

SÕNAVORMIDE ESINEMUS EESTIKEELISES TEKSTIS

J. Tuldava

Mitmesuguste lingvistilis-pedagoogiliste ülesannete lahendamisel (näit. õppevahendite koostamisel) on kasulik arvestada sõnade esinemissagedust vormide tasandil. Sel juhul selguvad kõige sagedamad käände- või pöördvormid ning muutumatute sõnade kasutusjuhud teatavas kindlas funktsioonis. Eesti keele puhul on võimalik kasutada nimetatud ülesannete lahendamisel ilukirjandusproosa autorikõne sagedussõnastiku esimest osa - sõnavormide sagedussõnastikku, mis avaldati kogumiku "Tõid keelestatistika alalt" esimeses väljaandes (Kaasik, Tuldava, Villup, Käremaa, 1976)^{*}. Alljärgnevalt analüüsime sõnavormide sagedussõnastiku mõningaid üldisi omadusi ja konkreetselt sõnavormide esinemust tekstis eri sõnaliikide kaupa ning käänete resp. ajavormide lõikes.

Sõnavormide sagedussõnastik kasutuskordaja alusel. Eesti tänapäeva ilukirjandusproosa autorikõne sagedussõnastik koostati 20 ilukirjandusliku teksti (osavalimi) põhjal, kusjuures koondvalimi üldmaht oli 100.000 sõnet (täpsemalt 99898 sõnet, mis selgus pärast töötlemist arvutil). Sõnavormide sagedussõnastik koostati täielikult elektronarvuti abil pärast tekstide eelnevat lingvistilist analüüsi ja homonüümsete sõnavormide indekseerimist. Arvuti registreeris sõnavormide esinemuse ka kõigis üksikutes tekstides ja seetõttu oli võimalik arvutada statistilise hajuvuse hinnangud, mis näitasid sõnavormi sageduse kõrval ka sõnavormi esinemuse stabiilsust (või ebastabiilsust) vaadeldavate tekstide lõikes. Sõnavarastatistika praktikas on tuntud nn. kasutuskordaja U, mis arvestab korruga nii sõna sagedust kui ka esinemise stabiilsust (vt. lähemalt Kaasik, Tuldava, Villup, Käremaa, 1976, 109-110). Eesti keele sõnavormide sage-

^{*} Eesti tänapäeva ilukirjandusproosa autorikõne sagedussõnastiku teine osa - lekseemide sagedussõnastik on avaldatud kogumiku teises väljaandes (vt. Kaasik, Tuldava, Villup, Käremaa, 1977).

dussõnastiku puhul võetigi aluseks kasutuskordaja U väärtused sõnavormide järjestamisel sageduse ja stabiilsuse koondnäitaja põhjal. Sõnastiku mahtu piirati sel teel, et sõnastikku lülitati ainult need sõnavormid, mille kasutuskordaja ehk "modifitseeritud sagedus" oli üle 1 ($U \geq 2$). Selliseid sõnavorme oli kokku ligikaudu 3000, mis tekstist katavad umbes 65 %. Praktiliselt tähendab see seda, et seades piiri $U \geq 2$, jäävad sagedussõnastikku ainult sellised sõnavormid, mis esinevad vähemalt kolmes tekstis kahekümnest ning mille üldine esinemissagedus koondvalimis peab olema vähemalt 4. Sel teel piiritletud sõnastiku osa vastab kõrg- ja kesksagedustsooni sõnavarale ja sagedussõnastikku võib vaadelda vastava allkeele põhisõnavara tselise esindajana (sõnavormide tasandil).

Lähtudes kasutuskordaja U abil moodustatud sõnavormide sagedussõnastikust, võtame esmalt vaatlusele **s õ n a s t i k u a l g u s o s a**, s.o. kõige sagedamate sõnavormide loendi.

100 sagedama sõnavormi hulgas leidub ainult üks nimi-sõnavorm (mees). See asub sagedusjärjestuses 60. kohal (astak $r = 60$). Verbivorme esineb saja esimese sõnavormi hulgas 14 (tuli, tuleb, läks, hakkas, jäi, võib, peab ja verbi olema vormid järjestuses on, oli, pole, polnud, oleks, olid (pl), olla). Ülejäänud 85 sõnavormi jagunevad järgmiselt:

- 34 määrsõna (ei, siis, ka, aga, nii, veel, ja, juba, nüüd jt.);
- 31 asesõna (ta, see, oma, mis, seda, ma, kes, kõik, nad jt.);
- 11 sidesõna (ja, et, kui, nagu, ning, aga, või, kuid jt.);
- 7 kaassõna (üle, pärast, vastu, all, poole, peale, juures);
- 2 arvsõna (üks, kaks).

100 sagedama sõnavormi hulgas pole ühtki omadussõna. Sagedaim adjektiivivorm hea (nom) on järjestuses 112. kohal. Määrsõnade hulgas on viis ühendverbi komponenti - prefiksaaladverbi (ära, välja, tagasi, maha, vastu). Iseseisvalt määrsõnadest on sagedamad siis, nüüd, kus, kuidas, siin, seal, kõige, palju, rohkem.

Kui vaadelda sõnavormide esinemise dünaamikat sõnaliikide järgi iga saja astaku kaupa sagedussõnastiku algusest arvates, võime täheldada peamiselt nimisõnade osatähtsuse järkjärgulist suurenemist (vt. tabel 1). Esimese saja vormi hulgas on nimisõnu 1 %, teise saja hulgas 10 %, kolmanda saja hulgas juba 21 % jne. Kümnenenda saja hulgas, s.o. astakute vahemikus 901 - 1000 on nimisõnavorme 31 %. Samas võib aga nentida määr-, ase- ja sidesõnade osatähtsuse vähenemist iga järgneva saja sõnavormi puhul. Hea ülevaate saame, kui esitame andmed kumulatiivsete sageduste (järgsummade) alusel (tabel 2).

