

ÕPPIJAKEELE KORPUSANALÜÜSI TÄIENDAVATEST MEETODITEST

Annekatriin Kaivapalu

Ülevaade

Õppijakeele korpusanalüüsi keskmeks on esmajoones õppija keeleline produktsioon, mida on valdavalt uuritud veaanalüüsi meetodi abil. Seevastu sihtkeele produtseerimisprotsessi kohta saadud teave on juhuslikum või puudub üldse. Artiklis keskendutakse veaanalüüsi peamistele probleemidele nagu vigade klassifitseerimine ja interpreteerimine ning vigade psühholingvistiline reaalsus. Samuti kirjeldatakse meetodeid, mis õppijakeele korpusanalüüsi selles osas täiendavad. *Eesti vahekeele korpuse* sõnajärjevigade näitel analüüsitakse tajutesti kasutamisevõimalusi vigade psühholingvistilise reaalsuse väljaselgitamisel nii emakeelse keelekasutaja kui ka keeleõppija seisukohalt. Lisaks kirjeldatakse õppijakeele korpusanalüüsi täiendavaid introspektiivseid meetodeid (valjusti mõtlemine ja intervjuu) ning tutvustatakse kirjutamisprotsessi fikseerivat arvutiprogrammi ScriptLog¹.

Võtmesõnad: korpuspõhine veaanalüüs, vigade psühholingvistiline reaalsus, eesti keele sõnajärjemallid, eestikeelne keele-

¹ Tööd on toetanud riikliku programmi „Eesti keele keeletehnoloogiline tugi (2006–2010)” projekt „VAKO: Eesti vahekeele korpuse keeletarkvara ja keeletehnoloogilise ressursi arendamine (2008–2010)” ja riikliku programmi „Eesti keel ja kultuurimälu” projekt „REKKi käsikirjaliste materjalide digiteerimine, Eesti vahekeele korpuse alamkorpuste loomine ja korpuse kasutusvõimaluste populariseerimine (2009–2013)”.

kasutaja, venekeelne eesti keele õppija, introspektiivsed meetodid

1. Õppijakeele korpusanalüüs

Õppijakeele uurimine mahukate elektrooniliste andmekogude, õppijakeele korpuste² abil on muutumas üha levinumaks meetodiks keeleõppealases uurimistöös. Õppijakeele korpuse koostatakse ja arendatakse mitmel pool maailmas³ ning järjest lisandub projekte, mille eesmärk on uurida korpusanalüüsi abil õppijakeele keeleomaseid ja universaalseid jooni nagu keeltevaheline mõju (Kaivapalu 2008; Tenfjord 2008), keelendite ebatüüpiline sagedus, lihtsustamine, ebatüüpiline konventsionaalsus jt (Jantunen 2007). Ka keeleoskuse omandamise ja mõõtmise uurimist ühendav mitteformaalne SLATE-võrgustik⁴ kasutab ainekuna esmajoonel õppijakeele korpuse, nii õppetöö käigus loodud kirjalike tekstide kui ka keeleoskuse tasemeid mõõtvate testide elektroonilisi kogusid (Tarnanen 2007).

Korpuspõhise lähenemisviisi eeliseks võrreldes traditsioonilise uurimisega on kahtlemata tulemuste parem üldistatavus: suurte andmehulkade analüüs annab objektiivsema pildi eri tasemel õppijate keeleoskuse arengust ning keele produtseerimise spetsiifikast. Samas on korpuspõhisel lähenemisviisil mitmeid piiranguid: korpusainestik ja -metodoloogia ei võimalda

² S. Grangeri (2003: 465) määratluse kohaselt on paralleelselt õppijakeelekorpuse mõistega kasutusel terminid *vahekeelekorpus* (ingl *interlanguage corpus*) ja *teise keele korpus* (ingl *L2 corpus*).

³ Ülevaadet õppijakeele korpustest vt Eslon & Metslang 2007, teoksil olevatest uurimisprojektidest vt <http://cecl.fltr.ucl.ac.be/SummerSchool2007/participants%20research%20interests.html> (19.8.2009).

⁴ <http://www.jyu.fi/hum/laitokset/kielet/cefling/suom> (19.8.2009).

uurida sugugi kõiki õppijakeele aspekte. Olulisim on asjaolu, et õppijakeele korpusanalüüs keskendub esmajoones õppija keelelisele produktsioonile, sihtkeele produtseerimisprotsessi kohta saadud teave seevastu on juhuslikum või puudub üldse. Kuigi protsessi uurimiseks tuleb eelnevalt põhjalikult tunda produkti (vt Odlin 1989) ja produkti vahendusel on mingil määral võimalik uurida õppijate keeleloome- ja mõistmisprotsesse, ei anna õppija keeleline produktsioon üksi siiski piisavalt informatsiooni sihtkeele omandamise ja produtseerimise käsitlemiseks. Käesoleva artikli eesmärgiks ongi analüüsida probleeme, mis kaasnevad õppijakeele korpusanalüüsiga ning kirjeldada korpusanalüüsi täiendavaid introspektiivseid meetodeid, mis võimaldavad lisaks produktile süveneda ka õppija keeleloomeprotsessi.

2. Veaanalüüs teooria ja meetodina

Valdav osa õppijakeelt käsitlevatest uurimustest lähtub nii teoreetilise tausta kui ka meetodi poolest veaanalüüsis. **Teooriana** erineb veaanalüüs ja sellega tihedalt seotud vahekeele analüüs neile eelnenud strukturalistlik-kontrastiivsest teooriast selle poolest, et vigade ainsaks põhjustajaks ei peeta mitte enam õppija lähtekeelt, vaid vigade allikana on hakatud nägema ka sihtkeele süsteemi ja keeleõpetust (Sajavaara 1999b: 116). Veaanalüüsi ja vahekeele analüüsi teooria järgi osutavad vead keeleõppija probleemidele, mida iseloomustab süstemaatilisus.

