

TALLINNA ÜLIKOOLI
EESTI KEELE JA KULTUURI INSTITUUDI
TOIMETISED 12

KORPUSUURINGUTE METODOLOOGIA JA MÄRGENDAMISE PROBLEEMID

Toimetanud Pille Eslon ja Katre Õim

Tallinn 2010

Tallinna Ülikooli Eesti Keele ja Kultuuri Instituudi toimetised 12
Publication of the Institute of Estonian Language and Culture 12
Tallinn University

Toimetajad / Editors:

Pille Eslon (Tallinn), Katre Õim (Tallinn)

Toimetuskolleegium / Advisory Board:

Lars Gunnar Larsson (Uppsala), Maisa Martin (Jyväskylä), Kaili Müürisep
(Tartu), Helle Metslang (Tartu/Tallinn), Meelis Mihkla (Tallinn), Renate Pajusalu
(Helsingi/Tartu), Helena Sulkala (Oulu), Urmas Sutrop (Tallinn/Tartu),
Maria Voeikova (Viin/Sankt-Peterburg)

Postiaadress / Contact information:

Eesti Keele ja Kultuuri Instituut / Institute of Estonian Language and Culture
Tallinna Ülikool / Tallinn University
Narva mnt 25
10120 Tallinn
ESTONIA

ekki.toimetised@tlu.ee

www.tlu.ee

Autoriõigus / Copyright: autorid ning Eesti Keele ja Kultuuri Instituut, 2010

ISSN 1736-4221

ISBN 978-9949-463-46-6

SISUKORD

Pille Eslon

Suundumustest eesti keele grammatiliste käänete kasutuses.....7

Ilmari Ivaska

Mitä syntaktisesti koodattu korpus voi kertoa sanajärjestyksestä? Predikaatti ja subjekti edistyneiden suomenoppijoiden eksistentsiaalisissa lauseissa 37

Erika Matsak, Pille Eslon, Jaagup Kippar

Eesti keele sõnajärje vealeidja prototüübi arendamine 59

Tuija Määttä

Menen seuraavaan kursiin. Korpuspohjainen tutkimus ruotsinkielisten suomenoppijoiden illatiivin ja allatiivin käytöstä kirjallisessa tuotannossa 101

Marianne Spoelman

The copula construction in Finnish learner language: Estonian, German and Dutch learners' use of partitive predicatives 127

SUUNDUMUSTEST EESTI KEELE GRAMMATILISTE KÄÄNETE KASUTUSES¹

Pille Eslon

Ülevaade

Artiklis arutletakse suundumuste üle eesti keele grammatiliste käänete kasutuses. Uurimuse aluseks on kirjakeele ja õppijakeele võrdleva korpuslingvistilise analüüsi tulemused ning andmed õppijakeele objektifraaside käändeasenduste kohta. Empiiriline teave viitab kahele tendentsile eesti õppijakeeles: ühelt poolt nominatiivse objektnoomeni kasutusala laienemisele partitiivi ja genitiivi arvelt ning teisalt genitiivse objektnoomeni kasutusala kitsenemisele valdavalt nominatiivi kasutusala laienemise tõttu. Esimene tendents tugineb verbi defiiniitsuse tuhmumisele ja aspektverbide objekti käändevariatiivsusele. Teine tendents on põhiliselt seotud partitiivverbide partsiaalobjekti ebareeglipärase asendamisega nominatiivse totaalobjektiga, mille on tinginud õppija emakeele mõju ning mis on analoogia põhjal laienenud ka genitiivse objektnoomeni

¹ Tööd on toetanud riikliku programmi "Eesti keele keeletehnoloogiline tugi (2006–2010)" projekt "VAKO: Eesti vahekeele korpuse keeletarkvara ja keeletehnoloogilise ressursi arendamine (2008–2010", riikliku programmi „Eesti keel ja kultuurimälu (2009–2013)“ projekt „REKKi käsikirjaliste materjalide digiteerimine, Eesti vahekeele korpuse alamkorpuste loomine ja korpuse kasutusvõimaluste populariseerimine (2009–2013)“ ja ETFi grant „Üle kantud tähenduses fraasid eesti keele korpustes (2010–2013)“.

kasutusaladele. Tegu on objektikäänete kasutusreeglite lihtsustamisega nominatiivi kasuks, kuna nominatiiv on markeerimata kääne, partitiiv aga kõikide muutmorfoloogiliste tunnuste suhtes negatiivne ning seetõttu kõige enam markeeritud grammatiline kääne. Tegu on keelesüsteemi sisemise kompensatoorsusega, mis kajastub eesti keele erinevates kasutusvariantides, sh õppijakeeles.

Märksõnad: objektifraas, grammatiliste käänete kasutamine

1. Uurimisandmed

Eesti õppijakeelele omaseid tendentse objektnoomeni käändevalikul olen uurinud *Eesti vahekeele korpuse* tekstide alusel. Valim koosneb verbifraasidest, mis on lausetes eristatud formaalselt. Valimi aluseks on võetud morfoloogiliselt analüüsitud tekstid (kokku rohkem kui 500 000 sõnet), millest on sõnaliigi märgendi järgi välja filtreeritud kõik verbivormid koos 2-3-sõnalise kontekstiga verbist paremale ja vasakule². Käesoleva uurimuse tarbeks olen kasutanud verbist paremale jääva kontekstiga fraase – kokku 38 251 verbifraasi, mille hulgas on üheselt interpreteeritavaid objektifraase viiendik ehk 7939. Objektifraaside kõige sagedasem käändevorm on partitiiv (5631 juhtumit), järgnevad nominatiiv (1799) ja genitiiv (927). Kuna eesti keeles on sage grammatiliste käänete vormihomonüümia ning valimis leidub kontekste, mille põhjal on käändevormi üle raske otsustada (*võtta eriala, viskaks prügi, viima klaastaara*), siis suurendab see objektifraaside hulka.

² Verbifraaside formaalse eraldamise *Eesti vahekeele korpuse* tekstidest viis läbi Kairit Sirts.

Nii selgub näiteks, et verbi *alustama* on valimis kasutatud vaid teatud vormides, mille esinemissagedus on järgnevalt sulgudes: *alusta* (1), *alustab* (9), *alustada* (11), *alustades* (3), *alustaksin* (1), *alustama* (1), *alustame* (3), *alustan* (6), *alustas* (15), *alustasid* (6), *alustasin* (3), *alustatakse* (2), *alustati* (3). Nende vormidega on valimis järgmised fraasid:

- (1) *alusta* – **alusta viisteistkümmesal detsembril*
- (2) *alustab* – *alustab, aga koor; alustab, siis inimesed; alustab 1991 aastal; alustab ajaleht; alustab Aleksandr lauset; alustab keel; *alustab koolid; alustab oma elu; alustab oma tööd*
- (3) *alustada* – *alustada, et igal suvel; alustada ehitamisega; alustada kaubavahetuse; alustada oma äri; alustada sellest, et eesti folkloor; alustada sellest, et me, venelased; alustada sellest, et millelegi vaatamata kooliaeg; alustada uue elu; alustada uut elu; alustada õpetust; alustada ühe tähega*
- (4) *alustades* – *alustades autogeensest treeningust; alustades elukutsete; alustades revolutsiooniga*
- (5) *alustaksin* – *alustaksin tööd*
- (6) *alustama* – *alustama, vara öö*
- (7) *alustame* – *alustame: vali oma telefonis; alustame pidu; alustame toa*
- (8) *alustan* – *alustan ajaloost; alustan elu; alustan kultuurist; alustan selle filmiga; alustan sellest, et lapseõlvest; alustan sellest, et minu ökoelu*
- (9) *alustas* – *alustas eeslaulja; alustas Eesti Reklaamfilm; alustas Hurt; alustas Kalevipoeg-projekti; alustas Kalevipoja projekti; alustas Kalevipoja turustamisega; alustas kontserdi; alustas Liivi sõda; alustas luuletusi; alustas läänelike saadete; alustas oma tegevust; alustas Regilaul; alustas uue poliitikat; alustas uut elu; *alustas õpima keemiat*

- (10) *alustasid – alustasid ka maja; alustasid nad vene keele; alustasid oma elu; alustasid regilaulu; alustasid sõda; alustasid vabadussõda*
- (11) *alustasin – alustasin ma tööd; alustasin veekogust; alustasin vestlust*
- (12) *alustatakse – alustatakse algust; alustatakse Sõle*
- (13) *alustati – alustati eestis esimesi ettevõtmisi; alustati kohe pärast registreerimist; alustati usuõpetuse*

Nagu näha, on formaalselt välja toodud fraasid struktuurilt erilaadsed ega ole sugugi mitte alati objektifraasid. Nii on verbi *alustama* kasutatud erinevates alltähendustes ja erinevate reksioonimallide alusel, millest paljud ei eelda objektiseoste olemasolu (**alusta viisteistkümmesal detsembril, alustada ehitamisega, alustades autogeenset treeningust, alustan ajaloost, alustame: vali oma telefonis* jne). Samuti on ühe või teise grammatilise käände vormi asemel ekslikult kasutatud semantiliste käänete vorme ja vastupidi. Näiteks seoses paronüümiaga on õppija segi ajanud kaks erinevat verbi ja nende reksioonimallid: **mõjutasid eesti muusikale millele? – mõjutasid eesti muusikat mida?, kuid mõjusid eesti muusikale millele?; *tutvusin välismaalast keda? – tutvusin välismaalasega kellega?, kuid tutvustasin välismaalast keda?* jne. Õppijakeele sõnajärje eripärast joltuvalt on verbist paremale jäävas kontekstis sageli kasutatud subjekti (*alustasid nad, alustas regilaul, alustab keel, kaasnes info-sulg* jne) ning koopula puhul on parempoolses kontekstis omal kohal öeldistäide (*on särtsakad ja vaprad inimesed, on ta süütu lapsuke* jne). Kõik loetletud juhtumid, sh öeldistäite sihitiskäändelised vormid olen analüüsist välja jätnud. Samuti olen välja jätnud taolised kasutused nagu *kaotame noorte inimeste*, kus verbile järgneva substantiivi ja selle täiendsõna käändevalik verbist ei sõltu: *kaotame kelle-mille (nt usu) > kaotame kelle usu kellesse-millesse – kaotame noorte inimeste usu õiglusse*.

Analüüsitud pole ka osalauseste piirile jäävaid ja muid ebaselgeid kooslusi nagu *alustab, aga koor; kaotada, vaid tema sõnad; alustama, vara öö; alustatakse Sõle; kaalusin: suhkur* jne.

Käesoleva töö uurimisobjekt on objektifraaside alarühm, kus õppijad on kasutanud üht grammatilise käände vormi teise asemel (vt tabel 1). Nominatiiviga on tehtud kokku 696 asendust, partitiiviga 170 ja genitiiviga 81 asendust.

Kõige sagedamini ja regulaarselt on nominatiiviga asendatud partitiivi (506 kasutusnäidet): **armastab jook, *õppisite seal aru-tiõpetus, väljendada emotsioonid, *võtan natukene lilled, *võrreldes teadmised* jne.

Tabel 1. Grammatiliste käänete asendused eesti õppijakeele objekti-fraasis

Asenduse suund	Sagedus
Nominatiiv > partitiiv	506
Nominatiiv > genitiiv	190
Partitiiv > nominatiiv	50
Partitiiv > genitiiv	120
Genitiiv > nominatiiv	19
Genitiiv > partitiiv	62

Vastupidine protsess ehk partitiivi kasutamine nominatiivi asemel jääb sageduselt selgelt alla nominatiivi allutusretkele partitiivi aladele: nominatiivi asendusi partitiiviga on valimi objektifraasis vaid 50 korral: **õppis pähi programmerimise keeleid, *vaikiti maha kogu pagulaskirjandust, *vahetada oma keemiatoodet, *teha hea jõulukungitust, *söö ära kõike marju, saada hea kõrghari-dusi* jne. Tasavägisem on seis partitiivi ja nominatiivi kasu-tamisega genitiivi asemel – vastavalt 120 (*võttis suunda, võttis mind tööle, veetsime oma vaheaega, *varastas kotti noormees, *vali-*

sin teist eriala jne) ja 190 korda (**õppis selgeks üks vana laul, võttis korv, valisin saksofon, unustasime meie rahvuskultuur, tegin pealkiri, sünnitama üks poiss* jne). Vastupidine ehk genitiivi kasutamine nominatiivi asemel pole olnud märkimisväärne – 19 juhtumit: **viidi läbi ürituse, taastati iseseisvuse, säilitada oma keele, sooritan kõike eksami* jne; veidi sagedamini on genitiiviga asendatud partitiivi – 62 korral: **armastab selle töö, tänasid puuraiujate ja kutsusid, tunnevad selle traditsiooni, tahaks õpida eesti keele* jne.

Need empiirilised andmed osutavad selgelt kahele tendentsile eesti õppijakeeles: 1) nominatiivse objektnoomeni kasutusala laieneb partitiivi ja genitiivi arvelt ning 2) genitiivse objektnoomeni kasutusala kitseneb valdavalt nominatiivi kasutusala laienemise tõttu.