T a b e l 2

Sõnavormide jaotumus (%) sõnaliikide järgi sagedussõnastiku algusosas (kasutuskordaja U alusel)

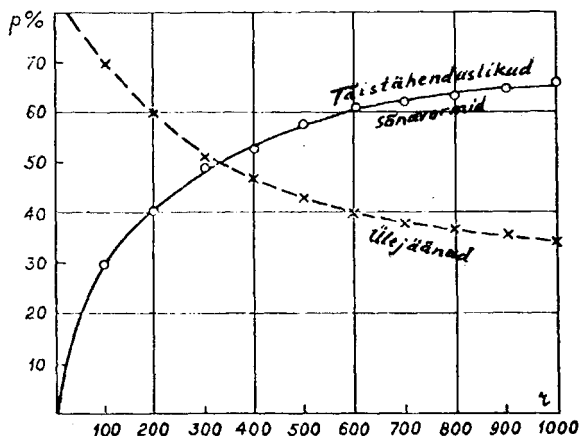
Astak Sõnaliik	1 ÷ 100	1 ÷ 500	1 ÷ 1000
Nimisõna	1	16	21
Tegusõna	14	23	26
Määrsõna	34	29	25
Asesõna	31	15	13
Sidesõna	11	3	2
Kaassõna	7	8	6
Arvsõna	2	2	2
Omadussõna	-	4	5
Kokku	100 %	100 %	100 %

Tabelist nähtub, et ka sageduste kuhjumisel ilmneb eriti nimisõnavormide osatähtsuse järkjärguline suurenemine. See tendents jätkub edaspidigi: kui tuhande sagedama sõnavormi seas on nimisõnavorme 21 %, siis teise tuhande seas on neid umbes 30 % ning kogu koondsõnastikus üle 40 % (vrd. Tuldava, Villup, 1976). Ka tegu- ja omadussõnade osas võib täheldada nende osakaalu suurenemist koos astaku suurenemisega, seevastu aga väheneb järk-järgult ase-, side- ja määrsõnade protsent. Tuhanda sagedama sõnavormi hulgas leidub ilukirjandusproosa autorikõnes ainult 3 hüüdsõna (0,3 %), mida tabelisse pole võetud.

Sõnavormide esinemise dünaamikat sagedussõnastikus võib vaadelda ka lähtudes sõnaliikide klassifitseerimisest täis-

tähenduslikeks ja mittetäistähenduslikeks.^{*} Kõigi täistähenduslike sõnavormide (sõnaliikide) suhtes võib konstateerida nende osatähtsuse suurenemist iga saja astaku kaupa. Graafilises kujutuses võib täheldada ühtlast esinemissageduse tõusu nii üksikute täistähenduslike sõnaliikide kui ka nende kogumi suhtes, kuna aga ülejäänud sõnaliikide puhul on märgata kasvutempo järkjärgulist vähenemist.

Eriti kujukalt tuleb erinevus täistähenduslike ja ülejäänud sõnavormide esinemisdünaamikas ilmsiks kumulatiivsete sageduste protsentuaalsel esitamisel (joon. 1).



Joon. 1. Täistähenduslike ja ülejäänud sõnavormide osatähtsus sagedussõnastikus (r - astak, p % - sõnastiku osa protsentides).

Jooniselt nähtub, et sagedussõnastiku algusosas domineerivad mittetäistähenduslikud sõnavormid, kuid nende suhteline osatähtsus väheneb järk-järgult. Umbes astaku $r = 350$ kohal on täistähenduslike ja ülejäänud sõnavormide osatähtsus võrdne (kumbagi 50 %). Siitpeale algab täistähenduslike sõnavormide ülekaal, mis järjest suureneb, kuni $r = 1000$ kohal on

* Antud juhul peame täistähenduslikeks nimi-, omadus- ja tegusõnu ning iseseisvaid määrsõnu.

vahekord 67:33 täistähenduslike sõnavormide kasuks.

Nimisõnavormide ja käänete sagedused. Nimisõnavorme on esimese tuhande sagedama sõnavormi hulgas 209, kusjuures esimeses ja teises sajas on neid kokku ainult 11. Kõige sagedam ja levinum nimisõnavorm ilukirjandusproosa autorikõnes on mees (kasutuskordaja U järgi koostatud sagedussõnastikus asub see vorm 60. kohal U väärtusega 120). Järgnevad ajal (astak r = 117 ja U = 53), aega (r = 135, U = 53), silmad (r = 147, U = 51), naine (r = 151, U = 48) jt. Alates kolmandast sajast tõuseb sagedussõnastikus järsult nimisõnavormide arv ja neid esineb edaspidi järjekindlalt 20-30 iga järgneva saja sõnavormi kohta (vt. tabel 1).

Käänete järgi leidub tuhande sagedama sõnavormi hulgas kõige rohkem nimisõnavorme nimetavas käändes (74 eri vormi, s.o. 35 % kõigist nimisõnavormidest). Sagedamad nimisõnavormid ainsuse nimetavas käändes on järgmised (sulgudes U väärtus)*: mees (120), naine (48), asi (42), isa (41), poiss (40), inimene (35), aeg (34), nägu (33), päike (33), ema (30), kord (29), tuul (28), meri (24), tunne (23), elu (20) jt. Mitmuse nimetavas käändes on sagedamad: silmad (51), mehed (35), inimesed (30), käed (30), poisid, naised, lapsed, asjad, mõtted, jalad jt.

