

ÕPPIJAKEELE KORPUSPÕHISEST ANALÜÜSIST

Helena Metslang

Ülevaade

Võõrkeele oskuse arengut suunavad mitmed faktorid, mis erinevad nii mõju iseloomult kui ulatuselt: sihtkeele ja emakeele sarnased ja erinevad struktuurid, siht- ja emakeele keelendite kasutussagedused; sõnade, sõnaühendite ja grammatika esitus õpikutes, keeleõpingute kestus, õppija mentaalsed protsessid sihtkeele kasutamisel jm. Millist rolli iga tegur õppijakeeles mängib, seda võib uurida õppijakeele kontrastiivanalüüsi abil. Niisuguste uuringute üheks olulisimaks ressursiks on saanud korpused, sest nad võimaldavad leida õppijakeeles nähtusi (sh kvantitatiivseid), mis jäävad traditsiooniliste meetoditega märkamatuks, ning empiiriliselt kontrollida seni väiksemahulisemal materjalil tehtud uuringute tulemusi. Käesolev artikkel tutvustab õppijakeele korpuspõhise vea- ja kontrastiivanalüüsi võimalusi¹.

Võtmesõnad: korpuspõhine veaanalüüs, korpuspõhine kontrastiivanalüüs.

¹ Tööd on toetanud riikliku programmi “Eesti keel ja rahvuslik mälu” (2004–2008) grant R05/01 “Koodivahetuse, vahe- ja lastekeele korpuste töötlemine ja haldamine” (2005–2008).

Korpuspõhine vea- ja kontrastiivanalüüs

Õppijakeele struktuuri ja olemust on võimalik uurida kui nähtust iseeneses (eelkõige veaanalüüsi abil) või kontrastiivselt. Sylviane Grangeri järgi tegeleb veaanalüüs märgenduse, veaotsingu ja -analüüsiga. Arvutipõhine veaanalüüs erineb varasemate kümnendite veaanalüüsist tänu suuremale standardiseeritusele ning vigaste kontekstide kõrvutamisele õigete kasutus-
tega (Granger 2002: 11–16).

Korpuspõhises veaanalüüsis kasutatakse peamiselt kaht meetodit:

- a) märgendamata korpuse läbivaatamine standardse tekstiotsingu tarkvara abil;
- b) veamärgendussüsteemi loomine ja õppijakeele korpuse märgendamine kas kõigi vigade või uuritavate veatüüpide osas.

Esimene võimalus on kiire ja teine töömahukas, kuid annab samas väga laiad võimalused edasiseks analüüsiks (Granger 2002: 14, vt ka Eslon & Metslang 2007: 106–112; Kitsnik 2007).

Veeanalüüs on saanud sageli negatiivsete reaktsioonide osaliseks: selles on nähtud tagasipöördumist õppijakeele analüüsi vanema lähenemise juurde, mis pidas vigu täielikult negatiivseks nähtuseks (Granger 2002: 14). Aluse selleks on andnud vähemalt viis põhilist puudust, millele on viidanud ka Lars Borin ja Klas Prütz: 1) veaanalüüs tugineb liiga heterogeensetel lähteandmetel, 2) piirdub ainult sellega, mida õppijad teha ei suuda, 3) käsitleb teise keele omandamist staatilisena, 4) ei suuda käsitleda vältimist jm nähtusi, 5) veakategooriad on laialivalguvad (Borin & Prütz 2004: 68). Samas näitavad Borin ja Prütz, et peaaegu kõigist neist probleemidest aitab üle saada lähte- ja sihtkeele kontrastiivanalüüs, Granger aga selgitab veaanalüüsi positiivset panust õppijakeele uurimisse: see suund aitab paremini mõista õppijakeele arengustaadiume ning omab

olulisi pedagoogilisi rakendusi. Kuna õpetajad ning õppematerjalide loojad soovivad pakkuda õppijale võimalikult kasutoovat keelesisendit, siis vajavad nad jätkuvalt infot selle kohta, mis etapiks ja millised keelendid ning keelestruktuurid peaksid õppijad olema omandanud.

Eesti õppijakeele arvutipõhiseks veaanalüüsiks annab võimaluse Tallinna Ülikooli üld- ja rakenduslingvistika õppetooli *Eesti vahekeele korpus* (EVKK)², mis sisaldab vene emakeelega õppijate eesti sihtkeele kirjalikke tekste.

