

## EESTI KEELE KASUTUSVARIANDID: KORPUSTEST TULENEV KÄANDEVORMIDE VÕRDLEV ANALÜÜS

Pille Eslon, Erika Matsak

**Ülevaade.** Eesti keele keeletehnoloogilise ressursi olemasolu annab võimaluse korpusest tulenevalt võrrelda kahte eesti keele aktiivselt kasutatavat varianti – kirjakeelt ja õppijakeelt. Tegemist on autentse keeleainesega, mille uurimisel saab rakendada tarkvara ning erinevaid kvantitatiivse ja kvalitatiivse analüüsi meetodeid, sh sõnavormide ja nende kasutuskontekstide sageduse automaatset leidmist. Saadud tulemustest peaks ilmnema, missugused vormid on erinevates keelevariantides sagedased; missuguseid tüüpilisi, ebatüüpilisi või spetsiifilist laadi kollokatsioonilisi üksuseid korpustes esineb; missugused grammatilised konstruktsioonid on erinevatele keelevariantidele tegelikult iseloomulikud ja tavapärased; missugust osa sõnavarast, vormidest ja konstruktsioonidest erinevad keelekasutajad tegelikult eelistavad.\*

**Võtmesõnad:** korpuslingvistika, korpustest tulenev võrdlev analüüs, käändegrammatika, käändevormide kasutuseelistused eesti keeles, eesti keel

### 1. Eesmärk

Uurimuse eesmärk on võrrelda käändekasutust kahes eesti keele variandis – kirja-keeles ja õppijakeeles. Aluseks on võetud Eesti Keele Instituudi (EKI) tekstikorpused<sup>1</sup> ja Tallinna Ülikooli eesti vahekeele korpus (EVKK)<sup>2</sup>. Nende ressursside põhjal tuuakse välja eesti keele nelja sagedasema nimisõna käändevormide sagedus ning kirjeldatakse nende sõnade morfoloogilist paradigmat. Statistikast peaks selguma, mil määral on sõna semantika seotud grammatiliste vormide moodustamise ning esinemissagedusega korpuses. Samas ei anna see teave tegelikult midagi olulist vormieelistuste kohta tekstiloomes. Küsimus aga on selles, missuguseid konstruktsioone ja käändevorme tekstiloomes eelistatakse ning miks. Seetõttu otsitakse

\* Tööd on toetanud riikliku programmi "Eesti keele keeletehnoloogiline tugi (2006 – 2010)" projekt "VAKO: Eesti vahekeele korpuse keeletarkvara ja keeletehnoloogilise ressursi arendamine".

<sup>1</sup> <http://www.eki.ee/corpus/> (3.09.2008).

<sup>2</sup> <http://evkk.tlu.ee> (3.09.2008).

korpusainesest välja kaks ning enam korda esinevad konstruktsioonid, milles on võimalik kasutada vaid kindlaid käändevorme ning mis on tekstiloomes olulised. Analüüsi tulemusena tuuakse välja kirjakeelele ja õppijakeelele omased kollokatsioonilised üksused, tüüpilised grammatilised konstruktsioonid ja nende leksikaalgrammatilised variandid, milles on kasutatud vaid teatud käändevorme. Siit peaks ilmnema, kuivõrd selgelt on eesti kirjakeeles ja õppijakeeles väljendunud seosed grammatiliste vormide ja sageli kasutatavate konstruktsioonide vahel, samuti see, mille poolest nende kahe eesti keele variandi käändekasutusmustrid sarnanevad või erinevad. Lahknevused kirjakeele ja õppijakeele käändekasutuse kontekstieelistustes annavad teavet ka võrreldavate keelevariantide diskursuserinevuste kohta ning toovad esile sel otstarbel kasutatud stereotüüpsed konstruktsioonid. See on aga oluline mitte üksnes pedagoogilistel ja leksikograafilistel eesmärkidel, vaid sünkroonsel tasandil keelesüsteemi sisemiste varjatult kulgevate arengutendentside väljatoomiseks.

## 2. Materjal, analüüsi suunad ja vahendid

Kirjakeele ja õppijakeele võrdlev analüüs piirdub sagedasemate nimisõnade käändevormide kasutamisega.<sup>3</sup> Sõnade sagedusandmed on võetud “Eesti kirjakeele sagedussõnastiku” tuhande sagedasema sõna ja sõnavormi järjendist (Kaalep, Muischnek 2002)<sup>4</sup> ning Eesti vahekeele korpuse sõna- ja vormisageduse statistikast. Nende allikate põhjal on kirjakeeles kümme sagedasemat nimisõna *aasta, aeg, mees, kroon, sõna, inimene, raha, elu, naine, ema* ja õppijakeeles *keel, inimene, aeg, elu, sõna, töö, kiri, kultuur, aasta, raha*. Et leida esikümne piires kirjakeele ja õppijakeele sõnavormide sagedusloendi ühisosa, mille all mõeldakse kokkulangeva või samalaadse sagedusega sõnapaare, siis tuleb kõigepealt välja jätta need sõnad, mida teise kasutusvariandi sagedusloendi esikümnes ei ole. Niisugune on õppijakeele sagedasem sõna *keel*, mis ei kuulu kirjakeele kümne ega kahekümnegi sagedasema hulka; samal põhjusel jäävad analüüsist välja kirjakeeles esikümnesse kuuluvad sõnad *mees, kroon, naine, ema* ja õppijakeeles *töö, kiri, kultuur, raha*. Sõna *aasta* on nii kirjakeele kui ka õppijakeele sagedasemate sõnade esikümnes (kirjakeeles sagedusrea eesotsas, õppijakeeles eelviimane ehk üheksas), kuid distants sageduses on selgelt märgatav. Seetõttu tasub kaaluda, kas sõna *aasta* on mõtet vaadelda sagedusloendite ühisosa all, olgugi et kuuluvus esikümnesse annab selleks formaalse aluse. Käesolevas uurimuses on otsustatud sõna *aasta* siiski analüüsist välja jätta. Ülejäänud sõnade sagedused on tunduvalt lähedased mad (*inimene* – kirjakeeles sageduselt kuues ja õppijakeeles teine, *aeg* – vastavalt teine ja kolmas, *elu* – kaheksas ja neljas), ühel juhul isegi kokkulangev (*sõna* – mõlemas sagedusjärjendis viiendal kohal). Niisiis on sõnade *inimene, elu, aeg* ja *sõna* kasutamine võrreldavates keelevariantides sageduse poolest analoogne ning need sõnad kuuluvad kirjakeele ja õppijakeele samalaadse sagedusega nimisõnade hulka. Nende vormisageduse uurimiseks ja kasutuskontekstide võrdlemiseks kahe korpusainese alusel tuleb eelnevalt tähelepanu pöörata mõningatele olulistele metodoloogilist laadi küsimustele, nagu valimite representatiivsus, võrreldavus, analüüsi suunad ja vahendid.

<sup>3</sup> Kuna sagedasemate nimisõnade väljatoomise protseduuri on kirjeldatud artiklis Esilon 2008: 33–35, siis sel küsimusel siinkohal pikemalt ei peatuta, antakse lühiülevaade ega laskuta metoodilist laadi üksikasjadesse.

<sup>4</sup> Vt ka <http://www.cl.ut.ee/ressursid/sagedused/> (4.09.2008).

## 2.1. Valimite representatiivsus ja võrreldavus

Kahe erineva korpusainestiku võrdlemine eeldab korpuste representatiivsust ja võrreldavust. Traditsiooniliselt on representatiivsuse tagatiseks peetud ühelt poolt korpuste ja teisalt nende uurimise mitmemõõtmelisust; võrreldavateks on tunnistatud samadel alusprintsipiidel rajanevaid korpusi (vt Atkins, Clear, Ostler 1992, Biber 1993, McEnery, Wilson 2001: 29–32). Samas pole reaalne, et ka kõige järjekindlamalt üles ehitatud tasakaalustatud korpus sisaldaks keele kohta absoluutselt kõike ning vastaks mis tahes uurimistöö eesmärkidele (vt nt Bergh 2005, Volk 2002), eriti puudub see suletud korpusi. Iseenesest on see küsimus muidugi oluline, ennekõike seoses mitmekeelsete ja paralleelkorpuste uurimisega (vt nt Baker 1995), kuid ühe keele erinevate kasutusvariantide võrdlemisel arvatavasti mitte niivõrd keskne. Näiteks on Ute Römer püüdnud välja selgitada, mis on inglise keele *if*-lausetate kasutamisel normipärane, sest tegelikult on normi ehk standardi mõiste kokkuleppeline ega pruugi kajastada loomulikku keelekasutust, mida haritud emakeelekõneleja aktsepteerib. Võrrelnud *if*-lauseid kolmes keeleandmestikus (inglise keele õpik, British National Corpus (BNC), saksa emakeelega õppijate inglise vahekeel ehk õppijakeel<sup>5</sup>), jõuab Römer järeldusele, et “kooli” inglise keel ei ole vastavuses loomuliku keelekasutusega. Seetõttu peab õpikutes ja keeleõppes ümber orienteeruma autentse kirjakeele sellele osale, mis on tavapärane Briti inglise keelele ja mida emakeelevaldaja loomulikus keelekasutuses eelistab (Römer 2007: 355, 358–363). Niiugusele järeldusele aitas U. Römeril tulla ühe keele kolme kasutusvariandi korpustest tulenev analüüs ja võrdlus, ent samas pole ei inglise keele õpik, BNC ega ka inglise õppijakeel samalaadsed ei mahult, adressaadi valikult, tekstiliikide ja registri ega millegi muu poolest. Teistsuguste uurimiseesmärkide puhul tuleb korpusainestiku tasakaalustatus uurimistulemuste adekvaatsusele kahtlemata kasuks, sest korpuspõhine mitmemõõtmeline keelekirjeldus näitab selgemalt keelendi(te) võimalikku varieerumist olenevalt sagedusest näiteks suulises kõnes ja kirjalikes tekstides või siis keelendi(te) diakroonilist ja regionaalset varieerumist, registrierinevusi, tekstiliikidest sõltuvat varieerumist jne. Inglise keele modaalverbide korpuspõhises analüüsis (läbi viidud BNC alusel) on seda kõike kujukalt demonstreerinud Graeme Kennedy (vt Kennedy 2002: 76–86). Sellele vaatamata ollakse seisukohal (nt Xiao, McEnery 2005), et korpuste mitmemõõtmelisuse ja tasakaalustatuse küsimus pole erinevate ainestike võrdlemisel ning uurimistulemuste adekvaatseks tunnistamisel omaette küsimus, sest keelt saab uurida mitte ainult lingvistiliselt ja korpuspõhiselt, vaid ka statistikal rajanevat automaatset formaalset analüüsi rakendades (nt WordSmith Tools, MonoConc, MonoConc Pro, WordCruncher, Raymond Hickey välja töötatud programmid, R Projecti programmide pakett jt). Korpusest otsitakse sõna- ja vormisagedust, statistilisi kollokatsioonilisi üksuseid, konstruktsioone, klastreid jms. Nii on Zhonghua Xiao ja Anthony McEnery väitnud, et üks WordSmith Toolsi programme KeyWords võimaldab võrrelda erinevaid korpusaineseid sama efektiivselt kui representatiivseks ja omavahel võrreldavaks tunnistatud mitmemõõtmeliste korpusuuringutele tuginevad laiahaardelised korpuspõhised uurimused (vt Xiao, McEnery 2005: 62–82). Kadri Muischnek näiteks on statistiliste kollokatsioonide automaatseks tuvastamiseks kasutanud keelest sõltumatu tarkvara SENTA modifi-

<sup>5</sup> Mõisteid *vahekeel* ja *õppijakeel* kasutatakse siinkohal sünonüümidena.

katsiooni SENVA, mis on kohandatud eesti keele verbikesksete ühendite leidmiseks (vt Muischnek 2006: 40).

Kuna nii EKI tekstikorpused kui ka EVKK on avatud ehk monitorkorpused, siis on sõnade arv neis erinev ning kasvab pidevalt. Järelikult on nende korpuste põhjal tehtavas keeleanalüüsis mõttekas alustada sellest, et määrata võrreldavate valimite suurus ja hinnata nende usaldusväärsust. Optimaalse suurusega valimi leidmiseks tehti mõlemas korpuses päring nelja sagedasema nimisõna vormikasutuse kohta ning leiti vastavalt 98928 ja 19433 erikäändevormi.

Järgmine samm optimaalse valimi leidmisel oli tekstilõikude arvu piiramine seitsme tuhandega. Kuna igal lõigul on põhimõtteliselt võrdne tõenäosus sellesse valimisse sattuda, siis eespool nimetatud kindla üldkogumi korral (98928 ja 19433 käändevormi) on EKI tekstikorpuse ja EVKK suhtes representatiivne niisugune valim ehk tekstilõikude kogum suurusjärgus 7000, millesse iga üldkogumi objekt ehk tekstilõik satub võrdse tõenäosusega. Kui see osutub võimalikuks, siis saab iga nimetatud suurusjärguga valimi tunnistada antud uurimuse jaoks sobivaks. Et veenduda kahe valimi alusel saadud päringute kokkulangevuses, tehti EKI tekstikorpusest ja EVKK-st kuuajalise vahega ühe ja sama sõna käändevormide ilmumise kohta kaks erinevat päringut (Päring 1 ja Päring 2). Siinkohal näide EKI tekstikorpuse põhjal tehtud kahest päringust, millest selgub, et sõna *inimene* käändevormide esinemissagedus kahes erinevas valimis langeb kokku (vt tabel 1).