Omastavas käändes kasutatud nimisõnu on tuhande sagedama sõnavormi hulgas 44 (21 % kõigist nimisõnavormidest). Ainsuses on sagedamad: mehe (37), ukse (30), elu (28), naise (27), inimese (26); mitmuses: inimeste (27), meeste (10), silmade (7).

Ka osastav kääne on esindatud sagedamate sõnavormide loendis küllaltki rohkearvuliselt: esimese tuhande hulgas leidub neid 41 (nimisõnadest 20 %). Sagedamad vormid ainsuses on aega (53), korda (37), aastat (35), meest (20), nägu (19), tööd (19) jt.; mitmuses: inimesi (22), sõnu (20), asju (18), silmi (17), aastaid (14), mehi (13), mõtteid (9). Osas-

* Kasutuskordaja U väärtus võib ideaalse esinemisstabiilsuse korral (kui antud sõnavorm esineb kõigis tekstides täpselt ühesuguse sagedusega) võrduda tegeliku esinemissagedusega tekstis (F). Üldiselt on aga kasutuskordaja väärtus alati mõnevõrra väiksem tegelikust sagedusest (U < F). Esitame artiklis U väärtused eri sõnavormide kvantitatiivsete osatähtsuste võrdlemise huvides.

tavas käändes kasutatud nimisõnade hulgas võib eristada väljendverbide komponente: aru (saama, pidama) - 34, juttu (ajama, puhuma) - 9, korda (saama, minema) - 8, osa (võtma, saama) - 8, sõna (kuulma) - 3 jt.

Teistest käänetest võib nimetada alalütlevat, sees- ja sisseütlevat, mille vorme esineb küllaltki palju tuhande sagedama sõnavormi loendis. Vaadeldes sagedamaid nimisõnavorme alalütlevas käändes, võime mentida, et need tähistavad peamiselt aega, näiteks ajal (58), õhtul (44), hommikul (34), päeval (34), hetkel (15), aastal (13), talvel (11), suvel (10), õöl (10), kevad (9). Sagedamate sõnavormidest tähistavad kohta laual (11), teel (8), tänaval (7). Seesütleva käände vormid on aga sagedamate sõnavormide loendis põhiliselt lokaalse tähendusega, näit. kodus (20), toas (15), met-sas (11), majas (10), koolis (7), nurgas (7). Küsimusele 'kus?' vastavad ka vormid käes (18), keeles (7), kuna aga elus (9) ja mõttes (9) võimaldavad mitmesugust tõlgitust. Sisseütleva käände vormidest on sagedamad tuppa (13), pähe näkku (9), suhu (8) - seega geminaat- või h-illatiivi vormid - ja mitmusevorm silmadesse (7).

Alalütlevas ja seesütlevas käändes esinevad mitmed adverbiaalselt kasutatud nimisõnavormid (pooladverbid), nagu rahal (4), kohal (2), seljas (13), peas (9), nõus (4), korras (4), käes (3). Näiteid:

Ma olen sisemiselt rahal, et Tallinna on otsustatud lõpuni kaitsta. (P. Kuusberg)

Või nüüdki Marel toitu pillata käes oli ...

(A. Hint)

Sisseütleva käände vorm on aga sageli väljendverbi komponendiks. Näiteks esinevad tuhande sagedama sõnavormi hulgas meelde (jätma, jääma) - 30, otsa (lõppema, saama) - 21, kätte (jõudma, saama, leidma) - 16, appi (minema, tulema) - 9, korda (saama, minema) - 8, silma (paistma, hakkama) - 8; peale selle veel lendu (tõusma) - 5, toime (tulema) - 5, pihta (saama) - 4, ellu (viima, rakendama) - 3, kaela (määrima) - 3.

Nimisõnavorme kaasaütlevas käändes leidub tuhande sagedama sõnavormi hulgas 6: käega (17), näoga (12), silmaga (11), peaga (10); mitmusevormid silmadega (12), kättega (7). Sagedamad kaasaütleva vormid on seega põhiliselt "instrumentaal-

Tabel 3

Nimisõna käändevormide esinemus tekstis ja sõnastikus üksikvalemi (mahuga $N = 5000$ sõnet) andmeil (A. Beekmani teose "Kartulikuljused" põhjal)

Kääne	Tekst		Sõnastik		Funktsionaalne koormus (m_t/m'_s)
	arv (m_t)	%	arv (m_s)	%	
Nimetav	597	35,0	318	27,0	1,88
Omastav	386	22,6	259	22,0	1,49
Osastav	250	14,7	189	16,0	1,32
Sisseütlev	62	3,6	51	4,3	1,22
Seesütlev	55	3,2	52	4,4	1,06
Seestütlev	82	4,8	78	6,6	1,05
Alaleütlev	80	4,7	71	6,0	1,13
Alalütlev	85	5,0	63	5,4	1,35
Alaltütlev	12	0,7	10	0,9	1,20
Saav	13	0,8	13	1,1	1,0
Rajav	3	0,2	3	0,3	1,0
Olev	11	0,6	11	0,9	1,0
Ilmaütlev	1	0,1	1	0,1	1,0
Kaasaütlev	69	4,0	59	5,0	1,17
Kokku	1706	100,0	1178	100,0	keskm. 1,45

se" tähendusega või väljendavad viisi. Alles teises tuhandes ilmuvad mõningad "puhtkomitatiivsed" sõnavormid, nagu mehega (5), isaga (4), emaga (3), naisega (3).

Ülejäänud käänded on tuhande sagedama sõnavormi hulgas esindatud vähesel määral, näit. alaleütleva vormid lauale (9), mehele (8), inimestele (7); seestütleva vormid käest (13), aknast (10); saava käände vormid lõpuks (22), hetkeks (18). Alaleütleva käände vormid esinevad ka väljendverbi komponendina, näit. tähele (panema) - 5, otsusele (jõudma) - 2, rahule (jätma, jääma) - 2.