Õppijakeele kontrastiivanalüüs (CIA ehk *contrastive interlanguage analysis*), võrdleb kvalitatiivselt ja kvantitatiivselt õppijakeelt siht- ja lähtekeelega ja / või erinevate õppijakeelte variantidega (Granger 1996, cit. Borin ja Prütz 2004: 69).

Õppijakeele kontrastiivanalüüsi valdkonda kuulub nelja tüüpi uuringuid.

Esimene – õppijakeele ning sihtkeele kui emakeele kõnelejate korpuse kontrastiivanalüüs – tegeleb nende kirjalike ja suuliste õppijakeelte joonte uurimise ja võrdlemisega, mis eristavad seda sihtkeelt emakeelena kõnelejate keelest. Võrdlusmaterjali valikul on seejuures oluline kontrollkorpuse meedium (suuline, kirjalik), sotsiolekt, formaalsuse aste ja žanr ning õppijakeele korpuse informantide keeleoskustase. Inglise sihtkeele uuringuid toetavad näiteks tudengite ja õpilaste esseede korpused LOCNESS³ ja USE⁴. Kasutusele on võetud meetodeid, mis pärinevad paralleelkorpustega tegelevast lingvistikast (Borin ja Prütz 2004: 69).

² <http://evkk.tlu.ee>, 2.10.2007

³ Louvain Corpus of Native English Student Essays, <http://www.fltr.ucl.ac.be/fltr/germ/etan/cecl/Cecl-Projects/Icle/locness1.htm>, 30.09.2007

⁴ Uppsala Student English Corpus, <http://www.engelska.uu.se/use.html>, 30.09.2007

Teine meetod integreerib õppijakeele analüüsi traditsioonilise kontrastiivanalüüsiga (õppija emakeele ehk K1 korpuse võrdlus sihtkeele ehk K2 kirjakeele korpusega). Traditsiooniline kontrastiivanalüüs (CA ehk *contrastive analysis*) toetab nende korpuste võrdlevanalüüsi just tulemuste interpreteerimisvõimaluste lisamise näol. Olles tuvastanud erinevuse siht- ja õppijakeele vahel, on korpusepõhine kontrastiivanalüüs abiks erinevuse põhjuste selgitamisel: näiteks, kas tegu võib olla ülekan- dega emakeelest, vältimisstrateegiaga või universaalse õppija- keele arengu nähtusega. Kontrastiivne õppijakeele analüüs sel- gitab K1, K2 ja õppijakeelevaheliste seoste iseloomu. Korpuste põhjal on tehtud palju sagedusuuringuid, kus õppijakeele viga- dele, üle- ja alakasutustele otsitakse seletust just kontrastiivana- lüüsi abil. Siia lisandub ka õppijakeele korpuste uurimine koos tõlke- ja paralleelkorpustega, mis võimaldab täpsemat võrd- lust tänu joondusele lause või lõigu tasandil. Bengt Altenberg näitab, et paralleelkorpuse uurimise eeliseks on ka kohmakate, kuid mittevigaste kasutuste süstemaatilise uurimise võimalus (Altenberg 2002: 39).

Kolmas meetod on õppijakeeltevaheline võrdlus. Granger nimetab erinevate õppijakeelte võrdleva analüüsi eesmärgina õppijakeele olemuse mõistmist: eri emakeeltega õppijate and- mete võrdlemisel püütakse eristada emakeelespetsiifilisi õppija- keele jooni ning õppijakeele arengu universaale – jooni, mis on ühised erinevatel õppijagrupidel (Granger 2002: 13). Teise ja kolmanda meetodi kombineerimine ehk kompleksvõrdlus aitab eristada intralingvaalseid nähtusi (nt liigne üldistamine, nor- mist kõrvalekalduv lihtsustamine, registri ülekanne) interling- vaalsetest (ülekanne ehk transfer). Kolmanda meetodi alla kuu- lub ka eri tasemel keeleõppijate korpuste võrdlus, mis on alu- seks õppijakeele kui protsessi diakroonilisele uurimisele ning informandikesksetele pikaajalistele uurimustele (Borin ja Prütz 2004: 69–70). Isiku keelearengu pikaajaline uurimine on eriti vaja-

lik näiteks stabiliseerumistendentside-kivinemisnähtuste ning positiivsete muutuste väljaselgitamiseks (Pool & Vaimann 2006: 214–215).

