

Suulise kõne sõnakasutuse sagedased erijooned

Käesoleva artikli eesmärgiks on tuua välja suulise spontaanse eesti keele sõnakasutuse statistiliselt kõige olulisemad erijooned, mis eristavad teda kirjakeelest. Töö aluseks on TÜ üldkeeleteaduse õppetooli arvutuslingvistika töörühma poolt tehtav suulise eesti keele korpus.

Tartu ülikooli suulise eesti keele korpus

TÜ eesti filoloogia osakonnas on alates 1980. aastate lõpust tehtud kirjaliku avaliku keelekasutuse korpuseid (korpuste kohta üldiselt vt McEnery & Wilson 1997; Muischnek 1998; tekstide valikuprintsiipide kohta vt Hennoste 1996; võrgulehekülg korpustega vt TÜ arvutuslingvistika lehekülg). Alates 1997. aastast on neile lisandunud suulise kõne korpus.

Korpus on planeeritud avatud korpusena, st selle piirsuurust ei ole määratud. Temasse on mõeldud koguda erinevat tüüpi suulist kõnet, nii argisuhtluse kui avaliku suhtluse keelekasutust, nii spontaanset kui ettevalmistatud kõnet, nii monolooge kui dialooge. Korpus on esialgu Wordi formaadis arvutis olev tekstipank. Praegu ei ole ta arvutivõrgus kättesaadav ning ei ole ka arvutiga töödeldav.

Korpuse litereerimisel kasutatakse T. Hennoste poolt G. Jeffersoni transkriptsioonisüsteemi alusel välja töötatud transkriptsioonisüsteemi. Igale lindile on lisatud samuti Tiit Hennoste poolt välja töötatud ja rühma poolt ühiselt viimistletud taustakirjeldus, milles on välja toodud kokku 44 situatiivset tegurit, mis on eri keelte analüüsimisel leitud mõjutavat keelekasutust (vt. Hennoste, ilmumas). Iga konkreetse lindi puhul on neid lisatud niipalju kui võimalik.

Lindistused on teinud suulise kõne töörühm (Tiit Hennoste, Liina Lindström, Andriela Rääbis, Piret Toomet, Riina Vellerind) ning T. Hennoste suulise kõne loengute kuulajad Tartu ülikoolis.

1999. aasta septembri seisuga oli lindikogus 222 linti, millest on litereeritud 287 teksti või tekstikatket kogupikkusega 184 602 tekstisõna/üneemi (keelestatistilise terminoloogia kohta vt Tuldava 1976; suulise kõne terminoloogia kohta vt Hennoste 1998, Hennoste & Vihalemm 1998). Selleks on lindistatud erinevaid suulise kõne situatsioone eeskätt Tartus ja Tallinnas, litereeritud neist valitud lõigud ning viidud arvu- tisse.

Tekstid jagunevad vahetuteks vestlusteks ja telefonidia- loogideks. Teisalt on nad jagatavad argivestlusteks ja avali- keks situatsioonideks. Ülivaldav osa litereeringutest on dia- loogid või polüloogid. Monolooge on vaid 3 jutlust.

	argised	avalikud
vahetud	70	107
telefoni	11	80

Avalike situatsioonide hulgas on kõige enam poedialooge (50) ja dialooge teenindus- ja ametiasutustes (25). Muid tekstitüü- pe on igaühte alla 10. Käesoleval aastal lindistavad üliõpi- lased saateid raadiost ja televisioonist.

Litereeritud tekstide pikkused on väga erinevad. A. Rää- bise andmetel on tekstide keskmine pikkus 632 tekstisõna. Argivestlused on pikkusega 200-2000 tekstisõna, enamik 600- 1200 tekstisõna. Avalikud vestlused 20-2500 sõna, tavaliselt küllalt lühikesed.

Käesoleva artikli statistika aluseks on korpuse varasem seis, nimelt korpus 1998. aasta aprillis. Sel ajal oli korpuses 141 teksti, kokku 106 451 tekstisõna/üneemi, mis andis kok-

ku 14 259 erinevat sõnavormi. Neist oli 53 argivestlust kogupikkusega 49 009 sõna ja 88 avalikku teksti, kogupikkusega 57 442 sõna. Seega jaguneb uurimise aluseks olev korpus sõnade hulgalt umbes pooleks argise ja avaliku suhtluse vahel, kuid avalikke tekste on enam ja nad on lühemad.

Käesoleva statistika aluseks olevate sõnavormide ja nende sageduste leidmiseks vajalikud programmid on koostanud Liina Lindström, sõnaloendid on üle kontrollitud rühma liikmete poolt.

Käesolevaid statistikaid tuleb võtta kui esialgseid. Sõltuvalt tekstide lindistajatest ei ole korpus esialgu tasakaalus. Nt on argivestlustes osalejate hulgas suures ülekaalus noored vähemalt keskharidusega naised. Mehi, vanemaid inimesi ja väiksema haridusega inimesi on vähe. Valdav osa lindistustest on tehtud Tartus ja kõnelejateks on Tartu ja Tallinna päritoluga inimesed.

Teiseks, statistika aluseks olevad sõnaloendid vajavad veidi puhastamist, sest nende sekka on sattunud ka tehnilist praaki. Kuid pisteline ülekontrollimine näitas, et muutused võivad olla 1–1,5%.

Kolmandaks, kuna eri tekstid on litereeritud erineva täpsusega, siis võib olla ka kõikumist litereeringutes. See tähendab, et mõni erijoon võib olla jäänud välja litereerimata või litereeritud ebatäpselt.

Neljandaks, korpuse transkriptsioon ei too loomulikult välja mitte kõiki võimalikke erijooni. Nt ei erista transkriptsioon teist ja kolmandat väldet ega too välja sõnaalgulise h tegelikku hääldust, vaid järgib neis küsimustes kirjakeele ortograafia reegleid.

Kuid ei ole usutav, et siin välja toodud kõige üldisemad statistilised tendentsid ka suurema ja enam tasakaalustatud korpuse puhul muutuksid.

Eesti keele suulise spontaanse dialoogilise keele erijooned

Järgnevalt kirjeldame eeltoodud korpuse põhjal eesti keele suulise spontaanse dialoogilise sõnavara ja sõnagrammatika statistiliselt sagedasi erijooni. Kui vaatame antud korpuses kasutatud sõnavorme, siis võime need jagada laias laastus kahte rühma:

- sõnavormid, mis on ühised kirjakeelega;
- sõnad või vormid, mis seal ei esine.

Meid huvitab siinjuures viimane rühm sõnavorme.

Sellest, mille poolest on suulise suhtluse sõnavara eripärane, olen ma kirjutanud ja omapoolse rühmituse andnud varem (Hennoste 1998; suulise kõne kohta üldisemalt vt nt Tiittula 1991). Praeguse uurimuse seisukohast jaotame need erinevused järgmiselt:

- kasutatakse selliseid tüvisõnu, mida kirjalikus avalikus valdavalt monoloogilises suhtluses ei kasutata (*loll, tere* jms);
- kasutatakse sõnu, mis on samad kui kirjakeeles, kuid millel on suulises või argisuhtluses omaette tähendus või mis kuuluvad seal teise sõnaliiki. Sellesse rühma kuulub hulk partikleid ja mitmed üldsõnad (*panema, tegema, lõpp* omadussõnana jms);
- kasutatakse samu tüvesid kui kirjakeeles, kuid nõ erinevalt töödeldud kujul: lühemal kujul (*sis, vata*), spetsiaalse te argiliidetega või ebatavaliselt liitsõnadeks tehtuna (*eksju*), eripäraste grammatiliste lõppude või tunnustega (*laudasi*);
- kasutatakse verbaalseid moodustisi, mis kirjakeele fonoloogisi piiranguid rikuvad ja seetõttu ei sobi hästi sõnade hulka (*mhmh, häh*) ja mida on nimetatud üneemideks või interjektsioonideks (vt nende kohta Hennoste 1997).

Osa neist võtetest on kasutusel eelkõige argikeeles, osa aga suulises kõnes laiemalt.

