

KEELEKORPUSED JA VÕORKEELEÕPE

MARE KITSNIK

Ülevaade. Eesti keele kui võorkeele õpe on alles noor ja arenev ala, seetõttu pole ka õppesüsteem nii hästi välja arendatud kui sooviksime. Et muuta eesti keele õpe mitte-eestlastele arusaadavamaks, efektiivsemaks ja atraktiivsemaks, tuleb leida vastused mitmele küsimusele. Kuidas esitada keelesüsteem võorkeeleõppijale mõistetavas ja õppimisprotsessi toetavas vormis? Kuidas koostada tõhusamaid ja õppijasõbralikumaid eesti keele kui võorkeele õpikuid, grammatikaid ja sõnastikke? Milline õppemetoodika vastab kõige paremini õppija vajadustele ja keele omandamise seaduspärasustele? Et leida neile küsimustele vastused, on eelkõige vaja läbi viia vastavad uuringud ja luua eesti keele pedagoogiline kirjeldus. Teiste võorkeelte ja eesti keele kui emakeele uuringud võivad olla siinkohal inspireerijaks ja meetoodiliseks eeskujuks, kuid nende tulemusi pole võimalik otse üle võtta.

Siiani on eesti keelt kui võorkeelt uuritud lünklikult: on kirjutatud mitmeid bakalaureuse- ja magistritöid ning avaldatud lühiaurimuste tulemusi, kuid eesti keel tervikuna on võorkeeleõppe seisukohalt kirjeldamata.

Kuigi näiteks inglise keele kui võorkeele õpe on võrreldes eesti keelega vägagi arenenud, toimub ka selles vallas pidev areng – lingvistide, psühholoogide, pedagoogide jt uuemate uurimuste põhjal täpsustatakse või muudetakse õppematerjale, õppekavu ja -metoodikat. Viimasel ajal on üheks olulisemaks keeleõppe arengu hoovaks saanud keelekorpused, mida tänu infotehnoloogia vahenditele on võimalik aina sügavamalt uurida ja paremini kasutada – ka võorkeeleõppes. Artiklis tutvustataksegi keelekorpuste osa võorkeeleõppe arendamisel ning vaadeldakse, kuidas keelekorpustest ka eesti keele kui võorkeele õppe arengule võiks tulu tõusta.*

Võtmesõnad: emakeelekorpus, õppijakorpus ehk vahekeelekorpus, sõnastikud, grammatika, õppemetoodika, andmepõhine õpe, eesti keel võorkeelena

* Artikkel on valminud riikliku sihtprogrammi "Eesti keel ja rahvuslik mälu (2004–2008)" grandi R 05/01 raames.

1. Keelekorpuste osa keeleuurimisel

Elektroonilised keeleandmekogud (korpused) hakkasid tekkima juba 1960-ndatel aastatel, aga viimase paarikümne aastaga on nende mahud ja kasutusvõimalused tunduvalt suurenenud. Arvutite mõju keeleuurimisele on võrreldud teleskoobi kasutuselevõtu mõjuga astronoomiale (Ranalli 2003: 2), sest nüüd on osutunud võimalikuks saavutada tulemusi, mis varasemate uurimismeetoditega jäid kättesaamatuks ja jõuda sel viisil lähemale keele tõelisele olemusele. Tänapäevane keelekorpused on polüfunktsionaalne elektrooniline tekstikogu, mis on koostatud kindlate reeglite järgi ja kuhu tekstid on valitud nii, et nad annaksid keelest tõepärase pildi. Suuremate korpuste maht ulatub miljardite sõnadeni. Korpused on muutunud asendamatuks nii keeleteaduses, leksikograafias kui ka automaatsete keeletöötlusvahendite väljatöötamisel. On tekkinud rakenduslingvistika haru – korpuslingvistika, mis pakub uut uurimismeetodit. Korpuslingvistika on teiste meetodite hindamatu abilise uuritava keeleainese (üksiksõnad, kollokatsioonid, grammatilised struktuurid jm) esinemissageduse määramisel autentses keeles. Korpusuuringud ongi juba viinud täiesti uue arusaamani keelest: sõnavara ja grammatika ei ole kaks eraldiseisvat, vaid omavahel lahutamatult seotud tekstiloome osa.

1990-ndate aastate algusest on hakatud huvi tundma ka korpuste kasutamise vastu võõrkeeleõppes. Praeguseks pole korpused küll võõrkeeleõpet revolutsiooniliselt muutnud, nende kasutamisevõimaluste üle käib aga elav arutelu. On ilmunud arvukalt publikatsioone autoritelt, nagu Ronald Carter, Douglas Biber, Tim Johns, John Sinclair, Antoinette Renouf, Henry Widdowson, Dave Willis, Chris Tribble, Guy Aston, Sylviane Granger, Melinda Tan, Nadja Nesselhauf, Ute Römer, Fanny Meunier. Teemaatika on aktuaalne erialaajakirjades (*Language Learning & Technology*, *International Journal of Corpus Linguistics*), toimuvad rahvusvahelised teadusüritused. Nii on alates 1994. aastast iga kahe aasta järel kokku kutsutud rahvusvaheline konverents “Teaching and Language Corpora” (TALC).¹ Tegeldakse nii emakeelekorpuste (emakeelekõnelejate tekstikogud) kasutusvõimalustega võõrkeeleõppes kui ka võõrkeeleõppespetsiifiliste õppijakorpustega.

2. Emakeelekorpused ja võõrkeeleõpe

2.1. Uus arusaam keelest

Keelekorpuste mõju võõrkeeleõppele algas emakeelekorpuste uurimistulemuste rakendamisest. Viimaste aastakümnete jooksul on korpuslingvistid teinud palju suuremahulisi uuringuid, mis on võimaldanud inglise keelt tervikuna palju täpsemalt kirjeldada ja toonud kaasa ka mitmeid põhimõttelisi muutusi arusaamades keele kohta.