Veeanalüüsi **meetodiks** on olnud vigade kogumine õppijate kirjalikest ja suulistest tekstidest. Läbi on viidud ka empiirilisi uurimusi veendumaks, kas õppijad tõepoolest teevad eeldatud vigu (Herranen 1978). Veaanalüüsis läbitakse tavaliselt järgmised etapid (James 1998): 1) vigade identifitseerimine,

mille käigus eristatakse kompetentsivead juhuslikest eksimustest; 2) vigade kirjeldamine; 3) vigade klassifitseerimine ja 4) vigade seletamine. Põhijoontes samadelt alustelt lähtub ka korpuspõhine veaanalüüs, mis erineb traditsioonilisest veaanalüüsist eelkõige suurema standardiseerituse ning normikohaste ja normivastaste kontekstide kõrvutamise poolest (Granger 2002: 11–16). Nii nagu traditsioonilises õppijakeele uurimises lähtub ka suurem osa õppijakeele korpuspõhiseid käsitlusi esmajoones veaanalüüsist. Seejuures on asjakohane rõhutada, et korpusanalüüsi kontekstis on veaanalüüs käsitletav meetodina, mitte aga teooriana (Hagen 2009). Viimasel ajal on siiski hakatud õppijakeelt korpuspõhiselt uurima muudiski teoreetilistes ja metodoloogilistes raamistiketes (nt Nesselhauf 2003, 2005; Rasier & Hiligsmann 2007; Jantunen 2007; Eslon & Matsak 2009).

Kuigi õppijakeele veaanalüüs aitab paremini mõista õppijakeele olemust ja arengustaadiume ning omab olulisi pedagoogilisi rakendusi (Granger 2002: 14), on seda meetodit õigusstatult kritiseeritud mitme metodoloogilise küsitavuse tõttu (vt Borin & Prütz 2004: 68; Ellis & Barkhuizen 2005: 70). Veaanalüüsile on ette heidetud lähteandmete heterogeensust, veakategooriate laialivalguvust, keeleomandamise käsitlemist staatilisena, piirdumist ainult sellega, mida õppijad teha ei suuda. Olulisemaid vajakajäämisi on väga juhuslik teave produtseerimisprotsessi ning õppijate strateegiate kohta. Samu probleeme ei ole suudetud vältida ka korpuspõhise veaanalüüsi juures. Järgnevalt leiavad käsitlemist mõned traditsioonilise ja korpuspõhise veaanalüüsi probleemid.

3. Korpuspõhise veaanalüüsi probleemid

Nii traditsioonilise kui ka korpuspõhise veaanalüüsi probleemid on eelkõige seotud vigade klassifitseerimise, vigade interpreteerimise ja vigade psühholingvistilise reaalsusega.

3.1. Vigade klassifitseerimine ja interpreteerimine

Õppijakeele korpuste veaklassifikatsioonid on reeglina lingvistilised: vigade klassifitseerimisel lähtutakse keelest kui paradigmaatiliste ja süntagmaatiliste seoste süsteemist. Paradigmaatilised seosed on hierarhilised: kõrgema tasandi keelend hõlmab madalamatasandilisi ja madalamatasandiline keelend sisaldub kõrgematasandilises. Süntagmaatilised seosed on linearsed, neis kombineerub leksikaalgrammatilise ja lause-tasandi reeglistik. Keelesüsteemi süntagmaatilise ja paradigmaatilise telje ristumisel tekib kahemõõtmeline lingvistiline veaklassifikatsioon (Hufheisen 1991; Michiels 1999). Samast põhimõttest on lähtunud ka *Eesti vahekeele korpuse* (EVKK) veaklassifikatsiooni loomisel (lähemalt vt Eslon 2007: 106–109). Vealiike ja nende alamliike on EVKK veaklassifikatsioonis kokku 170, vealiikide määramise aluseks on ligi 300 tunnust.

Ükskõik kui detailne ei oleks mis tahes õppijakeele korpuse veaklassifikatsioon, ei ole see kunagi lünkadeta: klassifikatsiooni loomisel ei ole võimalik ette näha kõiki potentsiaalseid vealiike ja alamliike, mis õppijakeeles võivad esineda. Vajalik vealiik võib klassifikatsioonist puududa, nagu näites (1) esitatud lause puhul, milles on tegemist vale käändevormi kasutamisega ja ühildumiseiga ühildumatuse asemel:

(1) *Kõikedele meile palvetele ja meelepahadele nad vastasid ...*

Seetõttu on töö veaklassifikatsiooni täiustamisel pidev, uusi vealiike ja alamliike tuleb korpuse märgendamisel juurde, osa olemasolevatest võib aga osutuda mittevajalikuks.

Tõsiasi, et interpretatiivsus kuulub vea olemusse, on üldteada (vt Corder 1981). Tavapäraselt saab üks keeleviga ikka mitu märgendit ning see ei tulene veaklassifikatsiooni puudulikkusest, vaid asjaolust, et interpretatiivsus on veale iseloomulik. Näiteks võib õppija produtseeritud keelendit *projekteerimine* tõlgendada EVKK veaklassifikatsiooni alusel nii veana internatsionalismide kasutamisel (vealiik 1.6) kui ka häälduspärase kirjaviisina (vealiik 1.7). Vea mitmemõõtmelisel kirjeldamisel toetub EVKK märgendussüsteem veaklassi, -liigi ja alamliikide vahelistele seostele, samuti arvestatakse lähte- ja sihtkeele vaheliste sümmeetria, asümmeetria ja analoogiaseoste olemasoluga. Väidetavalt annab veaklasside, -liikide ja alamliikide väljatoomine ning alamliikide jagunemine omakorda kitsama teks vea ilminguteks võimaluse rääkida veast mitte ainult kui tagajärjest, vaid välja tuua ka vea tekkepõhjusti (vt Eslon 2006, 2007: 107). Tuleb siiski rõhutada, et vaid korpuspõhise veaanalüüsi alusel ning introspektiivsete meetodite abil õppija produtseerimisprotsessi uurimata saab rääkida vaid vigade oletatavatest, mitte aga tegelikest tekkepõhjustest. Samuti jääb pelgalt veaanalüüsi põhjal selgusetuks, milliseid keelendeid õppija veana tunnetab ja kas need vead on häirivad ka emakeelse keekekasutaja jaoks, teisisõnu, vigade psühholingvistiline reaalsus.

3.2. Vigade psühholingvistiline reaalsus

Keele omandamine ei ole kunagi lingvistiline probleem. Seda mõjutavad nii psühholoogilised kui ka sotsiaalsed tegurid, mis on tihedalt seotud lingvistiliste teguritega. Lingvistiliste, psühholoogiliste ja sotsiaalsete tegurite osakaal keele omandamisel ja omandamisprotsessi mõjutamisviis on eri tingimustes erinev ning kõiki lõpptulemust mõjutavaid tegureid veel ei tuntagi, kuigi sihtkeele omandamist on uuritud juba nelja aastakümne vältel (Sajavaara 1999a: 76). Seetõttu on vigade klassi-

fitseerimise ja interpreteerimise kõrval korpuspõhise veaana-lüüsi olulisim probleem vigade psühholingvistiline reaalsus⁵ (Kaivapalu 2004; Sajavaara 2006): kas ja kuidas viga tunnetatakse, kelle arvates ja mille suhtes on tegemist veaga.