Kokku on selliseid eripäraseid vorme uuritud korpuses 16 991 tekstisõna ehk 16% kõigist tekstisõnadest. Need jagunevad järgmiselt:

5415 eripärase tüvega tekstisõna (5%)

491 eripärase liittekstisõna (0,5%)

4607 häälduslikult eripärase tekstisõna (4,5% kõigist tekstisõnadest)

418 eripärase grammatilise vormiga tekstisõna (0,4%)

2793 üneemikasutust (2,6%)

3283 suhtlusinterjektsiooni kasutust (3%)

Sagedasimad tekstisõnad korpuses

Erinevaid sõnavorme ja üneeme, mille esinemissagedus korpuses oli vähemalt 10 korda, oli kokku 1078. Eripäraseid sõnavorme on sellest 142 ehk ca 14% sõnavormidest. Kokku annavad need 13 972 tekstisõna ehk 13% kõigist tekstisõnadest.

TABEL 1

Sagedasimad eripäraseid sõnavormid

1371	jah	293	a/a:	135	mm
1056	noh	286	vä	125	nimodi
889	sis	266	nüd	120	ahaa
751	ee	232	õ/õ:	114	m/m:
717	no	210	tere	105	i/i:
608	e/e:	203	ütleme	103	nojah
498	se	187	sin	103	ned
468	mhmh	183	onju	102	ku
396	jaa	166	ahah	94	ni
371	mhm	162	aa	94	aitäh
344	he	149	eks	94	ah
317	hh	136	heh	92	siss

85	ha	33	viiskend	18	nh
83	vel	32	jaja	18	mmh
83	mhõh	31	tegelt	17	siuksed
79	nägemist	31	lähme	17	ot
76	pole	30	eh	17	omale
76	kule	29	okei	17	nelikend
73	ä/ä:	29	mõtsin	17	nagunii
71	mh	29	issand	17	kudas
69	täitsa	28	oo	17	jube
63	kuskil	28	mhh	17	hehehe
59	vot	28	jajah	17	hah
57	vat	26	öö	16	öheksa
57	kah	26	hehe	16	seline
56	siuke	25	ütsin	16	põhimõtselt
52	saand	25	sda	16	oh
51	noo	25	präegu	16	hõh
50	tändab	25	niuke	16	hm
50	nigu	24	jaah	16	ennem
50	eksole	23	nüt	15	ähäh
49	õõ	22	õudne	15	tändap
47	võ	22	ota	15	siukene
47	aint	22	näind	15	seles
46	võibola	21	vata	15	kuulmiseni
46	olnd	21	sinn	14	siiamaani
46	kakskend	21	sele	14	niiöelda
44	oi	21	seitsekend	14	kakskümend
38	kuradi	21	mhhh	14	eee
37	ää	21	hhh	14	aah
37	mhmh	21	eksju	13	õh
35	üts	20	sab	13	vaatsin
35	kurat	20	läind	13	teind
34	ö/ö:	19	old	13	raisk
34	tähendap	18	onn	13	mismoodi
34	h/h:	18	nooh	13	jajaa

12	ülesse	12	aind	11	khm
12	tüüp	11	võbola	11	jõle
12	siinpol	11	vabandust	11	einoh
12	praegult	11	selepärast	11	no
12	maru	11	san	10	söönd
12	las	11	misasi	10	sihuke
12	kõvasti	11	loll	10	kuuskend
12	kasvõi	11	lahe	10	kaa
12	jama	11	kuda	10	heheh
12	hõ	11	kuat	10	haa
12	hallo	11	kolmkend		

Eripärased tüved

Ühe rühma moodustavad sellised omaette tüvisõnad, mida kirjalikus avalikus mittespontaanses suhtluses ei kasutata või kasutatakse vaid kindlates retoorilistes funktsioonides, eelkõige ilukirjandusdialogis. Põhjuseks võib olla see, et kirjalik tüüpiliselt monoloogiline tekst lihtsalt ei vaja selliseid sõnu (nt dialoogisõnu) või see, et mitmesugustel, põhiliselt sotsiaalsetel põhjustel on need sõnad avalikus suhtluses tõrjutud (vandesõnad jms). Selle sõnarühma piiritlemine on alati veidi udune ning siin saab toetuda eelkõige uurijate enda keeletajule. Selliseid sõnu on kokku 5415 tekstisõna.

TABEL 2

Sagedased eripärased tüvisõnad

Dialoogipartiklid

1371	jah	51	noo	12	hallo
402	jaa	29	okei	11	vabandust
210	tere	24	jaah		
94	aitäh	18	nooh		
79	nägemist	15	kuulmiseni		

Muud partiklid	Omadussõnad/ partiklid	Asesõnad
1056 noh		56 siuke
727 no	22 õudne	25 niuke
149 eks	17 jube	17 siuksed
69 täitsa	12 maru	15 siukene
63 kah	12 jama	10 sihuke
59 vot	11 loll	
57 vat	11 lahe	Adverbid
38 kuradi	11 jõle	
35 kurat		12 kõvasti
29 issand		
13 raisk		Nimisõnad
12 las		
10 kaa		12 tüüp

Väiksema sagedusega (vähemalt 4 korda) esinevad veel järgmised sõnavormid:

- nimisõnad: *kutt, musi, issi, värgid, kutid, jubin, emmele, doss;*
- omadussõnad: *lolli, nõme, kihvt;*
- asesõnad: *nisukene, nihuke, siukseid, niuksed, siukse, siukest, nisuke;*
- määrsõnad: *ilgelt, õudsalt;*
- partiklid: *tshau, voh, allo.*

Lisaks kasutatakse ka võõrkeelseid toorlaene, mida on kokku 85 eri sõnavormis, millest igaüks esineb 1–3 korda.

Sõnade saamine

Suuline kõne, eelkõige argikeel tunneb kirjakeelest erinevaid sõnade saamise viise ja võtteid, mille omapoolse tüpoloogia olen andnud varem (vt Hennoste 1998). Need võtted jaotu-

vad 3 põhirühma: leksikaalsed, semantilised ja grammatilised saamisviisid.

Leksikaalsetest sõnade saamisviisidest tulevad korpuses välja võõrkeelsed toorlaenud ja sõnade ebatavaline liitmine.

Semantiliste sõnasaamisviiside all on tegu sõna tähenduse kontekstuaalse kõikumise ja kontekstisidususega.

Grammatilised sõnasaamisviisid on grammatilise kateooria muutmine, afektiivne geminatsioon, raie, püsiv hääliku või silbikadu, mõnede üldverbide laialdane kasutamine (*panema* jms), eraldi argikeelsed liited.

Käesolev sõnaloend ei luba uurida semantilisi sõnasaamisviise ja üldverbide konkreetseid tähendusi. Lühenemisi ja kadusid käsitleme aga omaette peatükis.

Liitsõnad ja tuletised

Käesolevas korpuses esinevad järgmised ebatavalised sõnasaamisviisid:

1) ebatavalised liitsõnad, mis on jagatavad kahte rühma:

- kirjakeelest erinevad liitsõnad. Neid on kokku 491 tekstisõna. Selles rühmas eristub omakorda sageda kasutusega sõnarühm, kuhu kuuluvad partiklid *onju*, *nojah*, *ekssole*, *jaja/jajah/jajaa*, *eksju*, *einoh*, *misasi*, mis esinevad vähemalt 10 korda ja annavad kokku 416 kasutust (87%). Neile võib veel lisada partiklid *odot*, *kesse*, *etet*, *nonoh*, *kuidasmoodi* ja *jadaega*, ning nimisõnad *vanaaasta* ja *ingliskeel*, mis esinevad vähemalt 4 korda. See on ainus sageda kasutusega eripärane sõnasaamisviis. Muud võimalused on harvad;

- ebatavalised liitumiskäänded liitsõnade esikomponentidel, mis esinevad üksikutel sõnadel ja väga harva: *nokamütsiga* < *nokkmütsiga*, *ketiülekande* < *kettülekande*, *sukapiüksid* < *sukkpüksid*;

2) ebatavalised tuletusvõtted, mis esinevad harva:

- -kas -liide: *telekas/telekat, pastakaga*;
- raie: *makk/maki* (< *magnetofon*), *konts* (< *konspekt*), *trenni/trennis* (< *treening*), *mersu* (< *Mercedes*);
- afektiivne geminatsioon: *emmele, issi*;
- ebatavalised lõpud adverbides: *kuidagid* < *kuidagi*, *jälegist* < *jällegi*, *ikkagist* < *ikkagi*, *ennemalt* < *enne*, *praegust* < *praegu*.