**Tabel 1.** Sõna *inimene* käändevormide sagedus EKI tekstikorpuse valimites

Sõnavorm	Päring 1	Päring 2	Kääne
<i>inimest</i>	1421	1421	SG.PART
<i>inimesed</i>	1164	1165	PL.NOM
<i>inimene</i>	960	964	SG.NOM
<i>inimese</i>	765	766	SG.GEN
<i>inimesi</i>	704	704	PL.PART
<i>inimeste</i>	598	600	PL.GEN
<i>inimestele</i>	230	232	PL.ALL
<i>inimesele</i>	176	177	SG.ALL
<i>inimestel</i>	137	137	PL.ADS
<i>inimesel</i>	125	123	SG.ADS
<i>inimestega</i>	107	107	PL.KOM
<i>inimestest</i>	106	106	PL.ELAT
<i>inimesega</i>	75	75	SG.KOM
<i>inimesest</i>	55	55	SG.ELAT
<i>inimestelt</i>	29	29	PL.ABL
<i>inimeseks</i>	27	27	SG.TRANS
<i>inimesena</i>	23	23	SG.ESS
<i>inimeselt</i>	19	19	SG.ABL
<i>inimeses</i>	17	17	SG.INES
<i>inimestes</i>	13	13	PL.INES
<i>inimeseni</i>	8	9	SG.TERM
<i>inimesteks</i>	8	8	PL.TRANS
<i>inimesteni</i>	5	5	PL.TERM
<i>inimestesse</i>	0	5	SG.ILL

Et olla veendunud, kas nii saadud tulemus on tõepoolest usaldusväärne, kontrolliti Päring 1 ja Päring 2 valimite sarnasust ka Pearsoni  $\chi^2$  valemiga:

$$\chi^2 = \sum_{i=1}^k \frac{(n_i - N_i)^2}{N_i} \underset{H_0}{\sim} \chi^2_{df}$$

$\chi^2$  valemi abil saab valimites hinnata vahemikus 0–1 seda, kas erinevus mingites tunnustes on oluline või mitte. Selleks arvutatakse välja teoreetiliselt võimalik vabadusaste (tähistatud suurusega  $df$  = ingl *degree of freedom*), mille alusel saab määrata valimi olulisustõenäosuse  $p$  (*probability level*), vt tabel 2.

**Tabel 2.** Valimite sarnasus arvatuna  $\chi^2$  valemi abil

Käänded	Päring1	Päring2	Kokku
SG.PART	1421	1421	2842
	1419,428	1422,572	
PL.NOM	1164	1165	2329
	1163,212	1165,788	
SG.NOM	960	964	1924
	960,9358	963,0642	
SG.GEN	765	766	1531
	764,6531	766,3469	
PL.PART	704	704	1408
	703,2212	704,7788	
PL.GEN	598	600	1198
	598,3373	599,6627	
PL.ALL	230	232	462
	230,7445	231,2555	
SG.ALL	176	177	353
	176,3047	176,6953	
PL.ADS	137	137	274
	136,8484	137,1516	
SG.ADS	125	123	248
	123,8628	124,1372	
PL.KOM	107	107	214
	106,8816	107,1184	
PL.ELAT	106	106	212
	105,8827	106,1173	
SG.KOM	75	75	150
	74,91703	75,08297	
SG.ELAT	55	55	110
	54,93915	55,06085	
PL.ABL	29	29	58
	28,96792	29,03208	
SG.TRANS	27	27	54
	26,97013	27,02987	
SG.ESS	23	23	46
	22,97456	23,02544	
SG.ABL	19	19	38
	18,97898	19,02102	
SG.INES	17	17	34
	16,98119	17,01881	
PL.INES	13	13	26
	12,98562	13,01438	

SG.TERM	8	9	17
	8,490597	8,509403	
PL.TRANS	8	8	16
	7,99115	8,00885	
PL.TERM	5	5	10
	4,994469	5,005531	
SG.ILL	0	5	5
	2,497234	2,502766	
<b>Kokku</b>	<b>6772</b>	<b>6787</b>	<b>13559</b>
Pearson's $\chi^2$ statistic	5,082593		
df	23		
p	0,999967		

Tabelis 2 esitatud andmed näitavad, et vabadusastme  $df = 23$  puhul on olulisustõenäosus  $p = 0,999967$ , mis tähendab, et kahe erineva päringu tulemused on sarnased. Seega on EKI tekstikorpuse 7000 lõigust koosnev valim ja EVKK 7000 lõigust koosnev valim nelja sagedasema nimisõna käändegrammatika võrdlevaks uurimiseks suurusjärgult piisavad.

## 2.2. Analüüsi suunad

Selle alusel, kuidas korpusi on keele uurimisel kasutatud, eristatakse kaht peamist suunda: ühelt poolt korpuspõhist keeleanalüüsi (ingl *corpus-based language analysis*) ja teisalt korpusest tulenevat uurimist (*corpus-driven research*). Kõige üldisema seletuse järgi on korpuspõhise analüüsi puhul korpus keeleaines uurija jaoks allikmaterjal, millest vastavalt püstitatud uurimiseesmärgile vajalikku teavet saada. Elena Tognini Bonelli on kirjutanud, et seejuures kasutab uurija korpusi kui keelekirjelduste ja teooriate paikapidavuse tõendusmaterjali: “.. use of a corpus as evidence for language description is usually referred to as corpus-based” (Tognini Bonelli 2002: 73). Selline lähenemisviis korpusetele on olnud traditsiooniline inglise keele sõnavara uurimisel nii leksikograafilistel kui ka peadagoogilistel eesmärkidel (vt Kennedy 1999: 108–121). Alates H. M. Palmerist, J. R. Firthist ja M. A. K. Halliday’st on sõna tähendust peetud sõltuvaks nii leksikaalgrammatilisest kui ka ekstralingvistilisest kontekstist (sotsiaalne keskkond, keelenormid jms), milles sõnakasutus ilmneb (vt McEnery, Wilson 2001: 23–24). John Sinclairi (1991) järgi on sõna tähenduse kujunemise aluseks semantika ja grammatika koosmõju kindlat tüüpi konstruktsioonides, fraasides ja lausetes. Seega on sõna tähendus olemuselt tekstuaalne nähtus, mis on kirjeldatav keelesüsteemi süntagmaatilisel tasandil. Analoogselt on uuritud ka grammatiliste vormide ja konstruktsioonide kasutust. Näiteks Sylviane Grangerit on huvitanud, kuidas inglased ja edasijõudnud inglise keele kui võõrkeele õppijad oma akadeemilist laadi kirjutistes partitsiipi kasutavad. Vastuse saamiseks on ta leidnud näitelauseid kahest erinevast korpusainestikust, seejärel neid lauseid lingvistiliselt analüüsinud ning saadud tulemusi võrrelnud (Granger 1997: 185–198).

Korpusest tulenev keeleainese uurimine esindab vastupidist suunda: korpus ei ole siin teooriate tõestamiseks mõeldud tekstikogu, vaid keeleaines, millel on võimalik rakendada erinevaid automaatseid programme ning statistilise analüüsi meetodeid ja saada põhimõtteliselt erinevat ning mõneti ootamatut teavet, mis

traditsioonilises korpuspõhises lingvistilises analüüsis ei eksplitseeru, kuid mis on omane loomulikule keelekasutusele ning just seetõttu huvipakkuv<sup>6</sup>. Uudsus seisneb selles, et traditsioonilist lingvistilist uurimisainest analüüsitakse mittelingvistiliselt ja formaalselt. Seejuures ei pea korpus olema märgendatud, piisab elektroonilisest tekstiarhiivist. Mittelingvistiline lähenemine keeleainesele on nii empiirika kui ka keeleteooria seisukohalt innovaatiline, kuna võimaldab lingvistil leida uusi lahendusi seni raskesti kirjeldatavatele keelesüsteemi nähtustele. Näiteks on Tognini Bonelli rõhutanud, et korpusest tulenev keeleuuring võimaldab ületada veelahkme leksika ja grammatika vahel, ühendades need kaks keelesüsteemi osa süntagmaatiliseks tervikuks (Tognini Bonelli 2002: 75). Nadja Nesselhauf kirjutab sagedusel põhinevatest erineva varieerumisvabadusega süntaktilistest ja semantilistest terviküksustest (Nesselhauf 2005: 21). Iga sõnavorm tuleb esile vaid talle omastes kasutuskontekstides, mille alusel saab välja tuua erineva pikkuse ja sünonüümireaga kollokatsioonilisi leksikaalseid üksuseid. Näiteks inglise keele sünonüümide *pretty* ja *handsome* kasutuspiirangud ehk tüüpilised kontekstid on ühelt poolt *pretty girl, boy, woman, flower, garden, colour, village* ja teisalt *handsome boy, man, car, vessel, overcoat, airliner, typewriter* (vt Leech 1981: 17). Neid kollokatsioonilisi üksuseid kasutatakse kui terviklikke konstruktsioone. Sellistena on nad tavapärased, üldkasutatavad, osati hinnangulisust edastavad ja pooleldi idiomaatilised.<sup>7</sup> Kui kasutuspiirangutest pole kinni peetud, siis on tegu ebatüüpiliste kollokatsiooniliste üksustega, mille ilmnmisel on oma kindlad põhjused: näiteks suhtlussituatsioonist tulenev või isikupärane keeleloome ja registrivalik, keeleõppija või tõlkija normivastane või laiemalt – mitteaktsepteeritav sõnakasutus jm. Sõnavormide kasutuskontekstide valik võib olla piiratud ka teatud leksikaalgrammatilise rühma, grammatiliste kategooriate ja vormide varieerumisega (süntaktilised sünonüümid).

Keeleüksuste automaatne väljatoomine korpusainestikust võimaldab leida konstruktsioone, milles üht ja sama vormi on vähemalt kahel korral samas ümbruses kasutatud. Nende võrdlemisel ilmnevad sageli kasutatud konstruktsioonide leksikaalsed ja leksikaalgrammatilised variandid. Inglisekeelsetes allikates on niisuguseid konstruktsioonide kogumeid nimetatud keelendite kasutusmustriteks või -mallideks (ingl *patterns*), mida on peetud olulisteks keeleõppe, sõnastike ja grammatikate tarvis (vt nt Granger 1998: 145, Hanston 2002: 167, Römer 2005, Nesselhauf, Römer 2007). Näiteks EVKK-s kasutatakse sõna *inimene* ainsuse nominatiiviga sagedasti loogilist implikatsiooni *kui inimene on*. Sel lingvistilisel konstruktsioonil on õppijakeeles hulk leksikaalseid (*kui inimene tahab, kogeb, elab, suhtleb*) ning leksikaalgrammatilisi variante (*et inimene peab ~ ja kui (et kui, sest kui) inimene (peab) või inimene peab olema (maksma, teadma) ~ inimene võib töötada ~ inimene ei saa (ei ole) või haritud (iga) inimene peab saab, võib, tahab*), mis kujutavad endast sõna *inimene* ühte eesti õppijakeelele omast ainsuse nominatiiviga kasutusmustrit. Kasutusmustrit dominandiks on konstruktsioon *kui inimene on*, mida selle leksikaalgrammatiliste variantidega võrreldes on korpuses kõige sagedamini kasutatud.

<sup>6</sup> Väga huvitavaid tulemusi on andnud teksti ideoloogia, suunitluse, allteksti mõistmise ning autori mõttelaadi ja eelistuste avamisel korpusest tulenev diskursusanalüüs (vt Stubbs 1996, Orpin 2005, Hoey jt 2007). Seda suunda on essee kui tekstitüübi määramisel rakendanud ka Heidi Meier (2003).

<sup>7</sup> Sünonüümide kasutuspiirangutest soome keeles vt Jantunen 2004; eesti keele verbikesksete püsiühendite – idioomide, poolidiomaatiliste ja kollokatiivsete mitmesõnaliste ühendite analüüsi vt Muischnek 2006; sageduse alusel esile tulevaid kollokatsioone loomulikule keelekasutuses ja inglise õppijakeeles on võrreldud monograafias Nesselhauf 2005.

### 2.3. Analüüsi vahendid

Kirjakeele ja õppijakeele nelja sagedasema nimisõna *inimene, elu, aeg* ja *sõna* käändevormide sageduse ja kasutuskontekstide väljatoomiseks ning võrdlemiseks on käesolevas uurimuses rakendatud programmi WordSmith Tools 5,0 (arendaja Mike Scott), mis on üks tunnustatumaid tekstitöötlusvahendeid maailmas (vt Scott, Tribble 2006).<sup>8</sup> Viimasel ajal propageeritakse John M. Chambersi (2007) arendatavat rikkalike visualiseerimisvõimalustega vabavaralist statistikapaketti “The R Project for Statistical Computing”<sup>9</sup>.

WordSmith Tools sisaldab kolme programmi, mille abil saab korpuse sõnavara sagedusloendi (Word List), konkordantsid ehk grammatiliste vormide kasutuskontekstid (Concord) ja võtmesõnade võrdluse (KeyWords). Kui tegu on lemmatiseerimata korpusega (nagu EKI tekstikorpuse ja EVKK), siis võib analüüsitava sõnade grammatilised vormid Word Listi alusel ühte koondada ja sel teel tuua välja lemmad.