EVKKs on keeleviga määratletud grammatikareeglile või kommunikatiivsele eesmärgile mittevastava keekekasutusena, mille hulka ei kuulu väsimusest ja hooletusest põhjustatud eksimused ning keelevääratused (Eslon 2007: 106). Grammatikareeglile või kommunikatiivsele eesmärgile vastavust ei tajuta aga alati sugugi üheselt, nagu ilmneb näites 2–4 esitatud sõnajärjevigade puhul:

- (2) 3 aastat tagasi **ma lõpetasin** 10 klassi Narva Hummanitaar gümnaasiumis.
- (3) Juba koolis **ma unistasin**, et tulevikus lähen TTÜ-sse õppima.
- (4) Üks kord nädalas **tüdruk sõidab** linna, et käia kinos.

Tegemist on struktuurilt sarnaste lausetega, milles kõigis on rikutud V2-reeglit: iseseisvas jaatavas väitlause on öeldisverb (täpsemalt öeldisverbi pöördeline osa) tavaliselt teisel kohal, teemana toimiva moodustaja järel (EKK 2007: sy93.html). Hoolimata lausestruktuuri sarnasusest suhtusid erinevad märgendajad analoogilistesse sõnajärjevigadesse erinevalt: kui esimeses lauses märgendasid normivastase sõnajärje määrus–alus–öeldis sõnajärjeveaks kõik kolm märgendajat, siis teises lauses kaks ja kolmandas lauses vaid üks märgendaja. Selline tulemus on seda üllatavam, et lausetes (2) ja (3) on aluseks rõhutu asesõna, mille puhul peetakse otsejärke suhteliselt tavaliseks (EKK 2007: sy93.html). Seega tekib küsimus, millest sõltuvad

⁵ Ligikaudu samas tähenduses on kasutatud ka terminit *psühholingvistiline reaalsus* (vt Eckmann 2004).

märgenduseelistused, ehk teisisõnu, kas kirjakeele normile vastav V2-reegel on psühholingvistiliselt reaalne ka tegeliku keelekasutuse seisukohalt. Keelend, mis ei vasta kirjakeele normile, ei tarvitse tingimata olla vastuolus emakeelse keelekasutaja keeletajuga. Keelendi loomulikkus või ebaloomulikkus võib sõltuda vanusest, soost, haridusest jt taustateguritest, keeleõppija puhul lisanduvad võimalike keeletaju mõjutajatenä veel lähtekeel ja sihtkeele oskuse tase.

4. Korpusanalüüsi täiendavaid meetodeid

Eelmises peatükis kirjeldatud veaanalüüsi probleemid on osaliselt lahendatavad introspektiivsete meetodite abil, mis täiendavad korpusanalüüsi nende uurimisküsimuste osas, mida viimane lahendada ei võimalda. Introspektiivsetest meetoditest on praktikas enim kasutatud psühholingvistilist tajutesti ja intervjuud, vähem valjusti mõtlemist ja õppija produtseerimisprotsessi fikseerivaid arvutiprogramme nagu ScriptLog. Järgnevalt kirjeldatakse lähemalt nende meetodite rakendusvõimalusi kombineerituna korpusanalüüsiga; põhjalikumalt peatutakse psühholingvistilisel tajutestil.

4.1. Psühholingvistiline tajutest

Lihtsaim võimalus uurida, milline keelend on emakeelse keelekasutaja või keeleõppija jaoks keeleomane ja loomulik, on küsida seda keelekasutajalt või -õppijalt endalt. Selleks kasutatakse ühe meetodina psühholingvistilisi tajuteste, mille käigus palutakse uurimuses osalejatel hinnata keelendi loomulikkust Likert-tüüpi skaalal *täiesti loomulik – pigem loomulik – pigem ebaloomulik – täiesti ebaloomulik*. Seejuures rõhutatakse, et lähtuda tuleb oma keeletundest, mitte aga keelendite normikohasusest.

Kuna EVKK sagedaseim veatüüp on sõnajärjevead, mille märgendamisel märgendajate hinnangud sõnajärjemalli normikohasusele ja normivastasusele sageli erinevad, siis kasutati sõnajärjemallide loomulikkuse uurimiseks psühholingvistilist tajutesti (vt lisa 1). Testilausete aluseks oli 14 sõnajärjevea tüüpi, mille osas märgendajad olid üksmeelsed, näiteks

(5) *Aasta jooksul nad kasvatavad põrsast.*

Õppija vigasest lausest moodustati „Eesti keele käsiraamatu“ (2007) alusel normikohane lause

(6) *Aasta jooksul kasvatavad nad põrsast.*

Normikohase lause (6) sõnajärge muudeti nii lause moodustajate ümberpaigutamise teel kui ka sõnavara varieerimise teel. Igast sõnajärjevea tüübist moodustati sel viisil 3–5 varianti, mille normikohasus võis olla varieeruv, nagu näiteks lause (6) variandid (7) – (9):

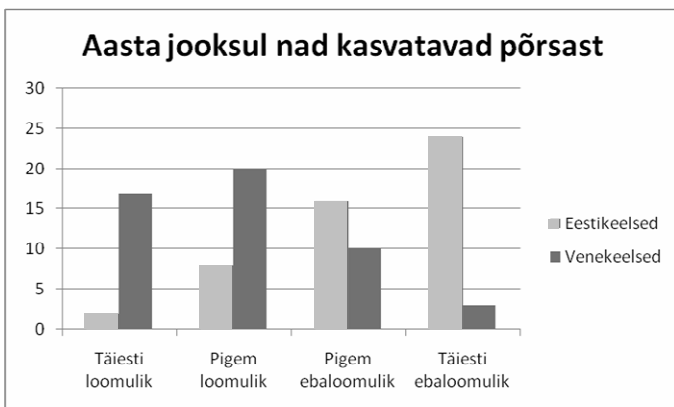
(7) *Nad õpivad terve nädala matemaatikat.*