Sealjuures on kahes rühmas olemas samatüübiline alarühm. Liitsõnade puhul on tegu sellega, et kirjakeeles nominatiivse liitumisega esikomponent liitub genitiivselt, mis on keskne ja prototüüpne liitumisviis eesti keeles. Adverbide -st lõpu puhul on tegu vormiga, millel on olemas taust ja analoogiad murretes. Uurimist vajab aga, kas on tegu vana vormiga, mis on jäänud kõnes käibele või käsitatakse kirjakeelset adverbiki tüve, millele lisatakse igaks juhuks liiasena juurde sufiks -st.

Grammatilise sõnaliigi muutmine

Suulises kõnes on sõnade sõnaliigiline kuuluvus küllalt kõikuv. Samuti võib kirjakeeles mingis sõnaliigis olev sõna olla suulises kõnes, eriti argikeeles teises sõnaliigis. Korpuses esinevad järgmised statistiliselt sagedased (üle 10 korra esinevad) sõnavormid, mida kasutatakse kirjakeelest erinevas grammatilises kategoorias:

- verb > partikkel: *ütleme/ütme, onju, kule, ota, vata/vat, tähendap/tändab/tändap, eksole*;
- adverb > partikkel: *vä/võ, nüüd*;
- sõnad, mis võivad olla nii partiklid kui ka nt omadussõnad: *õudne, jube, maru, jama, loll, lahe, jõle, kuat/kurat, issand*.

Morfoloogia

Suuline kõne, eriti argikeel kasutab mitmesuguseid kirjakeeles keelatud morfoloogilisi vorme. Korpuse põhjal saame välja tuua järgmised morfoloogiliselt eripärased jooned (418 tekstisõna):

- *-nd/-d* kasutamine *-nud* -kesksõnas (vaheldumise kohta vt Keevallik 1994);
- *-da* -tunnusega kaudne kõneviis;
- kirjakeele normingutest erinevad türevokaalid mõnedel sõnadel: *rämpsum* < *rämpsem*, *kihvt* < *kihvtim*, *õudsam* < *õudsem*;
- muutelõpp *-si* mitmuse partitiivis: *nägusi* < *nägusid*; *nastoikasi* < *nastoid*; *vahendisi* < *vahendeid*;
- teatud *-gi/-ki* -sõnade lõpusilpide asukohavahetus: *kellelegi* < *kellele*; *mõnegiga* < *mõnegi*;
- muud ebatavalised käändevormid: *pudelde* < *pudelite*; *meiele* < *meile*; *teiel* > *teil*; *väikset* < *väikest*; *pittu* < *peole*; *juure* < *juurde*, *töösid* < *töid*; *ülesse* < *üles*;
- ebatavalised verbivormid: *kaudata* < *kaotada*; *polla* < *ei olevat*, *ärme* < *ärgem*, *üteldasse* < *öeldakse*, *käse* < *käsi*, *kusesid* < *kusi*.

Need jagunevad statistiliselt kahte selgesse rühma. Ainus statistiliselt sage erijoon on *-nd* -lõpp: kokku 101 sõnavormi, 368 tekstisõna, neist üle 10 korra esinevad *saand*, *olnd*, *näind*, *läind*, *teind*, mis koos mitmesuguste hääldusvariantidega annavad kokku 178 kasutust (48% kõigist *-nd* kasutustest). Ülejäänud eripärased vormid on haruldased, st nad esinevad vaid mõnedes sõnades ning nende kasutussagedus on minimaalne (kokku ca 50 korda). Lisaks võime öelda, et selgelt eelistatakse nendegi vormide puhul kirjakeelega sarnaseid vorme. Nt *juure* esineb 1 kord, *juurde* 91 korda, *meele* 3 korda, *meelde* 24; *-si* -partitiiv esineb 4 sõnas kokku 4 korda.

Häälalusvariandid

Suulises kõnes on rida sõnu, mille häälalus erineb samade sõnade standardsest häälalusest, foneetikast ja fonoloogiast. Sõnal võib olla lühem ja pikem variant (*milla/millal*), eri variantides võivad mõned häälikud olla erinevad (*üheksa/öheksa*) jne. Järgnevalt vaatleme selliseid häälaluslikult erinevaid variante käesolevas korpuses.

Ajutine ja alaline erinevus

Häälaluslikult erinevad sõnad jagunevad kahte suurde rühma vastavalt sellele, kas need häälaluserinevused on ajutised või püsivad. Püsivad variandid on omakorda jagatavad kahte rühma, vastavalt sellele, kas neil on omaette tähendus või nad on kirjakeelse vormi sünonüümid.

Seega jagunevad eripärased vormid tegelikult kahte ossa, mille piir on ebaselge. Üht osa tuleks käsitleda häälalusjoonega, teist osa aga eripärase sõnavarana. Nende rühmade piiri on praegu raske paigale panna. Esiteks, on teoreetiliselt usutav, et see piir ongi udune, st osa eripäraseid vorme on üleminekuseisundis ajutise ja alatise vormi vahel ehk leksikaliseerumas. Teiseks on meie korpus liiga väike ning osade sõnade esinemissagedus nii madal, et me ei saa midagi järeldada nende kasutuse leviku kohta. Praegu on piiri tõmbamisel arvestatud vormi sagedust korpuses ning uurijate eneste keeleteadvust.

Eripäraste vormide põhirühmad

Eripärase häälaluse tüüpe on tavaliselt käsitletud foneetika terminites (vt Ariste 1977; Brown 1990, 57-88; korpusel baasil on valmimas Marju Mikkeli bakalaureusetöö). Me kasutame siinkohal lihtsamat lähenemist ja jaotame eripärasused kolme põhirühma: lühenemised, teised häälikud ja pikenemised:

1) Sõna võib olla mõne hääliku või ka terve silbi võrra lühem kui vastav kirjakeelne sõna. Transkriptsioon lubab välja tuua järgmised nähtused:

a) mingi vokaal või konsonant on hääldamata (*milla* < *mil-lal*);

b) 2./3.-värtelise vokaali või konsonandi asemel on esma-värteline (*koli* < *kooli*);

c) mingi silp on puudu (*mõtsin* < *mõtlesin*).

Transkriptsioon ei võimalda praegu välja tuua juhtumeid, kus kirjakeele kolmanda värtelise asemel on teine värtelise arvatud klusiilide puhul.

2) Sõnas võivad olla teised häälikud kui normatiivne hääldus ette näeb:

a) kirjakeelse sõna diftongi asemel on pikk või lühike üksikvokaal (*pääv* < *päev*);

b) konsonantühendi asemel on pikk või lühike konsonant (*sellast* < *seljast*);

c) kirjakeelse sõna vokaali asemel on mingi teine vokaal. Siin võib olla mitmeid variante: vokaal on kõrgem või madalam (*öheksa* < *üheksa*), järgsilbi vokaali asemel on redutseerunud vokaal jms.

Praegune transkriptsioon ei luba välja tuua järgsilbi vokaalide redutseerumist.

3) Mingid häälikud või silbid hääldatakse pikemalt või tugevamalt kui on tavaline:

a) mingi konsonant hääldatakse pikemalt kui kirjakeele normhäälduses. Meie korpus peegeldab praegu ainult teatud klusiilide tugevamat hääldust. Muid võimalusi transkriptsioon ei näita;

b) osaliselt kuulub siia ka kokutamisel tekkiv silbi või häälikukordus sõna alguses (*ko-ko-kohustuslik*);

c) häälik on venitatud. Venitada võib põhimõtteliselt kõiki häälikuid ning igasuguses positsioonis (*ko:ju*, *koju:*, *k:ju*).

Venitamine on statistiliselt väga sage nähtus, kuid pole seotud konkreetsete sõnade, vaid eelkõige teatud süntaktiliste positsioonidega: venitatakse eelkõige lausete piiridel olevaid sõnu. Sealjuures kannavad need venitused mitmeid funktsioone, millest keskne on ajavõitmine, hesitatsioon. Lisaks on venitamine oluliseks stiilimarkeriks mõnede registrite puhul (jutlus, lasteaiatädi jutt, teismeliste tüdrukute vinguv kõnelemisviis jms).