Elkõige on korpuste abil võimalik uurida sõnade ja struktuuride tegelikku esinemissagedust keelekasutuses, mida intuiitiivselt pole võimalik õigesti hinnata. Kõige suurem avastus, milleni korpuste uurimise abil seni on jõutud, seisneb arusaamas, et keele kasutamine pole nii originaalne looming, nagu võiks arvata. Tekstid ei koosne mitte üksiksõnadest, vaid pigem 2–7-sõnalistest ühikutest,

¹ Infot 2004. aasta konverentsi kohta vt aadressilt <http://www.ugr.es/~talc6/> (15.09.2005).

kusjuures sõnavara ja grammatika on omavahel tihedalt seotud. Keele olemus on modelleeritav, mis avaldub nii leksikaalsel tasandil (kollokatsioonid, st teatud sõnade sagedasem koosinemine, kui seda võimaldaks lihtne tõenäosus) kui grammatiliselt. Grammatilisi malle võib leida alates lihtsatest leksikaalgrammatilistest kombinatsioonidest (näiteks verb ja tema laiendid) kuni komplekssete süntaktiliste mallideni (vt Partington 1998). Douglas Biber (1988) algatas grammatiliste mallide uurimise seoses erinevate tekstitüüpide analüüsiga. Tulemuseks on kindlustus lingvistide arusaam sellest, et igale tekstitüübile on iseloomulik spetsiifiline grammatiliste struktuuride läbipõimumine. Need struktuurid võivad ulatuda üksikutest sõnadest (nt isikulised asesõnad, siduvad asesõnad) komplekssete mallideni (nt passiivikonstruktsioonid, kompleksne nimisõnafaas). Nii ei nähta inglise keele grammatikat enam monoliitse tervikuna nimega “haritud kirjalik inglise keel”, vaid pigem eraldi grammatikate (kirjaliku, suulise, jutustava teksti, argumenteeriva teksti jne) koguna (Meunier 2002: 122).

2.2. Autentne keel ja võõrkeeleõpikud

Võõrkeeleõpet puudutavad eelpool nimetatud uurimused nii palju, et kui korpus- te abil olid välja selgitatud autentsele keelele iseloomulikud jooned, siis hakati võrdlema võõrkeeleõpikutes kasutatavat ja autentset keelt. Selle tulemusel selgus, et õpikutekstides kasutatakse spetsiifilist nn õpikukeelt, mis erineb autentsest keelest nii sõnavara kui grammatiliste struktuuride poolest ja mida väljaspool keeletundi ei tarvitata (vt Römer 2004). Samuti põhineb õpikutes käsitletavate grammatiliste konstruktsioonide valik tavaliselt traditsioonidel (õpetatakse seda, mida on alati õpetatud) ja intuitsioonil (õpetatakse seda, mida autor õigeks peab), mitte empiirilistel uurimistulemustel. On leitud, et kõige suuremad erinevused õpikukeele ja autentse keele vahel kuuluvad suulise keele diskursusesse – kõnekeel on tunduvalt erinev kirjalikust keelekasutusest. Paljude õppematerjalide loomisel pole seda üldse teadvustatud. Võõrkeeleõpikutes esitatu järgi võib teha järelduse, et suuline suhtlemine on alati probleemitu, inimesed vestlevad viisakalt, kõik saavad kõigest ja kõigist aru, kasutatakse täislauseid, keegi ei katkesta teist, inimesed ei räägi mitmekesi korraga, küsimused ja vastused järgivad kindlat struktuuri, keegi ei takerdu ega eksi. Tegelikku suulist kõnekeelt iseloomustavad aga pigem fraasid kui täislaused, kordused, grammatilised kõrvalkalded, ellipsid, täitesõnad, informaalne kõnekeelsus. Tavavestluses katkestatakse rääkijat tihti, mistõttu laused jäävad lõpetamata; sageli räägitakse mitmekesi korraga. Iseloomulik on ka kiirete kõrvalpõigete tegemine teistesse teemadesse (vt Tan 2003).

2.3. Kui palju saab korpusi keeleõppes kasutada?

Kuigi üldiselt ollakse ühel meelel selles, et korpus- te andmed avavad huvitavaid perspektiive nii keelekirjeldusele kui keeleõpetusele (Granger 2002: 21), jätkuvad arvamustevahetused küsimuses, mil määral peaks korpusuuringuid keeleõppes rakendama. On välja kujunenud kaks peamist suunda – nn mõõdukas ja radikaal-

ne. Mõõduka lähenemise pooldajad (R. Carter, H. Widdowson jt) hoiatavad korpuste võimaluste ülehindamise eest. Korpuste uurimistulemusi ei saa pidada ainumääravaks. Üks asi on sõna või konstruktsiooni esinemissagedus tekstides, teine selle pedagoogiline olulisus. Puudusena nähakse ka asjaolu, et korpusepõhised õppematerjalid võivad osutada õppijaile üle jõu käivalt raskeks. H. Widdowson väidab, et juhul, kui tekst on õppija jaoks liiga keeruline, pole tegemist autentse materjaliga, sest puudub võimalus autentseks reageeringuks. Mõõduka lähenemise pooldajate arvates tuleb korpuste uurimisest saadud teadmiste põhjal täiustada õppekavu ja õppematerjale headest keeleõppepõhimõtetest lähtuvalt, näiteks keerukaid nähtusi lihtsustades. Õpikud peaksid modelleerima autentset keelekasutust. Edukaid näiteid võib leida näiteks järgmistest populaarsetest inglise keele kui võõrkeele õppesarjadest “Headway”, “Cutting Edge”, “Innovations” (Ranalli 2003: 5–7).

Radikaalse suuna esindajate (D. Willis, J. Sinclair jt) arvates aga ei tuleks piirduda nii vähesega – parim õppematerjal on korpused ise ja ilma nendeta pole üldse mõtet rääkida sisulisest keeleõppest.

2.4. Korpuste mõju sõnastikele

Kõige suuremat, võib isegi öelda revolutsioonilist mõju, on korpused seni avaldanud sõnastike koostamisele. On tekkinud uued korpusepõhised õppijasõnastikud, kus esitatakse lisaks tavapärasele leksikaalsele ja grammatilisele infole ka teavet tähenduste piirangute, kollokatsioonide, grammatiliste mallide, stiili, registri ja kasutussageduse kohta ning luuakse seosed sõna struktuuri, kasutamise ja tähenduse vahel. Sõnastikus “Longman Dictionary of Contemporary English” (1995) esitatakse sõna koos korpuse põhjal välja selgitatud tüüpiliste kollokatsioonidega ja leksikaalgrammatilise info (nt sõna *information* on esitatud kujul *information about, information on, provide information, additional information, gather/collect information* jne) (Meunier 2002: 125). Samuti antakse soovitusi sõna kasutamiseks: millal kasutada ja millal mitte. Nii selgitatakse sõnaga *that* seoses, et seda kasutatakse suulisel kõnes umbes kaks korda sagedamini kui kirjalikus, et vestluses jäetakse *that* tihti ära pärast kõige tavalisemaid verbe, nagu *think, say, see*, et kirjalikus keelekasutuses oleneb sõna *that* funktsioneerimine stiilist, näiteks akadeemilises kirjutises seda ära ei jäeta (Meunier 2002: 127).

2.5. Korpuste mõju grammatikatele

Ka inglise keele kui võõrkeele grammatikad on saanud kasu korpuste andmetest, kuid revolutsioonilist läbimurret pole seni veel toimunud. Kõige märgatavam muutus on olnud leksikaalgrammatilise info lisandumine. Näiteks on korpusepõhiselt koostatud “Longman Grammar of Spoken and Written English” (Biber jt 1999), mis rajaneb nii kirjalikul kui suulisel autentsetel keelekasutusel. Sõnastikus antakse iga grammatilise malli kirjeldus nelja registri seisukohalt (vestlus, ilukirjandus, uudised, akadeemiline proosa) näidates seoseid tüüpiliste ning muude grammatiliste joonte vahel. Sellised grammatikad sobivad edasijõudnud õppijaile,

kellel on algteadmised lingvistikast. Palju autentseid, korpustest pärit näiteid on kasutatud ka grammatikaõpikus “Exploring Grammar in Context” (Carter jt 2000).