(8) *Pool päeva lugesime me kriminulle.*

(9) *Luuletust kirjutas ta päev läbi.*

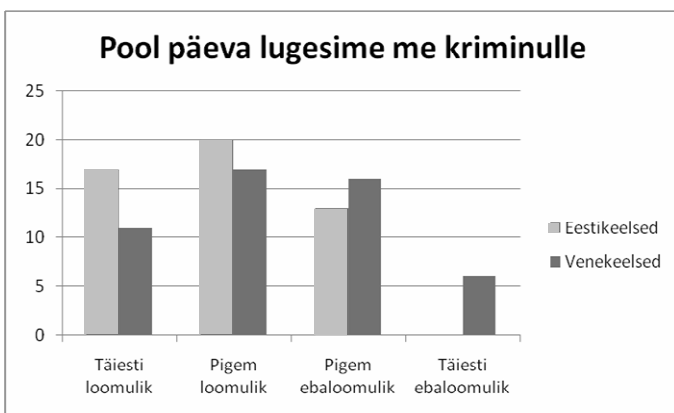
Uurimuses osalesid Tallinna Ülikooli üliõpilased vanuses 19–37 aastat. Osalejate hulgas oli 50 eestikeelset keelekasutajat ja 50 venekeelset eesti keele õppijat, kes olid koolis õppinud eesti keelt 6–12 aastat. Uurimuses ei kasutatud nende informantide vastuseid, kes olid õppinud eestikeelses koolis, kuid kelle emakeel oli vene keel ja kodune keel eesti keel või kelle emakeel oli eesti keel ja kodune keel vene keel. Uurimuse tulemused on esitatud joonistel 1–4.

Testivastustest selgus, et V2-reeglit eiravat sõnajärge määrusalus-öeldis (vt joonis 1) tajusid täiesti ebaloomulikuna või pigem ebaloomulikuna 80% eestikeelsetest keelekasutajatest. Seevastu enamiku venekeelsete keeleõppijate jaoks (37 informanti 50-st) oli selline sõnajärg täiesti loomulik või pigem loomulik.



Joonis 1. Eestikeelsete keelekasutajate ja venekeelsete eesti keele õppijate hinnang lause *Aasta jooksul nad kasvatavad põrsast* loomulikkusele

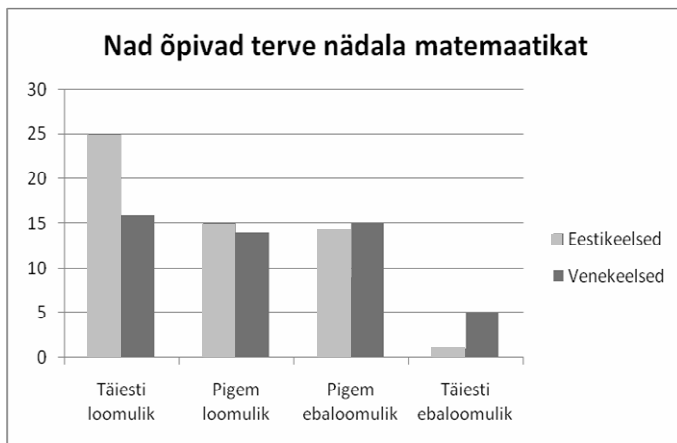
V2-reeglit järgivat sõnajärge (vt joonis 2) tajus täiesti loomulikuna või pigem loomulikuna 37 informanti 50 eestikeelsest keelekasutajast. Asjaolu, et 13 eestikeelse keelekasutaja jaoks



Joonis 2. Eestikeelsete keelekasutajate ja venekeelsete eesti keele õppijate hinnang lause *Pool päeva lugesime me kriminulle* loomulikkusele

tundus V2-sõnajärg olevat pigem ebaloomulik, võib muuhulgas olla tingitud mitmuse esimese pöörde lõpu ja personaalpronoomeni kõrvuti asetsemisest. Viimane oletus vajaks aga täpsustamist intervjuu või valjusti mõtlemise abil. Ka pisut enam kui poolte venekeelsete õppijate jaoks (28 informanti 50-st) oli antud sõnajärg täiesti või pigem loomulik.

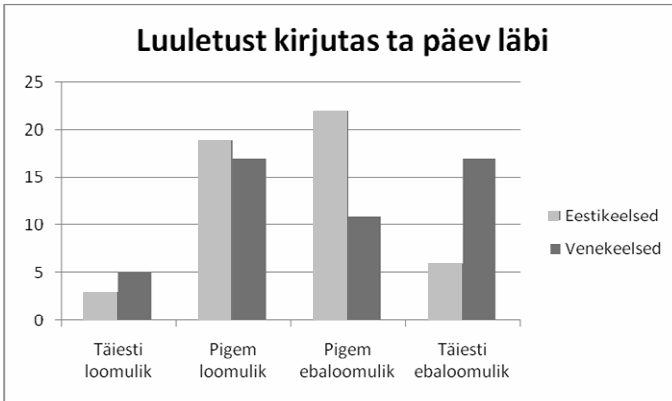
Tavapärane normaallause sõnajärjemall alus-öeldis-määrusihitis (vt joonis 3) oli täiesti loomulik või pigem loomulik 80% eestikeelsete keelekasutajate ning 60% venekeelsete keeleõppijate jaoks. 40% keeleõppijatest tajus antud sõnajärge siiski pigem või täiesti ebaloomulikkuna.



Joonis 3. Eestikeelsete keelekasutajate ja venekeelsete eesti keele õppijate hinnang lause *Nad õpivad terve nädala matemaatikat* loomulikkusele

Järgmises lauses (joonis 4) on tegemist nn fokuseeritud sõnajärgiga: rõhutatav lauseliige on paigutatud lause algusesse markeerimaks, et just luuletust, aga mitte romaani, novelli või näidendit kirjutati päev läbi. Testivastustest selgus, et kon-

tekstivaba fokuseeritud sõnajärjega lauset tajusid nii eestikeel-
sed keelekasutajad kui ka keeleõppijad ühtviisi pigem eba-
loomuliku kui loomulikuna: 28 keelekasutaja ja keeleõppija
jaoks oli antud sõnajärg pigem või täiesti ebaloomulik.