Sõnade venitamine ja kokutamine jäävad siinkohal käsitlemata, sest seda ei peegelda praegune sõnavormide nimestik tehnilistel põhjustel.

Lühemad tekstisõnad

Suulises tekstis on hulk tekstisõnu, mis on tüvelts samasugused kui kirjasõnad, kuid mõne hääliku või silbi võrra lühemad.

Lühemad sõnad annavad korpusesse 314 erinevat sõnavormi, kokku 4120 tekstisõna, neist üle 100 korra esinevad *sis*, *vä*, *nüd*, *nimodi*, *se*, *sin*, *ned*, *ku*, *mis* annavad kokku 2456 kasutust, lisaks arvud (*kolmkend* jms), kokku 231 kasutust, ühtekokku 2687 kasutust (65%).

TABEL 3

Lühemad sõnad

Partiklid

889	sis	76	kule	22	ota
286	vä	50	tändab	21	vata
266	nüd	47	võ	17	ot
125	nimodi	47	aint	16	põhimõttelt
92	siss	31	tegelt	15	tändap
83	vel	23	nüt	11	kuat

Asesõnad		Adverbid		Verbid	
498	se	187	sin	35	üts
103	ned	102	ku	29	mõtsin
56	siuke	94	ni	25	ütsin
21	sele	46	võibola	20	sab
17	siuksed	21	sinn	13	vaatsin
16	seline	17	kudas	11	san
15	siukene	12	siinpol		
15	seles	12	aind	Arvsõnad	
		11	võbola		
		11	selepärast		
		11	kuda		
				46	kakskend
				33	viiskend
				21	seitsekend
				17	nelikend
				14	kakskümend
				11	kolmkend
				10	kuuskend

Lisaks esinevad vähemalt neli korda veel: *selest, nisuke, ni-sugune, niivisi/nivisi, võibla/ võipola, mõtes, kudagi, peagu, niöelda, mite, küll, ütsid/ütme/üteld, län/lähvad/läeb, mõts/mõtsid, pand, kolgend, üheksakend, viiskümend, kaheksakend, seitsekümend, kolmkümend*.

Lühemad sõnad on jagatavad kahte rühma: omaette sõnad ja lühemad ajutised tekstisõnad.

Alguses vaatame sõnu, mida oleme tõlgendanud omaette sõnadena. Nende loend on järgmine (siin on toodud sõnavormid, mida esineb vähemalt 5 korda):

1) omaette tähendusega sõnad, millel on lühem tüvi

partiklid: *sis, vä/võ, nimodi, kule, tändab, kuat, ota/ot, vata, las;*

2) kindla kasutuskontekstiga püsivalt tarvitataavad lühemad sünonüümid

- partiklid ja adverbid: *nüüd/nüüd, veel, aint/aind, tegelt, põhimõttelt, tänapäev, ütme, kudas/kuda, sin, võibolla/võibolla/võibla/võbla, ku, ni, siinpol, seepärast, kudagi, peagu, niivisi;*
- asesõnad: *ned, siuke, niuke, se, sda, sele/seles, seline,*
- verbid: *üts, mõtsin, ütsin, vaatsin, län, sab, san;*
- numeraalid: *kakskend/kakskümend, viiskend, seitsekend, nelikend, kuuskend, kolmgend.*

Lühemad sõnavormid erinevad pikkadest mitmesuguste kaadude poolest. Statistiliselt sagedased kaod on järgmised:

- pika vokaali asemel lühike (12 sõna): *nimodi, kule, vata, ota, sis, nüüd, veel, san;*
- 2. silbi puudumine (8): *mõtsin < mõtlesin, tänapäev < tähendab, aint < ainult, ütsin < ütlesin;*
- diftongi asemel üksikvokaal (6): *vä, pä/bä < päev, kudas, võibolla*
- lõpukonsonandi või vokaali puudumine (6): *milla < milal, kuidas/kuda < kuidas, justku < justkui, mutku < muudkui, ku < kui;*
- *h* puudumine teise silbi algusest (5): *niuke, kaeksa.*

Muud võtted on haruldased:

- silbi puudumine liitsõna teisest poolest: *-kend < küm-mend;*
- 3. silbi puudumine: *tegelt < tegelikult;*
- konsonandi puudumine 2. silbi algusest: *kuat < kurat;*
- 1. silbi vokaali puudumine: *sda < sada/seda;*
- konsonandi lühem variant 1.-2. silbi piiril: *sele < selle, seline < selline, mite < mitte.*

Sõnaliigiliselt on lühemad sõnad eelkõige partiklid, adverbid, asesõnad, *-küm-mend*-lõpulised numbrid, *-päev*-lõpulised

nädalapäevad. Lisaks on lühemad mõned verbid, mida kasutatakse eelkõige et-algusega referaadi puhul: *mõtsin=et*, *ütsin=et*, *vaatsin=et*.

Suurem osa selliseid sõnu on grammatiliselt muutumatud sõnad, väike osa muutuvad sõnad. Viimaste puhul on võimalik väita ainult seda, et lühemad vormid on olemas neist muudetest, mis korpuses esindatud on. Näiteks arvude *-kend*-lõpp esineb vaid nominatiivis, muudes käänetes on tegu tavalise pika lõpuosaga.

Omaette probleemi moodustab *h* sõna algul, mida üldlevinud arusaama järgi ei hääldata. Sellele vastavalt ei olnud korpuse transkriptsioonis ka *h* häälduse erinevaid võimalusi välja toodud. Vaid mõned litereeringud peegeldavad juhuslikult *h* mittehääldamisi (*omme*, *akkama*, *elistama*, *ommik* jne). Kaily Cui bakalaureusetöö põhiliselt sellesama korpuse materjalide põhjal aga näitab, et tegelikkus on erinev: argisituatsioonides hääldatakse *h* sõna algul 63% juhtudest ja avalikes situatsioonides 96% juhtudest. *h* hääldamist või mittehääldamist argisituatsioonides konkreetse sõna alguses mõjutavad mitmesugused keelelised, situatiivsed ja sotsiaalsed tegurid (Cui 1999).

Ajutised kaod ja lühenemised võime jagada kahte rühma. Esimese moodustavad sagedased võimalused, mis kokku annavad 107 kasutust (66%):

- 2. silbi vokaali kadu (26 kasutust) *näitks* < *näiteks*; *vastvalt* < *vastqvalt*; *üldselt* < *üldiselt*; *näits* < *näitqs*;
- konsonandi lühenemine 1. ja 2. silbi piiril (24): *selealast* < *sellealast*; *hakate* < *hakkate*; *telitakse* < *tellitakse*;
- 1. silbi vokaali või diftongi lühenemine (20): *möda* < *mööda*; *säl* < *sääl*; *sovin* < *soovin*.

Teise rühma moodustavad harvemad lühenemised:

- 3. silbi vokaali kadu (12): *teenindse* < *teeninduse*; *koolits* < *koolitus*;
- klusiili nõrgem hääldamine (8): *täidsa*, *masinade*, *abostel*, *migrofoni*, *vaadage*, *üigsgord*, *kaugede*, *ehitatude*;
- konsonantühendist ühe konsonandi kadumine (5): *konreetselt* < *konkreetselt*, *aksiooni* < *aktsiooni*; *aksiaselts* < *aktsiaselts*; *türuku* < *tüdruku*.

Eelnevast võib näha, et keskseks lühenemisevõtteks eesti kõnes on 1. silbi pika vokaali lühenemine. Samal ajal võib vähemalt hüpoteesina pakkuda välja, et ajutine lühenemine eelistab teistsuguseid lühemaid kohti kui püsivad lühemad sõnad. Kuid see vajab uurimist suuremal materjalil.

Suureks probleemiks on lühemate ja pikemate vormide omavahelised seosed.

