilung des Stils. - Mathematik und Dichtung. München, 1965, 171-183.
- F r i e s , Ch. The Structure of English. New York, 1952.
- G u i r a u d , P. Problèmes et méthodes de la statistique linguistique. Dorrecht, 1959.

⁺ Eesti keele sõnaliikide faktoranalüüsi tulemusi tutvustab autorite ettekanne IV rahvusvahelisel fennougristide kongressil Budapestis 1975. (Faktoranalüüsi kohta lähemalt vt. Түлдана, 1976).

- K e l e m e n , B.** A propos des caractéristiques des styles de la langue à la lumière de la statistique linguistique. - Revue roumaine de linguistique, t. 9, No. 6. Bucarest, 1964, 621-624.
- K e s k ü l a , E.** Adverbi kui sõnaliigi kvantitatiivne esinamus eesti kaasaegse ilukirjandusliku proosa keeles. TPedI lõputöö, juh. A. Villup. Tln., 1972.
- K r a u s , J.** On the Stylistical-Semantic Analysis of Adjectives in Journalistic Style (A Quantitative Approach). - Prague Studies in Mathematical Linguistics. Vol. 4. Prague, 1972, 95-106.
- K õ i v a , S., R a a d i k , E.** Adverbi süntaktilised ja semantilised funktsioonid eesti kaasaegse ilukirjandusproosa autorikõnes. TPedI lõputöö, juh. A. Villup. Tln., 1974.
- Latviesu valodas biežuma vārdnīca.** Atb. red. T. Jakubaite. 3. sējums. Dailliteratūra. 1. daļa. Rīgā, 1972.
- L a u g a s t e , E.** Sõnaalguline ja sisealliteratsioon eesti rahvalauludes. Eesti rahvalaulu struktuur ja kujundid I. - Tõid eesti filoloogia alalt II. TRÜ Toimetised, 234. kd. Tartu, 1969.
- L y o n s , J.** Towards a 'Notional' Theory of the 'Parts of Speech'. - Journal of Linguistics, 1966, No. 2, 209-236.
- M i h k l a , K. jt.** Eesti keele lauseõpetuse põhijooned, I. Lihthause. Toim. E. Nurm. Tln., "Valgus", 1974.
- M u l l e r , Ch.** Etude de statistique lexicale. Le vocabulaire du théâtre de Corneille. Paris, "Larousse", 1967.
- P i l l e r , L.** Über die Häufigkeit der Wortarten im Deutschen. - Linguistica, III. Tartu, 1971, 179 - 189.
- R e n s k ý , M.** The Noun-Verb Quotient in English and Czech. - Philologica Pragensia, VIII. Prague, 1965, 289-302.
- Z s i l k a , T.** Štylistický rozbor básnickej zbierky štatistickou metódou. - Sborník pedagogickej fakulty v Nitre, 11/1967. Spoločenské vedy (jazyk, literatúra). Bratislava, 1967, 143-162.

- Z s i l k a , T. A stilus hirértéke. Bratislava, "Mádách Könyvkiadó", 1973.
- Z s i l k a , T. Stilizatika és statisztika. Budapest, "Akadémiai Kiadó", 1974.
- T ő š i t e l o v á , M. On the So-Called Vocabulary Richness. - Prague Studies in Mathematical Linguistics. Vol. 3. Prague, 1972, 103-120.
- T i i t , E. Matemaatiline statistika, I. Tartu, TRÜ, 1971.
- T i i t , E. Matemaatilise statistika tabelid, II. Tartu, TRÜ, 1972.
- T i i t s , M., V e i l e r , L. Sõnade ja sõnaliikide sagedusest A.H. Tammsaare romaani "Tõde ja õigus" I köites (II ja III osasõnastik). TPedI võistlustöö, juh. A. Villup. Tln., 1974.
- T u l d a v a , J. Statistiline väljavõttemetod keeleteaduses. - Linguistica, I. Tartu, 1969, 5-49.
- T u l d a v a , J. Statistilised testid keeleteaduses. - Linguistica, II. Tartu, 1970, 125-196.
- T u l d a v a , J. Sõnavara statistilisest struktuurist. - Linguistica, III. Tartu, 1971, 211-248.
- T u l d a v a , J. Korrelatsioonanalüüs keeleteaduses (1). - Linguistica, IV. Tartu, 1972, 163-198.
- V a l g e , J. Über die Frequenz der Wortarten in drei funktionalen Stilen. - Generatiivse Grammatika Grupi aastakoosolek. Teesid. Tartu, 1970, 44-45.
- V a l g e , J. Poliitika- ja spordialaste ajalehetekstide statistiliste struktuuride erinevusest. - Keel ja Struktuur, 7. Tartu, 1972, 28-38.
- V a l g m a , J., R e m m e l , N. Eesti keele grammatika. Käsiraamat. Tln., "Valgus", 1968.
- V i l l u p , A. Adverbide esinemissagedusest. - Linguistica, IV. Tartu, 1972, 221-246.
- V i l l u p , A. Adverbaalse sõnavara statistilisest struktuurist. - Emakeele Seltsi aastaraamat, 19-20. Tln., 1975, 73-90.
- А н д р е е в Н.Д. Статистико-комбинированные методы в теоретическом и прикладном языковедении. М.-Л., "Наука", 1967.