Joonis 4. Eestikeelsete keelekasutajate ja venekeelsete eesti keele õppijate hinnang lause *Luuletust kirjutas ta päev läbi* loomulikkusele

Psühholingvistiline tajutest näitis niisiis V2-reegli psühholingvistilist reaalsust eestikeelse keelekasutaja jaoks: V2-reegli rikkumist tajuti pigem ebaloomuliku kui loomulikuna. Seevastu venekeelsete keeleõppijate jaoks on V2-reegli rikkumine pigem loomulik kui ebaloomulik. Samas tajusid pisut enam kui pooled keeleõppijatest V2-reeglile vastavat sõnajärge täiesti loomulikuna või pigem loomulikuna. Seega ei ole V2-reegli järgimine või rikkumine venekeelse õppija keeletaju seisukohalt distinktiivne: loomulikuna tajuti nii kirjakeele normi kohast sõnajärge määrus-öeldis-alus kui ka normivastast sõnajärge määrus-alus-öeldis. Eestikeelse keelekasutaja jaoks väljendab aga V2-reegli järgimine loomulikku ning selle rikku-

mine ebaloomulikku keelekasutust. Järelikult on põhjendatud V2-reegli rikkumise märgendamise sõnajärjeveana: emakeelse keelekasutaja keeletaju toetas antud juhul õigekeelsusreeglit ning samas ilmnes, et keeleõppija ei taju selgelt nende sõnajärjemallide erinevust.

Kirjeldatud psühholingvistiline tajutest võimaldab uurida, milline keelend on keelekasutaja ja/või -õppija jaoks loomulik ja milline mitte, ehk teisisõnu, kaardistab olukorda. Samas ei anna tajutest teavet selle kohta, miks keeleõppija keelendit loomulikuna või ebaloomulikuna tajub, millised tegurid selle tajumist põhjustavad. Keelekasutaja ja/või -õppija keeletaju põhjuste väljaselgitamiseks sobivad valjusti mõtlemine ja intervjuu, mida kasutatakse sageli kombineerituna (nt Kaivapalu 2005).

4.2. Valjusti mõtlemine

Valjusti mõtlemine (ingl *thinking aloud*, sm *ääneenajattelu*) on meetod, mille eesmärgiks on pääseda ligi mentaalsetele protsessidele, mille kohta muul viisil informatsiooni ei saa (Swain, Lapkin 1995). Valjusti mõtlemine seisneb selles, et keelekasutaja või -õppija kommenteerib testi sooritamise või ülesande täitmise ajal valjusti oma tegevust ja põhjendab oma tegutsemisviisi. Kommentaarid lindistatakse ja litereeritakse. Valjusti mõtlemine võimaldab uurida keeleloome- ja mõistmisprotsessi mehhanisme ning iseärasusi. Meetodit on kritiseeritud, väites, et tegemist on kunstliku olukorraga, mitte spontaanse keelekasutusega, samas aga ei ole paremat lahendust mentaalsete protsesside uurimiseks pakutud. Enam on valjusti mõtlemist kasutatud teksti mõistmise uurimiseks (vt Block 1992), kuid see on andnud väga häid tulemusi ka õppijate produtsiooniprotsessi iseärasuste väljaselgitamisel (vt Kaivapalu 2005). Näited (10) ja (11) kirjeldavad eestikeelsete soome keele

õppijate kommentaare oma käänamisprotsessi kohta (Kaivapalu 2005: 269–270):

- (10) *toinen toisia*, see võiks nüüd olla selle järgi, et eesti keeles on *teine* ja *teisi*. Sisseütlev on *toisiin* ja seestütlev on *toisista*
- (11) *pankki* See on teistsugune kui *herkku* ja *kirkko*. *pankkeja* *pankkeja pank/kei/ta* ei *pan-* *pankkei/siin* *pan-* *pankei-* *pan/koista* või *pan/keista* *pankeista* siiski.

Esimene kommentaar annab tunnistust lähtekeele (eesti keel) analoogia kasutamisest sihtkeele (soome keel) mitmuse partitiivi produtseerimisel, teises näites püüab õppija erineva tüvevokaaliga sihtkeele sõnu omavahel võrreldes ja analoogiaseost otsides leida õiget käändeparadigmat.

Valjusti mõtlemine sobib korpusanalüüsi täiendama eelkõige heitmaks valgust keekekasutaja ja -õppija kirjutamisprotsessile: kuidas sõnu valitakse, käände- ja pöördevorme produtseeritakse, lauseid moodustatakse ja tekstiks seotakse. Samuti sobib meetod suurepäraselt teksti mõistmise uurimiseks. Seejuures tuleb arvestada, et kõik informandid ei ole ühtviisi aldis oma tegevust kommenteerima, samuti ei suuda kõik inimesed ühtviisi hästi keskenduda samaaegselt teksti kirjutamisele ja sellest rääkimisele. Valjusti mõtlemise kui uurimismeetodi kasutamise tulemusel selgusetuks jäänud küsimusi võib täpsustada retrospektiivse intervjuu abil.

4.3. Retrospektiivne intervjuu

Retrospektiivne intervjuu viiakse läbi vahetult valjusti mõtlemise järel, et intervjuuerija ei jõuaks unustada küsimusi, mida ta tahab täpsustada ja süvitsi uurida, ega intervjueeritav oma tegutsemise põhjendusi. Retrospektiivsel intervjuul nagu igal teist tüüpi intervjuulgi on meetodina mitmeid eeliseid (vt Hirsjärvi & Remes & Sajavaara 2005: 191–194; Hirsjärvi & Hurme

2000): intervjuueerija on intervjuueeritavaga vahetus keelelises interaktsioonis; ainekogu kogumine on paindlik; intervjuueeritav saab ennast vabalt väljendada ning ta on tähendusi loov ja seega aktiivne osapool. Retrospektiivse intervjuu küsimused kujunevad otseselt uurimuses osalejate testi sooritamise või ülesande lahendamise ja valjusti mõtlemise põhjal. Nii on näites (12) täpsustatud, kas õppija toetub sihtkeele sõnavormide produtseerimisel lähtekeelele või mitte (Kaivapalu 2005: 210):

(12) A.K.: *Kui sa mõtled, et sõna on nagu eesti keeles, kas sa siis moodustad eesti keele järgi ka või?*

R.: *Ei, eesti keele järgi ma ei moodusta.*

A.K.: *Miks sa ei moodusta, oskad sa öelda?*

R.: *Ma ei ole hakand nagu moodustama, proovin ikka mõelda, kuidas soome keeles on. Need on ikka üpris erinevad ka, eesti keeles moodustamine ja soome keeles moodustamine. Mulle on jäänud selline mulje, siis ei hakka.*

A.K.: *Kardad, et ajab segadusse?*

R.: *Nojah.*

A.K.: *Ei julge nagu eesti keelt appi võtta?*

R.: *Ei.*

Nii nagu valjusti mõtlemise puhul, on ka retrospektiivse intervjuu eesmärgiks korpusanalüüsi täiendamine õppija produtseerimisprotsessi käsitleva teabega.