Teise keele omandamise ja kasutamise uurimisel tuleks kindlasti tähelepanu pöörata ka sisendkeelele (*input*) ehk kogu õpitava keele ainesele, millega keeleõppija kokku puutub (Pool 2007: 17). **Neljas õppijakeele kontrastiivse uurimise valdkond ongi õpikutekstide korpuse võrdlus õppijakeele ja / või sihtkeelega**, et tuvastada õpikutest pärineva sisendkeele mõju sihtkeelele. Tunnis on üheks suurimaks õppijakeele mõjutajaks õpikutekstide keel (*school English* – nn kooliinglise keel, mida seatakse õpilasele eeskujuks). Õpikutekstide korpuse põhjal saab arvutada lekseemide, fraaside ja mitmesõnaliste üksuste esinemissagedusi, analüüsida kontekstinähtusi, võrrelda õpikutekstides ja sihtkeelt emakeelena kõnelejate tekstides kollokatsioone, polüsemseid sõnu ja struktuure. Õpikutekorpused aitab ka jälile saada leksika ning grammatika esituse järgnevusele, annab uut ainet arutelule autentse keelesisendi võimalikkuse ja vajaliku määra üle keeleõpetuses. (Römer 2004: 152)

Järgnevalt lähemalt kahest suunast kontrastiivanalüüsis: õppijakeele ja traditsioonilise kontrastiivanalüüsi integreerimine ning õpikutekstide korpuse võrdlus õppijakeele ja / või sihtkeelega.

Yukio Tono on jaapani õpilaste ja tudengite inglise õppijakeele uurimises täpsustanud, mis põhjustab kirjalikes tekstides tüüpverbide argumentstruktuuride üle- ja alakasutust ning muid vigu. Ta vaatleb ka, mis on kaasa aidanud grammatikanähtuste omandamisele. K1, K2 ja õppijakeelekorpuste statistiline analüüs näitab, et mõjurid on erinevad lausa ühe grammatikavaldkonna – verbi argumentstruktuuri – ulatuses. (Tono 2002: 45–65)

Tono analüüsib Jaapani inglise keele kui võõrkeele korpuse JEFLL⁵ kolme allkorpust: õppijakeelekorpus sisaldas eri tasemel keeleõppijate loomingulisi kirjutisi (500 000 sõnet), sihtkeelekorpuses on õpikutekstid (650 000 sõnet) ning K1 korpuses – ajalehetekstid (11 milj sõnet). Tono võrdles neis allkorpustes loglinearsel meetodil⁶ BNC (*British National Corpus*)⁷ kümne sagedasema verbi (*bring, buy, eat, get, go, like, make, take, think, want*) argumentstruktuuride ja nende vaheldustüüpide⁸ kasutust. Siinkohal mõningad uurimistulemused, nii üldisemat laadi kui argumentstruktuuritüübiti eraldi.

Tono K2 ja K1 korpuste võrdlus õppijakeelega näitab, et nende verbimustrite vale kasutust põhjustavad faktorid, mis peituvad sihtkeele verbide leksikaalsemantikas ning kasutusageduse sarnasustes ja erinevustes. Keeleõpingute kestus ei mõjuta oluliselt nende verbide argumentstruktuuride omandamist. Allkorpustest leitud argumendimustrite võrdlev analüüs näitas, et verbide kasutussagedus õpikutekstides mõjutas nende kasutussagedust õppijate poolt, kuid mitte oluliselt nende kasutamise õigsust.

Nimetatud faktorid mõjutavad sihtkeele omandamist erinevalt lausa argumentstruktuuritüübiti. Keeleõpingute kestus mõjutas K2s enim benefaktiivseid (verbid *buy, get, make, take*) ja kvantifitseerimist sisaldavaid argumentstruktuuritüüpe (ver-

⁵ <http://leo.meikai.ac.jp/~tono/jefll.html>, 2.10.2007

⁶ „Loglineaarne analüüs on Eestis seni üsna vähe kasutatud, kuid sotsiaalteadustes suurepäraselt sobiv andmeanalüüsi meetod. Loglineaarne analüüs annab võimaluse mitmemõõtmeliselt uurida tunnustevaheliste seoste struktuuri.“ (Pals 1998: 17)