Grammatiliste vormide kasutuskontekstid annab programmi Concord rakendamine. Konkordantside alusel võib välja tuua sõnavormide kasutusmuustrid ja saada andmed kindlate grammatiliste konstruktsioonide tavapärasuse ning nende võimalike leksikaalgrammatiliste variantide kohta korpusainestikus. Programm Concord lubab automaatselt rakendada ka tekstilõigu pikkuse piirangut, mis on oluline eesti keele grammatilise homonüümia probleemi lahendamisel, eriti ühendverbi ja nimisõna käändevormide eristamisel. Artiklis analüüsitava sõnade puhul on vormihomonüümia seotud sõnadega *elu, aeg* ja *sõna*: näiteks sõnavormi *elu* puhul on küsimus grammatiliste käänete eristamisest, vorm *elus* on määratav kas nimisõna ainsuse inessiivi või omadussõnana, mille vastandsõna on *surnud*; sõnavorm *aja* on kas ainsuse genitiiv või käskiv kõneviis verbidest *ajama, naerma ajama, minema ajama* jt; *ajad* on kas nimisõna mitmuse nominatiivi vorm või indikatiivi oleviku ainsuse 2. pööre (*mis juttu sa ajad*); *sõnas* – kas ainsuse inessiivi vorm või imperfekti ainsuse 3. pööre. EKI tekstikorpuse ja EVKK valimite vormihomonüümia väljaselgitamiseks oli Concordi tekstilõigu pikkus täiesti piisav. Siinkohal lühike näide EKI tekstikorpuse konkordantsist sõnavormiga *ajas*; näidetes (3) ja (5–7) on kasutatud nimisõna *aeg* ainsuse inessiivi vormis ja ülejäänud juhtudel on tegu verbiga indikatiivi imperfekti ainsuse 3. pöördes:

- |                                        |              |                                    |
|----------------------------------------|--------------|------------------------------------|
| (1) nagu poleks midagi juhtunud,       | <b>ajas</b>  | käed ja jalad laiali ning kordas   |
| (2) voolas särki alla, krae vahele,    | <b>ajas</b>  | vastikult lõdisema                 |
| (3) Ma ei oska asju                    | <b>ajas</b>  | ritta panna.                       |
| (4) natuke suurem kui vaja ja see      | <b>ajas</b>  | natuke hirmu nahka                 |
| (5) arhitektuur elab muutuste          | <b>ajas.</b> | Lätlaste Sandra Levane             |
| (6) ja kiirelt muutuv                  | <b>ajas.</b> | Arvestatavaks ehituskunsti         |
| (7) “Felixi” toodang asendab praeguses | <b>ajas</b>  | Eesti lastele Campbelli supipurke. |

Analüüsis loobuti automaatselt morfoloogilisest ühestamisest, sest käsitsi ühestamine, olgugi et töömahukam, annab täpsema tulemuse. Kui aga eesmärk oli sõnavormide kasutusmuustrite esiletoomine, siis piirati tekstilõik kolme üksusega, nt *tol ajal oli, sellel ajal oli, sellel ajal on, sel ajal tekkis* jne. Antud juhul on

<sup>8</sup> Kasutada võib ka teisi samalaadseid programme, nt MonoConc, MonoConc Pro (Randi Reppeni võrdlust WordSmith Toolsiga vt <http://lt.msu.edu/vol5num3/review4/default.html>, 12.10.2008).

<sup>9</sup> Vt <http://www.r-project.org/> (12.10.2008).



konstruktsiooni *ajamäärus + verbivorm indikatiivis* leksikaalgrammatilised variandid reastatud esinemissageduse alusel, mis neil on EVKK-s (vastavalt 18, 12, 6 ja 5 korda). Konstruktsiooni identifitseerimiseks kasutati eesti keele süntaksianalüsaatorit, mille väljundit töödeldi spetsiaalselt programmeeritud makrode abil. Tänu sellele oli võimalik üles leida samalaadselt analüüsitud konstruktsioonid, milleks antud juhul on `//_P_dem sg ad #cap // **CLB @ NN> //_S_com sg ad #cap // @ ADVL //_V_main indic imp ps3 sg ps af #cap #FinV #Intr // @+FMV`.<sup>10</sup> Niisugune lähenemisviis võimaldas sorteerida nelja sagedasema nimisõna kolmest üksusest koosnevaid konstruktsioone käändevormide alusel ning siduda käändevormide kasutamise kindlat liiki konstruktsioonide ja nende leksikaalgrammatiliste variantidega. Kirjakeele ja õppijakeele korpusainestiku alusel saadud andmete edasine võrdlus toob esile teatud käändevormide kasutuseelistused kindlat liiki konstruktsioonides.

Programmi KeyWords rakendamine annab sõnade olulisuse nende sageduse põhjal ning sobib seetõttu kahe keeleainese võrdlemiseks. Kuna programm arvestab keeleüksusi, mille minimaalne sagedus korpusel on kaks korda, siis on programmi abil välja toodud keeleaines mahukas. Seetõttu piirduakse valimiga 50% materjalist + 1 (vt Berber-Sardinha 1999). Kuna sinne uurimus on esialgne, siis piirduakse “Eesti kirjakeele sagedussõnastiku” ja EVKK statistikamooduli alusel kindlaks määratud nelja sagedasema nimisõna analüüsiga, mistõttu kahe valimi käändevormide statistilise sarnasuse või erinevuse esiletoomiseks on otstarbekam kasutada alternatiivset võimalust –  $\chi^2$  testi 2 x 2 tabelis. See tähendab, et iga käände kohta koostatakse tabel, mis näitab otsitava vormi sageduse suhet kahes erinevas valimis ülejäänud käändevormide arvuga, osutades käändevormi esindatusele (vt tabel 3).

**Tabel 3.** 2 x 2 tabeli näide

	Otsitava käände sagedus	Ülejäänud käändevormide arv valimis
EKI tekstikorpus	a	b
EVKK	c	d

$\chi^2$  arvutamiseks kasutatakse valemit

$$\chi^2 = \frac{(n-1)(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)},$$

kus  $n = a + b + c + d$ .

$\chi^2$  kriitiline väärtus on 5,02, mis antud juhul määrab käändevormide esinemise sarnasuse/erinevuse kahes valimis tõenäosusega 5%. Kui käändevorm pole valimis esindatud, siis  $\chi^2$  ei arvutata. Kui saadud  $\chi^2$  on kriitilisest väärtusest väiksem, siis on käändekasutus kahes valimis sarnane (vt tabel 4). Mida suurem on  $\chi^2$  kriitilisest väärtusest, seda suurem on kahe valimi erinevus käändevormi kasutamisel.

<sup>10</sup> Eesti keele kitsenduste grammatika süntaksianalüsaatori märgendite kohta vt <http://math.ut.ee/~kaili/papers/syntax.html> (10.03.2009).

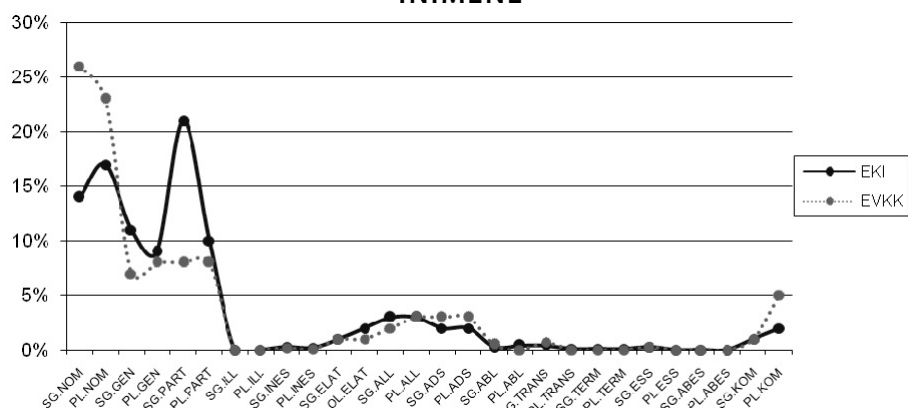
**Tabel 4.** Sõna *inimene* käändekasutuse sarnasus/erinevus EKI tekstikorpuses ja EVKK-s

Käanded	EKI tekstikorpus		EVKK		$\chi^2$
	a	b	c	d	
SG.NOM	960	5812	1257	3583	254,20
PL.NOM	1164	5608	1117	3723	62,03
SG.GEN	765	6007	359	4481	48,58
PL.GEN	598	6174	363	4477	6,58
SG.PART	1421	5351	397	4443	349,15
PL.PART	704	6068	374	4466	23,86
SG.ILL	0	6772	0	4840	–
PL.ILL	0	6772	0	4840	–
SG.INES	17	6755	11	4829	0,07
PL.INES	13	6759	7	4833	0,37
SG.ELAT	55	6717	43	4797	0,20
PL.ELAT	106	6666	40	4800	12,41
SG.ALL	176	6596	90	4750	6,89
PL.ALL	230	6542	150	4690	0,79
SG.ADS	125	6647	137	4703	12,41
PL.ADS	137	6635	155	4685	16,02
SG.ABL	19	6753	22	4818	2,43
PL.ABL	29	6743	0	4840	20,78
SG.TRANS	27	6745	34	4806	4,98
PL.TRANS	8	6764	0	4840	5,72
SG.TERM	8	6764	0	4840	5,72
PL.TERM	5	6767	0	4840	3,57
SG.ESS	23	6749	16	4824	0,01
PL.ESS	0	6772	0	4840	–
SG.ABES	0	6772	0	4840	–
PL.ABES	0	6772	0	4840	–
SG.KOM	75	6697	43	4797	1,35
PL.KOM	107	6665	225	4615	95,70
<b>Kokku</b>	<b>6772</b>		<b>4840</b>		

Tabelis 4 toodud andmete alusel on EKI tekstikorpuse ja EVKK suurem erinevus sõna *inimene* käändevormide kasutuses seotud järgmiste käännetega: ainsuse partitiiv 349,15 ja nominatiiv 254,20 – mitmuse komitatiiv 95,70 ja nominatiiv 62,03 – ainsuse genitiiv 48,58 – mitmuse partitiiv 23,86 ja ablatiiv 20,78 jne. Suurem sarnasus ilmneb ainsuse essiivi 0,01, inessiivi 0,07 ja elatiivi 0,20 – mitmuse inessiivi 0,37 ja allatiivi 0,79 – ainsuse komitatiivi 1,35 ja ablatiivi 2,43 – mitmuse terminatiivi 3,57 ja ainsuse translatiivi 4,98 kasutuses. Sagedusandmete statistiliseks töötlemiseks rakendati Exeli tabelarvutust ja graafilisi võimalusi, milles kahe korpuse vaheline käändevormide dünaamika kujukalt esile tuleb (vt joonis 1).

Statistikast nähtub, mis on eesti keele käändevormide sagedusele iseloomulik ning missuguseid leksikaalsemantilisi ja paradigmaatilisi arenguid erinevate keelevariantide vahel võib käändegrammatikas täheldada. Analüüsi tulemused nagu ka erinevused EKI tekstikorpuse ja EVKK käändevormide sageduses on toodud protsentides tabelis 5.

## INIMENE



Joonis 1. Sõna *inimene* käändevormide sagedus EKI tekstikorpuses ja EVKK-s

Tabel 5. Erinevused sõna *inimene* grammatiliste käänete sageduses

Kääne	EKI tekstikorpuses	Sagedus	EVKK	Sagedus	Erinevus
SG.NOM	14%	960	26%	1257	12%
PL.NOM	17%	1164	23%	1117	6%
SG.GEN	11%	765	7%	359	4%
PL.GEN	9%	598	8%	363	1%
SG.PART	21%	1421	8%	397	13%
PL.PART	10%	704	8%	374	2%
SG.ILL	0%	0	0%	0	0%
PL.ILL	0%	0	0%	0	0%
SG.INES	0,3%	17	0,2%	11	0%
PL.INES	0,2%	13	0,1%	7	0%
SG.ELAT	1%	55	1%	43	0%
PL.ELAT	2%	106	1%	40	1%
SG.ALL	3%	176	2%	90	1%
PL.ALL	3%	230	3%	150	0%
SG.ADS	2%	125	3%	137	1%
PL.ADS	2%	137	3%	155	1%
SG.ABL	0,3%	19	0,5%	22	0%
PL.ABL	0,4%	29	0,0%	0	0%
SG.TRANS	0,4%	27	0,7%	34	0%
PL.TRANS	0,1%	8	0%	0	0%
SG.TERM	0,1%	8	0%	0	0%
PL.TERM	0,1%	5	0%	0	0%
SG.ESS	0,3%	23	0,3%	16	0%
PL.ESS	0%	0	0%	0	0%
SG.ABES	0%	0	0%	0	0%
PL.ABES	0%	0	0%	0	0%
SG.KOM	1%	75	1%	43	0%
PL.KOM	2%	107	5%	225	3%

### 3. Käänevormide statistika eesti kirjakeeles ja õppijakeeles

#### 3.1. Sõna *inimene* käänevormide statistika

Sõna *inimene* käänevormide statistikas ilmneb ühelt poolt eesti keele grammatiliste ja semantiliste käänete vastandatus ja teisalt semantiliste käänete kokkulangev või samalaadne kasutus nii kirjakeeles kui ka õppijakeeles.

Erinevus grammatiliste ja semantiliste käänete vastandatuses seisneb vaid selles, et õppijakeele sagedasemad vormid on ainsuse (26%) ja mitmuse nominatiiv (23%), samas kui kirjakeeles on sagedasem ainsuse partitiiv (21%). Kõige suurem erinevus EKI tekstikorpuse ja EVKK valimite vahel ongi grammatiliste käänete sageduses: ainsuse nominatiivi puhul 12% ning partitiivi puhul 13% (vt tabel 5 ning joonis 1). Võrreldes kirjakeelega on õppijakeeles ainsuse nominatiivi selgelt ülekasutatud.

Semantilistest käänetest pole nii kirjakeeles kui ka õppijakeeles kasutatud ainsuse ja mitmuse illatiivi, mitmuse essiivi, ainsuse ja mitmuse abessiivi. See võib olla seletatav teatud loogilise vastuoluga sõna semantika ja vormi grammatilise tähenduse vahel. Õppijakeeles puuduvad ka mitmuse translatiivi, ainsuse ning mitmuse ablatiivi ja terminatiivi vormid. Kirjakeeleski tulevad need käänevormid äärmiselt harva ette (vaid 0,1% valimi kohta). Nii kirjakeeles kui ka õppijakeeles oli väga harva ainsuse (vastavalt 0,3% ja 0,2%) ja mitmuse inessiivi (0,2% ja 0,1%), ainsuse ablatiivi (0,3% ja 0,5%), translatiivi (0,4% ja 0,7%) ja essiivi (0,3% ja 0,3%). Harva esines ainsuse (1% ja 1%) ja mitmuse (2% ja 1%) elatiivi; ainsuse (3% ja 2%) ja mitmuse (3% ja 3%) allatiivi; ainsuse (2% ja 3%) ja mitmuse (2% ja 3%) adessiivi; ainsuse komitatiivi (1% ja 1%). Toodud statistilised andmed näitavad, et semantiliste käänete sagedus langeb kirjakeeles ja õppijakeeles kokku või on samalaadne. Mõneti eraldiseisvaks jääb mitmuse komitatiiv (vastavalt 2% ja 5%), kuid ilmselt on siin tegu eesti keelele omase nähtusega, sest ka nimisõnade *elu* ja *sõna* käänevormide statistika näitab (vt allpool), et tavapäraselt on komitatiivi sagedus ülejäänud semantiliste käänete sagedusest kõrgem.