2. Uurimiseesmärk

Õppijakeele uurijad ning keeleõpetajad peavad käändesendusi tavaliselt veaks (vt nt Metslang jt 2003; Pool, Vaiman 2005; Krall 2008; Pool 2007; Pool 2010). Käesolevas uurimuses olen veaanalüüsi suunast distantseerunud ning võtnud aluseks nende lingvistide arvamused, kes vaatlevad sedalaadi juhtumeid keeleomase või keeltevahelisest mõjust tingitud variatiivsuse ja näevad nende protsesside tekkepõhjust kas analoogias (vt nt Aikhenvald 2002, 2007; Heine, Kuteva 2008; Ehala 2000, 2009b; Eslon, Öim 2010) või keeltevahelises mõjus (vt nt Sajavaara 2006; Kaivapalu 2005). Varieerumine on keelekasutusele omane tunnusjoon, mis kõneleb keele arengupotentsiaalset, sageduse ja produktiivsuse dialektikale tuginevatest keelemuutustest (vt nt Kaalep 2010, 2009; Nikolaev, Niemi 2006; Lüdeling, Evert 2003; Lüdeling jt 2000) jm.

Eesti õppijakeele objektifraasi analüüsitulemuste interpreteerimisel ja objektnoomeni käändeasenduse mõtestamisel olen aluseks võtnud keelesisese ja keeltevahelise analoogia ning keeltevahelise mõju mõisted. Püüan näidata, millistest suundumustest eesti keele arengus võib kõnelda eesti õppijakeel: partitiivse ja genitiivse objektnoomeni asendumine nominatiivsega ning genitiivi kasutusala kitsenemine õppijakeeles pole pelk keeltevahelisest mõjust tulenev keeleviga, vaid selles kajastub eesti keele käändekasutusele omaseid üldiseid arengutendentse. Määravat rolli mängib siin loomulik lihtsustumine ja keelesisene analoogia, mis tuleb eriti ilmekalt esile siis, kui võrrelda omavahel eesti keele erinevaid kasutusvariante, nt õppijakeelt ja kirjakeelt. Analüüsitulemuste põhjal väidan, et õppijakeeles kajastuvad eesti kirjakeelele tüpoloogiliselt omased üldised arengusuunad, mis on ühelt poolt seotud nominatiivi kasutusala laienemise ja teisalt genitiivi taandumisega ning vastanduse nominatiiv/partitiiv kujunemisega.

Järgnevalt käsitlen lähemalt partitiivi ja nominatiivi kasutusala laienemist ning genitiivi kasutusala kitsenemist objektkäände funktsioonis. Küsimus on eesti grammatiliste käänete vaheldumises ja selle protsessi tekstisageduses, mis räägib partsiaal- ja totaalobjekti käändegrammatika võimalikest arengutest.

3. Küsimusi ja tähelepanekuid seoses grammatiliste käänete asendustega

3.1. Kas partitiivi kasutusala laieneb?

Diakroonilisest aspektist võib objektkäänete kasutuses täheldada lainelisi arenguid. Näiteks Aet Lees (2004) on välja toonud muutused objekti käändekasutuseelistustes 17.–18. sajandi

eesti- ja soomekeelsetes piibli tõlgetes, millest nähtub, et eesti-keelsetes tõlgetes on sel perioodil olnud valdav kääne partitiiv, millele Lees on vastandanud akusatiivi, s.o mittepartitiivi. Annika Kilgilt saadud andmetele tuginedes võib 1715. aasta Uue Testamendi tõlkes ja 1739. aasta põhjaeestikeelses piibli tõlkes märgata objektikäänete kasutuses rida asendusi, mis ei näita sugugi partitiivi kasutusala laienemist. Pigem jäävad silma just need juhtumid, kus partitiiv on asendatud nominatiivi ja genitiiviga. Valter Tauli on F. Tuglase, A. Mälgu ja T. Künnapase teoste esimese viiekümne lehekülje käändestatistika alusel leidnud, et sel perioodil kasutati eesti keeles enamasti partitiivobjekti ning totaalobjekt oli erand (Tauli 1968: 87).

Merilin Miljani objektnoomeni käändevahelduse testid näitavad, et tänapäeval eelistab eesti keelt emakeelena kõneleja partitiivsele objektile järjest enam genitiivset (eriti *da*-infinitiivi sisaldavas lauses) või nominatiivset ning vastupidi (Miljan 2008: 135–136). Informantide enesehinnangud näitavad, et objektikäände valik langetatakse enamasti konteksti ja mitte keelereegli põhjal, samas pole mõned informandid enesehinnagut kirjalikult üldse fikseerinud. Järelikult on nad leidnud objektnoomeni käändevaliku niivõrd loomuliku olevat, et ei ole vajalikuks pidanud oma eelistusi seletada. Pean seda Miljani testiainesest saadava teabe tugevaks küljeks: eesti keelt emakeelena kõnelevad informandid on tunnistanud eestlasele omast loomulikku keelekasutust, mida iseloomustab nominatiivse objektnoomeni intuiitiivne eelistamine partitiivsele.

Teiselaadse testidel põhineva objektikäände uurimuse autor on Martin Ehala. Eesti ja vene koolide õpilaste hulgas läbi viidud vormimoodustuse ja tajutesti analüüs näitab, et mitte-eesti emakeelega kõnelejal on objektnoomeni käändevaheldus avaram kui eestlastel, et mitte-eestlased kasutavad üle nominatiivi ja genitiivi, et nii eestlased kui mitte-eestlased eelistavad kasutada partitiivi ühtmoodi agarasti genitiivi ja nomina-

tiivi asemel (vt Ehala 2009b: 191–192). Käändevaliku suundade määramiseks on Ehala kasutanud Cronbachi testi ning erinevaid statistilisi teste, mis kinnitavad partitiivi ülekasutust enamasti genitiivses, vähemal määral nominatiivses kontekstis. Kuna see nähtus on samaväärselt omane nii mitte-eestlastele kui ka eestlastele, siis on Ehala järeldanud, et tegu on keelekontaktidest sõltumatu süstemaatilise kõrvalekaldega (Ehala 2009b: 193). Seega on kinnitust leidnud fakt, et partitiiv allutab genitiivse ja nominatiivse objektnoomeni alasid ning emakeelekõneleja ei taju siin normivastasust. Ega peagi tajuma, sest tegu on eesti keelele iseloomuliku soomeugrilise tendentsi, mitte normivastasusega, millest on põhjalikult kirjutanud juba Karl Kont (1963), taunides partsiaalobjekti ületaotluse vastu alustatud võitlust.

Konti arvates teeb objekti käändevormi valiku kirjeldamise keeruliseks see, et valik ei ole üheselt seotud põhisõna semantika ja grammatikaga või objektnoomeni semantikaga. Ta märkib, et objekti totaalsuse ja partsiaalsusega on tihedalt seotud põhisõna määratus/määramatus (definiitsus/indefiniitsus), mis on „kaasnähtuseks totaalsuse ja partsiaalsuse vaheldumisel kui kompleksnähtusel. Liivi ja eesti keeles kasutatakse partsiaalobjekti palju sagedamini kui teistes läänemeresoome keeltes ja sageli seal, kus ootuspärane oleks totaalobjekt (tegevus näib olevat terminatiivne ja objekti mõiste terviklikult allutatud verbi tegevusele)“, nt *Millega olete seda ära teeninud; Jätsin seda tegemata*. „Partitiiv viitab siin peamiselt põhisõna tegevuse määramatusele, mittekonkreetsusele .. Hoopis vähem oluline on asjaolu, kas kõnesolev tegevus on tulemuslik ja ajalisel lõpetatud või mitte. .. Eesti keeles on paljudel juhtudel seda laadi indefiniitsust väljendavat partsiaalobjekti peetud keeleveaks või taunitavaks võõrapärasuseks. Eesti vanemas kirjakeeles .. esineb partsiaalobjekti massiliselt seal, kus see pole millegagi õigustatud. Võitlust partsiaalobjekti ületaotluse

vastu alustati soome grammatika positsioonidelt“. Kont jätkab Johannes Aavikuga diskuteerides, et nii kuulutati vigasteks ka rahvapärased ja õiged partsiaalobjekti kasutused, mis kuuluvad enamasti fraseoloogilise objekti alla (Kont 1963: 98–99), nt *kes teisele auku kaevab, see ise sisse langeb*. Partitiivi leidub hulgaliselt tänastes analüütilistes verbides ja verbikesksetes püsiühendites, nt *nalja viskama, ettevalmistusi tegema* mis on leidnud kinnitust ka *saama*-verbi objekti käändekasutuse võrdluses ajakirjandustekstide ja *Eesti kõnekäändude ja fraseologismide andmebaasi* fraseologismide põhjal (vt Esilon, Öim 2010: 13 jj).

Partsiaalobjekti ületaotlusega seoses on Kont rõhutanud August Saareste seisukohti, kes ei pidanud seda nähtust mitte võõrapärasuseks, vaid eesti keele erijooneks: tegu pole ei objekti osalisusega ega tegevuse kursiivsuse ja irresultatiivsusega, vaid tegevuse määramatuse (indefiniitsuse) märkimisega. Samas võivad kõik need funktsioonid olla omavahel seotud, mis raskendab partsiaalobjekti kasutuse analüüsi, sidudes selle suuresti uurimiseesmärkide ja interpreteerimisega (vt Kont 1963: 100). Kont rõhutab, et fraseoloogiline partsiaalobjekt nagu ka personaalpronoomenite partsiaalobjekt on tuntud kõikides läänemeresoome keeltes (ennekõike liivi ja eesti keeles) ning järeldab, et tegu on oma, algupäralt puht läänemeresoomelise nähtusega, mitte „võõrmõjudel tekkinud keeletunde nürinemise tagajärg“, mida võib näha vene keele mõjul vadjal, karjala ning vepsa keeltes ilmnevas akusatiivistumises, kus partsiaalobjektiga on asendatud totaalobjekt, kui tegu on elusolendiga (vt Kont 1963: 101; akusatiivistumise kohta karjala murretes praegusajal vt Naumova 2006). Kui eesti keeles leiduks venemõjulist akusatiivistumist, siis peaks see hakkama regulaarselt ilmema kõigepealt vene emakeelega õppijate keelekasutuses, ent õppijakeele objekti-fraaside andmed seda ei kinnita. Partitiivi kasutamine geni-

tiivi ja nominatiivi asemel pole seotud referendi elusa/elutu tunnusega. 120 – genitiiviasenduse hulgas oli vaid 14 elusa referendiga objekti, 50 nominatiiviasenduses oli elusaid referente 7. Seega on tegu üksikjuhtumitega. Peale selle on genitiivi asendamisel partitiiviga elusat objekt eelistatud tähistada personaalpronoomeniga, mille partitiivset kasutust ei saa käsitleda asendusena. Konti andmeil on läänemeresoome keeltes personaalpronoomeneid ikka partsiaalobjektina kasutatud: *Sai sind sellesse komiteesse pandud* (vt Kont 1963: 103). Võimalik on ka genitiivivorm (nt *võtsite mind/mu tööle*), kuid siingi pole tegu mitte kõikidele personaalpronoomenitele omase nähtusega, vaid ainsuse esimese ja teise isikuga seotud võimaliku käändevaheldusega. Vastupidine asendus – genitiivi kasutamine traditsioonilise partitiivse personaalpronoomeni asemel – on harv. Analüüsitud eesti õppijakeele objektifraaside hulgas avaldub see vaid paaris kirjakeele seisukohalt ebareeglipärasas kasutuses: <ei> *näe *sinu ~ sind* – eitus nõuab partitiivi.

Mis puudutab elutu referendiga partitiivse objekti kasutamist õppijakeeles genitiivse asemel, siis selgub, et see on väga sageli seotud aspektverbidena kasutatud semantiliselt tuhmunud verbidega: *andma* (28 korda), *valima* (11), *võtma* (9), *leidma* (6), *saama* (6), *kaotama* (4), *kirjutama* (4), *lõpetama* (4), *saavutama* (4), *tegema* (4), *arutlema* (2), *neelama* (2), *ohverdama* (2), *ostma* (2), *otsustama* (2), *panema* (2), *pidama* (2), *sööma* (2), *veetma* (2). Seetõttu on loogiline, et 120-st genitiivi asendusest partitiiviga on rohkem kui pool sellised, milles pole tegu ebareeglipärasusega, vaid eesti keelele omase käändevariatiivsusega. Objektifraasi põhjaks olevad tuhmunud tähendusega verbid võimaldavad situatsiooni interpreteerida nii definiitse kui indefiniitsena, nt *andis võimaluse ~ võimalust, põhjuse ~ põhjust, vastuse ~ vastust, annab tulemise ~ tulemust, õiguse ~ õigust; valib elukutse ~ elukutset, teise eriala ~ teist eriala; võtavad suure koormuse ~ suurt koormust, võttis suuna ~ suunda, suure hoo ~*

suurt hoogu; leidsin ühise keele ~ ühist keelt; said agressiivse vormi ~ agressiivset vormi; kaotas oma armastuse ~ armastust; kirjutas oma arvamuse ~ arvamust, kirjutas alla lepingu ~ lepingut; lõpetab õppimise ~ õppimist, lõpetan selle ülikooli ~ seda ülikooli jne. Põhjalikku ülevaadet eesti keele perfektiiv- ja partitiivverbide tähendusnihetest ning nende üleminekust aspektverbide klassi (vt Vaiss 2004), mille tulemusel totaal- ja partsiaalobjekti käändevaheldus laieneb. See protsess on seotud nii objekti kvantifitseerimise kui ka tegevuse terminatiivsuse, kursiivsuse ja resultatiivsusega.

Ebareeglipärane genitiivse objekti asendamine partitiivsega on seotud objektnoomeni ja täiendsõna ühildumisega. Näited viitavad sellele, et eesti keele õppija ei oska rakendada ühildumisreegleid ja/või kahtleb objektikäände valikul: **annab hea võimalust*, **kirjutab sobivaid punkti*, **saab niisuguse elukutset*, **saada parema töökohta*, **teatada sulle ühe hea uudist*, **tuua soovimatu tagajärge*, **valisin just selle ametit*, **võib tekitada katastroofilise tagajärge*, **võita mingi suurema ja tugevama riiki*, **võttis tol ajal suure hoogu* jne. Samas ei ole õppijatele muret teinud rektsioon.