Esiteks, ajutiste lühenemiste puhul võime väita, et tegu on pika vormi lühenemisega. Klassikaliselt on selliseid lühenemisi nimetatud allegrosõnadeks ja nähtud nende põhjuseks kiiret kõnetempot. See on kindlasti üks lühenemise põhjusi, kuid mitte ainus ja mitte ka peamine. Sõnavormi lühenemise määrab enamasti tema asend kõnevoolus ning tema informatsiooniline staatus: esialgsete vaatluste põhjal võib väita, et vana info, teema, jutu pealiinist kõrvalepõikav info, kiilung jms kalduvad lühenema.

Sünonüümide puhul jääb lühemate ja pikemate vormide omavaheline suhe aga ebaselgeks. On kindlasti võimalik, et tegu on ka siin kunagi toimunud ja leksikaliseerunud lühenemisega. Kuid see ei saa olla ainus võimalus.

Asi on selles, et lühemad ja pikemad vormid iseenesest on universaalne nähtus, mis esineb erinevates keeltes. Pikki ja lühikesi vorme kasutatakse erinevatel kontekstuaalsetel tingimustel, mis on üldjoontes samad kui ajutiste lühenemiste puhul. Kuna pole mingit alust eeldada, et varasemal ajal oleks eestlased kuidagi teisiti kõnelnud kui teised rahvad, siis võib

eeldada, et lühemad ja pikemad sünonüümsed vormid on kogu aeg olemas olnud. Samuti on alati keeles olnud olemas sagedased sõnad, mille puhul võime eeldada, et nende sõnade lühemate vormide kasutus on olnud nii sage, et neid sõnu on käsitletud omaette sünonüümsete sõnadena. Seega: põhimõtteliselt peavad keeles olema ja peavad olema kogu aeg olnud sellised sünonüümsed paarikud. Küsimus on ainult selles, kas sellise paari on moodustanud ka varem needsamad sõnad või mingid muud sõnad. Nt teoreetiliselt võib olla nii, et varem oli pikem vorm *sis* ja lühem *s*, nüüd aga on pikem *siis* ja lühem *sis*. Sellega tahaks ma vastu astuda seletusele, et üks vormidest on lühenenud või et üks vormidest on ühest murdest ja teine teisest murdest. Selline seletus on võimalik, kuid üksnes muu seletuse kõrval.

Mis puutub omaette tähendusega lühematesse vormidesse, siis nende puhul on võimalik, et nad on pärit pikematest vormidest ning saanud lühenedes omaette tähenduse. Kuid seegi vajab veel uurimist.

Teised häälikud

Teiste häälikute kasutamisel on kaks keskset rühma, millest mõlema sagedus on väike:

1) diftongide asemel vokaali või konsonantühendi asemel konsonandi kasutamine, 16 erinevat sõna, 33 sõnavormi, kokku 392 kasutust:

- *vä/võ < vôi*;
- *kudas < kuidas, kääst < käest, präägu < praegu, nääb < näeb, pääval < päeval*;
- *välla < välja, tõmmad < tõmbad, annud < andnud, tunnud < tundnud, sellas < seljas*.

Sellest rühmast moodustab 333 (85%) kasutust *vä/võ*. Teise rühma moodustab *äe > ää* hääldus, kolmandasse kuuluvad *lj > ll, mb > mm, nd > n*.

2) vokaalide madalam ja kõrgem hääldamine, mida on kokku 18 sõna, 39 erinevat sõnavormi, 66 kasutust:

- *öheksa* < *üheksa*, *sehuke* < *sihuke*, *häste* < *hästi*, *egasugust* < *igasugust*, *asperiini* < *aspiriini*, *taldreku* < *taldriku*, *teenendajaks* < *teenindajaks*, *kahekeste* < *kahekesti*;

- *lueme* < *loeme*, *üelnud* < *öelnud*, *esimises* < *esimeses*, *nindest* < *nendest*, *ükspäiv* < *ükspäev*, *ütlis* < *ütles*, *sisimise* < *sisemise*, *tiame* < *teame*, *kaub* < *kaob*, *präegu* < *praegu*.

Kaks madaldumise põhivarianti on *ü* > *ö*, *i* > *e*; kõrgenemise põhivariandid on *o* > *u*, *e* > *i*, lisaks *ö* > *ü*, *a* > *ä*. Sealhulgas *öheksa* ja tema variandid annavad kokku 40 madaldumist (61% kasutustest).

Muud rühmad on statistiliselt haruldased:

- *e* > *ä*: *teläfon* < *telefon*, *lähäte* < *lähete*, *vähä* < *vähe*;
- *o* > *u*: *autu* < *auto*, *maud* < *maod*;
- *perastpoole* < *pärastpoole*, *keige* < *kõige*, *ikke* < *ikka*, *põle* < *pole*;
- *orantsi* < *oranzhi*, *muusjomi* < *muuseumi*;
- *eila* < *eile*, *praega* < *praegu*.

Praegune transkriptsioon ei kajasta järgsilbi vokaalide redutseerumist, seetõttu ei ole ka *e* > *ä* rühm ilmselt korrektselt kajastatud. Omaette probleemirühma moodustab siin ka võõrsõnade hääldamine, milles on palju idiolektilisi kaldeid.

Valdav osa selle rühma sõnadest on tõlgendatavad kui kirjakeelsete sõnade sünonüümsed paralleelvormid. Ainult *vä*/*võ* on omaette tähendusega sõna ning üksikute sõnade puhul võib kahtlustada, et tegu on ajutiste hääldusjoontega (*nindest*, *sisimise*, *orantsi*, *muusjomi*).

Nende vormide taust on erinev. Osalt on tegu murdetaus- taga, osalt ilmselt kiirkõnes toimuva häälduse lihtsustumise- ga.

Tugevam hääldamine

Tugevama hääldamise praegu välja toodud rühmad on kõik väikesed ja seisnevad klusiilide tugevamas hääldamises (meenutame, et muude konsonantide tugevamat hääldust praegune transkriptsioon ei kajasta):

- klusiili tugevam/geminatiivne hääldamine (13 sõna): *alkkirja, ringittegemine, sisseppidi, võipolla, neljappäeval, lauppäeval, kolmappäev, peitab, õikemini, pantud, ekspertid, sõitab, tekelevad*. Põhiliselt esineb selline hääldus liitsõna keskel;
- klusiili lisandumine: *küündlad, kõikse, enamb*.

Tugevama hääldamise üheks juhuks on võõrsõnade sõnaalgulise nõrga klusiili (*g, b, d*) hääldamine tugevana, mis on eestlaste puhul tavaline. Seetõttu on see jäetud transkriptsioonis eraldi välja toomata.

Praegune tugevama hääldamise rühm jaguneb kahte ossa. Esimese moodustab liitsõna keskel geminaadina hääldatud klusiil, mis on murdeliselt tuntud Lääne-Eestis. Kas siin on tegu murdetaustaga või on sellisel hääldusel mingid kontekstuaalsed kasutuspiirid, pole veel uuritud. Sõna sees kasutatud klusiilide tugevnemised on valdavalt ajutised hääldusjooned, mis võivad olla seotud sellega, et lausungirõhuliste sõnade rõhulised silbid on tavaliselt veidi pikemad kui samade sõnade rõhulised silbid siis, kui sõna on lauses rõhutu. See vahe võib mõnikord olla ka selline, et teisevärteline sõna või nõrk häälik kuuldu kolmandavärtelise või tugevana.

Eelnevale lisaks võib öelda, et suuremal osal teistmoodi hääldatud sõnadel ja tugevnemistel on olemas murdetaust, st selliseid vorme on registreeritud erinevatest murretest. See ei tähenda, et tegu oleks alati murdemõjudega vastava inimese kõnes. Nimelt jagunevad need vormid kahte küllalt hästi eristatavasse rühma. Ühtede puhul tajume, et tegu on murdelise vormiga, teistel juhtudel aga me sellist murdelisust ei

taju vaid sõna on tavaline argikõne sõna. Ülivaldava osa selle korpuse sõnade puhul me mingit murdelisust enam ei taju, murdelisena on tajutavad ehk sellised sõnad kui *välla, üks-päiv, küündlad, enamb*, mis on lõunaeesti sõnad.