#### 3.2. Sõna *elu* käänevormide statistika

Sõna *elu* käänevormide suuremad erinevused EKI tekstikorpuses ja EVKK-s on seotud semantiliste käänete ainsuse vormidega: komitatiiv 111,17 – inessiiv 93,88 – terminatiiv 61,10 – allatiiv 34,52 – elatiiv 32,81 – abessiiv 24,39 – illatiiv 13,15 jt. Sarnasus on ainsuse genitiivi 1,10 ja mitmuse partitiivi 4,98 puhul (vt tabel 6). Seda ei saa tõlgendada nii, nagu võiks sõna *elu* käänevormide statistika põhjal rääkida grammatiliste ja semantiliste käänete selgest vastandatusest; rääkida saab kirjakeele ja õppijakeele käände paradigmade sünkroonsusest.

**Tabel 6.** Sõna *elu* käändekasutuse sarnasus/erinevus EKI tekstikorpuses ja EVKK-s

Käänded	EKI tekstikorpus		EVKK		$\chi^2$
	a	b	c	d	
SG.NOM	184	1028	225	979	5,28
PL.NOM	8	1204	0	1204	7,97
SG.GEN	159	1053	141	1063	1,10
PL.GEN	0	1212	0	1204	
SG.PART	81	1131	122	1082	9,34
PL.PART	5	1207	0	1204	4,98
SG.ILL	0	1212	13	1191	13,15
PL.ILL	0	1212	0	1204	–
SG.INES	324	888	550	654	93,88
PL.INES	0	1212	0	1204	–
SG.ELAT	206	1006	110	1094	32,81
PL.ELAT	0	1212	0	1204	–
SG.ALL	58	1154	10	1194	34,52
PL.ALL	0	1212	0	1204	–
SG.ADS	12	1200	0	1204	11,98
PL.ADS	0	1212	0	1204	–
SG.ABL	8	1204	0	1204	7,97
PL.ABL	0	1212	0	1204	–
SG.TRANS	0	1212	9	1195	9,09
PL.TRANS	0	1212	0	1204	–
SG.TERM	60	1152	0	1204	61,10
PL.TERM	0	1212	0	1204	–
SG.ESS	0	1212	0	1204	–
PL.ESS	0	1212	0	1204	–
SG.ABES	0	1212	24	1180	24,39
PL.ABES	0	1212	0	1204	–
SG.KOM	107	1105	0	1204	111,17
PL.KOM	0	1212	0	1204	–
<b>Kokku</b>	<b>1212</b>		<b>1204</b>		

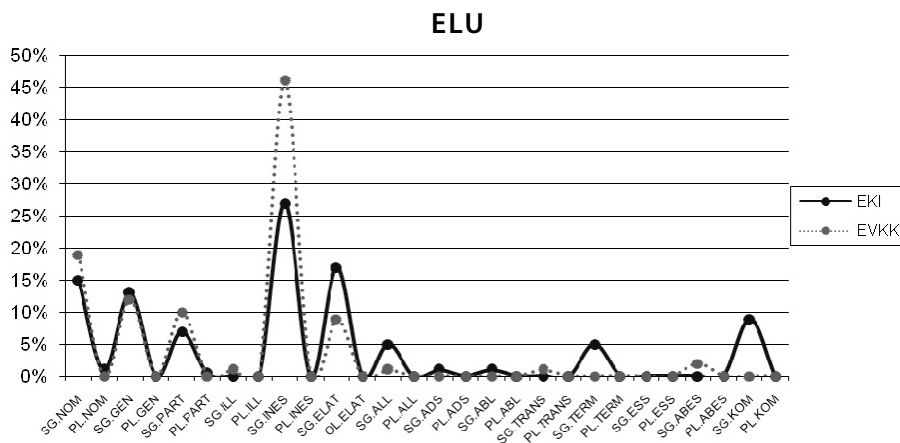
Esiteks on sõna *elu* käändevormide moodustamises teatud loogilis-grammatilised piirangud, mis tulenevad sõna semantikast. Ilmselt seetõttu puuduvad mõlemas eesti keele kasutusvariandis mitmuse genitiivi, illatiivi, inessiivi, elatiivi, allatiivi, adessiivi, ablatiivi, translatiivi, terminatiivi, abessiivi, komitatiivi ning ainsuse ja mitmuse essiivi vormid (vt tabel 7). EVKK-s pole kasutatud ka sõna *elu* mitmuse nominatiivi, ainsuse adessiivi, ablatiivi, terminatiivi ja komitatiivi, samas kui EKI tekstikorpuse valimis oli nimetatud vorme vastavalt 1%, 1%, 1%, 5% ja 9%. Seega on õppijakeeles alakasutatud ainsuse komitatiivi ja terminatiivi.

**Tabel 7.** Erinevused sõna *elu* käändevormide sageduses

Käänded	EKI tekstikorpus	Sagedus	EVKK	Sagedus	Erinevus
SG.NOM	15%	184	19%	225	4%
PL.NOM	1%	8	0%	0	1%
SG.GEN	13%	159	12%	141	1%
PL.GEN	0%	0	0%	0	0%
SG.PART	7%	81	10%	122	3%

PL.PART	0,4%	5	0%	0	0%
SG.JLL	0%	0	1%	13	1%
PL.ILL	0%	0	0%	0	0%
SG.INES	27%	324	46%	550	19%
PL.INES	0%	0	0%	0	0%
SG.ELAT	17%	206	9%	110	8%
PL.ELAT	0%	0	0%	0	0%
SG.ALL	5%	58	1%	10	4%
PL.ALL	0%	0	0%	0	0%
SG.ADS	1%	12	0%	0	1%
PL.ADS	0%	0	0%	0	0%
SG.ABL	1%	8	0%	0	1%
PL.ABL	0%	0	0%	0	0%
SG.TRANS	0%	0	1%	9	1%
PL.TRANS	0%	0	0%	0	0%
SG.TERM	5%	60	0%	0	5%
PL.TERM	0%	0	0%	0	0%
SG.ESS	0%	0	0%	0	0%
PL.ESS	0%	0	0%	0	0%
SG.ABES	0%	0	2%	24	2%
PL.ABES	0%	0	0%	0	0%
SG.KOM	9%	107	0%	0	9%
PL.KOM	0%	0	0%	0	0%

Sünkroonsus avaldub ka käändevormide sageduste samalaadses kõikumises (vt joonis 2). Lahknevused on seotud kindlate vormidega, mis on esindatud mõlema korpuse valimites, kuid mida õppijakeeles on kas üle- või alakasutatud. Nii on ainsuse inessiivi selgelt ülekasutatud (vahe kirjakeelega 19%), järgnevad ainsuse nominatiiv (vahe 4%), partitiiv (vahe 3%) ja abessiiv (vahe 2%), samas kui ainsuse elatiivi (vahe kirjakeelega 8%) ja allatiivi (vahe 4%) on alakasutatud.



Joonis 2. Sõna *elu* käändevormide sagedus EKI tekstikorpuses ja EVKK-s

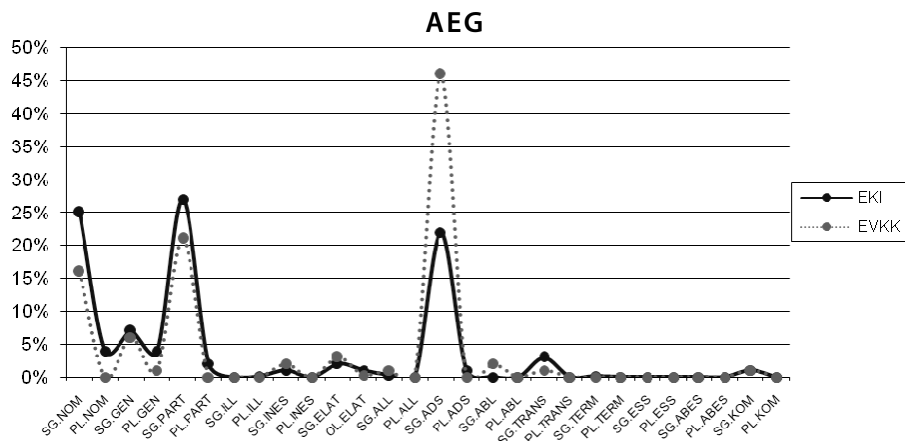
### 3.3. Sõna *aeg* käändevormide statistika

Sõna *aeg* käändevormide kasutuse erinevused EKI tekstikorpuses ja EVKK-s on seotud valdavalt ainsuse adessiiviga 528,81 – vähemal määral mitmuse nominatiivi 138,78 – ainsuse nominatiivi 85,96 – ablatiivi 81,97 – mitmuse genitiivi 81,56 – partitiivi 58,80 – ainsuse partitiivi 45,81 – translatiiviga 26,46 jt, sarnasused aga ainsuse genitiivi 0,16 – allatiivi 1,54 – komitatiivi 2,36 – elatiivi 2,57 – mitmuse elatiivi 2,84 – illatiivi 4,43 ja ainsuse terminatiiviga 4,43 (vt tabel 8).

**Tabel 8.** Sõna *aeg* käändekasutuse sarnasus/erinevus EKI tekstikorpuses ja EVKK-s

Kääne	EKI tekstikorpus		EVKK		$\chi^2$
	a	b	c	d	
SG.NOM	1168	3518	567	2888	85,96
PL.NOM	184	4502	0	3455	138,78
SG.GEN	310	4376	221	3234	0,16
PL.GEN	184	4502	25	3430	81,56
SG.PART	1268	3418	710	2745	45,81
PL.PART	114	4572	11	3444	58,80
SG.ILL	0	4686	0	3455	–
PL.ILL	6	4680	0	3455	4,43
SG.INES	41	4645	64	3391	14,92
PL.INES	0	4686	0	3455	–
SG.ELAT	109	4577	100	3355	2,57
PL.ELAT	27	4659	11	3444	2,84
SG.ALL	16	4670	18	3437	1,54
PL.ALL	0	4686	0	3455	–
SG.ADS	1020	3666	1583	1872	528,81
PL.ADS	43	4643	16	3439	5,71
SG.ABL	0	4686	60	3395	81,97
PL.ABL	0	4686	0	3455	–
SG.TRANS	144	4542	46	3409	26,46
PL.TRANS	0	4686	0	3455	–
SG.TERM	6	4680	0	3455	4,43
PL.TERM	0	4686	0	3455	–
SG.ESS	0	4686	0	3455	–
PL.ESS	0	4686	0	3455	–
SG.ABES	0	4686	0	3455	–
PL.ABES	0	4686	0	3455	–
SG.KOM	46	4640	23	3432	2,36
PL.KOM	0	4686	0	3455	–
<b>Kokku</b>	<b>4686</b>		<b>3455</b>		

Sõna *aeg* käändevormide sagedusandmed kirjakeeles ja õppijakeeles näitavad kahe keelevariandi vahelist sünkroonsust (vt joonis 3). See väljendub ühelt poolt üksikute käändevormide kõrges sageduses ja teisalt paljude käändevormide samalaadses mittekasutamises, mis tuleneb sõna semantikast (vt tabel 9).



**Joonis 3.** Sõna *aeg* käändevormide sagedus EKI tekstikorpuses ja EVKK-s

**Tabel 9.** Erinevused sõna *aeg* käändevormide sageduses

Käänded	EKI	Sagedus	EVKK	Sagedus	Erinevus
SG.NOM	25%	1168	16%	567	9%
PL.NOM	4%	184	0%	0	4%
SG.GEN	7%	310	6%	221	1%
PL.GEN	4%	184	1%	25	3%
SG.PART	27%	1268	21%	710	6%
PL.PART	2%	114	0%	11	2%
SG.ILL	0%	0	0%	0	0%
PL.ILL	0,1%	6	0%	0	0%
SG.INES	1%	41	2%	64	1%
PL.INES	0%	0	0%	0	0%
SG.ELAT	2%	109	3%	100	1%
PL.ELAT	1%	27	0,3%	11	1%
SG.ALL	0,3%	16	1%	18	1%
PL.ALL	0%	0	0%	0	0%
SG.ADS	22%	1020	46%	1583	24%
PL.ADS	1%	43	0%	16	1%
SG.ABL	0%	0	2%	60	2%
PL.ABL	0%	0	0%	0	0%
SG.TRANS	3%	144	1%	46	2%
PL.TRANS	0%	0	0%	0	0%
SG.TERM	0,1%	6	0%	0	0%
PL.TERM	0%	0	0%	0	0%
SG.ESS	0%	0	0%	0	0%
PL.ESS	0%	0	0%	0	0%
SG.ABES	0%	0	0%	0	0%
PL.ABES	0%	0	0%	0	0%
SG.KOM	1%	46	1%	23	0%
PL.KOM	0%	0	0%	0	0%



Kasutatud ei ole järgmisi käändevorme: ainsuse illatiiv; mitmuse inessiiv, allatiiv, ablatiiv, translatiiv, terminatiiv; ainsuse ja mitmuse essiiv ning abessiiv; mitmuse komitatiiv. Õppijakeele korpuses puudusid ka mitmuse illatiiv ja ainsuse terminatiiv; kirjakeeleski oli nende vormide sagedus vaid 0,01%. Samuti polnud õppijakeeles esindatud mitmuse nominatiiv, mis kirjakeeles moodustas sõna *aeg* käändevormidest 4%. Kõige sagedamini esines nii EKI tekstikorpuse kui ka EVKK valimites kolm käänat: ainsuse nominatiiv (vastavalt 25% ja 16%), partitiiv (27% ja 21%) ja adessiiv (22% ja 46%). Tunduvalt harvemini oli kasutatud ainsuse genitiivi (vastavalt 7% ja 6%). Võrreldes kirjakeelega oli õppijakeeles selgelt ülekasutatud käändevorm ainsuse adessiiv (erinevus 24%), alakasutatud aga ainsuse nominatiiv (erinevus on 9%) ja partitiiv (erinevus 6%), vähemal määral mitmuse nominatiiv (erinevus 4%) ja genitiiv (erinevus 3%).