2004. aasta TALC-i konverentsil rääkis Ute Römer, et võõrkeeleõppijad omandavad kehvasti inglise keele progressiivi (TO BE + V-ing) ning arutles, kas see teema on raskem kui teised või on seda võõrkeeleõppijaile lihtsalt halvasti kirjeldatud ja õpetatud. Tema arvates on õppijad väärt, et neile esitatakse ja selgitatakse tegelikku keelt ja räägitaks tõtt selle kohta, mida võib leida suurest korpusest, mis koondab tuhandete inimeste keelekasutust (Römer 2004).

2.6. Korpuste mõju õppemethodikale

Kui traditsioonilistes õpikutes esitatakse enamasti grammatikareeglid deduktiivselt järjestuses *esitamine-harjutamine-kasutamine* (*present-practice-produce*), siis Michael Lewis pakkus välja uue õppimismudeli *jälgimine-hüpoteeside tege- mine-katsetamine* (*observe-hypothesize-experiment*). See tähendab, et õppijad peavad saama võimaluse kuulda ja lugeda õpitavat keelt, et kogeda keelevormide tegelikku kasutamist ja luua ettekujutus toimivast keelest (Lewis 1993). Sellele mudelile vastab ka uudne korpusepõhine õppemethodika *Data-Driven Learning* ehk DDL (andmepõhine õpe), mille pakkus esimesena välja Tim Johns (1991). DDL-methodika tähendab seda, et õppijatest saavad uurijad, kes korpuse ja konkordantsileidjaid kasutades vaatlevad ise keelt ja avastavad selles esinevaid malle. Kõige tavalisem konkordantsivorm on praegu KWIC-indeks (*the Keyword-in-Context*), mille puhul iga näide esitatakse kindla pikkusega rea (näiteks 80 tähemärki) keskel. Joonisel 1 on esitatud sõna *make* esimesed konkordantsid 808-st, mis on leitud Interneti-keskkonna Virtual Language Centre konkordantsileidjaga.²

1 ly a cabinet department in order to make a “down payment” of \$400 billion (

2 fault, as I was sometimes scared to make a break. I’ve gone through that ph

3 s denying them. If you’re having to make a choice between crack-dealing in

4 ly was more impartial, declining to make a definitive judgement on who won

5 has so far resisted the pressure to make a permanent move to London, largel

6 less data network. The grandparents make a videophone call from a cruise sh

7 sign that the administration would make a better offer because it had simpl

8 rce said the Government intended to make a decision before Friday so that pa

9 concerned, it is only reasonable to make a formal announcement after a conse

10 sary for the Chinese authorities to make a matching change in the Basic Law.

Joonis 1. Sõna *make* konkordantsid

Konkordantsidega võib õppija töötada õpetaja juhendamisel või iseseisvalt. Kogemus on näidanud, et parim on töö korraldada teatud etappide kaupa, et kindlustada edu ja jätkuvat innukust. Alguses sobib töö õpetaja juhtimisel (näiteks väljatrükitud töölehtedega, õpetaja poolt ettevalmistatud ülesannete ja harjutustega), hiljem võivad õppijad töötada juba ka iseseisvalt korpuse toorandmetega. Konkordantside kasutamise plussideks võib pidada järgmisi asjaolusid: korpuse

² Vt <http://vlc.polyu.edu.hk/> (15.09.2005).

kasutades saab tutvuda autentsete keelenäidetega; tõuseb õppijate ja ka õpetajate keeleanalüüsivõime; kasvab õppijate omavaheline suhtlemine ja suhtlemine õpetajaga; areneb õppijate iseseisvus, nad saavad jätkata õpet väljaspool klassiruumi. Chris Tribble aga väidab, et lisaks suurte korpuste kasutamise eelistele keeleõppes on neil ka puudusi: kõik ei pääse korpustele ligi; kogemamatutel kasutajatel on oht uppuda andmeookeani; ka väga suur andmehulk ei anna garantiid, et õppija saaks sealt stiimuli, mis vastab tema vajadustele ja huvidele; mõnikord võivad emakeelekõneleja tekstinäited olla keeleõppija jaoks liiga raske. C. T. Tribble arvab, et keeleõppija ei vaja mitte üht suurt korpust, vaid pigem mitut väikest spetsialiseeritud korpust vastavalt õppe-eesmärgile. Selliseid väikesteid (25 000–30 000 sõna) korpuseid võib õpetaja ka ise koostada (Tribble 1997). Näiteks pidas TALC-i 2004. aasta konverentsil Francesca Bianchi ettekande erialakeele õpetamisest spetsialiseeritud korpuse abil, mis sisaldas psühholoogiaalaseid teadusartikleid. Esimesel etapil analüüsiti mõne psühholoogiaartikli sõnavara ja tähendusfunktsioonide (kausaalsuse ja tulemus) väljendamist, teisel etapil analüüsiti teadusartiklit (struktuur, leksikaalne ja grammatiline sidusus, metakommunikatiivsed funktsioonid) ja loodi ühiselt psühholoogiaartiklite korpust. Kolmandal etapil uuriti korpust konkordantside abil ja kirjutati iseseisev teadusartikkel (Bianchi 2004).

Et korpuseid saaks kasutada ka nõrgematel keeletasemetel, on välja pakutud leksikaalse filtri idee, mis töötaks sagedussõnastiku andmete abil ja filtreeriks korpusest välja õppija keeletasemele vastavad andmed (Wible jt 2000).

Konkordantside kasutamisel on esile kerkinud rida probleeme.

Esiteks. Kas uurija saab seda, mida küsib. Tegelikult ei pruugi see alati nii olla. Analüüsida saab ainult neid näiteid, mis kätte saadakse. Mida ei osatud mõelda ega küsida, seda sa ei saa ka analüüsida.

Teiseks. Analüüsida saab seda, mida on kerge leida. Mida ei ole, seda ei saa ka leida. Samas on “mitte midagi” (nt grammatilised nullid) paljude paradigmade liige ja kui neid ignoreerida, ei saa õiget tulemust. Üldse on grammatilisi konkordantse keeruline leida ja interpreteerida, seda enam nullmorfeemide ja -vormide puhul.

Kolmandaks. Uurimise alt jäävad välja mitteelektroonilised tekstid.

Neljandaks. See metoodika sobib ainult analüütilise stiiliga õppijaile.