Sõnade kokkuhääldamine

Väga oluline eesti keele hääldamise erijoon on sõnade kokkuhääldamine. See tähendab, et omaette sõnad hääldatakse teatud tekstuaalsetes tingimustes kokku ja nende intonatsioon on pigem liitsõnale sarnane. Kokkuhääldatud tekstisõnu on 6050. Neist 3463 kokkuhääldust (57%) on üks kord esinevad, 2587 tarvitust esinevad enam kui kord. Ülivaldav osa kokkuhääldusi on kahesõnalised, vaid üksikud on pikemad sõnaread. Kokkuhäälduste sagedusrea algus näeb välja järgmine:

TABEL 4

Sagedasemad kokkuhääldatud sõnapaariid

91	ja=ja	20	ei=ole	12	onju=et
75	et=et	18	nimodi=et	12	jaa=ja
61	se=on	18	ma=ei	11	ütleb=et
48	nii=et	16	ma=i=tea	11	mul=on
42	ja=sis	16	ja=ja=ja	11	mt=hh
39	ei=no	16	eks=ju	11	ei=tea
32	ni=et	15	üts=et	10	ütles=et
30	eks=ole	15	niimodi=et	10	ära=ja
26	see=on	14	mis=sa	10	nii=palju
26	ma=i	14	ja=jaa	10	kui=ta
25	on=ju	13	jaa=jaa	10	kogu=aeg
20	ma=ei=tea	13	et=ee		

Neile lisaks on 3 sagedast tagasisidepartikli kasutust, milles järgneva kõnevooru alustaja jätkab teksti, hääldades oma voo-
rualguse vahetult partiklile otsa: 36 *ja=*; 13 *mhm=*; 12 *mhmh=*.
Kokkuhääldatud paarikuid saab kõige paremini rühmitada
viimase komponendi järgi. Sellisel juhul eristub 14 sagedast
rühma kokkuhääldusi, kokku 2522 paarikut ehk 42% kokku-
hääldustest.

TABEL 5

Kokkuhääldatud sõnapaarid viimase komponendi järgi

ja	740	kasutust / tekstisõna
vä (+variandid)	335	
et	305	
sis (siis)	122	
on (o)	99	
aga	77	
ju	64	
noh	58	
jah	46	
nagu	23	
üneemid	235	
isikulised ja näitavad asesõnad	274	
seal / siin (variandid)	65	
arvud	79	

Sellele lisaks saab välja tuua *ei*-sisaldavate paarikute rühma,
mis osalt eelnevatega kattub, kokku 104 tarvitust.

TABEL 6

Kesksete kokkuhääldatud rühmade kesksed kokkuhäälduspaarid

<i>ja</i> - lõpulised		<i>et</i> -lõpulised		<i>sis</i> -lõpulised
ja=ja	92	et=et	75	ja=sis 42
ja=ja=ja	16	nii=et	48	
ja=jaa	14	ni=et	32	<i>on</i>-lõpulised
jaa=jaa	13	nimodi=et	18	
jaa=ja	13	niimodi=et	15	se=on 61
ära=ja	11	üts=et	15	see=on 26
		onju=et	12	mul=on 11
		ütleb=et	11	
		ütles=et	10	

ja-rühmas on kaks kesket alarühma: esimene on sidendi *ja* kordus ning teine on tagasiside *jaa* kordus. *et*-lõpuliste paarikute üks oluline rühm on samuti sidendi *et* kordus. Teise suure rühma moodustavad siin aga otsese kõne saatelause ja refereadi piirid (*ütles=et*, *mõtsin=et*), mida kokku on enamgi kui sidendikordust. Kolmanda suure rühma moodustab *nii=et*. *sis*-rühmas on keskne *ja=sis* sidend. *on*-lõpulises rühmas on kesksed *see=on* ning isikuline asesõna + *on* (*mul=on*, *ta=on* jms). *aga*-, *ju*- ja *noh*-lõpuliste seas sagedasi pole.

Eelnevad paaride lõpusõnad jagunevad nelja rühma vastavalt sellele kui tihti neid hääldatakse kokku nende ees olevate erinevate sõnadega:

- *vä/võ* hääldatakse praktiliselt alati eelneva sõnaga kokku;
- *ja*, *ju* hääldatakse eelneva sõnaga kokku 20% juhtudest;
- *et*, *sis*, *aga* hääldatakse kokku ca 10%;
- muud hääldatakse kokku maksimaalselt 4%.

vä on suulises kõnes kasutusel küsiva partiklina, mis on aja jooksul lühenenud sõnast *või* ning saanud kuju *vä/väh/veh/võ/võ?*. Toimunud on sõna lühenemine ning mõnikord hääldatakse sõna lõppu *h* või hõngushäälik *?*. Praktiliselt sajab protsendiliselt hääldatakse see sõna talle eelneva sõnaga kokku. Seega on kujunenud eesti keeles välja lisaks *-ki/gi* kliitikule teine selge kliitikpartikkel *-vä*.

või küsiva partiklina on vana nähtus (kasutusel juba nt Tammsaarel ja Lutsul), selle esinemine *vä* kujulisena on dokumenteeritud 1972. aastal ilmunud B. Shaw Pygmalioni tõlkes, kus lilleneiu Eliza, kes muidu kõneleb üpris eklektilist ja imelikku keelesegu, kasutab nt lauset: *Sinu poig veh?* (Shaw 1972, 280).

Teine selline sõna, mida palju kooshääldatuna kasutatakse, on *ja*, mida eriti tihti hääldatakse kokku lausete piiril lõppeva lause viimase sõnaga. Kui palju lausepiiri *ja*-sid hääldatakse kokku, vajab eraldi uurimist. Kuid paistab, et *ja* kui kliitik on välja kujunemisel. Tema kasutusala saab olema ilmselt lausepiiril teksti jätkuvuse osutamine.

Miks hääldatakse sõnu kokku? Üheks oluliseks põhjuseks on ilmselt kiirem tempo teatud tekstiosade hääldamisel:

- tavalisim positsioon on lausete või osalausete piiridel (sidendid, saatelause verb ja referaadi algus);
- teine oluline positsioon on takerdumine ja tekstitegemise edasilükkamine (sidendikordused, üneemid, *noh*, *nagu* jms partiklid).

Üneemid ja interjektsioonid

Kui me vaatleme suulise kõne teksti, leiame sealt lisaks sõnavarale mitmesugused häälitsused, mida kirjalik kõne ei kasuta või kasutab üksnes kindlates stilistilistes funktsioonides. Neid häälitsusi ei nimetata lingvistikas tavaliselt sõnadeks.

Ma olen kasutanud nende tarvis soome keelest laenatud terminit üneem.

Üneemid on suhteliselt püsivad / leksikaliseerunud sõnasarnased häämitsused, millel on vastava keele sõnades esinevatele häälikutele ja häälikukombinatsioonidele sarnane kuid mitte täpselt sama foneetiline olemus ning mis on mõeldud eelkõige teksti tootmisel planeerimisaja võitmiseks ning kõne sujuvuse tagamiseks. Foneetiliselt on üneemid kas vokaalid (*a, e, õ, ö, ä*), nasaalid (*m, n*), *h*-häälikud või mitmesugused imihäälikud. Lisaks kasutatakse üneemides häämitsusi, mis pole häälikud, vaid füsioloogiliste reaktsioonide väljendused (köhatused, üminad jms).

Omaette rühma moodustavad sellised keelelised üksused, mis jäävad oma foneetiliselt struktuurilt kuhugi üneemide ja sõnade vahele. Osalt paigutatakse nad tavaliselt hüüdsõnade / interjektsioonide hulka, osalt onomatopoeetilis-deskriptiivsete sõnade hulka, osalt aga jäävad lihtsalt kuhugi sõnavara ja häämitsuste vahelisele eikellegimaale (vt eesti interjektsioonide kohta Palmeos 1982, 34-53).

Üneemidega ühendab neid eeskätt foneetiline koostis. Ka nemad rikuvad keele foneetilisi ja fonoloogilisi piiranguid.

Erinevalt klassikalisest üneemidest on nende kasutamisel aga mitmesugused suhtluseesmärgid (vastus küsimusele, reaktsioon väitele või sündmusele, tähelepanu äratamine, kutsumine, oma emotsioonide väljendamine jne), neil on olemas oma suhteliselt püsiv semantiline sisu ning nad on kasutaja poolt teadlikud ja suunatud kuulajale.