- А н д р е е в а Л.Д. Статистико-комбинаторные типы словоизменения и разряды слов в русской морфологии. Л., "Наука", 1969.
- Б е к т а е в К.Б., Л у к љ я н е н к о в К.Ф. О законе распределения единиц письменной речи. - Статистика речи и автоматический анализ текста. Л., "Наука", 1971, 47-112.
- Б е к т а е в К.Б., П и о т р о в с к и й Р.Г. Математические методы в языкознании, ч. 2. Алма-Ата, 1974.
- Б е л о н о г о в Г.Г. Распределение частот появления флективных классов русских слов. - Проблемы кибернетики. Вып. II. М., 1964, 189-198.
- В и н о г р а д о в В.В. Современный русский язык. М., 1938.
- В у к о в и ч Й. К проблеме классификации частей речи. - "Вопросы языкознания", 1972, № 5, 49-61.
- Г а л ь п е р и н И.Р. Информативность единиц языка. Пособие по курсу общего языкознания. М., "Высшая школа", 1974.
- Г о л о в и н Б.Н. Язык и статистика. М., "Просвещение", 1971.
- Г о л о в и н Б.Н. Опыт применения корреляционного анализа в изучении языка. - Вопросы статистической стилистики. Отв. редакторы Б.Н.Головин и В.И.Перебийнос. Киев, "Наукова думка", 1974, 5-16.
- Ж и р м у н с к и й В.М. О природе частей речи и их классификации. - Вопросы теории частей речи. На материале языков различных типов. Л., "Наука", 1968, 7-32.
- З о р е ф М.Г. Машинные основы и машинная морфология в немецко-русском автоматическом словаре. Автореф.канд. дисс. Кишинев, 1972.
- К л о ч к о в а Э.А. О распределении классов слов в некоторых функциональных стилях русского языка. - Вопросы славянского языкознания. Саратов, Изд. СГУ, 1968.
- К л я в и н я С.П. О нормальности распределения частей речи в текстах (на материале современной латышской публицистики). - Latviešu valodas salīdzināmā analīze. LVU Zinātniskie raksti. 118. sējums. Rīgā, 1969.