50 korda on partitiivi kasutatud nominatiivse objekti asemel, milleks on 7 juhul –personaalpronoomen (vt eespool). Valdavalt on see seotud partitiivi/nominatiivi käändevariatiivsusega, mille puhul saab personaalpronoomenit kasutada mõlemas vormis, nt *kinnitati*, *pahteldan*, *unustan need* ~ *neid*. Elutu objekt esines kõige sagedamini koos semantiliselt tuhmunud verbiga *saama* (5), *andma* (3), *tegema* (3), *võtma* (3), harvem verbidega *kirjutama*, *meeldima*, *ostma*, *pakkuma* ja *sooritama*: *saada võimalus* ~ *võimalust*, *anti preemia* ~ *preemiat*, *kirjutada vastus* ~ *vastust*, *osta endale auto* ~ *autot* jne. Mõnel juhul on nominatiivi asendamine partitiiviga seotud ühildumisreegli vastu eksimisega: **saada hea haridust*, *kõrgharidusi*, *tulemuste*; **meeldib uued sõpru*; **sooritada kõik eksameid*.

Nominatiivi ja genitiivi asendamine partitiiviga toob järjekordselt esile, kuivõrd situatsiooni- ja kontekstitundlik ning interpretatiivne on totaal- ja partsiaalobjekti valik, eriti koos sageli kasutatavate semantiliselt tuhmunud verbidega. Kuna õppijakeeles eelistatakse eesti keele sagedasemat sõnavara (sh verbe), siis peegeldab õppijakeel kõige muu hulgas ka kirja-keelele omaseid suundumusi verbi semantikas ja leksikaalgrammatilises perifeerias.

3.2. Kas nominatiivi kasutusala laieneb?

Nominatiivi kasutusala laienemine ei ole olnud nii laialt diskuteeritud kui partitiiviga seotud küsimused. Enamasti on sellest räägitud seoses grammatiliste käänete variatiivsusega. Willem Grünthali (1941) nominatiivuurimus sisaldab ülevaadet eesti objektikäänete keerukusest, vormi ja funktsioonide lahususest. Erinevalt partitiivist märgib soomeugri line nominatiivne objekt tegevuse määratust, konkreetset. Arutlusainet on andnud küsimus, kas verbi definiitsus avab tee käände markeerituse juurde, kas eesti ja teistes soomeugri keeltes on verbi definiitsuse ja objektikäände markeerituse seosed samalaadsed, kas ja kuidas see on ühendatav verbi transitiivsuse kaoga, mil määral on need protsessid mõjutanud subjekti-objekti staatust, kas eesti keeles on akusatiiv jne (ammendava ülevaate saamiseks vt Hiietam 2003; Tamm 2004; Miljan 2008). Totaal- ja partsiaalobjekti käändevariatiivsust on seostatud edastatava tähendusega (sh aspektiliste tähendustega), lause pragmaatika ja teema-reema liigendusega, põhisõna semantika ja grammatikaga, verbiklassidega (nt Kont 1963; Pihlak 1985; Eslon, Pihlak 1993; Vaiss 2004; Tamm 2003, 2004) jm. Kuna senistes uurimustes on ühel või teisel teel jõutud tõdemuseni grammatiliste käänete hajususest, nende erisuse neutraliseerumisest ning kaost (nt Rajandi, Metslang

1979; R. Grünthal 2003; Eslon, Õim 2010), siis on kohane Tauli soovitus järgida elavat keelekasutust (st loomuliku morfoloogia põhimõtteid) ning kallutada grammatiliste käänete valik nominatiivi kasuks (Tauli 1968: 63), mis nt eesti-inglise kakskeelsete laste keelekasutuses (Torn 2003), täiskasvanute eesti õppijakeeles ja argisuhtluses (vt Pool 2007, 2010; Eslon, Matsak 2009; Eslon 2010) realselt toimibki. On täiesti tavaline, et keelekasutajad (sh keeleõppijad) püüavad grammatilist süsteemi enda jaoks selgemaks ning loomulikumaks muuta (Habicht jt 2006: 610), sest kõik loomulik on keelele tüpoloogiliselt omane, seotud sagedasemate ja markeerimata keeleüksustega.

Ehala vormimoodustuse ja tajutesti analüüsitulemused näitavad, et nominatiivi ülekasutus genitiivi funktsioonis on iseloomulik vene emakeelega kooliõpilastele; eestlaste keelekasutuses on see juhuslik ega moodusta statistiliselt ühtset suundumust (vt Ehala 2009b: 196). Samas on nominatiivi eelistamine partitiivi funktsioonis eesti emakeelega inimeste arvates loomulik. Ehala väitel on tegu „trendiga, mille levik on otseselt seotud keelekontakti tugevusega. Seosed ei ole küll väga tugevad ega erinevused eri kasutajate vahel väga suured, kuid need on läbivalt ühesuunalised ja enamikul juhtudest ka statistiliselt usaldusväärsed” (Ehala 2009b: 197).

Ehala analüüsis pean oluliseks seda, et vormimoodustus- ja tajutestiga on õnnestunud välja tuua emakeelsete kõnelejade nominatiivikasutuse avardumine partitiivi aladele, mis on läbivalt ühesuunaline. Samas kahtlen selle tendentsi võrnmõjulisuses, sest nagu Ehalagi märgib, ei ole seosed väga tugevad ja erinevused eri keelekasutajate vahel on väga suured. Järgnevalt püüan oma seisukohta põhjendada. Ühelt poolt toetun eesti kirjakeele ja õppijakeele üldise käändesageduse ning kasutuseelistuste võrdleva korpusanalüüsi

tulemustele (vt Eslon, Matsak 2009; Eslon 2010), samuti *saama*-verbi objektikäände käsitlusele (Eslon, Öim 2010), ning teisalt vene emakeelega inimeste eesti objektifraasi käändeasenduste analüüsi tulemustele.

Käändekasutusanalüüs tugineb *Eesti Keele Instituudi tekstikorpuse* ja *Eesti vahekeele korpuse* tekstidest saadud kahele valimile, mille võrreldavus oli eelnevalt statistiliselt tõestatud (vt Eslon, Matsak 2009: 81–84, 87). Tekstisagedus kinnitab veel kord, et grammatilised käänded on eesti keele käände paradigmas tavapärased ja kõige sagedasemad, et nii nominatiivi- genitiivi-partitiivi kui ka semantilisi käändeid kasutatakse eesti kirjakeeles ja õppijakeeles analoogselt ning sõltuvuses substantiivi semantikast. Kui rääkida vaid grammatilistest käänetest, siis esineb nominatiivi ja partitiivi enam-vähem proportsionaalselt nii kirjakeele ja õppijakeele valimis (vt Eslon, Matsak 2009: 90–97) kui ka *saama*-verbi objektikäände andmete hulgas (Eslon, Öim 2010: 13–15). Sagedaseim kääne on partitiiv, mis edestab nominatiivi, kuid mitte oluliselt. Olenevalt sõna semantikast võib kõige sagedasem grammatiline kääne olla hoopis nominatiiv. Näiteks eesti kirjakeele sagedasemaid nimisõnu *inimene*, *sõna*, *elu* ja *aeg* on kasutatud nominatiivis ja partitiivis kas võrdselt (*inimene* esineb nominatiivis 2124 korda ja partitiivis 2125 korda, *aeg* vastavalt 1352 ja 1382 korda) või edestab nominatiiv partitiivi enam kui poole võrra (*sõna* – vastavalt 750 ja 320 korda, *elu* – 192 ja 86 korda) (vt Eslon 2010: 47).

Veelgi selgemalt tuleb nominatiivi valdavus esile siis, kui uurida, missuguseid grammatilisi konstruktsioone eelistavad oma tekstiloomes emakeelekõneleja ja keeleõppija ning missuguseid grammatiliste ja semantiliste käänete vorme neis konstruktsioonides reaalselt kasutatakse (vt Eslon, Matsak 2009: 86–87). Keelekasutuseelistused näitavad eesti keele kasutus-

variantide morfosüntaktilise arengu ühis- ja erijooni, viidates võimalikele suundumustele käändevariatiivsuses ning paradigma lühenemisele. Morfosüntaktilise ja klasteranalüüsi tulemusel õnnestus ajakirjandus- ning õppijakeele tekstidest leida kolmest komponendist koosnevaid konstruktsioonimustreid, mis jagunevad kolme gruppi: stereotüüpsed, tavapärased ja ajakirjandus- või õppijakeelele tüüpilised konstruktsioonid ning nende leksikaalgrammatilised variandid (vt lähemalt Eslon 2009: 38–45). Selgus, et õppijakeele ja kirjakeele ühisjooneks on see, et mõlemas eelistatakse tavapäraseid ehk samalaadseid konstruktsioone, milles neli sagedasemat nimisõna *inimene*, *sõna*, *elu* ja *aeg* on valdavalt nominatiivis (vt Eslon 2009: 40–42).

Järgnevalt vaatleme grammatiliste käänete kasutamist eesti õppijakeele objektifraasis. Käändeasenduse analüüsi tulemused toetavad eespool kirjeldatud nominatiivi kasutusala üldist avarustumistendentsi: kõige sagedamini asendab nominatiiv partitiivi (506 juhtumit) ning oluliselt vähem genitiivi (190 korda), vt tabel 1. Arvestades ka eespool kirjeldatud uurimistulemusi, paistab vastupidiselt Ehala tulemustele, et vene emakeelega õppija asendab nominatiiviga pigem partitiivi, mitte genitiivi. Samas näitab objektifraasi üldvalimi vormisagedus, et nominatiivset (1799 kasutusnäidet) ja genitiivset (927 kasutusnäidet) objekti on selgelt vähem kui partitiivset (5631 kasutusnäidet), mis on eesti keelele iseloomulik asjade seis. Miks kasutatakse nominatiivi niivõrd laialdaselt partitiivi asemel, mis selle on tinginud? Analüüsitulemustest nähtub, et partitiivi asendamine nominatiiviga on õppijakeele objektifraasis olnud väga selgepiiriline.

1. Asenduste puhul on enamasti tegu ebareeglipärasustega, valdavalt verbirektsiooni vigadega (320 korda 506-st partitiivi asendusest): **armastab meie maa, oma abikaasa*; **armas-*

tame kingitused, kokteilid; *armastan ilus muusika, kiirus; *jõime ainult mahl, õlu, vein, õlud, *jõid limonaad; *kasutan mobiiltelefon; *kasutasin sõnaraamat; *kirjeldab oma emotsioonid; *kirjutan kiri; *kuulame laulud, rokk; *küllastasime ka teised riigid; *loevad ilukirjandus; *lugesime ajaleht jne. Rektsioonivigadega kaasneb ühildumisvigu (39 juhtumit): märkama *olulisemaid probleemid, näeb ainult *halb asjad, näitas *tõelist patriotism, saavutasid *häid tulemus, tahad *uue töökoht, tahavad *samasugune õigused jne. Eksitakse kvantorraasi kasutuses (14 korda): võtan *natukene lilled, õppima *palju sõna, müüakse *palju toidukaubad jne, samuti üldeituse puhul (8 korda): *ei unusta oma laulud, *ei armasta reeturlus ja valetamine, *ei kujuta ette hommikusöök, *ei mäleta eesti keel jne. Esines ka muid eksimusi (25 juhtumit), mille seas hakkavad silma vales sõnajärjest (*säilitama mina see kiri, *tegeme me gruppitöö jne) või rektsioonivigadest tulenevad (*armastasid lilled, *laulsid rahvalaulud, *luges raamat, *mängib roll, *toob rikkus jne) subjekti ja objekti piirimail olevad juhtumid. Ka polnud õppijad selgeks saanud mõningate verbide tähenduserinevusi (nt verbide oskama ja teadma, tundma ja teadma, vahtima ja vaatama eristamine; muutuma ~ muutma, otsima ~ ostma ja omama ~ omandama paronüümia), sõnade semantilise seose seaduspära (nt *näen soov, *omama riided, *valmistada paanika) jm.

2. Ebareeglipärasused partitiivse objekti asendamisel nominatiivsega ilmnevad vaid teatud verbide puhul: õppima (42 korda), armastama (32), vaatama (32), teadma (23), tegema (21), tahtma (19), lugema (19), nägema (16), kasutama (15), näitama (12), kuulama (12), laulma (11), küllastama (10), mäletama (8), andma (8), jooma (7), tundma (7), arendama (6), kirjutama (6), sööma (6), võrdlema (5), tähistama (5), omama (5), märkama (5), kontrollima (4), korraldama (4), müüma (4), ostma (4), otsima (4), saama (4), vajama (3), valima (3),

unustama (3), *pidama* (3), *oskama* (3), *mängima* (3); kaks korda on kasutatud verbe *analüüsima*, *ehitama*, *hoidma*, *kandma*, *kartma*, *kirjeldama*, *ette kujutama*, *kuulma*, *meenu-tama*, *sisaldama*, *tellima*, *toetama*, *valmistama*, *veetma*, *võtma*, *väljendama*, ülejäänud verbid olid ainukordsed. Tegu on põhiliselt partitiivverbidega, mille venekeelsed vasted nõuavad objekti kasutamist akusatiivis ehk *винительный-кäändes*³. Elusa referendi märkimisel langeb see objekti-kääne kokku vene *родительный*-käändega ja elutu puhul nominatiiviga (*именительный*-kääne), mis kajastub ka rektsioonimallis, vrld: *õppima* mida? – *учить, изучать* что?; *armastama* keda-mida? – *любить* кого-что?; *vaatama* keda-mida? – *смотреть* на кого-что?, что?; *teadma* mida? – *знать* кого-что?; *tegeta* mida? – *делать* что? jne. Objektifraasi ebareeglipärased partitiivi asendused nominatiiviga on eranditult seotud elutu referendiga. Tegu on emakeele mõjuga ja sellest tingitud objektireeglite lihtsustamisega. Nagu olen varem märkinud (vt Eslon 2008: 60–62), võib siin oma osa mängida eksitav õpetamis-/õppimisstrateegia, mille kohaselt samastatakse elutu referent käändeküsimuste kokkulangemise tõttu automaatselt eesti keele nominatiivse totaalobjektiga. See on mugav ja samas kooskõlas emakeelse otsesihitise rektsioonimalliga.