Korpustega saab töötada otse võrgus. On olemas palju Interneti-lehekülgi, kus saab kasutada konkordantsileidjat, näiteks Cobuild³, Virtual Language Centre⁴, Chemnitz Internet Grammar⁵. Kasutada saab ka CD-ROM-ide abi ja programmi WordSmith.

Vaatamata mõnede puudustele on konkordantside kasutamisel õppeprotsessi tõhustamisel ja elavdamisel kindlasti tulevikku.

2.7. Korpuste mõju õppekavale ja õppematerjalidele

Kõige vähem on korpusuuringute tulemused jõudnud õppekavadesse ja õpikutesse, kuigi esimesed muutused on juba näha. Kõige märgatavam on olnud mõju leksika osas, kuna korpuste andmed võimaldavad objektiivsemat sõnavara valikut. Rohkem tähelepanu on hakatud pöörama eri tüüpi sõnakombinatsioonidele

³ Vt <http://web.quick.cz/jaedth/Introduction%20to%20CCS.htm> (15.09.2005).

⁴ Vt <http://vlc.polyu.edu.hk/> (15.09.2005).

⁵ Vt <http://www.tu-chemnitz.de/phil/InternetGrammar/frontend/index.html> (15.09.2005).

(püsiühendid, kollokatsioonid, fraasikatked jms) ja eri žanride keeleerinevustele. Populaarses inglise keele õppesarjas “Headway” on käsitletud deleksikaliseerunud verbe *make, do, put, take*, mida soovitasid John Sinclair ja Antoinette Renouf (vt 2.8.).

Mitmed ettepanekud õppekava muutmiseks pole aga rakendunud: näiteks on veel soovitatud õpetada irregulaarseid verbe lähtudes nende tegelikust kasutusagedusest (Grabowski, Mindt 1995), asendada lausepõhine grammatika diskursusepõhiselega, mis peaks paremini rahuldama kommunikatiivse kompetentsi vajadusi (Byrd 1997) jne.

2.8. Leksikakeskne õppekava

Ainuke praeguseks välja töötatud korpusepõhine õppekava on nn leksikakeskne lähenemine (*lexical approach*). Selle pakkusid 1988. a välja John Sinclair ja Antoinette Renouf (vt Sinclair, Renouf 1988), kes uurisid korpust konkordantside abil ja koostasid andmebaasi, mis kirjeldab iga käsitletava sõna semantilist välja, diskursuse sagedust ja süntaktilist käitumist. Samuti töid nad välja fraasid, mis esinesid nii sageli, et funktsioneerisid nagu leksikaalsed ühikud. Leksikakeskset lähenemist arendasid edasi Dave ja Jane Willis, kes kirjutasid 1989. aastal uudse õpikuarja “Collins Cobuild English Course”, mille eesmärk oli näidata, et ei saa tõmmata selget piiri grammatika ja leksika vahele ning sõnu tuleb õpetada mallide või fraaside koosseisus. D. Willise arvates on võõrkeele õppimiseks vaja näha suurt hulka keelenäiteid reaalses kontekstis ning saada üksikasjalikku infot tavaliste kollokatsioonimallide, idioomide, kinnistunud ja poolkinnistunud väljendite ning nende erinevate varieerumisastmete, suhtelise esinemissageduse, esinemiskoha ja formaalsusastme kohta. Õppija ei pea oskama iga ühikut grammatiliselt analüüsida, et seda veatult kasutada. Kui õppijad omandavad kinnistunud väljendeid, saavad nad sellest implitsiitset kasu ka oma grammatikale (vt Willis 1990).

Tõenäoliselt on ees ootamas palju uurimusi, mille käigus inglise keele kui võõrkeele grammatika vaadatakse põhjalikult üle diskursuse perspektiivist lähtudes. Samuti muutuvad õppekavad pidevalt ülesandepõhisemaks ja žanrikeskemaks. Kuid siinkohal tuleb meeles pidada, et töö korpustega on lihtsalt üks meetodikatest ja autentsus ei garanteeri sugugi alati pedagoogilist edu. Kõike (näiteks ajavormide kasutamist) ei saa konkordantsidega õpetada (Meunier 2002: 138). Kui lähtuda pedagoogilise olulisuse seisukohast, siis jääb piirdumine emakeelekorpuse uurimisega väheseks ja võib mõnel juhul viia isegi valedele järeldustele õppimiseks vajalike teemade valikul.

Siin tuleb appi uut tüüpi korpus – õppija- ehk vahekeelekorpus.

3. Õppijakorpused

Teise keele omandamist on traditsiooniliselt uuritud kolmel viisil: kasutades keelekasutuse andmeid (mis on kas loomulikud või saadud eksperimendi teel), metakeelelisi kommentaare või eneseanalüüsi. Enamasti on uurimuste jaoks olnud tõhusad eksperimendi ja eneseanalüüsi andmed. Loomuliku keelekasutuse and-

metest pole märkimisväärsed kasu saadud. Üheks põhjuseks on kindlasti raskus keelevariantide kontrollimisel. Et eksperimendi läbiviimiseks on olnud raske palju informante kokku saada, ongi teise keele omandamise uuringud peamiselt tuginenud väga väikese andmehulga uurimisele, mis tekitab küsitavusi üldistuste tegemisel (Granger 2002: 6). Nüüd on korpuslingvistika ka siin olukorda muutmas. 1980-ndate aastate lõpus hakati looma esimesi korpusi, mis sisaldavad keeleõppijate loodud tekste. Neid korpusi nimetatakse õppijakorpuks (*learner corpus*) või vahekeelekorpuks (*interlanguage corpus*) ning nende abil on õppijate keeleandmete kogumine ja analüüs muutunud palju kergemaks, täpsemaks ja võimalusterohkemaks.

Lancasteri ülikoolis 2003. a toimunud rahvusvahelise korpuslingvistika sümpoosiumi eel toimunud õppijakorpu tööpaia korraldajad Yukio Tono ja Fanny Meunier märkisid, et õppijakorpus ei ole enam lapseas, vaid teismeliseas – äärmiselt lootusrikas ja plaane täis, aga veel mitte täiesti küps (Xiao, McEnery).

Tabelis 1 on Norma Praveci (2002: 82, 83) järgi esitatud kümme maailmas tuntud õppijakorpu.