### 3.4. Sõna sõna käändevormide statistika

Sõna sõna käändevormide kasutuses on EKI tekstikorpuse ja EVKK vahelised lahknevused seotud nii grammatiliste kui ka semantiliste käänetega (vt tabel 10): mitmuse elatiiv 37,45 – ainsuse genitiiv 32,20 – mitmuse nominatiiv 27,31 – ainsuse nominatiiv 25,43 – ainsuse allatiiv 21,80 – mitmuse komitatiiv 20,81 – mitmuse partitiiv 20,03 – mitmuse adessiiv 17,91 – ainsuse komitatiiv 17,55 – ainsuse partitiiv 16,90 – ainsuse ablatiiv 12,23 – mitmuse inessiiv 12,05. Sarnaselt oli kasutatud semantilisi käänedeid: mitmuse translatiivi 2,54 ja genitiivi 2,67 – ainsuse inessiivi 2,90, adessiivi 3,17 ja elatiivi 3,18 – mitmuse ablatiivi 4,00 ja allatiivi 4,38.

**Tabel 10.** Sõna sõna käändekasutuse sarnasus/erinevus EKI tekstikorpuses ja EVKK-s

Käänded	EKI tekstikorpus		EVKK		$\chi^2$
	a	b	c	d	
SG.NOM	397	2212	82	863	25,43
PL.NOM	353	2256	196	749	27,61
SG.GEN	233	2376	31	914	32,20
PL.GEN	293	2316	88	857	2,67
SG.PART	248	2361	49	896	16,90
PL.PART	72	2537	56	889	20,03
SG.ILL	0	2609	0	945	–
PL.ILL	19	2590	0	945	6,92
SG.INES	8	2601	0	945	2,90
PL.INES	78	2531	9	936	12,05
SG.ELAT	49	2560	27	918	3,18
PL.ELAT	51	2558	56	889	37,45
SG.ALL	24	2585	29	916	21,80
PL.ALL	40	2569	6	939	4,38
SG.ADS	38	2571	22	923	3,17
PL.ADS	342	2267	75	870	17,91
SG.ABL	13	2596	16	929	12,23
PL.ABL	11	2598	0	945	4,00
SG.TRANS	0	2609	0	945	–
PL.TRANS	7	2602	0	945	2,54

Käänded	EKI tekstikorpus		EVKK		$\chi^2$
	a	b	c	d	
SG.TERM	0	2609	0	945	–
PL.TERM	0	2609	0	945	–
SG.ESS	0	2609	0	945	–
PL.ESS	0	2609	0	945	–
SG.ABES	0	2609	0	945	–
PL.ABES	0	2609	0	945	–
SG.KOM	116	2493	76	869	17,55
PL.KOM	217	2392	127	818	20,81
<b>Kokku</b>	<b>2609</b>		<b>945</b>		

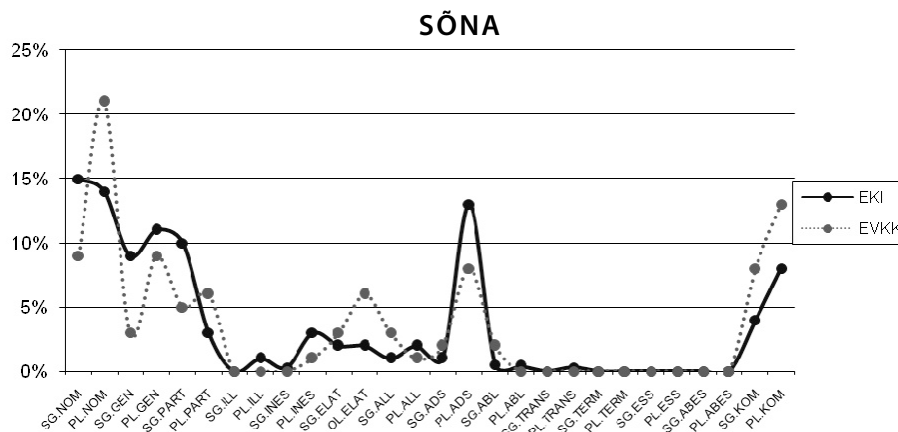
Nii EKI tekstikorpuses kui ka EVKK-s polnud üldse tarvitatud kolme käändevormi: ainsuse illatiivi, translatiivi ja terminatiivi (vt tabel 11). Lisaks puudusid õppijakeeles ka mitmuse illatiivi, ainsuse inessiivi, mitmuse ablatiivi ja translatiivi vormid, mida kirjakeeles küll leidis, kuid niivõrd harva, et väike erinevus 1% ilmnis vaid mitmuse illatiivi sageduses. Kirjakeele sagedasem käänne sõnaga *sõna* oli nominatiiv: ainsuse 15% ja mitmuse nominatiiv 14%. Järgnesid mitmuse adessiiv 13% ja genitiiv 11%, ainsuse partitiiv 10% ja genitiiv 9%, mitmuse 8% ja ainsuse komitatiiv 4%, mitmuse partitiiv 3% ja inessiiv 3%, ainsuse elatiiv 2%, mitmuse allatiiv 2%.

**Tabel 11.** Erinevused sõna sõna käändevormide sageduses

Käänded	EKI tekstikorpus	Sagedus	EVKK	Sagedus	Erinevus
SG.NOM	15%	397	9%	82	6%
PL.NOM	14%	353	21%	196	7%
SG.GEN	9%	233	3%	31	6%
PL.GEN	11%	293	9%	88	2%
SG.PART	10%	248	5%	49	5%
PL.PART	3%	72	6%	56	3%
SG.ILL	0%	0	0%	0	0%
PL.ILL	1%	19	0%	0	1%
SG.INES	0,3%	8	0%	0	0%
PL.INES	3%	78	1%	9	2%
SG.ELAT	2%	49	3%	27	1%
PL.ELAT	2%	51	6%	56	4%
SG.ALL	1%	24	3%	29	2%
PL.ALL	2%	40	1%	6	1%
SG.ADS	1%	38	2%	22	1%
PL.ADS	13%	342	8%	75	5%
SG.ABL	0,5%	13	2%	16	2%
PL.ABL	0,4%	11	0%	0	0%
SG.TRANS	0%	0	0%	0	0%
PL.TRANS	0,3%	7	0%	0	0%
SG.TERM	0%	0	0%	0	0%
SG.KOM	4%	116	8%	76	4%
PL.KOM	8%	217	13%	127	5%

Võrreldes kirjakeelega oli õppijakeeles kõige rohkem ülekasutatud mitmuse nominatiivi – 21%; erinevus EKI tekstikorpuse vormisagedusest on 7%. Ülekasutatud oli ka mitmuse (erinevus 5%) ja ainsuse komitatiivi (erinevus 4%). Alakasutatud käändevormid olid: ainsuse nominatiiv ja genitiiv (mõlemal juhul erinevus 6%), ainsuse partitiiv ja mitmuse adessiiv (erinevus 5%).

Kui võrrelda *sõna* käändevormide sageduserinevusi kirjakeeles ja õppijakeeles, siis hakkab silma, et märgatavamad vormisageduse kõikumised on seotud grammatiliste käänedega, eriti mitmuse nominatiivi, ainsuse nominatiivi ja genitiiviga, samuti ainsuse partitiivi, mitmuse elatiivi, adessiivi ja ainsuse ning mitmuse komitatiiviga (vt joonis 4).



Joonis 4. Sõna *sõna* käändevormide sagedus EKI tekstikorpuses ja EVKK-s

### 3.5. Kokkuvõtvalt käändevormide statistikast

Võrrelnud nelja sagedasema nimisõna (*inimene, elu, aeg, sõna*) käändevormide sagedust, võib väita järgmist.

- Üheselt ei saa öelda, et eesti keele nimisõnade käändeparadigma iseloomulikuks jooneks on grammatiliste ja semantiliste käänede vastandatus, sest käändevormide esinemine või mitteesinemine sõltub sõna semantikast, seoste loogikast semantika ja grammatika vahel (vt *elu, aeg, sõna* käändevormide sagedus paradigmas).
- Eesti kirjakeele ja õppijakeele käändevormide sageduse võrdlus näitab nende kahe keelevariandi käändeparadigmade samalaadsust ja sünkroonsust, mis ei tulene sellest, et õppijad on nimisõnade käändeparadigma hästi omandanud; tendentsi põhjus peitub keele sisemises loogikas, grammatiliste vormide moodustamise olenevalt sõna semantikast.
- Võrdlusest kirjakeelega ilmneb, et õppijakeeles on suundumus ülekasutada nominatiivi (*inimene* – ainsuses ja mitmuse vormis, *sõna* – mitmuses, *elu* – ainsuses), kuid samas on ka kirjakeeles sel käändel kõrged sagedusnäitajad (*sõna* – ainsuse ja mitmuse nominatiiv on käändevormide sagedusreas esikohal, *inimene* – mitmuse nominatiiv on ainsuse partitiivi järel sageduselt teine vorm ja ainsuse nominatiiv – kolmas, *aeg* – ainsuse nominatiiv on

ainsuse partitiivi järel sageduselt teisel kohal, *elu* – ainsuse nominatiiv on ainsuse inessiivi ja elatiivi järel käänete sagedusrea kolmas vorm); järelikult ei saa nominatiivi suurt sagedust pidada üheselt õppijakeele spetsiifiliseks jooneks, sest ka kirjakeeles on nominatiiv sage kääne ning ainsuse partitiiv on vaid kahe sõna puhul (*inimene, aeg*) kerges ülekaalus; seetõttu ei saa analüüsitud keeleainese alusel väita, et eesti keele käändesüsteemis kehtib partitiivi ja nominatiivi vastandus, mille markeerimata liige on partitiiv.

- Õppijakeeles on üle kasutatud mitmuse (*inimene*) ning ainsuse ja mitmuse komitatiivi (*sõna*); jällegi pole tegu õppijakeele spetsiifikaga, sest ka kirjakeeles on komitatiivi sagedus ülejäänud semantiliste käänete sagedusest kõrgem (*inimene* – mitmuse vormis, *elu* – ainsuses ja *sõna* – ainsuses ja mitmuses), järelikult on tegu samalaadse nähtusega, mis erinevates keelevariantides ilmneb erineva sagedusega.
- Õppijakeeles on üle kasutatud ainsuse inessiivi (*elu*) ja adessiivi (*aeg*), kuid ka kirjakeeles on sõna *elu* sagedasem vorm inessiiv ning sõna *aeg* adessiivi vormi kasutatakse ainsuse partitiivi ja nominatiivi järel kolmandana; tegu on samalaadse nähtusega.
- Õppijakeeles on alakasutatud ainsuse genitiivi ja partitiivi (*inimene, aeg*), ainsuse nominatiivi ja partitiivi (*sõna*), ainsuse nominatiivi (*aeg*), ainsuse elatiivi ja allatiivi, mitmuse terminatiivi ja komitatiivi (*elu*), mitmuse adessiivi (*sõna*), kuid ka selles osas on märgata sünkroonsust kirjakeelega, kus neid käändeid on teistega võrreldes vähem kasutatud.

Küsimused tekivad siis, kui õppijakeele käändevormide üle- ja alakasutus hakkab regulaarselt koonduma samade vormide ümber ning kaob vormisageduse sünkroonsus loomuliku keelekasutusega. Sel juhul võib olla tegu õppijakeelele omase ebatüüpilise vormikasutusega, mille põhjusi tasuks uurida. Lisaks õppija mitteaktsepteeritavale vormikasutusele võib osutuda ka võimalikuks, et õppijakeel kajastab loomulikus keeles varjatult kulgevaid keelemuutusi. Kadri Sõrmuse arvamuse järgi on keeleõppija vastuvõtlik kõigele, mida ta erinevates suhtlussituatsioonides kogeb, eriti sellele, mida pakuvad erinevad meediakanalid ja tavasuhetus (Sõrmus 2008: 35). Et nendele protsessidele jälile saada, ei piisa nelja sagedasema nimi-sõna käändevormide statistikast – vaja oleks läbi viia tunduvalt avaramapõhjaline uurimus, mis tugineks kahe keeleainese võtmesõnaanalüüsile (KeyWords analysis), vt eespool punkt 2.3. Teisalt ei saa statistilisi erinevusi ehk käändevormide üle- ja alakasutust õppijakeeles siduda automaatselt ebatüüpilise vormikasutusega. Selleks on vaja võrrelda kirjakeele ja õppijakeele tekstiloomes eelistatud konstruktsioone, milles ilmnevad ainult teatud kindlad sõnad ja käändevormid.

## 4. Tekstiloomes eelistatud konstruktsioonid ja käändevormid

### 4.1. Sõna *inimene*

Kuna selle sõna eri käändevormidel on rohkesti kasutusmustreid, siis kirjeldatakse konstruktsioone ja nende leksikaalseid ning grammatilisi variante käänete kaupa.

#### 4.1.1. Ainsuse nominatiiv – *inimene*

EKI tekstikorpuses on eelistatud kasutada substantiivset konstruktsiooni *inimene ja seadus ~ inimene ja loodus*, loogilist implikatsiooni *kui inimene on ~ et inimene on* ja eitust *inimene ei ole (ei saa)*. EVKK-s on eelistatud nii loogilist implikatsiooni *kui inimene on* (variandid: *kui inimene tahab, kogeb, elab, suhtleb*) ~ *et inimene peab ~ ja kui (et kui, sest kui) inimene* kui ka modaalseid lausekonstruktsioone *inimene peab olema (maksma, teadma) ~ inimene võib töötada; haritud (iga) inimene peab (saab, võib, tahab)* ja eitust *inimene ei saa (ei ole)*.