⁷ <http://www.natcorp.ox.ac.uk/>, 30.09.2007

⁸ Käsitlus põhineb B. Levini semantilis-süntaktilisel verbiklassifikatsioonil (Levin 1993). Levin eristab verbi(klassi) argumentstruktuuri ning verbi(klassi) argumentstruktuuritüüpi. Verbiklassi argumentstruktuuritüüpi võib kuuluda enam kui üks argumentstruktuur (nt *Ta käis mööda teed* (PP) ja *Ta käis pika tee* (NP) *maha*).

bid *buy, get, make*) – vastavalt näited (1) ja (2). Emakeele mõju oli enam nähtav daativsete (*bring, make, take, want, think*) ning lokatiivsete argumentstruktuuritüüpide (*take, go*) kasutamisel – vastavalt näited (3) ja (4). Täpsemalt töid daativsed ja lokatiivsed argumentstruktuuritüübid esile seose lähte- ja sihtkeele grammatikanähtuse struktuuri ja sageduse sarnasuse / erinevuse vahel. Õpikutekstide keel mõjutas enim näiteks *blame*-argumentstruktuuritüpe (*like, want*) – vt näide (5).

- (1) *"She opened the door for Tom"* – Ta avas **Tomile** ukse.
- (2) *You can make 1000 dollars now* – Sa võid praegu teenida **1000 dollarit**.
- (3) *Maria gave her the book* – Maria andis **talle** raamatu.
- (4) *Maria will go to park* – Maria läheb **parki**.
- (5) *I like Maria for a variety of reasons* – Mulle meeldib Maria **erinevatel põhjustel**. Vrd: *I blame Maria for a variety of reasons* – Ma süüdistan Mariat erinevatel põhjustel.

Saksa emakeelega õppijate inglise sihtkeele probleemsetest aladest on kirjutanud Ute Römer. Tema arvates peab hea keeleõpik olema õppurile jõukohane ning peegeldama võimalikult autentset keelt nii sõnade ja sõnaühendite, grammatiliste konstruktsioonide ja kontekstinähtuste kasutussageduse kui keelekasutussituatsioonide osas. Õpikutekstide kõrge autentsusaste aitab emakeelekõnelejate tekste ladusamalt mõista, annab suurema enesekindluse õpitava keelega kokku puutudes, loomulikuma väljendusoskuse ning keelendite kasutamise neile omases kontekstis. Püüdlust õpikutekstide autentsuse poole on ka kritiseeritud, sest tunnisituatsioon on paratamatult kunstlik ning seda ei saa muuta. Õpikulaused on küll välja mõeldud, kuid ikkagi emakeelekõnelejate poolt ning kindla eesmärgiga (Römer 2004: 151–155). Samas võib nende autentsus õpikutekste liialt keerukaks teha. Näitena toob Römer saksa emakeelega õppijatele

mõeldud inglise keele õpikutest madala kvaliteediga sisendkeele, dialoogi progressiivse *ing*-vormi õpetamiseks: „/-/ are they playing?“ – „No, they aren't, Ron. They are packing.“ Vrd näidetega BNC-korpuse suulisest osast: *What's happening, does anybody know?* (Kas keegi teab, mis toimub?) *Are you listening to me?* (Kas sa kuulad mind?)

Et vältida kommunikatiivselt väherelevantseid näitelauseid ja dialooge, soovib Rõmer võrrelda seni avaldatud keeleõpikute tekste loomuliku keelekasutusega ning kasutada grammatika esitamiseks näiteid inglise kirjakeele korpusest. Ta võrdleb saksa emakeelega õppuritele probleeme tekitavate konditsionaalsete *if*-lausete esitust BNC-korpuse suulisest osas ja inglise keele kui võõrkeele õpikutekstide korpuses GEFL TC (*German English as a Foreign Language Textbook Corpus*), mis koosneb kahest õpikusarjast kogumahuga 108 000 sõnet. Uuritavateks joonteks olid

- konditsionaallause osalause järjekord (kas *if*-kõrvallause on esimesel või teisel kohal);
- ajavormide järjestus (nt kõrvallausel lihtminevik + pealause olevikuline modaalverb + infinitiiv: *If I studied, I would pass the exams* – *Kui ma õpiksin, saaksin eksamitel läbi*);
- kollokatsioonid *kui*-sõnaga, mille pikkus on 1–2 kohta *kui*-sõnast vasakule ja paremale.