Seega võiks nad oma kuju poolest paigutada pigem üneemide hulka, oma kasutuse poolest aga partiklite hulka. Kummalgi juhul ei saa me ühest jaotust. Üneemide hulka paigutades (nagu Hennoste 1997) saame kaks teineteisest tähenduselt ja kasutuselt erinevat üneemide rühma. Partiklite hulka paigutades aga satuvad sellised üksused ühte rühma klassikaliste sõnadega (nagu Hakulinen 1998).

Käesolevas jaotuses on esmalt lähtutud foneetikast. St siia on kogutud sellised keelelised moodustised, mida me foneetiliselt ja fonoloogiliselt klassikalisteks eesti keele sõnadeks pidada ei saa. Seejärel on need aga jaotatud kahte rühma vastavalt oma funktsioonile. Üneemide hulka on arvatud sellised üksused, mille põhifunktsiooniks on hesitatsioon, ajavõitmine teksti tegemisel. Partiklite rühma on arvatud sellised moodustised, mille keskseks eesmärgiks on suhtlemine. Need on sellised üksused, mille abil kõneleja annab tagasisidet suhtlusel ning väljendab eeskätt oma erinevaid emotsioone. Statistiliselt kesksed üneemid on järgmised:

TABEL 7

Kesksed üneemid ja interjektsioonid

1377	ee/e/eee
463	aa/a/aaa
286	õõ/õ
257	mm/m
107	ää/ä
105	ii/i
60	öö/ö

Kokku esineb 2793 üneemikasutust, neist neli keskset on *ee*, *aa*, *õõ*, *mm*, mis annavad kokku 2358 kasutust (86%).

Interjektsioone on kokku 3283 kasutust. Interjektsioonide liigitamisega on praeguse materjali põhjal suuri raskusi. Põhiprobleemiks on see, et korpuse tegemise algajal ei olnud sellistest üksustest midagi täpsemat teada. (Millised on erinevad häälduslikud ja tähenduslikud variandid? jne.) Seetõttu on siin kõikumisi märkimisväärsed, mis ei luba esialgu samade interjektsioonide variante ning interjektsioonide põhitähen-dusi kuigivõrd välja tuua. Esitame siinkohal ainult kesksete interjektsioonide loendi.

TABEL 8

Kesksed interjektsioonid praeguses transkriptsioonis

970	mhmh ja selle variandid (mhõh, ähäh jms)
372	hh ja variandid (hhh jms)
344	he
286	ahah/ahaa
138	mh ja variandid
136	heh
108	ah/aah
95	ha/haa
44	oi
43	häh/hä/hah
35	eh / ehh
29	hõh/õh
28	oo
27	hm/hmm
22	n
18	nh
16	oh
12	hõ
11	khm
10	heheh

Lisaks on vähemalt 4 korda kasutatud järgmisi interjektsioone: *ai, näh, fuh, fui, mnja, bla, hä, tjah, oih, hõk*.

Kõige sagedasem interjektsioon on tagasisidepartikkel *mhmh* ja selle variandid, mis esinevad 970 korda. *mhmh* annab teada, et kuulaja on kuuldel ja lubab kõnelejal jätkata. Sellel on kinnise suuga variandid (*mhmh*) ja lahtise suuga hääldatud variandid (*mhõh/ähäh* jms). 286 korda esineb teine tagasisideinterjektsioon *ahah/ahaa*, mis märgib arusaamist. Lisaks saab välja tuua kõige sagedasema hüüatuse, milleks on *oi* (44 korda).

Kokkuvõte

Mis iseloomustab kokkuvõttes eesti suulist spontaanset dialoogilist kõnet selle korpuse põhjal? Me saame välja tuua rea jooni, mis on statistiliselt levinud ning seetõttu määravad selle registri mündi kuulaja jaoks. Lisaks on just need kõige olulisemad erijooned, mis eristavad seda allkeelt kirjalikust redigeeritud monoloogilisest keelekasutusest ja kitsamalt kirjakeelest.

Nende joonte puhul ei ole põhimõttelist vahet argikeele ja avaliku keelekasutuse vahel. Statistiline erinevus ilmselt on, kuid see vajab veel eraldi uurimist.

Eesti keele suulise kõne statistiliselt kesksed erijooned on selle korpuse põhjal järgmised:

1) Suulise kõne kõige olulisemaks leksikaalseks erijooneks on omaette sõnarühm, mida nimetatakse pragmaatilisteks partikliteks ja mille kasutussagedus on väga suur (vt võrdluseks kirjalike tekstide sõnavormide sagedussõnastike algusi Kaasik et al 1976; Tuldava 1978; TÜ arvutuslingvistika lehekülg). Neil sõnadel on üsna üldine semantiline sisu, kuid küllalt selgelt piirtletavad pragmaatilised rollid (vt nende kohta pikemalt Stenström 1990; Hakulinen 1998; Hakulinen et al 1989; Hennoste 1994a, 1994b).

2) Teiseks sagedaseks suulise kõne erivahendirühmaks on üneemid ja interjektsioonid, mida kirjalik tekst kasutab väga vähe spetsiifilistes retoorilistes funktsioonides (vt Hennoste 1997; Hakulinen 1998).

3) Kolmandaks oluliseks erijooneks on see, et suulises kõnes esinevad samadel sõnadel mitmed paralleelsed hääldusvariandid, millest üks on tavaliselt kirjakeelne ja muud on kasutusel valdavalt ainult suulises kõnes. Eripäraseid hääldusvariandid on kõige lihtsamalt jagatavad kolme põhirühma: häälikute/silpide lühenemine või kadu, teistmoodi häälikud, häälikute pikenemine. Lühemad hääldused on

jagatavad omaette sõnadeks ja ajutisteks hääldusvariantideks. Omaette sõnad omakorda jagunevad omaette tähendusega sõnadeks (*vä, sis*) ja pikemate sõnade sünonüümideks. Viimased ning ajutised lühemad hääldusvariandid on kasutusel teatud tekstuaalsetes tingimustes, eelkõige ebaolulise info edastamisel ja kiiremas kõnes. Teistmoodi hääldatud sõnad on valdavalt kirjakeelsete sõnade murdetaustaga variandid, mida võibolla kasutatakse ka kiirkõnes. Tugevama hääldusega variandid paistavad olevat eelkõige ajutised hääldusvariandid, mille kasutustingimused on esialgu teadmata.

4) Sõnu hääldatakse palju kokku. Kõige sagedasem on mitmesuguste sidendite ja partiklite kokkuhääldamine neile eelneva sõnaga (*ja, vä, et, sis*), eriti lausepiiril ning mitmesuguste takerdumiste puhul. Sealjuures on eesti suulises kõnes olemas küsiv partikkel *vä*, mis hääldatakse normaalselt eelneva sõnaga kokku ja mis seetõttu on pigem kliitik kui sõna.

Muud erijooned on palju haruldasemad. Neist tõusevad esile järgmised:

5) Suulises kõnes kasutatakse tüvisõnu, mida kirjalik keelekasutus kas ei vaja (dialoogisõnad) või ei luba sotsiaalsel põhjustel kasutada (ropendamine jms).

6) Suuline kõne kasutab kirjakeelest erinevaid sõnade saamise viise, millest selle korpuse põhjal on suure sagedusega ebatavaline sõnade liitmine ja grammatilise kategooria muutmine uute partiklite saamiseks. Ka tulevad välja mõned eripärased liited (*-kas*), raie ja kalduvus kasutada mõnedel juhtudel prototüüpseid sõnasaamisvõtteid ebatavaliste asemel (genitiivne liitumine nominatiivse asemel). Eelnevast on näha, et oluline ja statistiliselt sage sõnade saamisviis on lühemate sõnade kasutamine kui kirjakeeles. Seda lubavad nii raie, sõnade lühenemine kui ka mõnede liidete ja lõppude lühemad variandid (*-nd, -kend*).