4.4. Keeletehnoloogilised vahendid: arvutiprogramm ScriptLog

Produtseerimisprotsessi uurimiseks on loodud ka keeletehnoloogilisi vahendeid. Enim on kasutamist leidnud arvutiprogramm ScriptLog⁶. Tegemist on programmiga, mis fikseerib

⁶ <http://www.scriptlog.net> (19.8.2009).

kirjutamisprotsessi: salvestab õppija peatumised, tagasipöördumised ja parandused. Kuna igasugune peatumine produtsereerimisel annab tunnistust õppija probleemidest, vajadusest järele mõelda, siis sisaldavad peatumised, tagasipöördumised ja parandused väärtuslikku teavet selle kohta, milliste raskustega õppija tekstiloomes kokku puutub ning kuidas ta neid lahendab. Nii on Scriptlogi kasutatud näiteks poolakeelsete õppijate, kelle teine keel on saksa keel, rootsi keele kirjutamisprotsessi (vt Kowal 2009) uurimiseks, et välja selgitada õppijate esimese ja teise keele vaheline tööjaotus kolmanda keele omandamisel. Uurimus näitas, et teise keele (saksa keele) interferents rootsi keele omandamisele ilmnes eelkõige sõnatasandil (sõnavalik, sõnamuutmine) ning esimese keele (poola keele) interferents süntaktilisel ja tekstitasandil. Nii nagu psühholingvistiliste tajutestide puhulgi, ei anna programm informatsiooni probleemide põhjuste kohta ning need tuleb välja selgitada valjusti mõtlemise meetodil või intervjuude abil.

Kuigi teadaolevalt ei ole ScriptLogi korpusainestiku kogumisel seni kasutatud, on sellel kirjutamisprotsessi uurimise seisukohalt ulatuslik perspektiiv. Seetõttu on kavas luua EVKKsse ScriptLogi allkorpus.

5. Kokkuvõtteks

Korpusanalüüs on õppijakeele universaalsete ja keelespetsiifiliste joonte kaardistamisel ning õppijakeeles esinevate suundumuste väljaselgitamisel asendamatu eelkõige tulemuste kõrge üldistusastme tõttu. Seda võimaldab ulatuslik andmesitik. Õppijakeele nähtuste olemuse ja põhjuste süvitsi uurimine eeldab aga korpusanalüüsi täiendamist introspektiivsete meetoditega, millest olulisemad on valjusti mõtlemine ja sellele järgnev retrospektiivne intervjuu testi sooritajaga. Samamoodi

on võimalik kombineerida teksti kirjutamist produtseerimisprotsessi fikseeriva arvutiprogrammi ScriptLog abil ja retrospektiivset intervjuud. Oluline on niisiis ainestiku triangulatsioon: ainult ühe meetodi kasutamine piirab oluliselt õppijakeele uurimisvõimalusi ega luba teha põhjanevaid järeldusi keele omandamise universaalsete ja keelespetsiifiliste joonte kohta. Korpuspõhist veaanalüüsi tuleks õppijakeele uurimisel käsitleda vaid ühe meetodina, millest lähtutakse õppijakeele probleemsete piirkondade kaardistamisel ja mida saab rakendada samuti keeletehnoloogiliste lahenduste leidmisel. Veaanalüüsi puhul ei tuleks kindlasti arvestada ainult kirjakeele normi, vaid veenduda ka normivastase keelendi psühholingvistilises reaalsuses nii emakeelse keeikasutaja kui ka keeleõppija seisukohalt.

Kirjandus

Block, Ellen L. 1992. See how they read: Comprehension monitoring of L1 and L2 readers. – *TESOL Quarterly* 26 (2), 319–343.

Borin, Lars & Prütz, Klas 2004. New wine in old skins? A corpus investigation of L1 syntactic transfer in learner language. – *Corpora and Language Learners* / Ed. by G. Aston & S. Berardini & D. Stewart. Amsterdam / Philadelphia: John Benjamin Publ. Co, 67–87.

Corder, Stephen Pit 1981. *Error Analysis and Interlanguage*. London: Oxford University Press.

EKK 2007 = Eesti keele käsiraamat. <http://www.eki.ee/books/ekk07/> (19.8.2009).

Eckman, Fred 2004. From phonemic differences to constraint rankings: Research on second language phonology. – *Studies in Second Language Acquisition* 26, 514–539.

Ellis, Rod & Barkhuizen, Gary 2005. *Analyzing learner language*. Oxford University Press.

Eslon, Pille 2006. Eesti vahekeele korpusest korrelatsioonigrammatikani. – Eesti Rakenduslingvistika Ühingu aastaraamat 2 / Toim. H. Metslang & M. Langemets. Tallinn: Eesti Keele Sihtasutus, 11–24.

Eslon, Pille 2007. Õppijakeelekorpused ja keeleõpe. – Tallinna Ülikooli keelekorpuste optimaalsus, töötlemine ja kasutamine. Tallinna Ülikooli eesti filoloogia osakonna toimetised 9 / Toim P. Eslon. Tallinn: TLÜ Kirjastus, 87–120.

Eslon, Pille & Matsak, Erika 2009. Eesti keele kasutusvariandid: korpusest tulenev käändevormide võrdlev analüüs. – Eesti Rakenduslingvistika Ühingu aastaraamat 5 / Toim. H. Metslang & M. Langemets & M.-M. Sepper, R. Argus. Tallinn: Eesti Keele Sihtasutus, 79–110.

Eslon, Pille & Metslang, Helena 2007. Õppijakeel ja eesti vahekeele korpus. – Eesti Rakenduslingvistika Ühingu aastaraamat 3 / Toim. H. Metslang & M. Langemets & M.-M. Sepper. Tallinn: Eesti Keele Sihtasutus, 99–116.

Granger, Sylviane 2002. A bird's eye view of learner corpus research. – Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching / Ed by S. Granger & J. Hung & S. Petch-Tyson. Amsterdam / Philadelphia: John Benjamin Publ. Co, 3–33.

Granger, Sylviane 2003. Error-tagged learner corpora and CALL: a promising synergy. – CALICO Journal 20(3), 465–480. <http://selene.lib.jyu.fi:8080/julpu/9513915425.pdf> (20.6.2009).