- Кожина М.Н. О речевой системности научного стиля сравнительно с некоторыми другими стилями. Учебное пособие. Пермь, 1972.
- Кривонос А.Т. Структура языка и система классов слов. - "Вопросы языкознания" 1973, № 4, 86-97.
- Дукьяненко К.Ф. Об одном способе автоматического выявления лексической однородности текста. - Лингвостатистика и автоматический анализ текстов (сборник теоретических статей). Минск, Изд. МГПИИЯ, 1974, 325-338.
- Дятлова А.М. Опыт статистического анализа языка писателя. Автореф. канд. дисс. Л., 1968.
- Мистрик Й. Математико-статистические методы в стилистике. - "Вопросы языкознания", 1967, № 3, 42-52.
- Пак Х.Я. О некоторых статистико-комбинаторных характеристиках функциональных классов (на материале эстонского языка). - Статистико-комбинаторное моделирование языков. М.-Л., "Наука", 1965, 483-489.
- Папина А.Ф., Иванова-Маркова Л. П. Борьба категорий качества и действительности в поэтической речи первой половины XIX века. - Вопросы статистической стилистики. Киев, "Наукова думка", 1974, 237-243.
- Раскина А.А., Чепиго Т.С. Фактографическая ИПС и система микроуниверсалий (на материале русских словоформ). - "Научно-техническая информация. Серия 2. Информационные процессы и системы". М., 1970, №12, 21-28.
- Ревзин И.И. Метод моделирования и типология славянских языков. М., "Наука", 1967.
- Сиротнина О.Б. Некоторые жанрово-стилистические изменения советской публицистики. - Развитие функциональных стилей современного русского языка. М., "Наука", 1968, 101-125.
- Стеблин-Каменский М.И. Спорное в языкознании. Л., Изд. ЛГУ, 1974.
- Сумарокова Л.Н. К вопросу о критериях простоты грамматических систем. - Логика и методология науки. М., "Наука", 1967, 86-91.
- Суник О.П. Общая теория частей речи. М., "Наука", 1966.

- С у н и к О.П. Вопросы общей теории частей речи. - Вопросы теории частей речи. На материале языков различных типов. Л., "Наука", 1968, 33-48.
- Т а р л и н с к а я М.Г. Вспомогательный метод установления авторства стихотворного текста. - Проблемы прикладной лингвистики. Тезисы межвузовской конференции. Вып. 2. М., 1969, 300-304.
- Т и щ е н к о В. Частота частей мови в різних функціональних стилях сучасної української мови. - Питання структурної лексикології. Киев, 1970.
- Т о б и а с Т. Части речи в эстонском языке. - Сообщения по машинному переводу. Вып. I. Таллин, 1962, 90-96.
- Т у л д а в а Ю. Опыт количественного анализа художественного стиля. - *Studia metrica et poetica*, 1. Tartu, 1976 (trütkis).
- Т у р ы г и н а Л.А. Использование статистических методов при определении изменений в грамматическом строе языка (на материале английских и американских публицистических текстов). - Научная конференция "Проблемы синхронного изучения грамматического строя языка". Тезисы докладов и сообщений. М., 1965, 190-192.
- Н и м Х. О лексических категориях в глубинной структуре. - Проблемы моделирования языка, 3.2. Учен. зап. ТГУ, вып. 228. Тарту, 1969, 197-208.
- Ф р у м к и н а Р.М. О законах распределения слов и классов слов. - Структурно-типологические исследования. М., "Наука", 1962.
- Ш а й к е в и ч А.Я. Опыт статистического выделения функциональных стилей. - "Вопросы языкознания", 1968, № I, 64-76.
- Ш т е й н ф е л ь д т Э.А. Частотный словарь современного русского литературного языка. Таллин, 1963.
- Я к у б а й т и с Т.А. Зависимость вида распределения языковых единиц от величины выборки. - Математические методы в языкознании. Рига, "Звайгзне", 1969, 53-58.
- Я к у б а й т и с Т.А. Части речи в художественных текстах. - Известия АН Латвийской ССР, 1974, № 10 (327), 116-126.
- Я к у б а й т и с Т.А., С т у р и т е Б.А. О статистической однородности текстов. - Вопросы статистической стилистики. Киев, "Наукова думка", 1974, 43-54.