Normkirjakeele seisukohalt on kõik need konstruktsioonid tavapärased – analüüsitavates keelevariantides on lihtsalt erinevad kasutuseelistused. Sama kehtib konstruktsioonide leksikaalgrammatilise varieeruvuse kohta. Õppijakeele konstruktsioonidel on tänu modaal-, olemis- ja kogemisverbide vaheldumisele päris palju leksikaalseid variante (verbid *tahtma, teadma, võima, saama, olema, elama, kogema, suhtlema, maksma* indikatiivi ainsuse 3. pöördes); konstruktsioonid varieeruvad ka grammatiliselt (sidendid *kui, et, ja kui, et kui, sest kui* ning jaatav / eitav kõneliik). Kirjakeelse substantiivkonstruktsiooni leksikaalne varieeruvus piirdub kahe võimalusega: *inimene ja seadus ~ inimene ja loodus*. Samas lubab keelenorm selle konstruktsiooniga vabalt kasutada muidki abstraktseid sõnu (nt *inimene ja kosmos, inimene ja ühiskond, inimene ja meri, inimene ja elu, inimene ja aeg* jne). Niisugused substantiivkonstruktsioonid on tavapärased teema sõnastamisel ja sobivad pealkirjaks. EKI tekstikorpuse analüüs näitas, et võimalikest teemasõnadest esines kolm: *inimene, seadus* ja *loodus*. Need märksõnad toovad esile praegusaja Eesti riigi ning ühiskonna ühe olulisema sotsiaalpoliitilise diskursuse, mida kajastavad EKI tekstikorpuse ajakirjandustekstid. Teisena esile tulnud verbikesksete konstruktsioonidega oli väljendatud tingimust või millegi lubatavust. EVKK-s on sõna *inimene* käändevormide kasutamisel eelistatud modaalset, tingimuslik-eksistentsiaalset ja põhjendavat diskursust. Erinevus valikutes tuleneb kirjakeele ja õppijakeele diskursuste spetsiifikast: esimeses on rõhutatud inimesega seotud aktuaalset teemat ja viidatud sellele, mida inimene ei saa ega tohi teha; teises tuginetakse abstraktsele arutelule inimese olemise, vajaduste ja võimaluste üle, püüdes neid põhjendada.

#### 4.1.2. Mitmuse nominatiiv – *inimesed*

Mitmuse nominatiivi vormiga on kirjakeeles eelistatud kolme konstruktsiooni ja nende variante: 1) *need inimesed kes (kellele)*, 2) *et inimesed on (ei ole) ~ et (ka) need inimesed* ja 3) *need inimesed on*. Õppijakeeles on kasutatud analoogseid konstruktsioone – erineb vaid konstruktsiooni kordumise sagedus ja leksikaalgrammatiline varieeruvus: 1) *kõik (need) inimesed on (hoidsid, tahavad)* ja 2) *inimesed kes on (ei ole)*. EVKK-s sisalduvate tekstide spetsiifikast johtub konstruktsiooni *noorkeraamika (kammkeraamika) inimesed sarnanesid* esile kerkimine, kuna korpuses on palju Eesti kultuuriloo teemalisi esseid ja kontrolltöid.

#### 4.1.3. Ainsuse genitiiv – *inimese*

EKI tekstikorpuses on ainsuse genitiivi kasutatud valdavalt kvantorfraasis: *üle 200 (5000, 300, tuhande, saja) inimese*. Konstruktsiooniga antakse hinnang meie elu negatiivsetele nähtustele: katastroofid, kollapsid, liiklusõnnetused jms, kus on surma saanud/hukkunud sadu ja tuhandeid inimesi. Õppijakeele tekstides pole see teema olnud aktuaalne. Esile kerkivad hoopiski arutlused selle üle, mis mõjutab / ei mõjuta inimese elu või mis on elus tähtis: *mõjutab (ei mõjuta) inimese elu ~ inimese elu mõjutab*. Selles lausekonstruktsioonis varieeruvad sõnajärg ja kõneliik, verbi on kasutatud indikatiivi oleviku 3. pöördes. Sama sagedusega oli kasutatud ka inimese elu kirjeldavat noomenikonstruktsiooni, mille mittegrammatilised sõnavormid võivad tekitada arusaamisraskusi: *inimese elu kõige ~ ?inimese elu haridus ~ \*inimese elu tema (?teema) ~ \*inimese elu ma (?maal)*.

#### 4.1.4. Mitmuse genitiiv – *inimeste*

Konstruktsioone, milles peaks olema kasutatud mitmuse genitiivi, eesti kirjakeeles ei olnud. Õppijakeeles esines verbikonstruktsioon *kahjustab inimeste tervist*. Selle leksikaalne variant *matkida inimeste magamisasendit* on spetsiifiline, st iseloomulik kultuuriloo kursuse kontrolltöödele (vt PL.NOM).

#### 4.1.5. Ainsuse partitiiv – *inimest*

EKI tekstikorpuse materjalides on seda käändevormi kasutatud kvantorfraasis, kus varieeruvad määr- ja arvsõnad: *ligi (veel, umbes) 200 (kolm, 800) inimest; hukkus kolm (kaheksa, kuus, viis, üksteist) inimest*. Ka see konstruktsioon iseloomustab sotsiaalset diskursust: rõhutatult on esile toodud katastroofides ja õnnetustes hukkunute arv. EVKK-s on samuti eelistatud kvantorfraasi: *ainult (umbes) kuus (sada) inimest; ning kaheksa (üksteist) inimest, kaheksa (kuus) inimest 32-st*, kuid õppijakeeles ei räägita hukkunute arvust – tavapäraselt on kvantorfraasi kasutatud verbi *kirjutama* imperfekti mitmuse 3. pöördega (*üksteist inimest kirjutasid*).

#### 4.1.6. Mitmuse partitiiv – *inimesi*

Kirjakeeles on seda vormi kasutatud konstruktsioonides 1) *inimesi kellel on ~ inimesi kes ei*; 2) *neid inimesi kes* ja 3) *on inimesi kes*, samas kui õppijakeele eelistus on rõhusõna sisaldav kvantorfraas *väga (nii) palju inimesi*. See näide toob esile täiesti erinevad valikud: kirjakeeles on eelistatud öelda, missuguseid inimesi üldse on, õppijakeeles rõhutatakse inimeste hulka, mis midagi teeb. Kvantoriga *väga (nii) palju* toonitatakse kõnealuse nähtuse positiivsust.

#### 4.1.7. Ainsuse adessiiv – *inimesel*

Sõna *inimene* ainsuse adessiivi vormi kasutamine on õppijakeelekeskne (nagu ka ainsuse nominatiivi eelistamine modaalsetes konstruktsioonides ja loogilistes implikatsioonides). Konstruktsiooni *kui inimesel on (ei ole)* on kasutatud jaatavas ja eitavas kõneliigis; sage on olnud ka modaalne konstruktsioon *igal inimesel on (peab olema)*. Oma tekstides on õppijad kirjutanud kohustustest, mis inimesel on / peavad olema, samas kui EKI tekstikorpuse materjalides pole see teema esile tõusnud.

#### 4.2. Sõna *elu*

Sõna *elu* käändevormide kasutus on selgepiiriline: kirjakeeles on eelistatud konstruktsioone ainsuse nominatiiviga ja õppijakeeles ainsuse inessiivi, nominatiivi, genitiivi ja partitiiviga. EKI tekstikorpuses on kõige enam kasutatud substantiivkonstruktsiooni *haridus ja elu ~ kultuur ja elu* ja lausekonstruktsiooni *tema (ta) elu on*, milles sõna *elu* on ainsuse nominatiivis. Esile on toodud meie elu iseloomustav ja oluline teema – haridus, elu ja kultuur. Vastupidiselt kirjakeelele on EVKK-s kirjutatud peamiselt iseendast ja oma elust, sellest, mis minu ja teiste inimeste elus on, oli või olema peab. Sel eesmärgil on õppijakeeles eelistatud hulgaliselt konstruktsioone ja nende leksikaalgrammatilisi variante, milles sõna *elu* saab kasutada 1) ainsuse inessiivi vormis – *minu elus on (oli) ~ meie elus on; elus on palju (väga)*; 2) ainsuse nominatiivis ning koos verbidega *olema, võima, sõltuma* ja *erinema* oleviku ainsuse 3. pöördes – *inimese elu on (võib, sõltub, erineb); elu sõltub elukohast, <meie> elu erineb Ukraina <elust>; elu on väga*; 3) ainsuse genitiivi vormis lausekonstruktsioonis *ja vanemate elu ~ et noorte elu ~ et inimese elu*; 4) ainsuse partitiivis ja koos verbiga *mõjutama* ainsuse/mitmuse oleviku 3. pöördes – *inimese elu mõjutab ~ mõjutavad inimese elu*. Hakkab silma, et õppijakeeles on eelistatud konstruktsioone, milles saab kasutada nii semantiliste kui ka grammatiliste käänete vorme, kirjakeeles seevastu on eelistatud vaid ainsuse nominatiivi.

#### 4.3. Sõna *aeg*

Sõna *aeg* on nii kirjakeeles kui ka õppijakeeles kasutatud suure hulga konstruktsioonide ja nende leksikaalgrammatiliste variantidega.

##### 4.3.1. Ainsuse nominatiiv – *aeg*

EKI tekstikorpuses on kõige rohkem korduv konstruktsioon *<on (käes)> teo-inimeste aeg*, mida võib pidada kirjakeelse esituslaadi stereotüübiks. Ka õppijakeeles on eelistatud samalaadseid konstruktsioone: 1) *see aeg on ~ mõni aeg oli*; 2) *muutub (on) kogu aeg*; 3) *on aeg kus*; 4) *lendab aeg väga*. Peale selle on kirjakeeles tava-pärased konstruktsioonid, mille abil saab hinnata indiviide, predikaate või ajahetke ning sel viisil kaudselt üleskutset väljendada: 1) lausekonstruktsioon *on ilus (paras, viimane, õige) aeg*; 2) kvantorfraasid *mõni aeg hiljem ~ mõni aeg pärast <mida> ja on (oli, olnud, ajanud) kogu aeg*; 3) lausekonstruktsioonid *aeg on kallis ~ aeg on möödas (läbi, käes) ja et viimane aeg <on mida teha> ~ oli aeg mil*.

#### 4.3.2. Ainsuse genitiiv – aja

Kirjakeele kasutusmallide hulgas näited puuduvad; õppijakeeles oli üks konstruktsioon: < mõne > aja pärast tuli (oli).

#### 4.3.3. Mitmuse genitiiv – aegade

Tavapäraseid konstruktsioone, milles peaks kasutama mitmuse genitiivi, kirjakeeles ei esinenud. Õppijakeele analüüsis tuli välja viis korduvat konstruktsiooni: 1) kõigi aegade suurim (parim, esimene, teine); 2) on läbi aegade ~ läbi aegade on; 3) on kõigi aegade; 4) ESTO-d läbi aegade ja 5) Eesti kõigi aegade. Kõiki neist konstruktsioonidest on võimalik kasutada positiivse hinnangu andmiseks, eriti kui jutt on saavutustest spordis, kultuuris, majanduses jne. Eesti kirjakeele kasutajale pole see teemaring olnud samavõrd oluline.

#### 4.3.4. Ainsuse partitiiv – aega

Selle käändevormiga on nii EKI tekstikorpuses kui ka EVKK-s rikkalik kogum konstruktsioone. Kirjakeele sagedasim oli ajatähenduslik kvantorfraas *kuu aega tagasi* (hiljem, enne). Järgnesid ligilähedast või pikka ajavahemikku tähistavad kvantorfraasid *umbes* (ligi, ja) *kuu aega*; *juba pikka aega* ~ *väga pikka aega*; *oli* (on) *pikka aega* ~ *pikka aega on*; *tükk aega tagasi* (pärast); *nädal aega tagasi*; *on veel aega* ~ *vajan veidi aega*. Nende kvantorfraasidega hinnatakse indiviide, predikaate või ajahetki.

EVKK-le on iseloomulik kasutada konstruktsiooni *vabadus* (*vajadus*, \**vabanduse*) *planeerida aega*, mis tänu kõrgele sagedusele on õppijakeele stereotüüpne konstruktsioon hinnangulisuse edastamisel. Väga sageli kordus ka modaalne kvantorfraas \**pidis* (\**pidab*, *pidada*, *pidama*) *mõnda aega*. Selle konstruktsiooni mittegrammatilised pöördvormid viitavad õppija jaoks ühele suuremale vormimoodustusraskusele: \**pidis* ~ *pidi*, \**pidab* ~ *peab*. Sageli on kvantorfraasi kasutatud selleks, et rõhutada, kas aega napib või on liialt palju: *väga vähe aega* ~ *nii palju aega* ~ *liiga palju aega*. Sageli korduv on olnud ka eitust sisaldav konstruktsioon *ei ole* (*ei olnud*) *aega*.

#### 4.3.5. Ainsuse adessiiv – ajal

Selle vormiga moodustatud konstruktsioonid ja nende leksikaalgrammatilised variandid on õppijakeele spetsiifiline nähtus. Kõige sagedamini tuli ainsuse adessiiv esile 1) modaalses konstruktsioonis *külmal ajal peab*; 2) lausekonstruktsioonides *et* (*kuid, sest*) *sel ajal* ~ *sest tol ajal* ~ *ja samal ajal* ja *tol* (*sellel*) *ajal tekkis* (*kasutati*) ~ *samal ajal läksid* ~ *praegusel* (*sel, sellel*) *ajal on* ~ *vabariigi* (*vene, \*tsari*) *ajal oli* ~ *sõja ajal polnud* ~ *vabal* (*vabariigi*) *ajal teha*; 3) noomenikonstruktsioonis *samal ajal ajaloolised* <*sünnimused*> ~ *tol ajal inimesed* ~ *sel ajal laulud*.