3. Kui eraldada 506 partitiivi asendusest nominatiiviga kõik ebareeglipärased juhtumid, jääb alles 100 eesti keele objekt-noomeni käändevariatiivsuse kasutusnäidet, mille puhul vene emakeelega õppija on eelistanud nominatiivi. Võrdluseks: vastupidiseid näiteid ehk partitiivi vormi eelistamist nominatiivile esines poole vähem – vaid 50 (vt tabel 1). See

³ Erand on verb *tahtma* 'хотеть', mille puhul on võimalik *винительный*- ja *родительный*-käände variatiivsus (*хотеть* кого-что/кого-чего).

viitab kõneks olnud keeltevahelise mõju ja lihtsustamise analoogiapõhisele avardumisele, sh ka elusale referendile: *sõber, inimene, rootslane, õpilane, õpetaja, naine, Mari, abikaasa* jne. Niisuguses analoogiapõhises lihtsustumises kajastub eesti keele käändekasutuse üldine omadus – nominatiivi eelistamine. Näiteks: *analüüsid faktid ~ fakte, analüüsitakse töö ~ tööd, annavad suured võimalused ~ suuri võimalusi, arendati rasketööstus ~ rasketööstust, *areneda <arendada> oma teadmised ~ <arendada> oma teadmisi, kasutame pesuvahendid ~ pesuvahendeid, kasvatan lilled ~ lilli, korraldatakse mitmesugused laulupeod ~ mitmesuguseid laulupidusid, leian uued sõbrad ~ uusi sõpru, *saata <saada> töökoht ~ töökohta, stimuleerib noored inimesed ~ noori inimesi, tahaksin õppitada teised inimesed ~ teisi inimesi, teatada sulle huvitav uudis ~ huvitavat uudist, teed vastavad järeldused ~ vastavaid järeldusi* jne. Mõningatel juhtudel on varieerumine võimalik kõigis kolmes grammatilises käändes nagu *teed vastavad järeldused ~ vastava järelduse ~ vastavaid järeldusi*.

4. Genitiiv on objektifraasis asendatud nominatiiviga 190 korral, vastupidiseid kasutusnäiteid esines vaid 19. Ka see eesti õppijakeelele omane nähtus on seotud kindlate verbidega, mis kuuluvad enamasti aspektverbide hulka: *lõpetama* (26 korda), *valima* (19), *andma* (16), *saama* (15), *tegema* (15), *murdma* (12), *kirjutama* (11), *leidma* (8), *ostma* (8), *võtma* (6), *panema* (4), *veetma* (3), *võitma* (3), *lugema* (3), *tahtma* (2), *tooma* (2), *unustama* (2), (*selgeks*) *õppima* (2); ülejäänud verbid olid ainukordsed. Olenevalt aspektverbi grammatilisest vormist ning kasutuskontekstist võib objekt olla nii nominatiivis, genitiivis kui partitiivis (nt *anna vastus ~ vastust, lõpetati töö ~ tööd, andis välja kalendri ~ kalendrit, tuleb leida vastus, leiab vastuse ~ vastust* jne), samas kui õppijakeeles on valdav nominatiiv, kusjuures 165 juhtumit 190-st on ebareeglipärased ja 21 seotud käände-

variatiivsusega. Selle asenduse tulemusena muutub subjekti-objekti vaheline piir ebakindlaks, nt **andis välja kalender*, **annab ülevaade*, **avab hing*, **kirjutab teos*, **loob maailm*, **tapsi mees*, **teeb esileht*, **tekitas hea ja sõbralik atmosfäär*, **tõmbaks kast*, **vahetaks nimi*, **õõttis korv*, **õpioad muud erialad*. Enamasti on siin tegu õppija emakeele mõjust tingitud ebareeglipärasustega nagu ka partitiivi asendamisel nominatiiviga, kuna sageli kasutatud aspektverbide venekeelseteks vasteteks on otsesihistist nõudvad verbid, mille puhul objektnoomeni kasutatakse akusatiivis, nt *lõpetama* – кончатъ что, *valima* – выбирать кого-что, *andma* – давать что, *saama* – получать что jne. Toimib samalaadne emakeelemõjuline lihtsustumine, mida eespool kirjeldasime. Seejuures väärib tähelepanu, et protsess avardub ja laieneb diferentseerimatult nii elusale kui ka elutule referendile ja genitiivi kasutusalaadele. Keeltevahelisest mõjust tingitud protsess poleks nii edukas, kui sellel puuduksid eesti keeles eeldused. Pean siin silmas interpretatiivsust objekti kvantifitseerimisel, tegevuse määratlemisel definiitse ja indefiniitsena, aspektiliste tähenduste markeerimist objektnoomeni käändega (vt Larjavaara 1991; Rajandi, Metslang 1979; Esilon, Pihlak 1993; Sulkala 1996; Huumo 2006), substantiividega seotud muutemorfoloogilisi protsesse, laadivahelduse kadu, rööpvormide tunnistamist normipärasuseks, käändeparadigma lühenemist ja muutkondade lihtsustumist (vt Lüdelling jt 2000; Lüdelling, Evert 2003; Nikolaev, Niemi 2006; Kaalep 2010), nominatiivi-genitiivi-partitiivi eristuse neutraliseerumist kas vormihomonüümia tõttu või sellest mitteleonevalt, nt sõltuvuses kõneakti pragmaatikast ja lause infostruktuurist, vormide markeerituse/markeerimatuse seotusest nende tekstisageduse ja kasutuseelistustega (vt Esilon, Õim 2010).

Ehala (1997; vt ka 2009a) järgi vastab eesti käändeparadigma algvormile omastele tunnusjoontele kõige paremini mitmuse genitiivi vorm, kuna see on võimalikult vähe markeeritud, määrab muuttüübi ja tüveteisendused ning on aluseks vormisünteesile. Genitiivist lähtudes „on võimalik tuvastada enamik tüve morfoloogiliselt olulisi tunnuseid, välja arvatud astme- ja kujuvahelduslikkus“. Samas pole „nimetava käände puhul aga .. ükski loetletud tunnustest tuvastatav ja selle vormi algvormiks valimise kasuks kõneleb vaid ta markeerimatus“. Kõige markeeritum eesti kolmest grammatilisest käandest on partitiiv, kuna on nimetatud tunnuste suhtes negatiivne (vt Ehala 1997: 372–375). Järelikult sobib eesti käändesüsteemi tuumelemendiks kõige paremini genitiiv ning seetõttu on see ka muutuste suhtes kõige tundlikum vorm. Pealegi pole genitiivi põhifunktsioon objektilisuse edastamine, vaid possessiivsus. Sama kehtib ka soome keele kohta (vt Jaakkola 2004). Esinemissageduselt jääb eesti keele genitiiv alla partitiivile ja nominatiivile, kasutuseelistuse seisukohalt on olulised aga nimelt need kaks käänat, kusjuures mõlemad laiendavad oma kasutusala genitiivi arvelt (vt eespool). On loogiline järeldada, et morfoloogiliselt markeerimata genitiiv on muutumas marginaalseks, mis räägib selle vormi algvormiks sobimise kasuks. Niisugune on süsteemne taust, mis loob eeldused õppija emakeele sedavõrd laialdaseks mõjulepääsemiseks. Paistab, et kui eesti genitiiv objektikäände funktsioonist taandub, siis ei hakka seda kohta täitma mitte partitiiv, vaid nominatiiv, kuna nominatiiv on markeerimata kääne, partitiiv aga kõikide muutmorfoloogiliste tunnuste suhtes negatiivne ning seetõttu kõige enam markeeritud grammatiline kääne. Järelikult toimib keelesüsteemi sisene ning keelele tüpoloogiliselt omane kompensatoorne mehhanism, mille kajastustest pole vaba ükski keele kasutusvariant, sh eesti õppjakeel. Riho Grünthal on välja toonud, et liivi keele genitiivi-nominatiivi eristamatuse ja käändeparadigma lühenemise tõttu võeti appi

muud keelevahendid, ennekõike sõnajärg, mis kinnistus genitiivi-nominatiivi tähenduste edastamisega. Grünthal peab seda grammatilist süsteemi säilitavaks nähtuseks (R. Grünthal 2003). Eesti keele puhul on analoogseks nähtuseks nominatiivi ja partitiivi vastanduse stabiliseerumine, nominatiivi kinnistumine opositsiooni markeerimata liikmena, millel on avaram funktsionaalne potentsiaal kui partitiivil. Partsiaal- ja totaalobjekti eristamine substantiivi käändegrammatika alusel tuhmub märkamatuks. Seda kinnitab praegune, eriti suuline keelekasutus (*Oleks tore neile pakkuda selline tore üritus ~ sellise toreda ürituse ~ sellist toredat üritust; Meie eesmärk on tösta inimene ~ inimest esile* jne). Eeldatavasti muutub selle tulemusena lause infostruktuur, subjekti-objekti vahekord ning totaalsuse/partsiaalsuse eristamisel hakkavad järjest suuremat rolli mängima sõnajärg ja kõneakti pragmaatika.

Analoogseid arengusuundi võib märgata ka slaavi keeltes. Näiteks Mark Eliot Nuckols on uurinud transitiiivverbide objekti käändevariatiivsust vene ja tšehhi keeltes, mis kuuluvad akusatiivkeelte alla. Nuckols juhib tähelepanu verbi aspekti varieerumisele, millega kaasneb objektikäände varieerumine: ühelt poolt on kasutusel otsesihitist märkiv akusatiiv ja teisalt kaudobjekti obliikvakäanded genitiiv, datiiiv ja instrumentaal. Küsimus on selles, kuidas lause semantika (nt individuaalne tahteavaldus jm) mõjutavad verbi aspektivormi valikut ning kuivõrd oluline on seejuures diskursus ja lause (mõnikord ka teksti) teema-reema liigendus (vt Nuckols 2007: 1).

4. Kokkuvõte

Eesti õppijakeele ja kirjakeele käändekasutuse võrdlev analüüs ning tekstiloomes eelistatud konstruktsioonide käändevalikud on näidanud, et sõltuvalt substantiivi semantikast võib nomi-

natiiv olla tekstis kõige enam kasutatud kääne. Samas on kõige sagedasem objektikääne partitiiv, nominatiivi ja genitiivi kasutatakse selgelt vähem. Ka õppijakeele objektifraasi analüüs näitab, et partitiivi eelistatakse teistele grammatilistele käänetele (vt tabel 1). Üldjuhul pole siin tegu õppijakeele ebareeglipärasusega, ka ei saa väita võõrmõjulise akusatiivisumise olemasolu. Objekti märkimisel on partitiiv eesti keeles kõige markeeritum kääne. See tuleneb verbisemantika tuhmumisest tingitud protsessidest ning on seotud aspektverbide klassi avarumisega perfektiiv- ja partitiivverbide arvelt. Protsessiga kaasneb eesti keelele omane totaal- ja partsiaalobjekti käände vaheldus, mille abil kvantifitseeritakse objekti ning tegevuse definiitsust/indefiniitsust või terminatiivsust, kursiivsust ja resultatiivsust. Et aga emakeelekõnelejad ja eesti keelt teise keelena õppijad eelistavad aspektverbidega partsiaalobjekti, siis räägib see osutamisest tegevuse indefiniitsusele. Tegu on soomeugrilise ja eesti keelele tüpoloogiliselt omase nähtusega, mis kajastub ka eesti õppijakeeles. Järelikult ei saa absolutiseerida totaal- ja partsiaalobjekti vastanduse seost verbi aspektiliste tähenduste edastamisega.

Samas leidub õppijakeele objektifraasis käändeasendusi, mis on seotud kahe tendentsiga: nominatiivse objektnoomeni kasutusala laienemine partitiivi ja genitiivi arvelt ning genitiivse objektnoomeni kasutusala kitsenemine valdavalt nominatiivi kasutusala laienemise tõttu. Need tendentsid on kooskõlas eesti keele käänete tekstisageduse ja tekstiloomes eelistustega, kus valdavaks on muutumas nominatiiv.

Õppijakeele objektifraasis on nominatiiviga asendatud kõige sagedamini partitiivverbide objekti, millel on elutu referent. Tegu on enamasti ebareeglipärasusega (verbirektsiooni viga), mis tingitud keeltevahelisest mõjust, kuna kõikide kasutatud partitiivverbide venekeelsed vasted nõuavad otsesihitist aku-

satiivis (винительный *кого-что*), mille vorm elutu referendi korral langeb kokku nominatiiviga (mis? – что?). Toimub loomulik objektireeglite lihtsustamine, mis võib olla tingitud eksitavast õpetamis- või õppimisstrateegiast, kui käände-funktsioonide õpetamisel võetakse aluseks eesti ja vene keele käändevastavused, nende esiletoomisel aga tuginetakse käändeküsimustele. Vähem tulevad õppijakeele objektifraasis esile partitiivi asendused nominatiiviga ja ebareeglipärased genitiivse objekti asendused nominatiiviga. Partitiivi asendused nominatiiviga on enamasti käändevariatiivsuse juhtumid, ent õppija on eelistanud siiski kasutada nominatiivi. Ebareeglipärased genitiivse objekti asendused nominatiiviga viitavad keeltevahelise mõju ja lihtsustamise analoogiapõhisele avarandumisele, sh ka elusa referendiga objektile. Käände-asendustega kaasnevatest nähtustest on olulisim piiri kadumine subjekti ja objekti vahel, mis tuleneb õppijakeele vigasest sõnajärjest – eksimine peamiselt V2-reegli vastu.