О ЧАСТОТНОСТИ ЧАСТЕЙ РЕЧИ В АВТОРСКОЙ РЕЧИ ХУДОЖЕСТВЕННОЙ ПРОЗЫ

Ю. Тулдава, А. Виллуп

Р е з ю м е

В статье приводятся данные о распределении частот частей речи в тексте и словаре эстонского языка. Анализу подвергаются 10 выборок по 5000 словоупотреблений (10 x 5000) из авторской речи современной эстонской художественной прозы. В работе рассматриваются общие и индивидуальные тенденции в употреблении отдельных частей речи (по тексту - см. табл. I-2, по словарю - табл. 7) и их группировок, таких как изменяемые/неизменяемые слова (табл. 5) и полнозначные/неполнозначные слова. Тексты сравниваются на основе индексов соотношений частот частей речи (табл. 9) и выявляются корреляции между частотами частей речи в тексте (табл. 10). Результаты сравниваются с соответствующими данными других подязыков эстонского языка (табл. 3 и 5) и других языков (табл. 4 и 6). На основе сопоставления частот частей речи в тексте и словаре определяются функциональная нагрузка ("индекс итерации", или средняя частота по формуле: t/v , где t - частота в тексте, v - частота в словаре) и информативная нагрузка частей речи в тексте (по формуле r_s/r_t , где r_s - относительная частота части речи в словаре, r_t - относительная частота в тексте). (См. табл. 8). Рассматривается вопрос о репрезентативности материала и статистической достоверности результатов. Статистические данные характеризуются статистическим рядом, арифметической средней, абсолютной и относительной ошибкой, доверительным интервалом и показателем однородности "хи-квадрат" (на уровне достоверности 95 %).

Суммируя основные результаты, приводим данные о частоте появления частей речи в авторской речи современной эстонской художественной прозы на уровне текста и словаря словоформ (данные рассчитаны в среднем на выборку с объемом 5000 словоупотреблений):

Части речи	Относительная частота		Информационная нагрузка P_s/P_t
	в тексте P_t (%)	в словаре P_s (%)	
Имя существительное	31,7	44,2	1,4
Имя прилагательное	6,0	9,5	1,6
Имя числительное	1,1	1,3	1,1
Местоимение	11,4	4,4	0,4
И м е н а вместе взятые	50,2	59,4	1,2
Глагол:			
инфинитивные формы	8,2	12,0	1,5
финитивные формы	14,3	14,6	1,0
Все формы глагола	22,5	26,6	1,2
И з м е н я е м ы е слова вместе взятые	72,7	86,0	1,2
Наречие	15,8	10,7	0,7
Послелог и предлог	3,1	2,2	0,7
Сюж	8,2	0,9	0,1
Междометие	0,2	0,2	1,0
Н е и з м е н я е м ы е слова вместе взятые	27,3	14,0	0,5

STATISTICAL ANALYSIS OF THE PARTS OF SPEECH
IN ESTONIAN FICTION

J. Tuldava, A. Villup

S u m m a r y

The article provides data on the frequency of the parts of speech of the Estonian language in text and vocabulary. 10 samples of 5,000 words each of Modern Estonian prose fiction (non-conversational material) are analyzed. General and individual tendencies in the use of the different parts of speech (on the text level - see Tables 1-2, on the vocabulary level - Table 7) and groups of them, such as inflected/uninflected words (Table 5) and "full-meaning" and form words, are examined. The texts are compared on the basis of the ratios of the parts of speech (Table 9) and the correlations between the parts of speech in a text are fixed (Table 10). The results are compared with corresponding data of other sublanguages of Estonian (Tables 3 and 5) and those of other languages as well (Tables 4 and 6). On the basis of contrasting the parts of speech in the text and in the vocabulary, their functional load (according to the formula t/s , where t - frequency in the text, s - frequency in the vocabulary) and informational load in the text (p_s/p_t - where p_s - relative frequency in the vocabulary and p_t - relative frequency in the text) are calculated (Table 8). The problem of the representativeness of the material and statistical significance of the results is treated. Statistical data are characterized by series of frequencies, arithmetical mean, standard error and confidence intervals on the 95 % level. Homogeneity is verified by the Chi-Square-Test.

In the following table data on the frequency of use of the parts of speech in Modern Estonian fiction (non-conversational material) are presented.

Frequency of Use of Forms in Text and Vocabulary

Parts of speech	Relative frequency		Informational load (p_s/p_t)
	in text p_t (%)	in vocabulary p_s (%)	
Nouns	31.7	44.2	1.4
Adjectives	6.0	9.5	1.6
Numerals	1.1	1.3	1.1
Pronouns	11.4	4.4	0.4
Totals	50.2	59.4	1.2
=====			
Verbs:			
non-infinite forms	8.2	12.0	1.5
finite forms	14.3	14.6	1.0
Totals	22.5	26.6	1.2
=====			
All inflected forms	72.7	86.0	1.2
=====			
Adverbs	15.8	10.7	0.7
Postpositions and prepositions	3.1	2.2	0.7
Conjunctions	8.2	0.9	0.1
Interjections	0.2	0.2	1.0
All uninflected forms	27.3	14.0	0.5