Analüüsist ilmnes, et õpikutekstide konditsionaallausetes oli üle 50% ajavormide kombinatsioonidest esindamata (BNC-s leidis 17 erinevat konditsionaallause ajavormide kombinatsiooni, õpikutekstide korpuses vaid 8). Kolmel ajavormide kombinatsioonil oli õpikutekstides mitmekordne alaeesindatus ning kolmel suur üleesindatus. *kui*-sõnast paremal kasutavad emakeelekõnelejad enim pronomeni *you* (*sina*), samas kui õpikutekstides olid tunduvalt sagedamad kollokatsioonid *if I* – *kui ma*, *if we* – *kui me*, *if they* – *kui nad*, *if (s)he* – *kui ta*. Ka konditsionaallause osalause järjestuses oli tuntav erinevus.

Inglise keele õpikutekstide korpuse analüüs näitas, et keeleisend, mida õppijad õpiku kaudu saavad, on autentsest keelest vähemalt konditsionaallausete osas üsna erinev. Siiski järelgab Römer, et ei ole mõtet õppijat vaevata rohkelt varieeruvate suulise kõne ajavormivahelduste tutvustamisega. Küll aga tuleks õppekirjanduses vältida ebatüüpiliste konstruktsioonide õpetamist ning selgitada, et lisaks paarile-kolmele tüüpreeglile leidub keeles ka teisi võimalusi. (Römer 2004: 156–160)

Vea- ja kontrastiivanalüüs eesti õppijakeele korpuspõhises uuringus

Eesti õppijakeele korpuspõhises veaanalüüsis saab tugineda *Eesti vahekeele korpuse* ressurssidele. On võimalik leida nii korrektseid kui ebakorrektsed sõna- ja vormikasutusi nii kontekstis kui sõnavormide sagedustabeli kujul (veaanalüüs märgendamata tekstikogu peal). Teine võimalus on esitada päringuid EVKK märgendatud osale. Märgendus põhineb veaklassifikatsioonil, mis hõlmab vigu grafeemist morfeemi, lekseemi ja süntaksini. Iga uurija võib korpuse veaanalüüsi vastavalt oma uurimisteele süvendada, luues EVKKsse isikliku märgenditepuu.

Eesti õppijakeele korpuspõhises kontrastiivaanalüüsis saab õppijakeelt võrrelda sihtkeelega. Materjali selleks annab *Eesti vahekeele korpuse* ja *Tartu ülikooli eesti kirjakeele korpuse*⁹ või *Eesti Keele Instituudi korpuse*¹⁰ andmete võrdlemine. Olemasolevate kirjakeele korpuste materjal ei ole võrreldav EVKK tekstidega stiili, žanrite ja keerukusastme erinevuse tõttu, kuid esmaseid järeldusi võimaldab see võrdlus kindlasti teha.

⁹ <http://www.cl.ut.ee/korpused/baaskorpus/>, 2.10.2007

¹⁰ <http://www.eki.ee/corpus/>, 2.10.2007

Teine võimalus korpuspõhise kontrastiivanalüüsi rakendamiseks on seotud **õppijakeele analüüsi ja traditsioonilise kontrastiivanalüüsi integreerimisega** (vt ülalkirjeldatud *teine meetod*). See võiks olla teostatav *Eesti vahekeele korpuse* materjalide uurimisega võrdluses *Multext-East projekti*¹¹ paralleelstatud tõlkecorpusega või eesti ja vene ükskeelsete kirjakeelekorpuste põhjal. Ideaalis annaks kõige usaldusväärsemaid uurimistulemusi eesti ja vene emakeelekõnelejate kontrollkorpuste olemasolu (nt eesti ja vene koolide lõpetajate esseed ja eksami-kirjandid).

Erinevate õppijakeelevariantide võrdlust saab ka EVKK põhjal läbi viia, kuna korpus sisaldab tekste, millele märgendajad on vastavalt Euroopa keeleõppe raamdokumendis kirjeldatud keeletasemetele ja omavahelise arutelu tulemusel lisanud määrangu alg- (A), kesk- (B) või kõrgtase (C). Mitmete õppijakeele korpuste ülesehitusel on järgitud osaskuste tasakaalustatuse põhimõtet. EVKK-sse kogutud kirjalikud tekstid peegeldavad õppurite retseptiivseid oskusi (nt ümberjutustused ja harjutused). Keeleõppijate kommunikatiivsete oskuste uurimiseks jääb aga kirjalikest loomingulist laadi tekstidest (esseed) väheks. Seetõttu on EVKKd vaja täiendada suulise eesti õppijakeele allkorpusega.