Keeltevahelisest mõjust tingitud protsessid ja nende analoogiapõhine laienemine ei saaks olla niivõrd selgepiiriline, kui selleks puuduksid eesti keeles sisemised eeldused: ühelt poolt genitiivi taandumine objektikäände funktsioonis, marginaalseerumine algvormina ja teisalt vastanduse nominatiiv/partitiiv kujunemine.

Kirjandus

Aikhenvald, Alexandra Y. 2007. Grammars in contact: A cross-linguistic perspective. – Grammars in Contact: A Cross-Linguistic Typology / Ed. by Alexandra Y. Aikhenvald, Robert M. W. Dixon. Explorations in Linguistic Typology Series 4. Oxford: Oxford University Press, 1–66.

Aikhenvald, Alexandra Y. 2002. *Language Contact in Amazonia*, Oxford: Oxford University Press.

Ehala, Martin 1997. Eesti morfoloogia olemus. – *Keel ja Kirjandus* 6, 370–383.

Ehala, Martin 2000. Second Language Learners' impact on the Structure of Estonian. – *Languages at universities today and tomorrow* / Ed. by Kiira Allikmets. Tartu: Tartu University Press, 20–32.

Ehala, Martin 2009a. Linguistic strategies and markedness in Estonian morphology. – *Sprachtypologie und Universalienforschung* 1/2, 29–48.

Ehala, Martin 2009b. Keelekontaktide mõju eesti sihitiskäänete kasutamisele. – *Keel ja Kirjandus* 3, 182–204.

Eslon, Pille, Ants Pihlak 1993. *Vid i vrem'a*. Tallinn: TPEDI kirjastus.

Eslon, Pille 2008. Käänevormide kasutussageduse võrdlus eesti õppijakeeles ja kirjakeeles. – *Õppijakeele analüüs: võimalused, probleemid, vajadused* / Toim. Pille Eslon (Tallinna Ülikooli eesti filoloogias osakonna toimetised 10). Tallinn: TLÜ kirjastus, 31–66.

Eslon, Pille 2009. Eestikeelses tekstiloomes eelistatud konstruktsioonid ja käänevormid. – *Korpusuuringute metodoloogia ja märgendamise probleemid* / Toim. Pille Eslon, Katre Õim (Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 11). Tallinn: TLÜ kirjastus, 30–53.

Eslon, Pille, Erika Matsak 2009. Eesti keele kasutusvariandid: korpusest tulenev käänevormide võrdlev analüüs. – *Eesti Rakenduslingvistika Ühingu aastaraamat 5* / Toim. Helle Metslang, Margit Langemets, Maria-Maren Sepper, Reili Argus. Tallinn: Eesti Keele Sihtasutus, 79–110.

Eslon, Pille 2010. Muutustest eesti keele grammatiliste käänete kasutamisel. – *Lähivõrdlusi* 19 / Peatoim. Annekatrin Kaivapalu, toim. Eve Mikone, Kirsti Siitonen, Maria-Maren Sepper. Tallinn: Eesti Rakenduslingvistika Ühing, 38–60. doi:10.5128/LV19.03.

Eslon, Pille, Katre Õim 2010. Objektikäänete kasutamisest sageduse ja markeerituse seisukohalt. – ESUKA 1–2, 69–89.

Grünthal, Riho 2003. Finnic adpositions and cases in change. *Suomalais-Ugrilaisen Seuran toimituksia* 244. Helsinki: Finno-Ugrian Society.

Grünthal, Willem 1941. Itämerensuomalaisten kielten yksikön nominatiivi objektin edustajana aktiivin yhteydessä. *Lauseopillinen tutkimus*. *Suomalaisen Kirjallisuuden Seuran toimituksia* 218. Helsinki: Suomalaisen Kirjallisuuden Seura.

Habicht, Külli, Leelo Keevallik, Ilona Tragel 2006. Keele muutumine kasutuskontekstis. – *Keel ja Kirjandus* 8, 609–625.

Heine, Berndt, Tanja Kuteva 2008. Constraints on contact-induced linguistic change. – *Journal of language contact – THEMA* 2. *Language Contact and the Dynamics of Language: Theory and Implications*, 57–89. <http://www.jlc-journal.org>, 8.07.2008.

Hiietam, Katrin 2003. Definiteness and Grammatical Relations in Estonian. PhD Thesis. University of Manchester.

Huumo, Tuomas 2006. Kvantiteetti ja aika I: Nominaalisen aspektin näkökulma suomen objektin ja subjektin sijamerkintään. – *Virittäjä* 4, 504–538.

Jaakkola, Minna 2004. Suomen genitiivi. *Suomalaisen Kirjallisuuden Seuran toimituksia* 995. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kaalep, Heiki-Jaan 2010. Mitmuse osastav eesti keele käändesüsteemis. – *Keel ja Kirjandus* 2, 94–111.

Kaalep, Heiki-Jaan 2009. Kuidas kirjeldada ainsuse lühikest sisseütlevat kasutamisandmetega kooskõlas? – *Keel ja Kirjandus* 52, 6, 411–425.

Kaivapalu, Annekatrin 2005. Lähdekieli kielenõppimise apuna (*Jyväskylä Studies in Humanities* 44). Jyväskylä: Jyväskylän yliopisto.

Kont, Karl 1963. Käändsõnaline objekt läänemeresoome keeltes. *Eesti NSV Teaduste Akadeemia Keele ja Kirjanduse Instituudi uurimused* IX. Tallinn: Eesti Riiklik Kirjastus.

Krall, Ingrid 2008. Eesti keele käändevormide moodustamise ja kasutamisega seotud probleeme vene üliõpilaste kirjalikes töödes. – Õppijakeele analüüs: võimalused, probleemid, vajadused / Toim. Pille Eslon (Tallinna Ülikooli eesti filoloogia osakonna toimetised 10). Tallinn: TLÜ kirjastus, 137–195.

Larjavaara, Matti 1991. Aspektuaalisen objektin synty. – Virittäjä 95, 4, 372–408. http://www.kotikielenseura.fi/virittaja/hakemistot/jutut/1991_372.pdf, 27.02.2010.

Lees, Aet 2004. The case of the object in Early Estonian and Finnish Texts. – Proceedings of the 2004 conference of the Australian Linguistics Society. <http://ses.library.usyd.edu.au/bitstream/2123/13/1/ALS-20050630-AL.pdf>, 6.11.2008.

Lüdeling, Anke ja Stefan Evert 2003. Linguistic experience and productivity: corpus evidence for finegrained distinctions. – Proceedings of the Corpus Linguistics 2003 Conference / Ed. by Dawn Archer, Paul Rayson, Andrew Wilson, Tony McEnery, 475–483. <http://purl.org/stefan.evert/PUB/LuedelingEvert2003.pdf>, 27.02.2010.

Lüdeling, Anke, Stefan Evert, Ulrich Heid 2000. On measuring morphological productivity. KONVENS-2000 Sprachkommunikation / Ed. by Werner Zühlke, Ernst G. Schukat-Talamazzini. VDE Verlag, 57–61. <http://purl.org/stefan.evert/PUB/LuedelingEvertHeid2000.pdf>, 27.02.2010.

Metslang, Helle, Ingrid Krall, Renate Pajusalu, Kristi Saarlo, Elle Sõrmus, Silvi Vare 2003. Keelehärm. Eesti keele probleemseid piirkondi. Tallinn: TPÜ Kirjastus.

Miljan, Merilin 2008. Grammatical case in Estonian. PhD Thesis. University of Edinburgh.

Naumova, Marina 2006. Glagolnoie upravlenie v livvikovskom narechii karelskogo iazyka. Avtoref. dis .. kand. filol. nauk. Petrozavodsk.

Nikolaev, Alexandre, Jussi Niemi 2006. Nominien paradigmaattistuminen suomessa. Millä rakenteellisilla ehdoilla kielenkäyttäjät sitovat potentiaaliset nominit taivutusluokkiin? – Virittäjä 1, 46–69.

Nuckols, Mark Eliot 2007. Case variation in Czech and Russian: Implications for the transitivity hypothesis. Dissertation. Presented in partial fulfillment of the requirements for the degree doctor of philosophy in the graduate school of the Ohio State University. <http://www.ohiolink>, 12.03.2010.

Pihlak, Ants 1985. Eesti ühendverbid ja perifrastilised verbid aspekti-tähenduse väljendajana. – *Ars Grammatica* 1985. Tallinn: Valgus, 62–93.

Pool, Raili 2010. Õppijakeele kontekstuaalne varieeruvus eesti keele täis- ja osasihitise kasutamisel. – *Eesti Rakenduslingvistika Ühingu aastaraamat 6 / Toim. Helle Metslang, Margit Langemets, Maria-Maren Sepper*. Tallinn: Eesti Keele Sihtasutus, 225–242.

Pool, Raili 2007. Eesti keele teise keelena omandamise seaduspärasusi täis- ja osasihitise näitel. *Dissertationes Philologiae Estonicae Universitatis Tartuensis* 19. Tartu: TÜ Kirjastus.

Pool, Raili, Elle Vaimann 2005. Vead kõrgtasemel eesti keele kõneleja kirjalikus keelekasutuses. – *Eesti Rakenduslingvistika Ühingu aastaraamat 1 (2004) / Koost. Margit Langemets, toim. Maria-Maren Sepper*. Tallinn: Eesti Keele Sihtasutus, 115–137.

Rajandi, Henno, Helle Metslang 1979. Määramata ja määratud objekt. Tallinn: Valgus.

Sajavaara, Kari 2006. Kontrastiivinen analyysi, transfer ja toisen kielen oppiminen. – *Lähivertailuja 17/ Toim. Annekatrin Kaivapalu, Külvi Pruuli (Jyväskylä Studies in Humanities 53)*. Jyväskylä: Jyväskylän yliopisto, 9–26.

Sulkala, Helena 1996. Expression of Aspectual Meanings in Finnish and Estonian. – *Estonian: Typological Studies I / Ed. by Mati Ereft (Publications of the Department of Estonian of the University of Tartu 4)*. Tartu: Tartu Ülikooli Kirjastus, 165–225.

Tamm, Anne 2004. Relations between Estonian verbs, aspect, and case. *Doktoral dissertation*. Budapest.

Tamm, Anne 2003. Estonian transitive verb classes, object case, and progressive. – Proceedings of the 19th Scandinavian conference of linguistics, vol. 31.4: Finnic, 639–653.

Tauli, Valter 1968. Keelekorralduse alused. Stockholm: Vaba Eesti.

Torn, Reeli 2003. Object Case Marking: English influence on the use of Estonian by three Estonian-English bilingual children. – *Kalbotyra/Linguistics* 53 (3), 126–134.

Vaiss, Natalia 2004. Eesti keele aspekti väljendusvõimalusi vene keele taustal. Magistritöö. Tallinn.

Tendencies in the use of grammatical cases in Estonian

Pille Eslon

Summary

This paper discusses some of the tendencies in the use of grammatical cases in Estonian. The discussion is based on a comparison of the results obtained from two corpora, one of which comprises standard literary Estonian and the other the Estonian learner language. The main focus is on the object marking in the learner language. The results indicate two main tendencies in the Estonian learner language: on the one hand, the nominative form of the object NP seems to be preferred to the partitive and genitive case; on the other hand, the distribution of the genitive case is likely to become more limited, i.e. mainly associated with marking the possession. The preference for the nominative on the object NP may be caused by bleached semantics of certain core verbs as well as by variation in case marking with accomplishment and

achievement verbs. The avoidance of the genitive case on the total object NP can be explained by analogy which learners seem to extend to instances where it is not appropriate, i.e. they tend to use nominative on the object of verbs which actually assign partitive case to its object (so called partitive verbs) and then appear to apply this pattern also to the marking of the total object. The use of nominative instead of the partitive may be due to the influence of the learners' first language (Russian) in which one of the factors for differential object marking is animacy so that inanimate nouns with the same type of verbs are used in an unmarked, nominative form. It appears then that the rules for object marking are being simplified in the Estonian learner language and as a result it is the nominative form which is preferred. These tendencies in the object marking are only possible because the Estonian language shows internal development in the same direction, specifically the marginalization of the genitive as the object case. The function of the genitive is being taken over not by the partitive case as often claimed in the literature, but by the nominative form instead, since nominative is an unmarked form while partitive is the most marked one. Thus the developments in Estonian appear to be reflected in learner language too.