Tuhande sagedama sõnavormi hulgas puuduvad nimisõnavormid alaltütlevas, rajavas, olevas ja ilmaütlevas käändes.

Eespool vaatlesime sagedamate nimisõnavormide jaotumust käänete kaupa. Järgnevalt käsitleme nimisõna k ä ä n e t e sagedusi kokkuvõtlikult nii sõnastikus kui ka tekstis, kusjuures võrdlemisel lähtume üksikvalimi (A. Beekmani teosest "Kartulikuljused"; vt. tabel 3). Selgub, et nii sõnastikus kui ka tekstis on kuus sagedamat käänet (nimisõnade puhul) järgmised:

	Sõnastik (%)	Tekst (%)
Nimetav	27,0	35,0
Omastav	22,0	22,6
Osastav	14,7	16,0
Seestütlev	6,6	4,8
Alaleütlev	6,0	4,7
Alalütlev	5,4	5,0
	<hr/>	<hr/>
	83,0	86,8

Kuue sagedama käände kontsentratsioon tekstis (86,8 %) on mõnevõrra suurem kui sõnastikus (83,0 %). Sõna käänete keskmine korduvus tekstis (tekstisageduse ja sõnastikusageduse suhe, mida käsitleme käänete funktsionaalse koormusena) näitab, et kõige suurem funktsionaalne koormus on nimetaval, omastaval ja alalütleva käändel (vt. tabel 3). See tähendab seda, et nimetatud käänetes kasutatud nimisõnad on keskmiselt suurema korduvusega kui teistes käänetes esinevad nimisõnad. Näit. alalütleva käände puhul on suure funktsionaalse koormusega sellised sõnavormid nagu ajal, õhtul, hommikul, päeval jt.

Kui vaadelda käänete sagedusi eraldi ainsuses ja mitmuses nii teksti kui ka sõnastiku tasandil (tabelid 4 ja 5), siis selgub, et sõnastikus moodustavad nimisõna ainsusevor-

Tabel 4

Nimisõna käändevormide esinemus tekstis
üksikvalimil (N = 5000) andmeil

Kääne	Ainsus		Mitmus		Kokku	
	arv	%	arv	%	arv	%
Nimetav	408	33,9	109	41,0	597	35,0
Omastav	342	23,8	44	16,5	306	22,6
Osastav	193	13,4	57	21,4	250	14,7
Sisseütlev	58	4,0	4	1,5	62	3,6
Seesütlev	46	3,2	9	3,4	55	3,2
Seestütlev	74	5,2	8	3,0	82	4,8
Alaleütlev	71	4,9	9	3,4	80	4,7
Alalütlev	75	5,2	10	3,7	85	5,0
Alaltütlev	11	0,8	1	0,4	12	0,7
Saav	12	0,8	1	0,4	13	0,8
Rajav	2	0,1	1	0,4	3	0,2
Olev	10	0,7	1	0,4	11	0,6
Ilmaütlev	-	-	1	0,4	1	0,1
Kaasaütlev	58	4,0	11	4,1	69	4,0
Kokku	1440	100,0	266	100,0	1706	100,0
Ainsus/ mitmus	84,4 %		15,6 %		100,0 %	

Tabel 5

Nimisõna käändevormide esinemus sõnasti-
kus üksikvalimil (N = 5000) andmeil

Kääne	Ainsus		Mitmus		Kokku	
	arv	%	arv	%	arv	%
Nimetav	239	24,6	79	38,0	318	27,0
Omastav	218	22,5	41	19,7	259	22,0
Osastav	149	15,4	40	19,2	189	16,0
Sisseütlev	47	4,8	4	1,9	51	4,3
Seesütlev	45	4,6	7	3,4	52	4,4
Seestütlev	74	7,6	4	1,9	78	6,6
Alaleütlev	62	6,4	9	4,3	71	6,0
Alalütlev	54	5,6	9	4,3	63	5,4
Alaltütlev	9	0,9	1	0,5	10	0,9
Saav	12	1,3	1	0,5	13	1,1
Rajav	2	0,2	1	0,5	3	0,3
Olev	10	1,0	1	0,5	11	0,9
Ilmaütlev	0	0	1	0,5	1	0,1
Kaasaütlev	49	5,1	10	4,8	59	5,0
Kokku	970	100,0	208	100,0	1178	100,0
Ainsus/ mitmus	82,3 %		17,7 %		100,0 %	

mid kokku 82,3 %, tekstis aga 84,4 %. Kõigis käänetes on ainsusevormid arvulises ülekaalus, kuid käänete suhteline osatähtsus ainsuses ja mitmuses eraldi võttes on erinev. Nii näiteks on mitmuse nimetav kääne nii sõnastikus kui ka tekstis suurema suhtelise sagedusega kui ainsuse nimetav kääne (tekstis vastavalt 41,0 ja 33,9 %, 38,0 ja 24,6 %; vt. tabelid 4 ja 5). Ka osastava käände suhteline osatähtsus mitmuses on suurem kui ainsuses, kuna aga omastav kääne on suurema funktsionaalse koormusega ainsuses.

T a b e l 6

Käänete suhtelised sagedused ilukirjandusproosa autorikõne tekstis kõigi käändsõnade ja nimisõnade alusel

Kääne	Kõik käändsõnad (%)	Nimisõnad (%)	Vahe
Nimetav	39,4	35,0	-4,4
Omastav	19,0	22,6	+3,6
Osastav	15,7	14,7	-1,0
Sisseütlev	2,1	3,6	+1,5
Seesütlev	4,1	3,2	-0,9
Seestütlev	3,9	4,8	+0,9
Alaleütlev	3,4	4,7	+1,3
Alalütlev	6,1	5,0	-1,1
Alaltütlev	0,3	0,7	+0,4
Saav	2,2	0,8	-1,4
Rajav	0,3	0,2	-0,1
Olev	0,5	0,6	+0,1
Ilmaütlev	0,1	0,1	-
Kaasaütlev	2,9	4,0	+1,1
Kokku	100,0	100,0	-

Nimisõnade käänete sagedusi võib võrrelda käänete sagedustega, mis on arvatatud kõigi käänduvate sõnade põhjal (viimaste kohta vt. Valge, 1970).