#### 4.4. Sõna sõna

Nii kirjakeeles kui ka õppijakeeles on sõna käändevormidel spetsiifilised kasutusmustrid. Neid ei ole palju ja seetõttu puudub vajadus esitada kirjeldus käänete kaupa. Näiteks EKI tekstikorpuse alusel tulid esile järgmised kirjakeele stereotüüpsed konstruktsioonid: 1) noomenikonstruktsioon ainsuse genitiiviga (*sõna otseses mõttes (tähenduses)*) ja selle leksikaalsed variandid *selle sõna otseses (kõige <otsesemas mõttes, tähenduses>);* 2) noomenikonstruktsioon ja selle leksikaalsed variandid mitmuse genitiiviga (*tema (oma) sõnade kohaselt (järgi, eest)*). Vähem kordusid hinnangulisust edastavad stereotüüpsed konstruktsioonid, kus oli kasutatud ainsuse nominatiivi ja genitiivi: *tehnika viimane sõna ~ oma viimase sõna ja oma sõna öelda (ütleva)*.

Õppijakeele spetsiifika selle sõna kasutuses tuleneb korduvatest tööjuhustest, mis EVKK-s pole tekstidest eraldatud: nt ainsuse ja mitmuse nominatiiv – *pane sõna(d) õigesse <vormi>*, ainsuse partitiiv – *umbes 80 (120, 160) sõna*, mitmuse komitatiiv – *moodustage etteantud sõnadega*. Ülejäänud juhtudel on ainsuse nominatiivi vormi kasutatud noomenikonstruktsioonides (*sõna eesti kultuur, väga levinud sõna ~ nõ esimene sõna*) ja lausekonstruktsioonis *esimene sõna mis*. Ka mitmuse genitiivi on EVKK-s kasutatud noomenikonstruktsioonides *hümni sõnade kirjutaja (autor), Eesti hümni sõnade* ja lausekonstruktsioonides *et hümni sõnade; sõnade autor oli*. Ainsuse partitiivi vorm esines tavapäraselt verbidega *kuulma* ja *mõtleva*: *kuulen sõna kultuur, ma kuulen sõna, kui kuulete sõna ja on mõeldud sõna ~ <on> mõeldud sõna neljakesi*. Veel on kasutatud ainsuse adessiivi (*sõnal võib olla*) ja mitmuse komitatiivi vorme (*<minu> jaoks seostub sõnadega*).

#### 4.5. Kokkuvõtvalt tekstiloomes eelistatud konstruktsioonidest ja käändevormidest

Kui nimisõnade vormistatistika andis ülevaate kirjakeele ja õppijakeele käände-paradigmade samalaadsusest ning sünkroonsusest, teatud vormide suhtelisest üle- ja alakasutusest õppijakeeles, siis käändevormide kasutuseelistused kaks ja enam korda ilmnenuid konstruktsioonides kitsendavad oluliselt nende käänete hulka, mida eesti keeles tegelikult tarvitatakse. Statistika järgi on EKI tekstikorpuses sõna *inimene* käände-paradigma kõige sagedasem vorm ainsuse partitiiv. Kirjakeelele iseloomulike konstruktsioonide kasutusmustrites on aga sellele vormile eelistatud ainsuse nominatiivi: *inimene ja seadus ~ inimene ja loodus; kui (et) inimene on; inimene ei ole (ei saa)*. Samalaadne nihe on iseloomulik ka ülejäänud sõnadele. Kirjakeeles on sõna paradigma sagedasemad vormid ainsuse ja mitmuse nominatiiv, mitmuse genitiiv, ainsuse partitiiv ja genitiiv, samas kui kirjakeele tekstiloomes on eelistatud konstruktsioone, milles on võimalikud vaid kolm käänet: 1) ainsuse nominatiiv (*tehnika viimane sõna*), 2) ainsuse genitiiv (*sõna otseses mõttes (tähenduses), selle sõna otseses, oma viimase sõna, oma sõna öelda*); 3) mitmuse genitiiv (*tema (oma) sõnade kohaselt (järgi, eest)*). Sõna *elu* puhul on kirjakeeles käändevormide sageduse ja nende kasutuseelistuste erinevus veelgi selgem. Paradigmas on suurem osakaal semantilistel käänitel (ainsuse inessiiv, elatiiv, nominatiiv, genitiiv, allatiiv, terminatiiv), samas kui kirjakeele tekstiloomes on eelistatud kasutada konstruktsioone, kus on võimalik vaid ainsuse nominatiiv (*haridus ja elu ~ kultuur ja elu*;

*tema (ta) elu on*). Sõna *aeg* vormistatistika alusel on kirjakeeles sagedasemad ainsuse partitiiv, nominatiiv, adessiiv ning mitmuse nominatiiv ja genitiiv, samas kui korduvalt kasutatud konstruktsioonides on aga eelistatud vaid kahte käändevormi: 1) ainsuse nominatiiv (<on> *teoinimeste aeg ~ käes <on> teoinimeste aeg; on ilus (paras, viimane, õige) aeg; mõni aeg hiljem ~ mõni aeg pärast <mida>; on (oli, olnud, ajanud) kogu aeg; aeg on kallid ~ aeg on möödas (läbi, käes); et viimane aeg ~ oli aeg mil*); 2) ainsuse partitiiv (*kuu aega tagasi (hiljem, enne); umbes (ligi, ja) kuu aega; juba pikka aega ~ väga pikka aega; oli (on) pikka aega ~ pikka aega on; tükk aega tagasi (pärast) ~ nädal aega tagasi; on veel aega ~ vajan veidi aega*). Eespool analüüsitud nelja sagedasema nimisõna käändevormide sageduse ja kasutuseelistuste võrdlusest tuleb selgelt esile, et tegelikult on eesti kirjakeele tekstiloomes eelistatuim vorm ainsuse nominatiiv.

Ka õppijakeele analüüs näitas teatud disproportsiooni vormi sageduse ja kasutuseelistuste vahel kindlat tüüpi konstruktsioonides, kuid need erinevused olid siiski väiksemad kui kirjakeeles. Näiteks sõna *inimene* puhul on nominatiivi vormisagedus nii paradigmas kui ka tekstiloomes eelistatud konstruktsioonides ühesugune (ainsuse nominatiiv – mitmuse nominatiiv, sageduselt kolmas on mitmuse genitiiv, konstruktsioonides aga ainsuse genitiiv). Eriti rikkalikult on õppijakeeles kasutatud konstruktsioone ja nende leksikaalgrammatilisi variante ainsuse nominatiiviga: *kui inimene on (tahab, kogeb, elab, suhtleb) ~ et inimene peab ~ ja kui (et kui, sest kui) inimene; inimene peab olema (maksma, teadma) ~ inimene võib töötada ~ inimene ei saa (ei ole); haritud (iga) inimene peab (saab, võib, tahab)*. Mitmuse nominatiiv on esindatud kahes sagedases konstruktsioonis ja nende leksikaalgrammatilistes variantides: *kõik (need) inimesed on (hoidsid, tahavad); inimesed kes on (ei ole)*. Niisugust kokkulangevust ainsuse ja mitmuse nominatiivi sageduse ja tegeliku vormikasutuse vahel võib pidada õppijakeelele omaseks, nominatiivi eelistamist kirjakeele ja õppijakeele tekstiloomes aga samalaadseks tendentsiks. Sõna *inimene* ülejäänud käändevormide sageduse (mitmuse genitiiv, ainsuse partitiiv, mitmuse partitiiv, ainsuse genitiiv) ja kasutuseelistuste vahel on erinevus kirjakeelest vaid selles, et õppija on rohkem kasutanud ainsuse genitiivi ja partitiivi (ainsuse genitiiv, mitmuse genitiiv, ainsuse partitiiv ja mitmuse partitiiv): 1) ainsuse genitiiv (*mõjutab (ei mõjuta) inimese elu ~ inimese elu mõjutab; inimese elu kõige ~ ?inimese elu haridus ~ \*inimese elu tema (?teema) ~ \*inimese elu ma (?maal)*); 2) mitmuse genitiiv (*matkida inimeste magamisasendit*); 3) ainsuse partitiiv (*ainult (umbes) kuus (sada) inimest; ning kaheksa (üksteist) inimest; kaheksa (kuus) inimest 32-st; üksteist inimest kirjutasi*); 4) mitmuse partitiiv (*väga (nii) palju inimesi*).

Ka teiste analüüsitud sõnade puhul on õppijakeeles tegu samalaadsete nähtustega. Näiteks sõna sagedasematest käänetest (mitmuse nominatiiv, komitatiiv ja genitiiv; ainsuse komitatiiv ning nominatiiv) on õppija eelistanud konstruktsioone, milles saab kasutada vaid kolme käänet: 1) ainsuse nominatiiv (*sõna eesti kultuur; väga levinud sõna ~ nõ esimene sõna; esimene sõna mis*); 2) mitmuse genitiiv (*hümni sõnade kirjutaja (autor); Eesti hümni sõnade; et hümni sõnade; sõnade autor oli*); 3) ainsuse partitiiv (*kuulen sõna kultuur; ma kuulen sõna; kui kuulete sõna; on mõeldud sõna ~ <on> mõeldud sõna neljakesi*). Sõna *elu* käändevormide sageduse ja kasutatavuse vahe on analoogne: sagedusreast ainsuse inessiiv, nominatiiv, genitiiv, partitiiv ja elatiiv on tekstiloomes eelistatud vaid kolme vormi – 1) ainsuse nominatiiv (*inimese elu on (võib, sõltub, erineb); elu sõltub elukohast; <meie> elu erineb Ukraina <elust>; elu on väga*); 2) ainsuse genitiiv (*ja vanemate*

*elu ~ et noorte elu ~ et inimese elu*); 3) ainsuse partitiiv (*inimese elu mõjutab ~ mõjutavad inimese elu*). Sõna *aeg* vormiparadigmas on õppijakeeles sagedasemad ainsuse adessiiv, partitiiv, nominatiiv ja genitiiv. Ent konstruktsioonides, mida õppija tekstiloomes on selle sõnaga eelistatud, ilmneb kõigepealt mitmuse genitiivi vorm (*kõigi aegade suurim (parim, esimene, teine); on läbi aegade ~ läbi aegade on; on kõigi aegade; ESTO-d läbi aegade; Eesti kõigi aegade*), järgnevad ainsuse partitiiv (*vabadus (vajadus, \*vabanduse) planeerida aega; pidis (pidab, pidada, pidama) mõnda aega; väga vähe aega ~ nii palju aega ~ liiga palju aega; ei ole (ei olnud) aega*), ainsuse nominatiiv (*see aeg on ~ mõni aeg oli; muutub (on) kogu aeg; on aeg kus; lendab aeg väga*) ja ainsuse genitiiv (<*mõne*> *aja pärast tuli (oli)*). Niisiis on õppijakeele tekstiloomes eelistatud konstruktsioonides kasutusel vaid kindlad käändevormid, mida vormisageduse statistika alusel käändeparadigmas oleks raske olnud eeldada. Õppijakeele vaieldamatult eelistatuim käänne on nominatiiv, samas aga olenevalt sõna semantikast on tarvitatud konstruktsioone, milles saab kasutada vaid semantilisi käändeid. Nende konstruktsioonide korduvus EVKK-s on hüppeliselt kõrge ja silmatorkava leksikaalgrammatilise varieeruvusega, kirjakeeles aga tunduvalt tagasihoidlikum või mitteoluline. Näiteks sõnaga *aeg* on õppijakeele tekstiloomes eelistatud konstruktsioonides kasutatud põhiliselt ainsuse adessiivi: *külmal ajal peab; et (kuid, sest) sel ajal ~ sest tol ajal ~ ja samal ajal; tol (sellel) ajal tekkis (kasutati) ~ samal ajal läksid ~ praegusel (sel, sellel) ajal on ~ vabariigi (vene, \*tsari) ajal oli ~ sõja ajal polnud ~ vabal (vabariigi) ajal teha; samal ajal ajaloolised <sündmused> ~ tol ajal inimesed ~ sel ajal laulud*. Teiste sõnadega esines ainsuse adessiivi küll harvem, kuid konstruktsioone, kus seda vormi on kasutatud, iseloomustab leksikaalgrammatiline varieeruvus: *kui inimesel on (ei ole); igal inimesel on (peab olema); sõnal võib olla*. Sõna *elu* tarvitati õppijakeeles aktiivselt konstruktsioonides, mis nõudsid ainsuse inessiivi (*minu elus on (oli) ~ meie elus on; elus on palju (väga)*), ning sõna konstruktsioonides, milles on vaja kasutada mitmuse komitatiivi (<*minu*> *jaoks seostub sõnadega*). Seega on õppijakeeles tihti korduvates konstruktsioonides lisaks grammatilistele käänetele esindatud ka semantilised, kuid kirjakeele tekstiloomele pole see omane.

Lahknevused kirjakeele ja õppijakeele käändekasutuse konstruktsioonielistustes annavad teavet ka võrreldavate keelevariantide diskursuserinevuste kohta ning toovad esile sel otstarbel kasutatud stereotüüpsed konstruktsioonid. EKI tekstikorpuses on esile tõstetud inimese, seaduse ja looduse ning elu, hariduse ja kultuuri küsimusi ühiskonnas. Oluliseks on peetud seda, missuguseid inimesi üldse on. Kirjakeeles on olnud tähtis teada anda, kui palju inimesi on õnnetustes või katastroofides viga saanud / hukkunud. Õppijakeelele seevastu on omane modaalsete hinnangute andmine, arutlused teemal, mida inimene oma elus teha võib ning mida ta teha ei tohi või tegema peab. Rõhutatult positiivselt on esile toodud teatud hulka inimesi, kes midagi teevad. Eelistatakse kirjutada neist, kes midagi on saavutanud (nt vabaduse, iseseisvuse, tulemise spordis, kultuuri alal, majanduses).

Kirjakeele iseloomulikuks jooneks on mõningate stereotüüpsete konstruktsioonide kasutamine teksti sidususvahendina või hinnangu edastamiseks: *on (käes) teo-inimeste aeg; sõna otseses mõttes (tähenduses); selle sõna otseses (kõige <otseemas mõttes, tähenduses>); tema (oma) sõnade kohaselt (järgi); tehnika viimane sõna ~ oma viimane sõna ~ oma viimase sõna; oma sõna öelda (ütleva)*. Need kirjakeeles stereotüüpsetena kasutatud konstruktsioonid pole õppijakeelele omased, v.a hinnangu väljendamiseks kasutatud stereotüüp *vabadus (vajadus) planeerida aega*.