Keywords: the Estonian language, the object phrase, the use of grammatical cases

Autor

PhD Pille Eslon, Tallinna Ülikooli eesti keele ja kultuuri instituudi vanemteadur, dotsent.

peslon@tlu.ee

MITÄ SYNTAKTISESTI KOODATTU KORPUS VOI KERTOA SANAJÄRJESTYKSESTÄ? PREDIKAATTI JA SUBJEKTI EDISTYNEIDEN SUOMENOPPIJOIDEN EKSISTENTIAALISISSA LAUSEISSA

Ilmari Ivaska

Abstrakti

Eksistentiaalilause kuuluu lausetyyppinä kaikkiin suomen kielen lausetyyppi-jaotteluihin. Yksi lauseen leimallisia piirteitä on sen sanajärjestys. Yleisesti suomen kielen lauseissa subjekti edeltää verbiä (SV-sanajärjestys), mutta prototyyppisessä eksistentiaalilauseessa lauseen subjektiksi kutsuttu jäsen on verbinjälkeisessä asemassa (VS-sanajärjestys). Tällainen ilmiö on hyvin kiinnostava kielen oppimisen näkökulmasta, ja sanajärjestyksen kaltaista kahden muuttujan välistä suhdetta on hyvin mielekästä tarkastella korpustutkimuksen keinoin. Tutkin tässä artikkelissa Edistyneiden suomenoppijoiden korpuksen (LAS2) avulla eksistentiaalisten lauseiden sanajärjestyksestä subjektin ja predikaatin osalta ja pohdin näiden varianttien tyypillisiä esiintymisympäristöjä. Vaikuttaa siltä, että SV-sanajärjestys on suhteellisesti todennäköisempi silloin, kun eksistentiaalinen lause on kielteinen. Huomio pätee niin edistyneiden suomenoppijoiden tutkimusaineistossa kuin

ensikielisten vertailuaineistossakin. On mahdollista, että sana-järjestyksen ja lauseen kielteisyyden välillä on edistyneiden suomenoppijoiden kielessä tiettyjä valintapreferenssejä nimenomaan eksistentiaalisten lauseiden kontekstissa, sillä kielteisissä lauseissa on suhteellisesti paljon useammin lausetyypille epäprototyyppinen SV-sanajärjestys. Vertailuaineistossa inversiolle on aina osoitettavissa jokin muu motivoija, kun taas tutkimusaineistossa tällaista motivoijaa ei aina näytä olevan.

Avainsanat: suomi toisena kielenä, korpuslingvistiikka, sana-järjestys, kiello, syntaksi¹

1. Johdanto

Eksistentiaalilauseeksi kutsuttu lausetyyppi on kuulunut suomen kielen syntaksin lausetyyppilähtöiseen tarkasteluun halkan historian. Suomen kielen eksistentiaalilauseella tarkoitetaan olemassaoloa ilmaisevaa rakennetta, jossa on prototyyppisesti VS-sanajärjestys, jonka subjektiksi kutsuttu jäsen² voi olla joko nominatiivissa tai partitiivissa ja jossa subjektin ja yksikön kolmannen persoonan näköisen predikaatin välillä ei ole kongruenssia (Hakanen 1972: 42–43). Usein eksistentiaalilauseessa on myös paikkaa ilmaiseva lokatiivadverbiaali (ISK: 851). Eksistentiaalilause eroaa tyyppipiirteidensä vuoksi suomen kielen ns. normaalilauseista niin sanajärjestyksen, kongruoimattomuuden kuin subjektin sijamerkinänkin osalta.

¹ Kiitän Helle Metslangia hänen esittämistään huomioista ja korjausehdotuksista.

² Kyseisen jäsenen subjektimaisuus on usein kyseenalaistettu. Esimerkiksi Marja-Liisa Helasvuo ja Tuomas Huumo ehdottavat jäsenen kutsumista neutraalisti E-NP:ksi (Helasvuo, Huumo 2010: 191–192).

Tästä syystä sen idiomaattinen käyttö vaatii monipuolista kielitaitoa, ja lausetyypin tarkasteleminen oppijansuomen viitekehyksessä on luontevaa. Tarkastelen tässä artikkelissa edistyneiden suomenoppijoiden eksistentiaalisten lauseiden sanajärjestystä ja pohdin, mitä syntaktisesti koodattu korpusaineisto voisi tästä kertoa. Tutkimus perustuu osin pro gradu -tutkielmaani (Ivaska 2009) ja se on osa Turun yliopistossa käynnissä olevaa Edistyneiden suomenoppijoiden korpus -tutkimushanketta (jatkossa LAS2, ks. esim. Ivaska, Siitonen 2009).

2. Eksistentiaalilauseen sanajärjestys ja toisen kielen korpustutkimus

Eksistentiaalilauseesta ja sen suhteesta muihin suomen kielen lauseisiin puhutaan useimmiten, kun tarkastellaan kielen syntaktisia lausetyyppejä. Kuhunkin lausetyyppiin liittyy kuitenkin aina niiden nimiä myöten presuppositioita myös semanttisesta sisällöstä. (Suomen eksistentiaalilauseesta käydystä keskustelusta ks. esim. Tiainen 1997; eksistentiaalilauseen luonteesta ja sen suhteesta muihin lausetyyppeihin ks. esim. Ivaska 2010: tulossa.) Käsite lainattiin alun perin Otto Jesperseniltä, joka käytti termiä *existential sentence* kuvaamaan sellaisia olemassaoloa ilmaisevia rakenteita, jotka vastaavat englannin *there is/there are* -tyyppisiä ilmaisuja ja joiden tarkoitteet ovat tyypillisesti epämääräisiä (Jespersen 1924: 154–156). Partitiivisubjektin ja olemassolon ilmaisemisen välisen yhteyden nosti Suomessa ensi kerran esiin Martti Airila todetessaan näiden olevan lauseita, joiden predikaatin toiminnallisuus on korkeintaan vain taka-alalla (Airila 1924: 17–18). Termiä *eksistentiaalilause* käytti ensi kerran Osmo Ikola pohtiessaan niin ikään partitiivisubjektin käyttöä (Ikola 1954: 226).

Suomen eksistentiaalilauseen VS-sanajärjestyksen voidaan katsoa juontavan lausetyypin normaalilauseista poikkeavasta informaatorakenteesta: kun normaalilauseen subjekti on oletusarvoisesti lauseen alussa – sen teemapaikassa – on eksistentiaalilauseen teemana neutraalisti paikallissijainen adverbiaali, ja verbinjälkeinen subjektijäsen käsitetään remaattiseksi (ISK: 1310, 1313). Tässä artikkelissa keskeistä on kuitenkin lausetyypin sisäinen variaatio ja sen tarkasteleminen korpus-tutkimuksen metodein, kuten erilaisten sanajärjestyksvarianttien konkreettisia esiintymisympäristöjä tutkimalla.

Eksistentiaalilauseesta puhuttaessa on mielestäni perusteltua erottaa kaksi rinnakkaista termiä. *Eksistentiaalilauseella* tarkoitan prototyyppistä ja abstraktia lausetyyppiä ja *eksistentiaalisella lauseella* tarkoitan todellisessa kielenkäytössä tuotettua tekstilauseetta, joka ainakin joiltakin osin vastaa eksistentiaalilauseelle asetettujen muotopiirteiden kriteereitä.

Kun tietyn rakenteen ilmenemistä oppijankielessä tarkastellaan korpuksen avulla, on yksi luonteva tarkastelutapa frekvenssianalyysi. Tällä tavoin voidaan tavoittaa oppijankielelle ominainen sisäinen variaatio ja jatkuva kehittyminen ja asettaa tilannesidonnainen variaatio osaksi koko käytetyn kielen kirjoa (Cancino jt 1978: 209–210; Ellis, Barkhuizen 2005: 94). Näin oppijan välikieltä (engl. *interlanguage*, ks. tarkemmin Selinker 1972) voidaan tarkastella omana kokonaisuutenaan ja välttää välikielen ja kohdekielen välisestä tarkastelusta syntyvä vertailuharha (engl. *comparative fallacy*, ks. tarkemmin Bley-Vroman 1983: 2).

Tämän artikkelin aineisto ei yksin riitä kvantitatiiviseen tarkasteluun, mutta tulokset voi nähdä pilottikokeina, ja ne voivat avata uusia näkökulmia korpuksella tehtävään oppijankielen tutkimukseen. Suomen eksistentiaalilause tuskin on omaa, muista poikkeavaa kognitiivista alkuperää, vaan kaikki lausetyyppeihin vaikuttavat tekijät ovat ennemminkin jokaiseen

tekstilauseeseen vaikuttava rihmasto. Eksistentiaalilauseen tunnuspiirteet – subjektin sijamerkintä, subjektin ja predikaatin välinen inkongruenssi ja VS-sanajärjestys – ovat kaikki osia syntaktisesti hyvin perustavanlaatuisista mekanismeista. Eksistentiaalilauseessa ne kuitenkin poikkeavat joiltakin osin kunkin muotopiirteen prototyyppisestä käytöstä, ja esimerkiksi eksistentiaalilauseen subjektin käyttöä määrittävät samat tekijät kuin passiivilauseen objektia (Hakanen 1972: 44). Analogia eli rakenteellinen samankaltaisuus on yksi kielen omaksumisen keskeisiä mekanismeja (ks. esim. Itkonen 2005: 93). Opittaessa uusia kielellisiä muotoja ne suhteutetaan jo tunnettujen rakenteiden suhdeverkostoon. Näin kielen osajärjestelmät vaikuttavat toisiinsa, vaikka niillä ei olisikaan suoraa yhteyttä (Goldberg 1995: 71–72).

Tässä mielessä oppijansuomen eksistentiaalisten lauseiden sanajärjestyksiä tarkasteltaessa on kiinnostavaa tarkastella eri sanajärjestysvalintojen myötäesiintymiä eli tekijöitä, joiden esiintyminen lisää jonkin toisen ilmiön tai elementin esiintymisen todennäköisyyttä. Erityisesti lauseiden sisäiset kolligaatiot eli leksikaalis-kieliopilliset suhteet ja näiden synnyttämät valintapreferenssit ovat seikkoja, joihin on syytä kiinnittää huomiota aineistoa analysoitaessa. (Myötäesiintymä, kolligaatio ja valintapreferenssi, ks. Jantunen 2004: 15, 29). Kuten Jarmo Jantunen toteaa, myötäesiintymien kaltaisen kontekstuaaliset epätyypillisyydet ovat piirteitä, joita voi tavoittaa välikielestä vain tarkastelemalla suuria aineistomassoja (Jantunen 2008: 83). Keskityn tässä artikkelissa yhteen eksistentiaalilauseelle ominaisena pidettyyn piirteeseen, sen sanajärjestykseen ja siinä esiintyvään variaatioon. Käsiteltävät esiintymisympäristöt ja myötäesiintymät ovat kuitenkin puhtaasti aineiston esiinnostamia kielenpiirteitä. Tarkastelussani on siis piirteitä niin korpuspohjaisesta kuin korpusvetoisestakin analyysistä (vrt. esim. Tognini-Bonelli 2001: 65–68, 84–87).

3. Aineiston eksistentiaaliset lauseet

Aineistoni 2575 virkeriviä jakautuvat yhteensä 4028 lauseeseen, ja se on kokonaisuudessaan osa LAS2-korpuksen tenttivastausaineistoa. Vertailuaineistona on 1310 virkeriviä ensikielisten suomenpuhujien tenttivastauksia yhteensä 27 informantilta (VA). Aineisto on edistyneiden suomenoppijoiden tenteissä kirjoittamaa suomea, eikä aineistoa ole valikoitu muin tavoin. Aineisto koostuu korpuksen vuonna 2007 tallennetuista tenttivastauksista. Tenttivastauksia pidetään oppijan kielen korpustutkimukseen hyvin soveltuvana aineistona, koska tällöin tuottamistilanne on hyvin samankaltainen keruuajankohdasta riippumatta. Tenttitapahtumaa voidaan myös pitää kohtalaisen autenttisenä kielentuottamistilanteena, sillä siinä informantti tuottaa kieltä jostain muusta syystä kuin kielen tutkimuksen aineistoksi. (Ks. esim. Granger 2002: 8–9.) Virkerivillä tarkoitetaan virkettä tai virkkeen kaltaista tekstuaalista kokonaisuutta.

Tätä artikkelia kirjoitettaessa korpuksen laaja syntaktinen ja morfologinen koodaus (ks. esim. Ivaska, Siitonen 2009) ei vielä ole käytettävissä. Analyysi onkin tehty digitalisoidusta raakatekstistä, mutta olen valinnut sovelletut metodit niin, että tarvittavat haut voitaisiin toteuttaa koodauksen avulla. Aineiston yhdeksän informanttia (jatkossa LAS2-1...LAS2-9) ovat hyvin heterogeeninen ryhmä³. Ensikielet ovat englanti (LAS2-8), komi (LAS2-6), liettua (LAS2-5, LAS2-9), ruotsi (LAS2-3), tšekki (LAS2-2), venäjä (LAS2-1, LAS2-7) ja viro (LAS2-4).

Korpuksen avulla tehtävä tutkimus vaatii määriteltävissä olevat hakuehdot, tässä tapauksessa eksistentiaalisten lauseiden muotopiirteisiin perustuvat kriteerit, joita vasten variaatiota

³ Informanteista tarkemmin ks. esim. Ivaska 2010: tulossa.

voidaan sitten tarkastella (ks. tarkemmin Ivaska 2010: tulossa). Pidän muodoltaan prototyypisenä eksistentialilauseena Hakasen (1972) määrittämää mallia. Prototyypisen lauseen kriteereitä – kutakin vuorollaan – väljentämällä voidaan edelleen tavoittaa niissä esiintyvä variaatio ja Tuomas Huumon ja Jari Perkon hahmottelema lauseiden eksistentialisuuden jatkumo (Huumo, Perko 1993: 389–391). Esimerkit (1) ja (2) ovat aineistoni sellaisia eksistentialisia lauseita, jotka täyttävät kaikki eksistentialilauseen tunnusmerkit.