Teksti tasandil on mõlemal juhul kolm sagedamat käänet: nimetav, omastav, osastav (vt. tabel 6). Ilmneb aga asjaolu, et nimetav ja osastav kääne esinevad nimisõnade puhul suhteliselt vähem kui käändsõnade puhul kokku, kuna aga omastav kääne on eriti omane nimisõnale (suhteliste sageduste vahe

+3,6 nimisõna kasuks). Tüüpilised "nimisõnalised" käänded on ka sisseütlev, alaleütlev ja kaasaütlev. Vastandi moodustab (nimetava ja osastava kõrval) saav kääne, mille suhteline sagedus tekstis on nimisõnade puhul ainult 0,8 %, kõigi käändsõnade puhul aga 2,2 %. See on seletatav selliste asesõnavormide nagu milleks, kelleks, selleks suhteliselt suure sagedusega. Nimisõnavorme saavas käändes esineb harva, sagedamatest nimetasime eespool lõpuks ja hetkeks.

Verbivormide ja aegade sagedused. Tuhanda sagedama sõnavormi hulgas on verbivorme 263 (vt. tabel 1), kusjuures nende esinemus alates kolmandast sajast võtab enam-vähem stabiilse ilme. Kõige rohkem on minevikuvorme (92), neist on sagedamad oli, olid (pl), tuli, läks, hakkas, jäi, sai, vaatas, ütles, tegi, võis, võttis, istus, käes, rääkis. Täismineviku eitavatest vormidest on sagedamad: polnud, (ei) saanud, (ei) olnud, (ei) teadnud, (ei) tahtnud.

Olevikuvorme on tuhande sagedama sõnavormi nimistus 59, neist on kõige sagedamad on, tuleb, võib, peab, saab, läheb, jääb, hakkab, teeb, teab. Tingiva kõneviisi olevikuvormidest võib nimetada vorme oleks, peaks, võiks, oleksid (pl), oleksid (pl), oleksin, saaks; eitav vorm poleks.

Tuhanda sagedama sõnavormi hulgas on da-infinitiive 42 ja ma-infinitiive 13. da-infinitiividest on sagedamad olla, teha, näha, minna, vaadata, teada, panna, võtta, saada, rääkida, mõelda, tulla, hakata, oodata; des-vormidest minnes, tulles, vaadates, arvates. ma-infinitiivide sagedamad vormid on minema, olema, tegema, seisma, vaatama, otsima, võtma, nägema, tulema, ütleva, rääkima, astuma ja mas-vorm olemas.

Küllaltki palju (22) esineb tuhande sõnavormi hulgas nud-partitsiipe: olnud, saanud, läinud, jäänud, näinud, teinud, tulnud jt. tud-vormidest esineb ainult antud. Sagedamate sõnavormide hulka kuuluvad üksikult ka vat-vorm olevat ja umbisikuline räägiti. Käskivat kõneviisi esindavad olgu, mine, ära. Peale selle on tuhande sagedama sõnavormi hulgas kaks verbivormi liitega -gi: oligi, ongi.

Eriuurimuse alusel (neli üksikvalimit üldmahuga N = 20.000 sõnet) oleme varem kindlaks teinud verbi a e g a d e sagedused ilukirjandusproosa tekstis (autori- ja tege-laskõne koos; vt. tabel 7).

T a b e l 7

Verbide ajavormid ilukirjandusproosa tekstis
(teksti üldmaht N = 20.000 sõnet)

Kõneviis Aeg	Kindel	Tingiv	Kaudne	Käskiv	Kokku		
					arv	%	
Olevik	1272	170	14	52	1508	42,6	
Lihtminevik	1724	-	5	-	1729	48,9	
Täisminevik	101	52	2	5	160	4,5	
Ennemindevik	141	-	-	-	141	4,0	
Kokku	arv	3238	222	21	57	3538	-
	%	91,5	6,3	0,6	1,6	-	100,0

Sageduselt esikohal on lihtminevik (48,9%), millele järgneb olevik (42,6 %). Täisminevik ja ennemindevik moodustavad kokku ainult 8,5 %. Kõneviisidest kuulub esikoht kindlale kõneviisile (91,5). Tingiv kõneviis moodustab kõigist teksti verbivormidest ainult 6,3 %, käskiv kõneviis 1,6 % ja kaudne kõneviis 0,6 %. Kui vaadelda toodud andmeid tegumoodide seisukohast, siis selgub, et isikuline tegumood moodustab teksti verbivormidest 96,5 % ja umbisikuline tegumood - 3,5 %. Kõigist umbisikulise tegumoe vormidest langeb 47 % olevikule, 22 % ennemindevikule, 19 % lihtminevikule ja 8 % täisminevikule.