Hagen, Jon Erik 2009. Strukturfeil og kateogorfeil i skriftlig andrespråksperformanse. Den 9. nordiske andetsprogskonference. Helsingør, 11.–13. juni 2009.

Herranen, Tauno 1978. Errors made by university students in the use of the English article system. – Further contrastive papers. Jyväskylä Contrastive Studies 6 / Toim. K. Sajavaara & J. Lehtonen & R. Markkanen. Jyväskylä: University of Jyväskylä, 74–95.

Hirsjärvi, Sirkka & Hurme, Helena 2000. Tutkimushaastattelu. Tee-mahaastattelun teoria ja käytäntö. Helsinki: Yliopistopaino.

Hirsjärvi, Sirkka & Remes, Pirkko & Sajavaara, Paula 2005. Uuri ja kirjuta. Tallinn: Medicina.

Hufheisen, Britta 1991. English als erste und Deutsch als zweite Fremdsprache. Empirische Untersuchung zur fremdsprachlichen Interaction. Frankfurt/M etc: P. Lang.

James, Carl 1998. Errors in language learning and use. Exploring error analysis. London: Longman.

Jantunen, Jarmo 2007. Oppijansuomen piirteitä korpusvetoisesti. – VIRSU III. Suomalais-ugrilaisia kohdekieliä ja kontakteja. Studies in Languages 42 / Toim. P. Muikku-Werner & O. Kokko & H. Remes. Joensuu: Joensuun yliopisto, 69–84.

Kaivapalu, Annekatrin 2004. Kui sarnane on sarnane? Eesti ja soome mitmusevormide psühholingvistilisest reaalsusest. – VIRSU II. Suomi ja viro kohdekielinä. Oulun yliopiston suomen ja saamen kielen ja logopedian laitoksen julkaisu 24 / Toim. H. Sulkala & H. Laanekask. Oulu, 62–71.

Kaivapalu, Annekatrin 2005. Lähdekieli kielenoppimisen apuna. Jyväskylä Studies in Humanities 44. Jyväskylä: Jyväskylän yliopisto.

Kowal, Iwona 2009. Svenska som L3. Var döljer sig L1 och L2? Den 9. nordiske andetsprogsconference. Helsingør, 11.–13. juni 2009.

Kaivapalu, Annekatrin 2008. Lähtekeele mõju korpuspõhine uurimine. – Õppijakeele analüüs: võimalused, probleemid, vajadused. Tallinna Ülikooli eesti filoloogia osakonna toimetised 10 / Toim P. Eslon. Tallinn: TLÜ Kirjastus, 93–119.

Michiels, B. 1999. Die Rolle der Niederländischkenntnisse bei Französischsprachigen Lernern von Deutsch als L3: Eine empirische Untersuchung. – Zeitschrift für Interkulturellen Fremdsprachunterricht 3(3). http://www.spz.tu-darmstadt.de/projekt_ejournal/jg-03-3/beitrag/mich1.htm (18.6.2009).

Nesselhauf, Nadja 2003. The use of collocations by advanced learners of English and some implications for teaching. – Applied Linguistics 24: 2, 223–242.

- Nesselhauf, Nadja 2005. Collocations in learner corpus. Amsterdam: John Benjamins Publishing Company.
- Oldin, Terrence 1989. Language Transfer. Cross-linguistic influence in language learning. Cambridge, USA: Cambridge University Press.
- Rasier, Laurent & Hiligsmann, Philippe 2007. Prosodic transfer from L1 to L2. Theoretical and methodological issues. – *Nouveaux cahiers de linguistique française* 28, 41–66.
- Sajavaara, K. 1999a. Toisen kielen oppiminen. – Kielenoppimisen kysymyksiä. Soveltavan kielen tutkimuksen keskus / Toim. K. Sajavaara & A. Piirainen-Marsh. Jyväskylä: Jyväskylän yliopisto, 73–102.
- Sajavaara, K. 1999b. Kontrastiivinen kielen tutkimus ja virheanalyysi. – Kielenoppimisen kysymyksiä. Soveltavan kielen tutkimuksen keskus / Toim. K. Sajavaara & A. Piirainen-Marsh. Jyväskylä: Jyväskylän yliopisto, 103–128.
- Sajavaara, Kari 2006. Kontrastiivinen analyysi, transfer ja toisen kielen oppiminen. – *Lähivertailuja* 17. Jyväskylä Studies in Humanities 53 / Toim. A. Kaivapalu & K. Pruuli. Jyväskylä: Jyväskylän yliopisto, 9–26.
- Swain, Merrill & Lapkin, Sharon. 1995. Problems in output and the cognitive processes they generate: A step toward second language learning. – *Applied Linguistics* 16 (3), 371–391.
- Tarnanen, Mirja 2007. Testiaineistosta kielenoppijakorpuksiksi. – *Kieli oppimisessa – Language in Learning*. AFinLAN vuosikirja 65 / Toim. O.-P. Salo & T. Nikula & P. Kalaja. Jyväskylä: Jyväskylän yliopisto, 197–214.
- Tenford, Kari 2008. ASKeladden – a corpus-based approach to L1 transfer in Norwegian learner language. <http://www.askeladden.uib.no/summary.page> (19.8.2009).

LISA 1.

Test eesti keele sõnajärjemallide loomulikkuse uurimiseks

Sõnajärje loomulikkuse uuring

Eesti keele sõnajärje loetakse võrreldes mõne teise keelega suhteliselt vabaks. Siiski tunduvad mõned sõnajärjemallid nii emakeelsele keelekasutajale kui ka keeleõppijale loomulikumad kui teised. Käesoleva uuringu eesmärgiks on uurida eesti keele teatud sõnajärjemallide loomulikkust. Seepärast ärge mõelge, kas Teie vastused on õigekeelsuse seisukohast õiged või valed, vaid vastake nii, nagu Teie enda keeletunne Teile ütleb. Täname Teid väga uuringule kaasaaitamise eest!

I. Taustandmed vastaja kohta

Vanus

.....

Sugu

.....

Emakeel

.....

Kodune keel

.....

Kus ja kui kaua eesti keelt õppinud

.....

.....

II. Sõnajärjemallide loomulikkuse taj

Palun hinnake järgnevaid lauseid oma keeletundest lähtuvalt ja tehke rist sellesse lahtrisse, mis vastab kõige täpsemini Teie keeletundele.