(1) suomalais-ugrilaiset kielet ovat agglutinoivia, *niistä suurimmassa osassa on SOV-sanajärjestys*⁴ (subjekti-objekti-verbi). (las2-4tt01te04lo02⁵)

(2) *Saamessa ei ole soinnillisia konsonantteja.* (las2-7tt01te04lo03)

Näiden kriteerien perusteella tutkimusaineistossa on 449 sellaista eksistentialiseksi luokiteltavaa lausetta, joissa voidaan tarkastella subjektin ja predikaatin välistä sanajärjestystä. Prosenttiosuuksina ilmaisten 11 %:ssa tutkimusaineiston lauseista on eksistentialisia ja 17 %:ssa virkeriveistä on eksistentialinen lause. Eksistentialisten lauseiden joukosta 407 lauseessa (10 % ; 90 %)⁶ verbi edeltää subjektia. 42 (1 % ; 10 %)

⁴ Olen kursivoivoinut eksistentialiseksi tulkitsemäni lauseen ja lihavoinut sen subjektiksi käsittämäni jäsenen. Sama merkintätapa on käytössä läpi artikkelin.

⁵ Kunkin esimerkin lopussa on LAS2-korpukseen viittaava tunnistetieto, joka kertoo informantin (las2-4), tekstilajin (tt01 eli tenttivastaus, tässä artikkelissa ei muita tekstilajeja), tekstikokonaisuuden numeron (te06) ja tekstikokonaisuuden sijoittumisen kyseisen informantin tarkastelujaksolle kuukausina jakson alusta (lo02).

⁶ Sulkumerkinnän ensimmäinen luku (10 %) kertoo prosenttiosuuden kaikista aineiston lauseista ja jälkimmäinen (90 %) prosenttiosuuden eksistentialisiksi luokiteltujen lauseiden joukosta. Sama merkintä on käytössä läpi koko artikkelin.

tapauksessa subjekti edeltää verbiä. Edelleen, näistä SV-tapa-
uksista 16 lausetta (0,4 % ; 4 %) on ns. kvanttorilauseita (ks.
esim. Vähämäki 1987, 185–186; ISK, 848), joissa subjektijäsen
on jo prototyypillisesti verbinetisessä asemassa. Vertailuaineis-
tossa puolestaan on 139 sellaista eksistentiaaliseksi luokitel-
tavaa lausetta, joista voidaan tarkastella subjektin ja predikaatin
sanajärjestystä. Näistä lauseista 121:ssä on prototyypinen VS-
sanajärjestys ja 18:ssa SV-sanajärjestys. Eksistentiaalisten lausei-
den jakaumaa sanajärjestyksen kannalta havainnollistaa
asetelma 1.

Asetelma 1. Tutkimusaineiston eksistentiaalisten lauseiden jakauma
sanajärjestyksen kannalta. Sulkeiden ensimmäinen luku kertoo pro-
senttiosuuden kaikista lauseista ja jälkimmäinen prosenttiosuuden
eksistentiaaliseksi luokitelluista lauseista.

	LAS2	VA
VS	407 (10 % ; 90 %)	121 (87 %)
myönteisiä	87 %	91 %
kielteisiä	13 %	9 %
SV	42 (1 % ; 10 %)	12 (13 %)
myönteisiä	62 %	64 %
kielteisiä	38 %	46 %
joista alisteisia sivulauseita	4 (0,1 % ; 1 %)	7 (5 %)
kvanttorilauseita	16 (0,4 % ; 3,9 %)	8 (6 %)
ed. lauseiden viittaussuhteiden määrittämiä	2 (0,05 % ; 0,5 %)	2 (1 %)
Kysymyssana motivoi sanajärjestyksen <i>on olemassa</i> -lauseita	2 (0,05 % ; 0,5 %)	
Sanajärjestykselle ei motivoijaa	6 (0,15 % ; 1,5 %)	1 (1 %)
Yht.	449 (11 % ; 100 %)	139 (100 %)

Voidaan sanoa, että tutkimus- ja vertailuaineiston jakaumat ovat verraten lähellä toisiaan, ja että sanajärjestyksen kannalta eksistentiaalisten lauseiden käyttö on edistyneillä suomenoppijoilla kohtalaisen idiomaattista. Korpusvetoisen lähestymistavan kannalta on kuitenkin mielenkiintoista tarkastella niiden eksistentiaalisten lauseiden esiintymisympäristöjä, joissa on epäprototyypinen SV-sanajärjestys.

On luonnollisesti mahdotonta osoittaa yksiselitteisesti, mitkä tekijät ovat laukaisseet epäprototyypisen SV-sanajärjestyksen niihin tutkimusaineiston lauseisiin, joista ei ole löydettävissä mitään kohdekielen mukaista inversion motivoijaa. Se, että lauseiden sanajärjestys on kohdekielen kannalta epäidomaattinen, voi kuitenkin olla osaltaan myötäesiintymien ja kolligaatiosuhteiden vaikutusta. Jantunen toteaa tällaisten ko(n)tekstuaalisten epätyypillisyyksien olevan sellaisia välikielen piirteitä, jotka on mahdollista tavoittaa vain suurilla aineistomassoilla tarkastelemalla (Jantunen 2008: 83). Jokin ilmiö saattaa analogisen omaksumisen myötä saada sille kohdekielen kannalta kuulumattomia koodausominaisuuksia, mikäli kyseisen ilmiön todennäköinen esiintymisympäristö kohdekielessä tukee tällaisen analogian syntyä. Tämän artikkelin aineisto ei riitä luotettavien päätelmien tekemiseen, mutta nämä huomiot saattavat silti antaa viitteitä tarkempaa käsittelyä kaipaavista ilmiöistä.

Yksi mahdollinen kolligaatiosuhde niissä eksistentiaalisissa lauseissa, joissa subjekti edeltää predikaattia, on negaatio. Merkillepantavaa on ennen kaikkea se, että SV-sanajärjestys on suhteellisesti huomattavasti todennäköisempi kielteisissä lauseissa kuin myönteisissä lauseissa. Tämä huomio pitää paikkansa niin tutkimus- kuin vertailuaineistossakin. ISK ei mainitse sanajärjestyksestä käsitellessään kiellon ilmaisukeinoja (ISK, 1534–1535). Kieltoa ja sen vaikutusalaa ilmaistaan muodon kannalta pääverbin erityisellä taiputusmuodolla sekä objektin,

eksistentiaalilauseen subjektin ja osman partitiivisijaisuudella. Näiden lisäksi kiellolla voi olla ns. saattoilmiöinä kieltokausia sananmuotoja ja sanoja. (Ks. mt., 1536.) Kieltolause voi kuitenkin usein olla luonteeltaan reaktiivinen, minkä vuoksi konteksti saattaa motivoida sanajärjestyksen muutoksen. Näin ollen negaatio ei itsessään motivoi epäprototyypistä sanajärjестystä mutta niiden yhteisesiintyminen on odotuksenmukaista. (Mt., 1326.)

Negaatio on epätyypillisen sanajärjестyksen yleinen myötäesiintymä, ja kieltolause onkin epätyypilliselle sanajärjестykselle suhteellisesti todennäköisempi esiintymisympäristö kuin myönteinen väitelause. Tällainen yhteisesiintyminen saattaa analogisen oppimisen myötä synnyttää oppijan välikieleen kohdekielen vastaisen koodausominaisuuden. Päätelmä ei ole tilastollisesti pätevä, mutta se esittää yhden mahdollisuuden kohdekielen myötäesiintymien potentiaalisesta vaikutuksesta oppijan välikieleen. Esimerkit (5) – (7) ovat tapauksia, joissa kieltolauseessa on eksistentiaalilauseelle epäprototyypillinen SV-sanajärjестys ilman kontekstuaalista motivoijaa tai sellaista reaktiivisuutta, että ratkaisu tuntuisi idiomaattiselta kohdekielen kannalta.

- (3) *Näissä sanoissa **astevaihtelua** ei ole*, eli geminaattoja esiintyy, koska ne kuuluvat sanoihin (sanojen vartaloihin) sellaisenaan. (las2-5t01te02lo02)
- (4) Suomeen on kehittynyt -is(i) konditionaali, ***-is(i) konditionaalina** ei ollut sgr. kantakielessä*. (las2-6tt01te03lo05)
- (5) linna: ei voi olla ikivanha sana, ***nn-geminaattaa** ei ollut* (las2-6tt01te02lo04)
- (6) Viime vuosisadalla eteläkarjalan murretta puhuttiin myös Valdain alueella, *mutta nykyään siellä **karjalaisia kyliä** ei ole*. (las2-7tt01te05loxx)

- (7) *Diftongia* kantakielessä ei ollut vaan oli syntynyt pitkistä vokaalista ē, ö ja myös esim. pohjoissaamessa pitkiä kantakielen vokaalia on jäänyt mutta suomessa on diftongutoineet {esimerkki} kieli~keel, nuori~noor. (las2-8tt01te02lo00)

Kiellon ja sanajärjestyksen vaikutusta toisiinsa nimenomaan eksistentiaalisissa lauseissa puoltaa se, että yleiskielessä näiden ilmiöiden välistä suhdetta ei näytä ainakaan pilottitarkastelun perusteella olevan. Tutkin sanajärjestyksen ja lauseen kielteisyyden välisen suhteen Lauseopin X-arkiston (LaX) yleiskielen aineistosta tietokirjallisuuden kielestä, tietosanakirjojen alaryhmästä. Aineistossa on 854 lausetta, joista 508:sta voidaan määrittää subjektin ja predikaatin välinen sanajärjestys (LaX-1). Valitsin aineiston, koska se on tekstilajina kohtalaisen lähellä tutkimusaineistona olevia tenttivas-
tauksia. Hakutulokset osoittavat, että tässä aineistossa sanajärjestyksellä ja kiellolla ei vaikuta olevan mitään vaikutusta toisiinsa, sillä kielteisten väitelauseiden suhteellinen määrä on niin SV- kuin VS-sanajärjestyksissäkin lauseissa 4 % kaikista väitelauseista (LaX-2...LaX-5). Asetelma 2 havainnollistaa jakaumaa tarkemmin.

Asetelma 2. LaX-arkiston yleiskielen aineiston tietokirjallisuuden kielen ja tietosanakirjojen alaryhmän sellaisten lauseiden esiintymät, joista voidaan määrittellä subjektin ja predikaatin välinen järjestys.

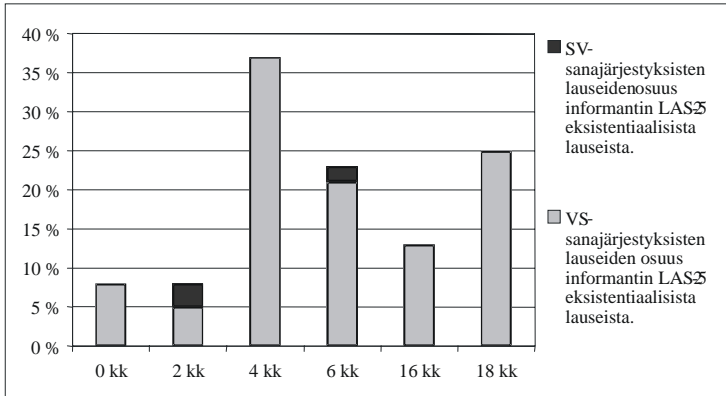
SV	376 kpl	74 %
myönteisiä (LaX-2)	361 kpl	96 % (SV:istä)
kielteisiä (LaX-3)	15 kpl	4 % (SV:istä)
VS	132 kpl	26 %
myönteisiä (LaX-4)	127 kpl	96 % (VS:istä)
kielteisiä (LaX-5)	5 kpl	4 % (VS:istä)
Yht.	508 kpl	100 %

Lauseen kielteisyys voidaan tulkita sanajärjestykseltään epä-prototyypin eksistentiaalisen lauseen myötäesiintymäksi sekä tutkimusaineistossa että vertailuaineistossa. Sanajärjestykseltään prototyypin eksistentiaalisten lauseiden jakauma tutkimusaineistossa vastaa likimain Hakulisen, Karlssonin ja Vilkun esittämää myönteisten ja kielteisten lauseiden jakaumaa suomen kielessä yleisesti. Kun tämän tutkielman aineiston sanajärjestykseltään prototyypin eksistentiaalisista lauseista 87 % on myönteisiä, on Hakulisen, Karlssonin ja Vilkun saama luku 91 %. (Hakulinen, Karlsson, Vilku 1980: 120.) Kuten aiemmin todettiin, ei kielteisyys yksinään riitä selittämään odotuksenvastaista sanajärjестystä. Ilmiöiden yhteisesiintyminen on kuitenkin odotuksenmukaista. Analogisen oppimisen myötä myötäesiintymästä voi nousta yksi tekijä, joka on edistänyt näiden tutkimusaineiston tekstilauseiden subjektijäsenten järjestymistä verbinetiseen asemaan juuri eksistentiaalisissa lauseissa.

4. Pitkittäismuutokset kahden informantin aineistossa

Aineiston koko ja tarkastelujakson pituus mahdollistaa pitkittäismuutosten tarkastelun informanttien LAS2-5 ja LAS2-8 aineistosta – molemmilla tarkastelujakso on 18 kuukautta. Tarkastelen eksistentiaalisten lauseiden esiintymistä niiden sanajärjестyksen kannalta tarkastelujakson eri vaiheissa.