Eraldi vaatleme veel partitsiibi- ja infinitiivivormide esinemissagedusi tekstis. Partitsiibid jaotuvad järgmiselt:

Olevik: isikuline (-y)	49,1 %
umbisikuline (-tav)	7,0 %
Minevik: isikuline (-nud)	19,3 %
umbisikuline (-tud)	24,6 %
	100,0 %

Infinitiivide hulgas on 35 % ma-infinitiive ja 65 % da-infinitiive. Eri vormide suhteliste sageduste jaotumus on järgmine (kõrvutatult andmetega A.H. Tammsaare romaani "Tõde ja õigus" I põhjal, vt. Kaljund, 1970):

Tänapäeva autorid A.H. Tammsaare
(keskmine)

- <u>ma</u>	29,5	31,5
- <u>mas</u>	0,9	2,5
- <u>mast</u>	0,1	0,5
- <u>mata</u>	4,5	4,2
- <u>da</u>	56,1	50,4
- <u>des</u>	8,9	10,9
	<hr/>	<hr/>
	100,0 %	100,0 %

Esitatud andmetest nähtub, et tänapäeva ilukirjandusproosa tekstis ilmneb tendents kasutada rohkem da-vorme, võrreldes A.H. Tammsaare keelepruugiga, kuna aga ma- ja des-vorme kasutavad tänapäeva autorid suhteliselt vähem. A.H. Tammsaarele on iseloomulik ka mas-vormide suhteliselt kõrge esinemissagedus tekstis.

Muud sõnaliigid. Ilukirjandusproosa autorikõne sõnavormide sagedussõnastikus on esindatud kõik tähtsamad s i d e s õ n a d. Neist kuuluvad tuhande sagedama sõnavormi hulka ja, et, kui, nagu, ning, aga, või, kuid, sest, vaid, kuigi, ent, kuni (antud järjestuses), kusjuures loetletud sidesõnade puhul on arvestatud ka nende esinemist ühendsidesõna osisena. Eraldi märgitakse sõnastikus ühendsidesõna esimese komponendi sagedus, nii näit. on kõige sagedam esikomponent nii sagedusjärjestuses (U alusel) 127. kohal. Järgnevad ei, enne, kas, ainult, mitte - esimese tuhande hulgas - ning olguigi, niipea, ilma ülejäänud sõnastikus. Võib nentida erinevusi sünonüümsete konjunktsioonide sagedustes (esimene arv tähistab astakut ja teine arv kasutuskordaja väärtust):

ja (1/3138) - ning (16/315);
aga (17/313) - kuid (41/156) - ent (275/27).

Tänapäeva eesti keelele on iseloomulik sidesõnade kasutamine ka rõhumäärsõna funktsioonis. Ilukirjandusproosa autorikõne andmeil on see funktsioon eriti omane sõnale aga, mis rõhumäärsõnana paikneb üldises sagedusjärjestuses 14. kohal (U = 359), sidesõna funktsioonis 17. kohal (U = 313). Tunduvalt rohkem esineb rõhumäärsõnana ehk (U väärtused vastavalt 35 ja 4). Umbes ühepalju kasutatakse kummaski funktsioonis sõnavormi ent (rõhumäärsõnana U = 23, sidesõnana U = 27). Ülejäänud juhtudel tuleb aga nentida sidesõna funktsiooni

siooni prevaleerimist (U väärtused vastavalt side- ja rõhu-
määrsõnana): ja (3138 - 225), et (1186 - 23), nagu (515 -
18), ning (315 - 24), või (292 - 29), kuid (156 - 92), sest
(126 - 12), ega (117 - 14), vaid (72 - 31). Mõned sõnad
võivad esineda niihästi sidesõnana kui ka iseseisva määrsõnana,
näit. kui (sidesõna funktsioonis U = 951, iseseisva
määrsõnana U = 36) ja kuigi (U väärtused vastavalt 52 ja 7).

K a a s s õ n a d e arv vaadeldavas sõnastikus läheneb
sajale, kusjuures esimese tuhande sõnavormi hulgas esineb 55
kaassõna. Kõige suurema esinemissagedusega on üle, pärast,
vastu, all, poole, peale, juures (esimese saja sõnavormi
hulgas). Järgnevad juurde, läbi, ees, vahel, ümber, eest,
mööda, taga, alla, poolt, enne, kohal (teise saja sagedama
sõnavormi hulgas). Mõned kaassõnad võivad teatavasti esine-
da üksikutes käändevormides, moodustades rühmad, näit. alla
- all - alt. Meie materjal võimaldab kindlaks teha ka rühma-
siseseid sagedussuhteid, näit. moodustuvad järgmised kolmi-
kud (üksikud rühmal liikmed on järjestatud kasutusordaja U
suuruse järgi):

all 101 - alla 43 - alt 21
ees 59 - eest 54 - ette 26
juures 69 - juurde 65 - juurest 8
järele 22 - järel 17 - järelt 0
(järgi 32)
hulgas 5 - hulka 2 - hulgast 1
kohal 37 - kohale 3 - kohalt 0
kõrval 34 - kõrvale 13 - kõrvalt 0
käes 7 - kätte 7 - käest 5
külge 10 - küljes 5 - küljest 4
otsa 8 - otsas 8 - otsast 3
peale 80 - peal 27 - pealt 12
poole 98 - poolt 38 - pool 17
sisse 11 - sees 4 - seest 3
taga 52 - taha 28 - tagant 18
vahel 58 - vahele 19 - vahelt 18
ääres 23 - äärde 6 - äärest 1

Suuremal osal juhtudel on sageduselt esikohal kaassõ-
nad sees- või alalütlevas käändes. Erandina on esikohal sis-
se- ja alaleütleva vormid järele (ja järgi), külge, peale,
poole, sisse. Seest- ja alaltütleva vormiga kaassõnad on
vastavas rühmas kõige väiksema sagedusega, välja arvatud
eest ja poolt.