Õpikutekstide korpuse võrdlus õppijakeele ja / või sihtkeelega pole hetkel reaalne, kuna sobiv ressurss puudub.

Suurte andmehulkade kontrastiivanalüüs aitaks aga leida vastuseid küsimustele, milliseid grammatika ja sõnavara kasutamise oskusi saab õpikutekstidega paremini mõjutada, milliste oskuste omandamine sõltub eelkõige keeleõpingute kestusest, milliste oskuste areng on paratamatult seotud õppiija emakeele ning eesti keele vaheliste sarnasuste ja erinevustega,

¹¹ <http://nl.ijs.si/ME/>, 2.10.2007

millised leksikaalsed ja grammatilised struktuurid omandatakse kõigepealt, millised – etapiti, kuivõrd loomuliku keele lähedane on eesti keele esitus olemasolevates õpikutes jne.

Kokkuvõttes on *Eesti vahekeele korpuse* näol olemas mahukas keeleressurss, mille alusel saab arendada teadmisi eesti õppijakeelest, panustada uuringute kaudu teise keele omandamise teooriasse, täiustada õppematerjale jm keeleõppevahendeid.

Kirjandus

- Altenberg, Bengt 2002. Using bilingual corpus evidence in 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's, 37–54.
- Aston, Guy & Berardini, Silvia & Stewart, Dominic (Eds.) 2004. *Corpora and Language Learners*. Amsterdam / Philadelphia: John Benjamin's.
- 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's, 67–87.
- Elson, Pille & Metslang, Helena 2007. *Õppijakeel ja eesti vahekeele korpuse*. – *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's, 3–33.
- Pals, Heili 1998. Koosmõju tõlgendusprobleeme loglineaarses analüüsis ja dispersioonanalüüsis. Bakalaureusetöö. Tartu Ülikool, <http://www.psych.ut.ee/esta/TARGET.EST/UURIMUS/publikatsioon/logdisper.html>, 2.10.2007.
- Kitsnik, Mare 2007. Õppijakeele uurimise ja arendamise võimalusi eesti vahekeele korpuse põhjal (eituse väljendamise näitel). Magistritöö. Tallinna Ülikool, http://evkk.tlu.ee/pdfs/mag_too_mare.pdf, 30.09.2007.
- Pool, Raili 2007. Eesti keele teise keelena omandamise seaduspärasusi täis- ja osasihitise näitel. Doktoriväitekiri. Tartu Ülikool.
- Pool, Raili & Vaimann, Elle 2006. Kivinemisnähtusi vene üliõpilaste kirjalikus eesti keeles. – Eesti Rakenduslingvistika Ühingu aastaraamat 2 / Toim. H. Metslang, M. Langemets. Tallinn: Eesti Keele Sihtasutus, 203–217.
- Römer, Ute 2004. Comparing real and ideal learner language input: The use of an EFL textbook corpus in corpus linguistics and language teaching. – *Corpora and Language Learners* / Ed. by G. Aston, S. Berardini, D. Stewart. Amsterdam / Philadelphia: John Benjamin's, 151–168.
- Tono, Yukio 2004. Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb subcategorization patterns by Japanese learners of English. – *Corpora and Language Learners* / Ed. by G. Aston, S. Berardini, D. Stewart. Amsterdam / Philadelphia: John Benjamin's, 45–66.

Kaudviited

Granger, Sylviane 1996. From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. – Languages in Contact / Ed. by K. Aijmer, B. Altenberg, M. Johansson. Lund University Press, 37-51.

Levin, Beth 1993. English verb classes and alternations: a preliminary investigation. Chicago: The University of Chicago Press.

Learner language corpus-based analysis

Summary

The methods of analyzing learner language are usually divided into two: contrastive interlanguage analysis and error analysis. Contrastive Interlanguage Analysis (CIA) compares interlanguage to the properties of source and target language or with other variants of interlanguage. CIA helps to identify the role of different factors that impact the development of IL: similar and different structures and frequencies in source and target language, the representation of lexemes, multi-word items and grammar in textbooks, duration of studies, students' mental processes, etc that are the cause for correct and native-like or incorrect language use. Differently from CIA, error analysis is focused to studying just learner language itself: corpus-based error analysis deals with error annotation and retrieval and with analyzing the errors. The article introduces the methods, possibilities and some drawbacks of different forms of CIA and error analysis and looks at what kind of research can be done in the field of Estonian interlanguage analysis (source language: Russian, target language: Estonian) with the tools available.