Tabel 1. Õppijakorpused

Korpuse nimi	Korpuse tüüp	Korpuse asukoht	Keeletaust
CLC (Cambridge Learner Corpus)	kommerts	Inglismaa, Cambridge'i ülikool	mitmekesine
HKUST (Hong Kong University of Science & Technology)	akadeemiline	Hong Kongi Teaduse ja Tehnoloogia Ülikool	hiina Kantoni murre
ICLE (International Corpus of Learner English)	akadeemiline	Belgia, Louvain-la-Neuve ülikool	mitmekesine
JEFLL (Japanese EFL Learner)	akadeemiline	Jaapan, Meikai ülikool	jaapani
JPU (Janus Pannonius University)	akadeemiline	Ungari, Pécsi ülikool	ungari
LLC (Longman Learner's Corpus)	kommerts	Inglismaa	mitmekesine
MELD (Montclair Electronic Language Database)	akadeemiline	USA, Montclairi Riiklik Ülikool	mitmekesine
PELCRA (Polish-English Language Corpus Research and Applications)	akadeemiline	Poola, Łódźi ülikool	poola
TSLC (TELEC Secondary Learner Corpus)	akadeemiline	Hong Kongi ülikool	hiina Kantoni murre
USE (Uppsala Student English)	akadeemiline	Rootsi, Uppsala ülikool	rootsi

3.1. Õppijakorpu koostamise kriteeriumid

Kuigi mõnikord defineeritakse õppijakorpu kui õppijalt saadud materjalide elektroonilist kollektsiooni, ei vasta igasugune elektrooniline keeleõppija tekstide kogu siiski õppijakorpu kriteeriumidele.

Õppijakorpus on autentsete võõrkeele tekstiandmete elektrooniline kogu, mis on koostatud vastavalt eksplitsiitsetele disainipõhimõtetele ja teise keele

õppimise/õpetamise eesmärkidest lähtuvalt. Korpuse materjalid on kodeeritud standardiseeritud ja homogeensel viisil ning dokumenteeritud olenevalt lähteko-
hast ja päritolust (Granger 2002: 7).

Õppijakorpuses peavad olema näited diskursusest, mitte üksikud sõnad või laused.

Seega ei ole õppijakorpus vigade korpus, vaid ta sisaldab nii ekslikku kui korrektset keelekasutust. Tavaliselt on õppijakorpustes esseed (keskmise pikkusega 500–1000 sõna). Üldised õppijakorpuse koostamise kriteeriumid on samad, mis emakeele korpuse puhul. Lisandub aga vajadus arvestada õppija emakeelt, keeletaset, õppimiskonteksti, ülesande täitmise ajalimiiti, abimaterjalide kasutamist, ülesande täitmise situatsiooni (näiteks eksam), teemat (määrab kasutatava leksika), teiste osaliste või pealtvaatajate/-kuulajate olemasolu (Granger 2002: 10).

3.2. Märgendamine

Et korpusest võimalikult palju kasu saada, tuleb see märgendada – lisada eri liiki lingvistilist infot. Õppijakorpuse puhul saab kasutada samu märgendusvõimalusi, mida emakeele korpuses, levinum on praegu sõnaliikide märgendamine (*part of speech tagging* ehk POS). Üldiselt võib öelda, et mida madalam on õppijate keeletase, seda väiksem on ka automaatsete märgenduste õigsuse protsent (automaatne märgendaja ei tunne vale sõna või vormi ära). Tõenäoliselt on see ka üheks põhjuseks, miks täielikult märgendatud korpustel põhinevaid uurimusi on veel vähe. Need aga, mis on, näitavad suurt potentsiaali keeleõppe tarvis (Granger 2002: 18).

Õppijakorpuse spetsiifiline märgendusvõimalus on vigade märgendamine, mis annab suured võimalused õppijakeele täpsemaks uurimiseks ja õppeprotsessi tõhustamiseks. Vigade märgendamise puhul on enne vaja otsustada, mida valida märgendamise aluseks, kas vigade liik (näiteks grammatilised, leksikaalsed jt vead) või vigade allikas (inter- ja intralingvaalsed vead). S. Grangeri arvates tuleks eelistada esimest tüüpi vigu, sest need on vähem subjektiivsed. Louvaini ülikoolis välja töötatud vigade märgendamise süsteem on hierarhiline: iga viga saab oma sarja koodi, mis läheb üldisemalt spetsiifilisema poole. Esimene täht koodis näitab vea ala (G – grammatiline, L – leksikaalne, X – leksikaalgrammatiline, R – register, W – süntaks ja S – stiil), järgmine täht täpsustab vea liiki (GV – verbi viga, GVM – morfoloogia viga jne) (Granger 2002: 19).

3.3. Mida uurida ja kuidas uurida?

Tihti kasutatakse õppijakorpust uurimustes, mille esimene faas on intuitsioonil põhinev hüpotees, teine korpuse uuring ja kolmas kontrollitud eksperiment.

Tavaliselt hõlmab õppijakorpuste lingvistiline uurimine üht kahest järgmisest meetodilisest lähenemisest: arvutipõhine veaanalüüs või kontrastiivne vahekeeleanalüüs.

Arvutipõhine veaanalüüs erineb varasemast veaanalüüsist mahuliselt. Samuti on võimalik uurida vigu täiskontekstis koos mittevigaste vormidega (Granger 2002: 14) või mingite grammatiliste vormide omandamist võõrkeeleõppijate poolt.

Näiteks Alex Housen püüdis teada saada, kuidas võõrkeeleõppijad omandavad inglise keele verbide põhivorme ja funktsioone, millised on omandamise etapid, kuidas õppijad ühendavad vormi ning õige ajalise, aspektilise või grammatilise tähenduse, millised etapid on vormi ja tähenduse vahekorras (Housen 2002). Palju on uuritud, mis vead on erineva emakeele, hariduse, keeletasemega rühmadel samad, mis erinevad. Vigade uurimise eesmärgiks on veatüübi identifitseerimine, seletamine ja veaohliku teema jaoks õppimisvõimaluste väljapakumise. Muidugi tuleb uurijale siinkohal kasuks mõlema keele oskus.

Kontrastiivne vahekeeleanalüüs on õppijakeele ja emakeelekõneleja keele või eri õppijakeelte omavaheline kvantitatiivne ja kvalitatiivne võrdlus. Eesmärgiks on võimalikult täpselt välja selgitada õppijakeelele iseloomulikud tunnused. Lisaks leksikaalsetele või grammatilistele vigadele uuritakse ka seda, milline on või ei ole õppija keel, miks mõjub õppijakeel tihti võõrapäraselt ka ilma otseste sihtkeelest kõrvalekalleteta. Näiteks on välja selgitatud, et õppijakeelt iseloomustab väiksem sõnavara kui emakeelekasutajal. Tavaliselt piirduvad keeleõppijad sagedussõnastiku esimese 1000 sõnaga ja kasutavad neid seetõttu tihedamini kui emakeelekõnelejad (Cobb 2003: 402). Vahekeelt iseloomustab samuti mõnede väljendite üle- ja teiste alakasutamine (näiteks inglise keele kui võõrkeele õppijad ülekasutavad väljendeid *nevertheless, on the one hand .. on the other hand*). Selle põhjuseks võib pidada ka õppematerjale, kus esitatakse tavaliselt pikk nimekiri sidendeid, mis on grupeeritud laiade semantiliste kategooriate kaupa (vastandamine, lisamine, tulemus jne), aga ei selgitata nende täpsemat süntaktilist, semantilist ja stilistilist kasutamist, vaid jäetakse õppijale mulje, justkui oleks need sõnad kõik omavahel sünonüümsed (Granger 2004: 135). Rootslaste inglise keele vahekeelt iseloomustab näiteks suur modaalverbide ülekasutamine ja teiste modaalsuse väljendamise võimaluste (määrsõnad, leksikaalsed modaaltähendusega verbid) alakasutamine; registri interferents, nagu suulise kõne eripära ülekandmine kirjalikku keelekasutusse (Ajmer 2004: 55). Sagedased on veel juhtumid ebatavaliste kollokatsioonidega, mis on edasijõudnud õppija üks põhivigu. Madalamal tasemel see nii tõsine probleem pole, tõenäoliselt piiratud sõnavara tõttu.