7) Suulises kõnes esinevad mõned grammatilised erijooned, millest ainus statistiliselt silmapaistev on *-nd/-nud* partitiivi vaheldumine. Muud on statistiliselt vähe levinud (*-da* kaudne kõneviis, eripärased tüvevokaalid, *-si* -partitiiv jms); Teiselt poolt vaadates saame välja tuua erijooned, mis iseloomustavad erinevaid sõnaliike:

1) Partiklite seas on nii omaette tüvesid, kirjakeelest erineva tähendusega tüvesid, eripäraseid lühenemisi, liitsõnu, sõnaliigi muutmisi kui ka võõrkeelseid toorlaene, näiteks:

- eraldi tüved: *jah, noh, no, jaa, tere, eks, aitäh, nägemist, täitsa, vot, kah, kurat, okei, issand, kuulmiseni, hallo, vabandust, tshau*;
- lühemad sõnad: *sis, vä, nüüd, nimodi, vel, aint, tegelt, põhimõtselt, tändap, kuat, ütme*;
- liitsõnad: *onju, nojah, eksle, jaja, eksju, jajaa, einoh, nojaa, jadaega*;
- verbivormidest tehtud partiklid: *ütleme, kule, tändab, ota, vata*;
- võõrkeelsed sõnad: *jes, vau, hellou, sorri*.

2) Omadussõnade seas on olulised omaette tüved, mis tavaliselt seostuvad argikeelega (*õudne, jube, maru, jama, loll, lahe, jõle, nõme, kihvt*). Need sõnad on kasutatavad nii omadussõnadena (*loll lugu*) kui ka partiklitena (*loll!*)

3) Asesõnade eripäraks on samuti mõned omaette tüved, mida käsitletakse tavaliselt argikeelsena (*siuke, niuke, nisuke, sihuke*), lisaks esinevad lühemad vormid (*se, ned, seline*).

4) Adverbide seas on kesksed lühemad tüved (*sin, ku, ni, kudagi*). Sellele lisaks on kasutusel ka ebatavalised lõpud (*jällegist*).

5) Verbide seas on kesksed lühemad tüved (*mõtsin, ütsin, sab*), mis tihti on hääldatud kokku järgneva sidendiga (*ütsin=et*). Verbide grammatilistest erijoontest on sagedane

nd -partitiiv. Muud eripäraseid verbivorme on statistiliselt haruldased.

6) Arvsõnade seas on keskseks *-kümme* -liide lühemal kujul nominatiivis (*-kend*, *-küme* jms) ning lisaks esinevad *h* kadu teise silbi alguses (*kaeksa*) ja *ü* > *ö* sõnas *öheksa*.

7) Eripäraseid nimisõnavorme on korpuse väiksuse tõttu vähesagedased. Praegu tõusevad oluliseks omaette tüved (*kutt*, *värk*, *jubin*), *-kas* -liide (*pastakas*, *telekas*) ja lühemad tüved (*makk*, *trenn*). Kõik need sõnad seostuvad tavaliselt argikeelega. Grammatiliselt sagedasi erijooni ei ole. Välja võib tuua mõned eripäraseid türevokaalid (*rämpsum*) ja *-si* -lõpu partitiivis (*nägusi*).

8) Keskseid üneemid on vokaalid (*ee*, *aa*, *õõ*). Keskseid tagasisidevahendid on interjektsioonid *mhmh* ja *ahah*. Muud sagedased interjektsioonid on eelkõige mitmesugused kõhatusesarnased häämitsused (*hh*, *he*, *mh*, *ah*, *ha*), kõige tavalisemad emotsiooni väljendavad hüüatused on *oi*, *eh*, *oo* ja *oh*.

Seega võime näha, et sõnaliigiti on kesksed suulise kõne erijooned erinevad. Partiklite puhul kasutatakse paljusid erinevaid sõnahankimisvõtteid. Omadussõnade, asesõnade ja nimisõnade seas on olulised omaette tüved, mis seostuvad eelkõige argikeelega. Adverbide, verbide ja arvsõnade seas on aga kesksed lühemad tüved, mida kasutatakse ka nimisõnade puhul. Keskseteks üneemideks on vokaalid, keskseteks interjektsioonideks aga mitmesugused kõhatusesarnased h-elementi sisaldavad häämitsused.

KIRJANDUS

Ariste, Paul 1977: Eesti keele foneetika. Tartu Riiklik Ülikool.

Brown, Gillan 1990. Listening to Spoken English. Patterns of simplification in informal speech. 2nd ed. Longman. 57–88.

Cui, Kaily 1999: Sõnaalguline *h* eesti keeles. Bakalaureusetöö. Tartu Ülikool.

- Hakulinen, Auli 1998: Partikkelit. Ilmub raamatus Suomen kielen deskriptiivinen kielioppi. Käsikiri.
- Hakulinen, Auli; Suojala, Marja; Sorjonen, Marja-Leena 1989: Partikkelit ja muut kiteymät vuoroissa. –Kieli 4. Suomalaisen keskustelun keinoja I. Toim. A. Hakulinen. 115–146
- Hennoste, Tiit 1994a: Prospektiivsed minimaalhesitatsioonid eesti keele suulises tekstis. – Lähivertailuja 7. Turun yliopiston suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 44. Toim K. Pajusalu ja V. Yli-Vakuri. Turku. 33–51.
- Hennoste, Tiit 1994b: Pragmaatiline partikkel NOH eesti keeles. Sissejuhatav ülevaade. – Lähivertailuja 8. Oulun yliopiston suomen ja saamen kielen laitoksen tutkimusraportteja 40. Toim H. Sulkala ja H. Laanekask. Oulu. 9–20.
- Hennoste, Tiit 1997: Planeerimisüneemid eestikeelsetes suulistes tekstides. – Lähivertailuja 9. Suomalais-virolainen kontrastiiviseminaari 3.–5.5. 1996 Lammi. Castrenianumin toimitteita 53. Toim. Riho Grünthal ja Reet Kasik. Helsinki. 53–67.
- Hennoste, Tiit 1998. Suulise kõne sõnavara. Sissejuhatav ülevaade. – Oe-keeta asijoo. Commentationes Fenno-Ugricae in honorem Seppo Suhonen. Suomalais-Ugrilaisen Seuran toimituksia 228. Toim. Riho Grünthal ja Johanna Laakso. SUS, Helsinki. 156–167.
- Hennoste, Tiit 1996: Tartu University Corpus of Written Estonian: A Survey of the Structure of Texts and Principles of Selection. – Estonian in the Changing World. H.Õim (ed.). University of Tartu. Department of General Linguistics. Tartu. 7–32.
- Hennoste, Tiit ilmumas: Suulise eesti keele uurimine: transkriptsioon, taust ja korpus. Ilmub ajakirjas Keel ja Kirjandus.
- Kaasik, Ülo; Tuldava, Juhan; Villup, Astrid; Ääremaa, K. 1976: Eesti keele ilukirjandusproosa autorikõne sõnavormide sagedussõnastik. Keelestatistika 1. TRÜ Toimetised 377. Tartu. 107–153.
- Keevallik, Leelo 1994: <nud> varieerumine tänapäeva eesti kõnekeeles. Tartu. MA thesis.
- McEnery Tony & Andrew Wilson 1997: Corpus Linguistics. Edinburgh Textbooks in Empirical Linguistics. Edinburgh UP.
- Muischnek, Kadri 1998: Korpuslingvistika. Keel ja Kirjandus 1/1998. 8–12.
- Palmeos, Paula 1982: Eesti keele grammatika II. Teine vihik. Sidesõna ja hüüdsõna. TRÜ. 34–53.
- Shaw, G. B. 1972: Pygmalion – Näidendid. Eesti Raamat. 279–360.

Stenström, Anna-Brita 1990: Lexical items peculiar to spoken discourse. The London-Lund Corpus of Spoken English. Description and Research. J. Svartvik (ed.). Lund UP. 137–176.

Tiittula, Liisa 1991: Puhuva kieli. Suullisen viestinnän erityispiirteitä. Finn Lectura.

TÜ arvutuslingvistika lehekülgi Internetis: <http://www.cl.ut.ee>.

Tuldava, Juhan 1978: Sõnavormide esinemus eestikeelsetes tekstides. – Eesti keele sõnavarastatistika probleeme. Töid keelestatistika alalt 3. TRÜ Toimetised 446. Tartu. 107–126.

Palju tänu abi eest Eesti linnakeelte uurimisrühma liikmetele, minu suulise kõne loengukursuste kuulajatele ja Karl Pajusalule.