## 5. Lõpetuseks

Eesti keele kahe kasutusvariandi võrdlemise tulemuste põhjal võib järeldada, et nii loomulikus keelekasutuses kui ka õppijakeeles on grammatiliste käänete puhul suundumus eelistada nominatiivi ja semantiliste käänete puhul inessiivi, adessiivi ning komitatiivi. Edaspidine laiemapõhjaline uuring peaks näitama, kas tegu on üldisema nominatiivistumisega, ning tõlgendama erinevaid asjaolusid, millest see protsess võiks olla tingitud.

Sõna käändevormide statistiline sagedus korpusaineses ei seleta, millised käändekasutusmustreid läheb vaja, et tekste produtseerida ning nendest aru saada. Vormikasutuse mõistmiseks oleks vaja teada, missugustes leksikaalsetes üksustes, konstruktsioonides ja nende variantides grammatilisi vorme reaalselt kasutatakse. Siit saaks teavet ka selle kohta, kuidas sõna semantika ja grammatika süntagmaatiliselt tasandil omavahel põimuvad. Samuti oleks mõttekas välja selgitada, milliste lingvistiliste ja ekstralingvistiliste asjaoludega on seotud teatud tüüpi konstruktsioonide korduv kasutamine erinevat liiki tekstide produtseerimisel (teema, eesmärgid, register, tekstitüüp, sotsiaalne interaktsioon jms).

Leksikaalsete üksuste, grammatiliste vormide ja konstruktsioonide statistiline väljatoomine võimaldab keeleainest analüüsides leida lingvistilist teavet nähtuste kohta, milleni traditsioonilise korpuspõhise lingvistilise analüüsi tulemusel ei pruugi jõuda, kuid mis tegelikult on keelekasutusele omased. Seetõttu on korpusel tulenevad, erinevat statistikal põhinevat keeletarkvara kasutavad uurimused perspektiivikad ning rakenduslikult (keeleõpe, tõlkimine, sõnastikud, õppematerjalid) olulised.

### Viidatud kirjandus

- Atkins, Sue; Clear, Jeremy; Ostler, Nicholas 1992. Corpus design criteria. – *Literary and Linguistic Computing*, 7 (1), 1–16. doi:10.1093/lc/7.1.1
- Baker, Mona 1995. Corpora in translation studies: An overview and suggestions for future research. – *Target*, 7 (2), 223–243.
- Berber-Sardinha, Tony 1999. Using Key Words in text analysis: Practical aspects. <http://www2.lael.pucsp.br/direct/DirectPapers42.pdf> (12.10.2008).
- Bergh, Gunnar 2005. Min(d)ing English language data on the Web: What can Google tell us? – *ICAME Journal. Computers in English Linguistics*, 29, 25–46.
- Biber, Douglas 1993. Representativeness in corpus design. – *Literary and Linguistic Computing*, 8 (4), 243–257. doi:10.1093/lc/8.4.243 [2nd ed. in: *Practical lexicography: A reader*. Ed. by Thierry Fontenelle. Oxford: Oxford University Press, 63–88.]
- Chambers, John M. 2007. *Software for Data Analysis: Programming with R*. New York: Springer.
- Eslon, Pille 2008. Käändevormide kasutussageduse võrdlus eesti õppijakeeles ja kirjakeeles. – Pille Eslon (toim.). *Õppijakeele analüüs: võimalused, probleemid, vajadused*. Eesti filoloogia osakonna toimetised 10. Tallinn: Tallinna Ülikooli kirjastus, 31–66.
- Granger, Sylviane 1998. Prefabricated patterns in advanced EFL writing: Collocations and formulae. – Anthony Paul Cowie (Ed.). *Phraseology: Theory, analysis, and applications*. Oxford: Oxford University Press, 145–243.
- Granger, Sylviane 1997. On identifying the syntactic and discourse features of participle clauses in academic English: Native and non-native writers compared. – Jan Aarts, Inge de Mönnink, Herman Wekker (Eds.). *Studies in English Language and Teaching*. Amsterdam: Rodopi, 185–198.

- Hanston, Susan 2002. Pattern grammar, language teaching, and linguistic variation: Applications of a corpus-driven grammar. – Randi Reppen, Susan Fitzmaurice, Douglas Biber (Eds.). *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins, 167–186.
- Hoey, Michael; Mahlberg, Michaela; Stubbs, Michael; Teubert, Wolfgang (Eds.) 2007. *Text, Discourse, and Corpora: Theory and Analysis*. With an Introduction by John Sinclair. London, New York: Continuum International Publishing Group.
- Jantunen, Jarmo Harri 2004. Synonymia ja käännösuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännöskielen leksikaalisiin erityispiirteisiin. *Joensuun yliopiston humanistisia julkaisuja* 35. Joensuu: Joensuun yliopisto.
- Heiki-Jaan Kaalep, Kadri Muischnek. *Eesti kirjakeele sagedussõnastik*. Tartu: Tartu Ülikooli Kirjastus, 2002.
- Kennedy, Graeme 2002. Variation in the distribution of modal verbs in the British National Corpus. – Randi Reppen, Susan Fitzmaurice, Douglas Biber (Eds.). *Using Corpora to Explore Linguistic Variation*. Amsterdam/Philadelphia: John Benjamins, 73–90.
- Kennedy, Graeme 1999. *An Introduction to Corpus Linguistics*. London, New York: Longman.
- Leech, Geoffrey 1981. *Semantic. The Study of Meaning*. London: Penguin Books.
- McEnery, Anthony; Wilson, Andrew 2001. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Meier, Heidi 2003. *Essee asend allkeelte tekstiüübivõrdluses*. Magistritöö. Käsikiri Tallinna Ülikooli eesti keele ja kultuuri instituudis.
- Muischnek, Kadri 2006. Verbi ja noomeni püsiühendid eesti keeles. *Dissertationes philologiae Estonicae Universitatis Tartuensis* 17. Tartu: Tartu Ülikooli Kirjastus.
- Nesselhauf, Nadja 2005. *Collocations in a Learner Corpus*. Amsterdam, Philadelphia: John Benjamins.
- Nesselhauf, Nadja; Römer, Ute 2007. Lexical-grammatical patterns in spoken English: The case of the progressive with future time reference. – *International Journal of Corpus Linguistics*, 12 (3), 297–333. doi:10.1075/ijcl.12.3.02nes
- Orpin, Debbie 2005. Corpus linguistics and critical discourse analysis: Examining the ideology of sleaze. – *International Journal of Corpus Linguistics*, 10 (1), 37–61. doi:10.1075/ijcl.10.1.03orp
- Römer, Ute 2007. Learner language and the norms in native corpora and EFL teaching materials: A case study of English conditionals. – Sabine Volk-Birke, Julia Lippert (Eds.). *Anglistentag 2006 Halle. Proceedings*. Trier: Wissenschaftlicher Verlag Trier, 355–363.
- Römer, Ute 2005. Progressives, Patterns, Pedagogy: A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics. *Studies in Corpus Linguistics* 18. Amsterdam: John Benjamins.
- Scott, Mike; Tribble, Christopher 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Philadelphia: John Benjamins.
- Sinclair, John 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Stubbs, Michael 1996. *Text and Corpus Analysis*. Oxford: Blackwell Publishing.
- Sõrmus, Kadri 2008. *Emakeeleõppija korpus*. Statistiline analüüs ja veamärgendussüsteem. Magistritöö. [http://dspace.utlib.ee/dspace/bitstream/10062/6217/1/sormus\\_kadri.pdf](http://dspace.utlib.ee/dspace/bitstream/10062/6217/1/sormus_kadri.pdf) (9.09.2008).
- Tognini Bonelli, Elena 2002. Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach. – Bengt Altenberg, Sylviane Granger (Eds.). *Lexis in Contrast. Corpus-based Approaches*. Philadelphia: John Benjamins, 73–95.
- Volk, Martin 2002. Using the Web as corpus for linguistic research. – Renate Pajusalu, Tiit Hennoste (toim.). *Tähdendusepüüdja*. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 3. Tartu: Tartu Ülikooli Kirjastus, 355–369.

Xiao, Zhonghua; McEnery, Anthony 2005. Two approaches to genre analysis: Three genres in Modern American English. – *Journal of English Linguistics*, 33 (1), 62–82. doi:10.1177/0075424204273957

**Pille Eslon** (Tallinna Ülikool). Viimase aja teadustegevus on seotud eesti õppija- ja kirjakeele korpusest tuleneva kõrvutava uurimisega. Teadustöö teoreetiline taust on funktsionaalgrammatiline: keelte uurimisel on lähtunud keelendite funktsionaalsest potentsiaalist ja kategooriate koosmõjust erinevates kontekstides. Kitsamad uurimisvaldkonnad on aspektuaalsus, temporaalsus, modaalsus ja grammatiseerimine. pille.eslon@tlu.ee.

**Erika Matsak** (Tallinna Ülikool). Uurimisvaldkonnad on loogilised konstruktsioonid eestikeelsetes tekstides; vahendite loomine tekstides sisalduvate leksikaalsete mallide ja grammatiliste konstruktsioonide automaatseks eraldamiseks ning grammatiliste reeglite esiletoomiseks; keele kasutusmallide kvantitatiivne analüüs. matsak@tlu.ee.

## CORPUS-DRIVEN COMPARATIVE ANALYSIS OF VARIANTS OF ESTONIAN

Pille Eslon, Erika Matsak

Tallinn University

The objective of the study was to compare the use of case forms in two variants of Estonian – standard language and learner language. The material was taken from the Standard Estonian corpus of the Institute of the Estonian Language and the Estonian interlanguage corpus of Tallinn University. The frequency of the inflectional forms of the four most common nouns in Estonian (*inimene* ‘person’, *sõna* ‘word’, *elu* ‘life’ ja *aeg* ‘time’) is found out and their morphological paradigms and synchronicity are described. The statistics showed to what extent semantics tells on the formation and frequency of grammatical forms.

Unfortunately the statistics of case forms gives us only the frequency of forms in a sample (e.g. the most frequent cases of the word *aeg* ‘time’ in the Standard Estonian corpus are SG.PART, SG.NOM, SG.ADS, PL.NOM, PL.GEN), but it does not say much about contextual preferences (e.g. in the Standard Estonian corpus the following constructions containing the word *aeg* ‘time’ in SG.NOM are preferred: <on> *teoinimeste aeg* ‘it is time for people of action’ ~ *käes <on> teoinimeste aeg* ‘the time has come for people of action’; *on ilus (paras, viimane, õige) aeg* ‘it is a beautiful (a good, high, the right) time’; *mõni aeg hiljem* ‘a little later’ ~ *mõni aeg pärast <mida>* ‘a little after <something>’; *on (oli, olnud, ajanud) kogu aeg* ‘all the time it is (was, had been, had made)’; *aeg on kallis* ‘time is valuable’ ~ *aeg on möödas (läbi, käes)* ‘the time is over (is up, has come)’; *et viimane aeg* ‘that <it is> high time’ ~ *oli aeg mil* ‘there was a time when’, and with SG.PART: *kuu aega tagasi (hiljem, enne)* ‘a month ago (later, before); *umbes (ligi, ja) kuu aega* ‘about (around, and) one month’; *juba pikka aega* ‘for a long time already’ ~ *väga pikka aega* ‘for a very long time’; *oli (on) pikka aega* ‘was (has been) for a long time’ ~ *pikka aega on* ‘has <done something> for a long time’; *tükk aega tagasi (pärast)* ‘a long time ago (after <something>)’ ~ *nädal aega tagasi* ‘a week ago’; *on veel aega* ‘there is time yet’ ~ *vajan veidi aega* ‘I need a little time’).

Analysis of contextual preferences is necessary for pedagogical and lexicographical purposes, and, importantly, also for making explicit the hidden tendencies that are working in the language system synchronically. For this purpose, the grammatical constructions and their lexico-grammatical variants (allowing only certain case forms) typical of Standard Estonian and learner Estonian were found out. This reduced significantly the number of the case forms eligible for further analysis. Mere statistics of case forms would never have provided the data on their use in constructions and on their preference order in text production.

From a comparison of the frequencies of the case forms in learner language versus Standard Estonian it appeared that there is an overwhelming tendency in learner language to use the nominative case above all (*inimene* ‘person’ – in singular and plural, *sõna* ‘word’ – in plural, *elu* ‘life’ – in singular). However, the nominative is a frequent form in Standard Estonian as well (*sõna* ‘word’ – in nominative singular and nominative plural holds the first place in the frequency list of

word forms, *inimene* 'person' – nominative plural holds the second place in the frequency list after partitive singular; nominative singular holds the third place, *aeg* 'time' – nominative singular holds the third place after inessive singular and elative singular). There are only two words in the standard language that slightly prefer the partitive (*inimene*, *aeg*). This evidence calls into question the statement that in the Estonian case system there is an oppositional pair consisting of the nominative and partitive cases, of which the partitive is the unmarked member.

In case the learner language and standard language lacked synchronicity, then it could be assumed that it is a manifestation of atypical use of word forms, which is worth studying. But the results of this study showed that in both language variants, the clearly preferred case is the nominative. Further research should reveal whether this phenomenon can be considered a process of nominativization, and what its possible motivations could be. The next most preferred semantic cases in both standard and learner language were the inessive, the adessive and the comitative. The frequency of case forms of a word does not explain, though, which patterns of case usage are necessary to actually produce and understand texts.

Statistical representation of lexical units, grammatical forms and constructions has enabled us to reveal such linguistic data that, despite being characteristic of language use, is not easily accessible by traditional corpus-based analysis. Therefore we believe that corpus-driven studies using statistical and language software have a long-term theoretical and applied value (in language teaching, translation, dictionaries and learning materials).

**Keywords:** corpus linguistics, corpus-driven comparative analysis, inflectional grammar, case preferences in Estonian usage, Estonian language