Kontrastiivne vahekeeleanalüüs aitab vältida valesid järeldusi vahekeele kohta, näiteks pidas Jacqueline Schachter (1974) hiinakeelsete õppijate vähest relatiivlauset kasutamist inglise keeles vältimisstrateegiaks. Tadamitsu Kamimoto jt aga näitasid, et see nähtus on põhjustatud relatiivlauset vähesusest hiina keeles (Kamimoto jt 1992, vt ka Altenberg 2002: 38).

Kontrastiivse vahekeeleanalüüsi probleemid on sageli põhjustatud sellest, et on raske leida õppijakorpusele analoogset emakeelekorpust ja seetõttu kasutatakse kirjakeele üldkorpust. Tegelikult aga ei ole päris õige võrrelda võõrkeeleõppijate (mitteelukutseliste kirjutajate) tekste emakeelsete professionaalide kirjatöödega. Kontrastiivanalüüs peab tuginema ka hästi läbitöötatud teooriale (näiteks edasijõudnud õppijate puhul on tõenäoline, et nende erinevus emakeelsest kõnelejust jääb suures osas väljapoole morfosüntaksit ja on seotud hoopis leksika, diskursuse ja pragmaatikaga, mille kohta on veel vähe andmeid). Probleeme tekitab ka küsimus, mis ikkagi on kõrvalekalle standardist (nt inglise keelel on mitu standardit) (Tan 2005).

3.4. Õppijakorpus ja andmepõhine õpe

Praegu on andmepõhise õppe ülesanded üldiselt seotud leksikaalsete ja leksikaalgrammatiliste teemade õppimisega emakeelekorpusse näidete järgi. C. Tribble on aga seisukohal, et kasulik oleks kombineerida õppijakorpuse ja emakeelekorpusse andmeid. Näiteks võib teha õppijate endi töödest anonüümse minikorpusse ja seda koos analüüsida. Võib koostada ülesandeid õppijatele raskusi valmistavate teemade kohta, kus võrreldakse kasutusnäiteid õppijakorpusest ja emakeelekorpussest (vt Meunier 2002: 133). See võimaldab õppijail endal näha, mis neid veel sihtkeelest lahutab ja märgata edusamme.

Õppijakorpuste uurimine ning analüüsimine võiks kindlasti olla ka üks osa õpetajakoolituses. Näiteks tegi Barbara Seidlhofer (2004: 218) tulevastele keeleõpetajatele spetsiaalse korpuslingvistika kursuse, mille käigus koostati nende endi töödest korpus ja esitati ise küsimused, millele hiljem korpusse vastused otsiti. Selline töövorm oli väga innustav ja tõhus.

Daniel Krieger (2003) arvab, et kuigi tuleb igati julgustada õppijaid ise keele kohta avastusi tegema, on korpuste suurim kasutuspotentsiaal keeleõppes siiski õppematerjalide arendamine: korpustest saavad autorid palju üliolulist infot.

4. Kokkuvõte

Kui võrrelda eesti keele kui võõrkeele õppe olukorda inglise keelega, siis on selge, et meil on palju arenguruumi. On küll käsitletud mitmeid küsimusi (vt nt Pajusalu 2004, Pool, Vaimann 2005, Verschik 2004, Rammo, Pill 2002), kuid eesti keel tervikuna on võõrkeeleõppe seisukohalt läbi uurimata ja kirjeldamata. Ilma korraliku keelekirjelduseta pole võimalik arendada ka keeleõppe materjale.

Viimastel aastatel on eesti keele kui võõrkeele õppe alane tegevus hoogustumas. Toimuvad teadusüritused: Tartu Ülikooli eesti keele võõrkeelena õppetool korraldab konverentse “Eesti keel ja teised keeled”⁶, rakenduslingvistika 2005. a kevadkonverentsi teema oli “Teine keel. Eesti keele õpetamine mitte-eestlastele”⁷. Nii Tartu kui Tallinna ülikool on avaldanud kogumikke, kus käsitletakse eesti keele kui võõrkeele temaatikat. Tänapäevastes eesti keele kui võõrkeele õppekomplektides on kasutatud autentseid lugemis- ja kuulamistekste ning nende grammatikateemad on püüdnud esitada lähtudes pedagoogilisest olulisusest ja kommunikatiivsetest vajadustest. Ilmunud on keeletasemetes esialgsed kirjeldused (Mall Laur “Esimene verstapost. Eesti keele algata” (1998), Martin Ehala jt “Suhtluslävi” (1997) ja Mai Loog ja Krista Kerge “Tuul tiibades. Eesti keele suhtluse kõrgtase” (1999)), mõned käsiraamatud õppematerjalide autorile ja keeleõpetajale (Krista Kerge “Eesti süntaks keeleõppe praktikule” (2001), Helle Metslang jt “Keelehärm” (2003) jt) ning õppijasõnastikud (Raili Pool “Eesti keele verbi- rektsioone” (1996)) jm. On tehtud ka esimene katse kaasata arvutuslingvistika õppematerjalide koostamisse. PHARE eesti keele õppe programmi algatusel koostatud muukeelse kooli 4.–9. klassi eesti keele õpikute sarjale on Tartu ülikooli arvutuslingvistika töörühm teinud arvutianalüüsi (Asser jt 2004), mille eesmärk on kaasa aidata optimaalse raskusastmega sarja loomisele. Samuti on valminud mõned interaktiivsed eesti keele õppematerjalid – “E-fant”⁸, “Kaunis külaline”

⁶ Vt <http://www.ut.ee/ekv/> (15.09.2005).

⁷ Vt <http://www.rakenduslingvistika.ee/main.php?mida=konverents> (15.09.2005).

⁸ Vt <http://www.efant.ee/student/> (15.09.2005).

(CD-ROM-il töötav interaktiivne eesti keele õppematerjal muukeelsetele koolidele, mis on loodud Ameerika Ühendriikides väljatöötatud keeleõppeprogrammi alusel).