	Lause	Täiesti loomulik	Pigem loomulik	Pigem ebaloomulik	Täiesti ebaloomulik
1	<i>Nad õpivad terve nädala matemaatikat.</i>				
2	<i>Teemat muudame me vabalt.</i>				
3	<i>Põhiteema oli elu planeedil Marss tõenäosus.</i>				
4	<i>Siidist õmmeldud kleidi kinkis ema tütrele.</i>				
5	<i>On vaja õpilastele raamatuid.</i>				
6	<i>Järelikult su teod paistavad praegu välja kentsakad.</i>				
7	<i>Teatud puudused on isegi pangas töötamisel.</i>				
8	<i>See on ainuke ülikool, mis üldse välja näeb nagu ülikool.</i>				
9	<i>Ta kirjutas lõpukirjandi paari aasta eest Tallinnas.</i>				
10	<i>Pärast seda ma kolisin Peterburi.</i>				
11	<i>Nad põnevalt jutustasid oma õppeainet.</i>				

12	<i>Ma saan uue võimaluse alles pärast vaheaega.</i>				
13	<i>Aasta jooksul nad kasvatavad põrsast.</i>				
14	<i>Vivian teeb kodutöid sageli rutakalt.</i>				
15	<i>Vestlusteema oli eksperimendi olulisus uuringus CEFR.</i>				
16	<i>Mind aitab see eriti tulevases töös.</i>				
17	<i>Esimest korda toimunud aastal 1869 laulupidu muutus ajaloolise tähtsusega sündmuseks.</i>				
18	<i>Zaura kinkis mulle siili tehtud paberist.</i>				
19	<i>Lastele oli tarvis vanemaid.</i>				
20	<i>Elamisel õpilaskodus on ka oma miinused.</i>				
21	<i>Mõne aasta pärast alustavad nad magistriõpinguid välismaal.</i>				
22	<i>Mõtles poiss pingsalt oma isast.</i>				
23	<i>Pärast kooli ma sain hea keeleoskuse.</i>				
24	<i>Aastal 2003 esimest korda ilmunud monograafia sai üldise tunnustuse osaliseks.</i>				

25	<i>See on ainus küsimus, mis tegelikult kerkib üles nagu probleem.</i>				
26	<i>Pool päeva lugesime me kriminulle.</i>				
27	<i>Ta alati riietub pidulikult.</i>				
28	<i>Olemise võimalikkus allveelaeval Nautilus on probleem.</i>				
29	<i>Mu elu praegu näeb välja järgmiselt.</i>				
30	<i>Nad lendasid mõne aja pärast Indiasse.</i>				
31	<i>Tehnikumi järel sai ta hea töökoha.</i>				
32	<i>Poiss kinkis tüdrukule puust meisterdatud kujukese.</i>				
33	<i>Energiat on vaja inimestele.</i>				
34	<i>See on rikkalik kingitus, mis kuldab mind täiesti üle nagu miljonäri.</i>				
35	<i>See väga aitas mind järgmises koolituses.</i>				
36	<i>Ta rääkis huvitavalt lossi ajaloo.</i>				
37	<i>Kutsekoolis õppimisel on samuti oma head küljed.</i>				
38	<i>Mõtlesid tüdrukud pärast trenni minna raamatukokku.</i>				
39	<i>Kergelt vahetavad nad värvi.</i>				

40	<i>Õpetajad kirjutasid ministee- riumi pärast seaduse vastu- võtmist.</i>				
41	<i>Aastal 2002 toimunud konverents muutus esimest korda tähelepanuväärseks sündmuseks.</i>				
42	<i>Edu ainult saatis teda kõikjal.</i>				
43	<i>Ülikooli lõpetamise järel sõitis Henrik Londonisse.</i>				
44	<i>Edasises elus teid juhib vaid südametunnistus.</i>				
45	<i>Nüüd läheb meie olukord ülemustele tõsiselt korda.</i>				
46	<i>Vanad raamatud saame üle anda pärast aastavahetust.</i>				
47	<i>Rahva tahtest kirjutas innustavalt ajaleht.</i>				
48	<i>Kolm aastat tagasi ma lõpetasin kümnenda klassi Narva Humanitaargümnaa- siumis.</i>				
49	<i>Luuletust kirjutas ta päev läbi.</i>				
50	<i>Ajaloolise tähtsusega sündmuseks kujunes esimest korda aastal 1900 läbi viidud rahvaloendus.</i>				

On some methods of contributing to corpus-based analysis of learner language

Annekatriin Kaivapalu

Summary

Most of corpus-based studies of learner language have been completed in the framework of error analysis and have focused mainly on the learners' products, while learners' production processes have been less researched. This article deals with some problems and limitations of corpus-based error analysis of learner language. The problems of classification and interpretation as well as the psycholinguistic reality of errors are discussed on the bases of word order errors of Russian learners of Estonian. Some complementary possibilities enabling the researcher to investigate learners' production process are proposed: introspective methods such as psycholinguistic perceptual tests, *thinking aloud*-method, retrospective interviews and language technology software such as *Script-Log*-program.

The article also presents some results of a preliminary study on psycholinguistic reality of some word order patterns of Estonian from the point of view of Russian learners of Estonian as well as from that of native Estonian speakers. The following word order patterns are discussed: 1) adverbial-subject-predicate-object, 2) adverbial-predicate-subject-object, 3) subject-object-adverbial-object, 4) object-predicate-subject-adverbial. The study concludes that the main differences in perception of the Estonian native speakers and the Russian learners of Estonian concerned V2-rule: the pattern adverbial-predicate-subject-object was perceived as natural by most of

Estonian native speakers and the pattern adverbial-subject-predicate-object was mostly perceived as unnatural. For most of the Russian learners the pattern adverbial-subject-predicate-object was quite natural or nearly natural. The pattern adverbial-predicate-subject-object was perceived as quite natural or nearly natural by a half of the Russian learners of Estonian.

Keywords: corpus-based error analysis, psycholinguistic reality of errors, word order patterns of Estonian, Estonian native speaker, Russian learner of Estonian, introspective methods

Autor

PhD Annekatrin Kaivapalu, Tallinna Ülikooli eesti keele ja kultuuri instituudi dotsent, riikliku „Eesti keele keeletehnoloogiline tugi (2006–2010)” projekti „VAKO – Eesti vahekeele korpuse keeletarkvarana ja keeletehnoloogilise ressursi arendamine” põhitäitja, kaivapa@tlu.ee