Ühed ja samad sõnavormid esinevad eesti keeles sageli
kaassõna ja määrsõna funktsioonis - nii abimäärsõnana
(ühend- või väljendverbi komponendina) kui ka iseseisva

määrsõnana. Näit. sõnavormi alla puhul on vastavad kasutus-
kordaja väärtused 43 - 26 - 10, s.t. vorm esineb kõige sa-
gedamini kaassõnana, seejärel abimäärsõnana ja lõpuks ise-
seisva määrsõnana. Toome veel mõned näited homonüümsete sõ-
navormide esinemuse kohta: ette (26-41-3), juurde (65-12-0),
järele (22-22-0), kõrvale (13-19-1), peale (80-44-0), peal
(27-1-3), sisse (11-45-9), taga (52-9-3). (Täielik nimistu
leidub lekseemide sagedussõnastikule lisatud loendis 12, kus
esinemissagedus on antud tavalise esinemissageduse F järgi,
vt. Kaasik, Tuldava, Villup, Käremaa, 1977, 138-139)

Mõned kaassõnad võivad kokku langeda ka nimisõnavormi-
dega, näit. kätte, käes, käest; otsa, otsas, otsast. All-
järgnevad näited illustreerivad niisuguste sõnavormide ka-
sutamist kaassõna funktsioonis:

Süüa sai ta talust pererahva käest.

(A. Hint)

Enne propsikoormaga Inglismaale jõudmist
sattus "Käina" tormi kätte. (A. Hint)

Naine tõusis, vinnas lüpsiku käe otsa ja
astus pika sammuga mehest mööda. (A. Beekman)

Ta võttis konksu otsast Joosepi sinise
treeningu ... (A. Beekman)

Nimisõnana tuleb käsitada järgmisi juhte:

Kukalt kattis triibuline sall ja ühes
käes kandis eit kirbukirja labakinnast.

(A. Beekman)

Iga kartul, mis käest lendas ...

(A. Beekman)

Küüni otsas lahtise kuuri all seisis tur-
bahunniku kõrval looreha. (A. Beekman)

Nimisõnalist kasutust esineb ülaltoodud vormide puhul
suhteliselt vähem, kuid küllaltki sageli kohtab adverbiaal-
selt kasutatud nimisõnavorme (pooladverbe), näit.:

Ja ära usu, et on käes see kõige hullem.

(R. Sirge)

Sagedamate määrõnade hulgas domineerivad rõhumäärõnad ja iseseisvatest määrõnadest nn. pronominaalsed adverbid (proadverbid), s.o. "abstraktse, neutraalse sisuga adverbid, mis väljendavad üldistatult aega, kohta või määrat" (Villup, 1975, 81). Saja sagedama sõnavormi hulka kuuluvad rõhumäärõnad ei, ka, aga, ja, ainult, kas, mitte, küll, isegi, ju, kuid, just, muidugi, vist ja iseseisvatest määrõnadest siis, nii, veel, juba, nüüd, enam, kus, kuidas, siin, seal, kõige, palju, jälle, väga, rohkem. Varem juba nimetasime, et sellised sõnad nagu ja, aga, ent, kuid esinevad ühtlasi sidesõna funktsioonis. Paljud määrõnad on mitmetähenduslikud selles mõttes, et neid kasutatakse nii iseseisva kui ka rõhumäärõnana. Tavaliselt on iseseisva määrõna suures ülekaalus, näit. nii (iseseisva määrõnana U = 308, rõhumäärõnana U = 16), siis (407-55), juba (225-11), veel (276-8), nüüd (204-5), jälle (88-2), kohe (51-4). Funktsioonide sagedused on lähedased sõnadel ikka (62-43) ja hoopis (32-30). Üksikuid sõnu kasutatakse valdavalt rõhumäärõnana, näit. otse (13-31) ja lihtsalt (5-28). Esitame mõningad näited ülalloodud sõnade kasutamise kohta rõhumäärõna funktsioonis:

See võttis juba märksa enam aega ning kestis kevadeni välja. Nii oli ta ikkagi varastel hommikutundidel koolitöö kõrval teist jumalat teeninud. (A. Hint)

Nõnda siis oli venna sissetulek majast üksinda juba igas kuus kuussada krooni ... (A. Hint)

Naised on juba kord niisugused, olgud nad kas vanad moorid või kooliplikad. (P. Kuusberg)

Ta näib olevat otse hullul kombel noolanud igasuguseid tarku raamatuid. (P. Kuusberg)

Piidlesin vargsi, kuidas ta rahulikult ja täpselt oma lihtsalt liiga ilusaid jalgu teineteise ette tõstis. (T. Kallas)

Pakub huvi võrrelda mõningate tähenduselt lähedaste määrõnade kasutatavust ilukirjandusproosa autorikõnes (arvud tähistavad kasutuskordaja U väärtust):

a) rõhumaärsõnad:

ja 225 - ning 24
aga 359 - kuid 92 - ent 23
vist 68 - võib-olla 66 - ehk 35 - või 29 - arvatavasti
 9 - vahest 3
ainult 196 - vaid 31 - üksnes 17

b) iseseisvad määrsõnad:

nüüd 204 - praegu 64
enam 152 - rohkem 69
nii 305 - nõnda 18 - niiviisi 18
jälle 88 - uesti 54 - taas 12
kiiresti 13 - ruttu 3
tihti 9 - sageli 7
mõnikord 32 - vahetevahel 11

A s e s õ n a d kuuluvad sagedate sõnade tsooni. Meie koondsõnastikus esineb neid esimese saja sõnavormi hulgas 31. Kõige sagedamad isikulise asesõna vormid on ta (nom), ma, nad, teda, tema (gen), ta (gen), tal, talle, tema (nom), meid, nende, me (nom), meie (gen), mulle, mu, mind. Esimese isiku asesõna vormid esinevad autorikõnes neil juhtudel, kui autor kasutab minavormi kogu teose ulatuses või osaliselt.