Korpusepõhist keeleuurimist võimaldavad Tartu ülikooli⁹ ja Eesti Keele Instituudi¹⁰ eesti keele korpused. Tallinna Ülikooli üld- ja rakenduslingvistika õppetooli juures on loomisel eesti vahekeelekorpus (vt Eslon 2006).

Et muuta eesti keele kui võõrkeele õpe lihtsamaks, tõhusamaks ja loomulikumaks, on vaja tõsiselt tegelda eesti keele uurimisega ning õppematerjalide ja õppemethodika arendamisega. Selleks oleks vaja uurida eesti vahekeelekorpus: 1) selgitada arvutipõhise veaanalüüsi abil välja eesti keele kui võõrkeele õppijate tüüpvead, luua nende põhjalik ja süsteemne klassifikatsioon, leida vigade sügavamad põhjused ja 2) tegelda kontrastiivse vahekeeleanalüüsiga – leida jooned, mis iseloomustavad eesti vahekeelt ja eristavad seda sihtkeelest. Tähelepanu vajaksid kindlasti sõnavara rikkalikkus, tüüpilised kollokatsioonid, tähendusfunktsioonide väljendamine, grammatiliste struktuuride ala- ja ülekasutus, registri- ja stiilitunnetus, suulise ja kirjaliku keele erinevused jms. Selliste keeleuuringute tulemuste põhjal oleks võimalik luua uusi metoodilisi lahendusi sihtkeele poole liikumiseks, arendada õppekavasid, õppematerjale ja õppemethodikat. Samuti saaks uuringute tulemusi kasutada keeleõppija vajadustest lähtuvate põhjalike sõnastike ja pedagoogiliste grammatikate kirjutamiseks.

Kirjandus

- Ajmer, Karin 2004. Modality in advanced Swedish learner's written interlanguage. – Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Ed. by Sylviane Granger, Joseph Hung, Stephanie Petch-Tyson. Amsterdam/Philadelphia: John Benjamins Publishing Company, 55–76.
- Altenberg, Bengt 2002. Using bilingual corpus evidence in learner corpus research. – Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching. Ed. by Sylviane Granger, Joseph Hung, Stephanie Petch-Tyson. Amsterdam/Philadelphia: John Benjamins Publishing Company, 37–54.
- Asser, Hiie; Kaalep, Heiki-Jaan; Linnas, Siret; Mikk, Jaan; Muischnek, Kadri; Songe, Merje; Uibo, Heli 2004. Õpikute keerukuse analüüs arvutitel. – Toimiv keel II. Töid rakenduslingvistika alalt. Toim Maria-Maren Sepper, Jane Lepasaar, koost Helle Metslang. Tallinna Pedagoogikaülikooli eesti filoloogia osakonna toimetised 3. Tallinna Pedagoogikaülikool, Eesti Keele Instituut. Tallinn: TPÜ Kirjastus, 72–84.
- Bianchi, Francesca 2004. Teaching how to write research articles in English to Italian psychology students: a case history. – TALC 2004. <http://www.ugr.es/~talc6/> (15.09.2005).
- Biber, Douglas 1988. Variation across Speech and Writing. Cambridge University Press.
- Biber, Douglas; Johansson, Stig; Leech, Geoffrey; Conrad, Susan; Finegan, Edward 1999. Longman Grammar of Spoken and Written English. Longman Publications Group.
- Byrd, Patricia 1997. Grammar FROM Context: re-thinking the teaching of grammar at various proficiency levels. – <http://jalt-publications.org/tlt/files/97/dec/byrd.html> (15.09.2005).
- Carter, Ronald; Hughes, Rebecca; McCarthy, Michael 2000. Exploring Grammar in Context. Cambridge University Press.

⁹ Vt www.murre.ut.ee ja www.cl.ut.ee (15.09.2005).

¹⁰ Vt www.eki.ee/corpus (15.09.2005).

- Cobb, Tom 2003. Analyzing late interlanguage with learner corpora. Quebec replications of three European studies. – http://www.er.uqam.ca/nobel/r21270/cv/593_Cobb.pdf (15.09.2005).
- Ehala, Martin; Saarsoo, Kristi; Vare, Silvi; Õispuu, Jaan 1997. Eesti keele suhtluslävi. Strasbourg: Council of Europe Publishing.
- Eslon, Pille 2006. Eesti vahekeele korpusest korrelatsioonigrammatikani. – (Käesolev kogumik.)
- Grabowski, Eva; Mindt, Dieter 1995. A corpus-based learning list of irregular verbs in English. – *ICAME Journal* 19, 5–22. <http://helmer.aksis.uib.no/icame/ij19/mindt.pdf> (15.09.2005).
- Granger, Sylviane 2002. A Bird's-eye view of learner corpora in SLA research and FLT. – *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Ed. by Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson. Amsterdam/Philadelphia: John Benjamins Publishing Company, 3–33.
- Granger, Sylviane 2004. *Computer Learner Corpus Research: Current Status and Future Prospects*. – <http://cecl.fltr.ucl.ac.be/Downloads/Indianapolis%20status%20&%20prospects.pdf> (15.09.2005).
- Housen, Alex 2002. A corpus-based study of the L2-acquisition of the English verb system. – *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Ed. by Sylviane Granger, Joseph Hung, Stephanie Petch-Tyson. Amsterdam/Philadelphia: John Benjamins Publishing Company, 77–119.
- Johns, Tim 1991. Should you be persuaded – two examples of Data-Driven Learning materials. – *English Language Research Journal* 4, 1–16.
- Kamimoto, Tadimitsu; Shimura, Akihiko; Kellermann, Eric 1992. A second language classic reconsidered – the case of Schachter's avoidance. – *Second Language Research* 8 (3), 251–277.
- Kerge, Krista 2001. Eesti süntaks keeleõppe praktikule. Tallinn: TEA Kirjastus.
- Krieger, Daniel 2003. *Corpus Linguistics: What It is and How It Can Be Applied to Teaching*. – <http://fis.ucalgary.ca/lrc/doc/krieger.pdf> (15.09.2005).
- Laur, Mall 1998. *Esimene verstapost. Eesti keele suhtluse algfase. Riiklik Eksami- ja Kvalifikatsioonikeskus*. Tallinn.
- Lewis, Michael 1993. *The Lexical Approach*. Hove: Language Teaching Publications.
- Longman Dictionary of Contemporary English. 1995. Harlow: Longman.
- Loog, Mai; Kerge, Krista 1999. *Tuul tiibades. Eesti keele suhtluse kõrgtase. Riiklik Eksami- ja Kvalifikatsioonikeskus*. Tallinn.
- Metslang, Helle; Krall, Ingrid; Pajusalu, Renate; Saarsoo, Kristi; Sõrmus, Elle; Vare, Silvi 2003. *Keelehärm. Eesti keele probleemseid piirkondi*. Tallinn: TPÜ Kirjastus.
- Meunier, Fanny 2002. The pedagogical value of native and learner corpora in EFL grammar teaching. – *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Ed. by Sylviane Granger, Joseph Hung, Stephanie Petch-Tyson. Amsterdam/Philadelphia: John Benjamins Publishing Company, 119–141.
- Pajusalu, Renate 2004. *Keeleõppija tähendusviga ja selle hindamine*. – *Toimiv keel II. Töid rakenduslingvistika alalt*. Toim Maria-Maren Sepper, Jane Lepasaar, koost Helle Metslang. Tallinna Pedagoogikaülikooli eesti filoloogia osakonna toimetised 3. Tallinna Pedagoogikaülikool, Eesti Keele Instituut. Tallinn: TPÜ Kirjastus, 218–235.
- Partington, Alan 1998. *Patterns and Meanings*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Pool, Raili 1996. *Eesti keele verbirektsioone*. Tartu: Tartu Ülikooli Kirjastus.
- Pool, Raili; Vaimann, Elle 2005. *Vead kõrgtasemel eesti keele kõnelejate kirjalikus keelekasutuses*. – *Eesti Rakenduslingvistika Ühingu aastaraamat 1 (2004)*. Koost