Saja sagedama sõnavormi hulgas on näitava asesõna vormid: see, seda, selle, sellest, need; omastav asesõna oma; küsiva ja siduva asesõna vormid mis, kes, mida; enesekohase asesõna vormid ise, end; umbmäärased asesõnad kõik, midagi, kogu, keegi.

Isikuliste asesõnade puhul eelistatakse nimetavas käändes lühemat vormi, näit. ta 1491 - tema 108; ma 305 - mina 49; me 91 - meie 29; nad 228 - nomad 15 (arvud tähistavad U väärtusi). Omastavas käändes on aga pikem vorm tavalisem: tema 182 - ta 154; meie 82 - me 1 (kuid minu 64 - mu 76). Teistes käänetes eelistatakse lühemaid vorme, näit. tal 154 - temal 10; talle 143 - temale 6; mul 52 - minul 3; mulle 77 - minule 3 (välja arvatud seest- ja alaltütlev käanne, näit. temast 18 - tast 1). Ka enesekohase asesõna lühemad vormid on sagedamad, näit. end 106 - ennast 12; enda 39 - enese 12.

Mõned asesõnavormid võivad kokku langeda määrsõnadega, näit. kõige (asesõnana U = 13, iseseisva määrsõnana U = 123), kõigest (asesõnana 4, rõhumäärsõnana 7), isegi (asesõnana 9, rõhumäärsõnana 113).

A r v s õ n a d e s t on saja sagedama sõnavormi hulgas üks ja kaks (üldises sagedusjärjestuses vastavalt 76. ja 96. kohal). Tuhande sagedama sõnavormi loendis esinevad veel nimetava käände vormid kolm, teine, paar, tuhat, esimene, neli. Teistest käändevormidest on sagedamad ühe, kahe, esimese, esimest, esimesel, ühel, ühest, teise, teist jt.

O m a d u s s õ n a on hüüdsõna kõrval ainuke sõnaliik, mis puudub saja sagedama sõnavormi loendist. Esimesed omadussõnavormid leiduvad alles teises sajas: hea (nom), suur (nom), suure, vana (nom). Tuhande sagedama sõnavormi hulka kuuluvad kokku ainult 51 omadussõnavormi.

H ü ü d s õ n a d e s t on esikohal oh (üldises sagedusjärjestuses 760. kohal kasutuskordaja väärtusega U = 9), järgnevad noh ja ah. Need kolm kõige tavalisemat hüüdsõna mahuvad tuhande sagedama sõnavormi loendisse.

.-.-.-.-.-.

Eespool vaadeldud kõrg- ja kesksagedusega sõnavormide esinemus tekstis võib anda mõningaid pidepunkte eesti keele morfoloogia kvantitatiivsel uurimisel ning eesti keele õpetekstide või harjutuste koostamisel. Vaatlesime peamiselt selliseid sõnavorme, mis esinevad tuhande sagedama sõnavormi hulgas tänapäeva ilukirjandusproosa autorikõne sagedussõnastikust lähtudes. Avaldatud sõnavormide sagedussõnastik (kogumiku "Tõid keelestatistika alalt" esimeses väljaandes) sisaldab ligi 3000 sõnavormi ja analüüsi võib vajaduse korral laiendada ka teisele ja kolmandale tuhandele. Stilistilise uurimise seisukohast võivad aga erilist huvi pakkuda harva esinevad sõnad, mis on avaldatud tänapäeva ilukirjandusproosa autorikõne lekseemide täielikus sagedussõnastikus (kogumiku teises väljaandes).

Viidatud kirjandus

Kaasik, Ü., Tuldava, J., Villup, A., Ääremaa, K. Besti ilukirjandusproosa autorikõne sõnavormide sagedussõnastik. - Tõid keelestatistika alalt, 1. TRÜ Toimetised, vihik 377. Tartu, 1976, 107-153.

Kaasik, Ü., Tuldava, J., Villup, A., Ääremaa, K. Besti tänapäeva ilukirjandusproosa autorikõne lekseemide sagedussõnastik. - Tõid keelestatistika alalt, 2. TRÜ Toimetised, vihik 413. Tartu, 1977, 5-140.

Kaljud, H. ma- ja da- infinitiivide ja nende käändeliste vormide esinemissagedus A.H. Tammsaare romaanis "Tõde ja õigus" I. TPedI lõputöö (5. ptk.). Tallinn, 1970.

Tuldava, J., Villup, A. Sõnaliikide sagedusest ilukirjandusproosa autorikõnes. - Tõid keelestatistika alalt, 1. TRÜ Toimetised, vihik 377. Tartu, 1976, 61-106.

Valge, J. Besti keele käänete sagedused kolmes funktsionaalses stiilis. - Keel ja Struktuur, IV. Tartu, TRÜ, 1970, 145-161.

Villup, A. Adverbaalse sõnavara statistilisest struktuurist. - Emakeele Seltsi aastaraamat, 19-20. Tallinn, 1975, 73-90.

ВСТРЕЧАЕМОСТЬ СЛОВОФОРМ В ЭСТОНСКОМ ТЕКСТЕ

Ü. Tuldava

Резюме

В статье рассматриваются частотность и функция отдельных словоформ в текстах современной эстонской художественной прозы. Анализ подвергаются в основном словоформы из списка трех тысяч наиболее частотных словоформ, распределенных по отдельным частям речи. Результаты исследования могут быть использованы при квантитативном исследовании морфологии эстонского языка, при составлении учебных пособий и т.п.

OCCURRENCE OF WORD FORMS IN ESTONIAN TEXTS

J. Tuldava

Summary

The article deals with frequency of occurrence and function of word forms in a sample of texts of contemporary Estonian prose fiction. The 3,000 most frequent word forms are analyzed according to various parts of speech. The results of the investigation may be used in resolving problems of the quantitative analysis of Estonian morphology, in compiling teaching aids etc.