- Margit Langemets, toim Maria-Maren Sepper. Tallinn: Eesti Keele Sihtasutus, 115–136.
- Pravec, Norma A. 2002. Survey of learner corpora. – *ICAME Journal* 26, 81–114.
- Rammo, Sirje; Pill, Julia 2002. Vene ja eesti keele verbirektsiooni vastastikused mõjutused. – *Emakeel ja teised keeled III*. Toim Liina Lindström, Oksana Palikova. Tartu Ülikooli eesti keele (võõrkeelena) õppetooli toimetised 1. Tartu: Tartu Ülikooli Kirjastus, 145–154.
- Ranalli, James M. 2003. *ELT Coursebooks in the age of corpus linguistics: constraints and possibilities*. – <http://www.cels.bham.ac.uk/resources/essays/Ranalli6.pdf> (15.09.2005).
- Römer, Ute 2004. A need for better descriptions. Teaching the truth about progressives. – *TALC 2004*. <http://www.ugr.es/~talc6/> (15.09.2005).
- Schachter, Jacqueline 1974. An error in error analysis. – *Language Learning* 24, 205–214.
- Seidlhofer, Barbara 2004. Pedagogy and local learner corpora: working with learning-driven data. – *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Ed. by Sylviane Granger, Joseph Hung, Stephanie Petch-Tyson. Amsterdam/Philadelphia: John Benjamins Publishing Company, 213–234.
- Sinclair, John; Renouf, Antoinette 1988. A lexical syllabus for language learning. – *Vocabulary and language teaching*. Ed. by Ronald Carter, Michael McCarthy. Harlow: Longman, 140–158.
- Tan, Melinda 2003. Language Corpora for Language Teachers. – *Journal of Language and Learning* 1 (2). http://www.shakespeare.uk.net/journal/jllearn/1_2/tan1.html (15.09.2005).
- Tan, Melinda 2005. Authentic language or language errors? Lesson from a learner corpus. – *ELT Journal* 59/2. Oxford University Press, 126–134.
- Tribble, Chris 1997. Improvising corpora for ELT: quick-and-dirty ways of developing corpora for language teaching. – *PALC '97 Proceedings*. Ed. by B. Lewandowska-Tomaszczyk, P. J. Melia. Łódź: Łódź University Press, 106–117. <http://www.tribble.co.uk/text/Palc.htm> (15.09.2005).
- Verschik, Anna 2004. Mõnda vahekeelest ja kontrastiivsest analüüsist. – *Eesti keel: võõrkeelest teiseks keeleks*. Koost Mart Rannut, toim Mart Rannut, Marju Kõivupuu, Tiit Päeva. Tallinna Pedagoogikaülikooli eesti filoloogia osakonna toimetised 1. Tallinn: TPÜ Kirjastus, 132–144.
- Wible, David; Chien, Feng-yi; Kuo, Chin-Hwa; Wang, T. C. (2000). Adjusting corpus searches for learners' level. Filtering results for frequency. – *TALC 2000*. <http://www-gewi.kfunigraz.ac.at/talc2000/Htm/index1.htm> (15.09.2005).
- Widdowson, Henry 1983. *Learning Purpose and Language Use*. Oxford: Oxford University Press.
- Willis, Jane; Willis, Dave 1989. *Collins COBUILD English course*. London: Collins COBUILD.
- Willis, Dave 1990. *The Lexical Syllabus*. – <http://www.cels.bham.ac.uk/resources/LexSyllabus/lexsch3.pdf> (15.09.2005).
- Xiao, Zhonghua; McEnery, Tony. *Corpora and language education*. – http://www.lancs.ac.uk/postgrad/xiaoz/papers/corpora_teaching.doc (15.09.2005).

LANGUAGE CORPORA AND FOREIGN LANGUAGE TEACHING

Mare Kitsnik

The field of teaching Estonian as a foreign language is relatively new and in order to develop it further the following questions should be answered: how to present the language system to the learners in a way that is comprehensible and supports learning the best? how to compile more effective and learner-friendly textbooks, grammars and dictionaries for students of Estonian as a foreign language? what kind of learning method corresponds best to the learners' needs and language acquisition principles?

In the development of foreign language learning the role of language corpora has recently increased significantly. Studying the first language corpora has helped to describe more precisely the uses of authentic language and it has also helped to find out how the language in foreign language learning materials differs from that.

So far the studies of corpora have influenced most of all the compilation of dictionaries. There have been created new corpus-based dictionaries: in addition to general lexical and grammatical information they include information about ranking of meanings, collocations, grammatical patterns, style, register and frequency of use. Grammars and textbooks have benefited from the data from corpora, too. There has emerged a new approach called *data driven learning*. In this method the learners act as researchers who observe the language and discover different patterns in it by using corpora and concordances.

In addition to native language corpora the researchers have started to compile learner corpora or interlanguage corpora, which contain texts created by language learners. Investigating interlanguage corpora makes it possible to identify the types of errors the learners make and other characteristic features of interlanguage (vocabulary, collocations, grammatical structures, etc.).

In Estonia the corpora have not been studied from the perspective of the foreign language learner's needs. There are several first language corpora, and Estonian interlanguage corpus is being created at the Chair of General and Applied Linguistics at Tallinn University.

Keywords: first language corpora, learner corpora or interlanguage corpora, dictionaries, grammars, teaching methodology, data-driven learning, Estonian as foreign language

Mare Kitsnik (1962) on Tallinna Ülikooli lingvistikamagistrant, mitmete eesti keele kui võõrkeele õpikute autor. Uurimisvaldkonnaks on vahekeel ja teised eesti keele kui võõrkeele omandamisega seotud küsimused.
Mare.Kitsnik@tlu.ee