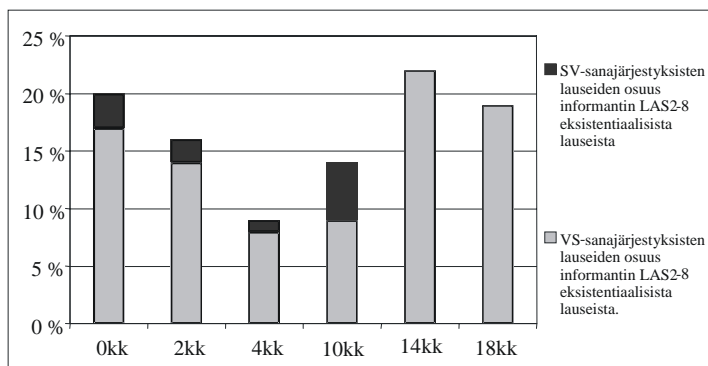
Informantin LAS2-5 aineiston 66 eksistentiaalista lausetta ovat lähes poikkeuksetta sanajärjестykseltään prototyypin VS-lauseita. Tendenssi pysyy samanlaisena koko tarkastelujakson ajan. Kuvio 1 havainnollistaa eksistentiaalisten lauseiden esiintymistä ja niiden sanajärjестystä tarkastelujakson aikana.



Kuvio 1. Eksistentialisten lauseiden osuus ja sanajärjestys subjektin ja verbin osalta informantin LAS2-5 aineistossa. Prosenttiluku kertoo eksistentialisten lauseiden osuuden saman informantin virkeriveistä kunakin tarkasteluajankohtana.

Eksistentialisten lauseiden osuus kaikista lauseista kasvaa ajan kuluessa. Tarkastelujakson alussa ja kahden kuukauden kohdalla eksistentialisten lauseiden määrä (8 %:ssa virkeriveistä on eksistentialinen lause) on sekä vertailuaineiston että tutkimusaineiston keskiarvoja pienempi ja 99 virkerivissä on yhteensä 8 eksistentialista lausetta. Neljän kuukauden kohdalla eksistentialisia lauseita on 37 %:ssa virkerivejä eli 13 lausetta 35 virkerivissä. Määrän nopea nousu voisi viitata tekstin sisällölliseen poikkeuksellisuuteen, mutta aineistojakso ei tunnu poikkeavan oleellisesti sitä edeltävistä tai seuraavista jaksoista. Tämän jälkeen eksistentialisten lauseiden suhteellinen määrä laskee hieman. Tarkastelujakson lopussa eksistentialisia lauseita on 44 virkerivissä 11 kappaletta. Neljän kuukauden kohdalla tapahtuneesta määrän kasvusta alkaen määrä on lähes kaksi kertaa niin suuri kuin vertailuaineistossa. Eksistentialiset lauseet ovat siis selvästi yliedustuneina informantin LAS2-5 aineistossa. Luvut eivät ole tilastollisesti merkitseviä, mutta niiden osoittama tendenssi on selvä.

Informantin LAS2-8 tarkastelujakso on niin ikään 18 kuukautta. Tarkastelujakson eksistentiaalisten lauseiden suhteellinen määrä ja jakauma sanajärjestyksen kannalta on osin yllättävä. Kaikkiaan 16 %:ssa virkeriveistä on eksistentiaalinen lause, ja määrä on siis 6 prosenttiyksikköä vertailuaineiston määrää suurempi. Tarkastelujakson alussa eksistentiaalisia lauseita on 20 % virkeriveistä ja jakson lopussa 19 %. Ääripäiden välissä eksistentiaalisten lauseiden määrä kuitenkin laskee jopa alle 10 %:iin virkeriveistä. Tarkastelujakson alkuvaiheessa informantin eksistentiaalisissa lauseissa on sekä VS-että SV-sanajärjystä, mutta lopussa eksistentiaalilauseelle prototyyppinen VS-sanajärjestys muuttuu yksinomaiseksi. Kuvio 2 havainnollistaa eksistentiaalisten lauseiden jakaumaa informantin LAS2-8 aineistossa tarkastelujakson eri vaiheissa.



Kuvio 2. Eksistentiaalisten lauseiden osuus ja sanajärjestys subjektin ja verbin osalta informantin LAS2-8 aineistossa. Prosenttiluku kertoo eksistentiaalisten lauseiden osuuden saman informantin virkeriveistä kunakin tarkastelujakohtana.

Sekä lauseiden väliaikainen väheneminen että VS-sanajärjestyksen muuttuminen yksinomaiseksi saattavat kuvastaa informantin välikielen muutosta. Informantti on käsittänyt

VS-sanajärjestyksen eksistentiaalilauseelle prototyypiksi ja saattaa pyrkiä välttämään virheitä. Voidaan olettaa, että kyse on eksistentiaalilauseen ja sitä määrittävien funktioiden uudelleenjärjestymisestä oppijan välikielessä. Uuden sanajärjestysvaihtoehdon tultua kieleen se alkaa saada yksiselitteisen muodon ja funktion välisen tehtävän. (Vrt. Ellis 1985: 95–96.)

5. Päätelmät

Edistyneiden suomenoppijoiden eksistentiaalisten lauseiden sanajärjestystä koskevat tulokset osoittavat, että lausetyypille epäprototyypinen SV-sanajärjestys on todennäköisempi silloin, kun lause on kielteinen. Tämä huomio pätee niin tutkimus- kuin vertailuaineistossakin. Vertailuaineiston kaikissa eksistentiaalisten lauseiden SV-tapauksissa inversiolla on kuitenkin osoitettavissa jokin muu motivoija, kun taas tutkimusaineistossa tällaista motivoijaa ei aina ole. Syynä saattaa olla pyrkimys osoittaa reaktiivisuutta edeltävään kontekstiin, jolloin lauseen informaatorakenne, ja sitä myötä sanajärjestys, voi muuttua. Korpusvetoisen tutkimusotteen kannalta inversion ja kieltolauseiden yhteisesiintyminen on niin ikään kiinnostava tulkintavaihtoehto. Informanttien välikielten kieliopeissa kielteisyyden voi analogian myötä olla myös itsessään epäprototyypisen sanajärjestyksen motivoiva tekijä, tai ainakin kielteisyyden voi jo itsessään lisätä inversion mahdollisuutta. Ajatus sopii tuloksiin, jotka Maija Ikonen on saanut pro gradu-tutkielmassaan tarkastellessaan vieraskielisten lukiolaisten kieltolauseita suomenkielisessä kirjoituksessa (Ikonen 2006). Ikonen mukaan kielteisissä lauseissa ainoa sanajärjestyksen kannalta poikkeamaksi tulkittava ero on tarpeeton inversio (mt., 70). Yleisimmin vieras sanajärjestys liittyy subjektin paikkaan lauseessa, mutta tämäkin on verraten harvinaista

(mt., 66).⁷ Analogian syntymistä puoltaa se huomio, että eksistentiaalisten lauseiden SV-sanajärjestys on suhteellisesti todennäköisempi kielteisissä eksistentiaalisissa lauseissa sekä tutkimus- että vertailuaineistossa. Ottamatta kantaa lauseiden informaattiorakenteisiin ja suomen kielen inversiota laukaiseviin mekanismeihin eksistentiaalisen lauseen epäprototyypin SV-sanajärjestyksen ja kiellon yhteisesiintymisen on siis tyypillistä ja odotuksenmukaista myös kohdekielessä ja siitä saatavassa syötöksessä. Kieltä opittaessa oppija muodostaa oppimastaan analogioita, joita soveltamalla hän käyttää kieltä. Nämä analogiat voivat syntyä syötöksessä havaituista tendensseistä, osin riippumatta siitä, noudattaako analogia kohdekielen kieliopin malleja. Ellis toteaa vertikaalista vaihtelua käsitellessään, että tuotokseen vaikuttaa aina myös se, missä määrin kielenoppija osaa soveltaa kieli-taitoaan (Ellis 1988: 168–171). On ilmeistä, että prototyyppisyyden vähetessä myös tämä soveltaminen vaikeutuu, mikä voi kasvattaa suomen kielessä yleisemmän SV-sanajärjestyksen esiintymistodennäköisyyttä.

Tulosten valossa analogisen oppimisen vaikutus välikielen kielioppijärjestelmään vaikuttaa suurelta. Tämä puoltaa tilastollisesti todennäköisten myötäesiintymien tutkimista ja vertailua ensikielisen suomen ja oppijansuomen välillä myös syntaksin osalta. Leksikaalis-kieliopillisten ja kieliopillis-kieliopillisten kolligaatiosuhteiden laajamittainen tarkastelu vaatii mittavia aineistoja ja valistuneisiin hypoteeseihin perustuvaa kohteen tarkkaa rajaamista. Lauseen kielteisyys ja kontekstiin viittaavat elementit ovat sellaisia tekijöitä, joiden vaikutuksen tutkiminen edistyneiden suomenoppijoiden sana-

⁷ Ikosen aineistosta oleva lause *jolloin ei lapsi ajattele paljon asioita* kuvaa tarpeetonta inversiota, jossa subjekti on siirtynyt kieltoverbin jälkeiseen asemaan (Ikonen 2006: 66).

järjestykseen vaikuttaisi mielekkäältä. Nimenomaan korpusvetoisen tutkimusotteen soveltaminen syntaksin tutkimukseen tuntuu mielenkiintoiselta. Tämä tarkoittaa esimerkiksi edellä kuvattujen kaltaisten, tilastollisesti todennäköisten yhteisesiintymien tarkastelua ja niiden vaikutusten arvioimista oppijankieleen. Tässä artikkelissa osoitettujen tendenssien yksityiskohtaisempi tarkastelu suurella korpusaineistolla on kiintoisa jatkotutkimuksen kohde. Kieliopillisten ilmiöiden esiintymisympäristöjen tarkastelu on mahdollista järkevimmin juuri syntaktisesti koodatun korpuksen avulla.

Kirjallisuus

Airila, Martti 1924. Kielellisten ilmiöiden perustelemisesta. – Kieli- ja kansatieteellisiä tutkielmia. Juhlakirja professori E. N. Setälän kuusikymmenvuotispäiväksi 27.2.1924 (Suomalais-ugrilaisen seuran toimituksia 52). Helsinki: Suomalais-ugrilainen seura, 15–26.

Bley-Vroman, Robert 1983. The comparative fallacy in interlanguage studies: The case of systemacity. – *Language Learning* 33:1, 1–17.

Cancino, Herlinda, Ellen J. Rosansky, John H. Scumann 1978. The acquisition of English negatives and interrogatives by native Spanish speakers. – *Second Language Acquisition: A Book of Readings* / Ed. by Evelyn Marcussen Hatch. Newbury House, Rowley (Massachusetts), 207–230.

Ellis, Rod 1985. *Understanding Second language Acquisition*. Oxford: Oxford University Press.

Ellis, Rod 1988. *Classroom Second Language Development*. London: Prentice Hall International (UK) Ltd.

Ellis, Rod, Gary Barkhuizen 2005. *Analysing Learner Language*. Oxford: Oxford University Press.

Golberg, Adele E. 1995. *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.

- Granger, Sylviane 2002. A Bird's-eye view of learner corpus research. – Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching / Ed. by Sylviane Granger, Joseph Hung, Stephanie Petch-Tyson. Amsterdam: John Benjamins Publ. Co, 3–33.
- Hakanen, Aimo 1972. Normaalilause ja eksistentiaalilause. – Suomen kielen seuran vuosijulkaisu 14, 36–76.
- Hakulinen, Auli, Fred Karlsson, Maria Vilkuna 1980. Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus (Department of General Linguistics Publications No. 6). Helsinki: University of Helsinki.
- Helasvuo, Marja-Liisa, Tuomas Huumo 2010. Mikä subjekti on? – Virittäjä, 165–194.
- Huumo, Tuomas, Jari Perko 1993. Eksistentiaalilause lokaalisuuden ilmaisijana. – Virittäjä, 380–399.
- Ikola, Osmo 1954. Suomen lauseopin ongelmia I–III. – Virittäjä, 209–245.
- Ikonen, Maija 2006. Kieltolauseiden kiellonalaiset poikkeamat vieras-kielisten lukiolaisten suomenkielisessä kirjoituksessa. Pro gradu -tutkielma. Turun yliopiston suomalaisen ja yleisen kielitieteen laitos, suomen kieli.
- ISK 2004 = Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen, Irja Alho 2004. Iso suomen kielioppi (Suomalaisen Kirjallisuuden Seuran Toimituksia 950). Helsinki: SKS.
- Itkonen, Esa 2005. Analogy as structure and process. Approaches in linguistic, cognitive psychology and philosophy of science. Amsterdam: John Benjamins Publ. Co.
- Ivaska, Ilmari 2009. Eksistentiaalilauseen ilmiöedistyneiden suomenoppijoiden kirjoituksessa. Pro gradu -tutkielma. Turun yliopiston suomalaisen ja yleisen kielitieteen laitos, suomen kieli.
- Ivaska, Ilmari 2010 (tulossa). Eksistentiaalilauseen subjektijäsen edistyneiden suomenoppijoiden kirjoituksessa. – Suomen kielen seuran vuosijulkaisu 52.

Ivaska, Ilmari, Kirsti Siitonen 2009. Syntaktisesti koodattu oppijan-kielen korpus: mahdollisuuksia ja ongelmia. – Korpusuuringute metodoloogia ja märgendamise probleemid / Toim. Pille Eslon, Katre Õim (Tallinna Ülikooli eesti keele ja kultuuri Instituudi toimetised 11). Tallinn: Tallinna Ülikooli kirjastus, 54–71.

Jantunen, Jarmo 2004. Synonymia ja käännösuomi: korpus-näkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännös-kielen leksikaalisiin erityispiirteisiin (Joensuun yliopiston huma-nistisia julkaisuja). Joensuu.

Jantunen, Jarmo 2008. Haasteita oppijankielen analyysille: oppijan-kielen universaalit. – Öppijakeele analüüs: Võimalused, probleemid, vajadused / Toim. Pille Eslon (Eesti filoloogia osakonna toimetised 10). Tallinn: Tallinna Ülikooli kirjastus, 67–91.

Jespersen, Otto 1924. *The Philosophy of Grammar*. London: George Allen & Unwin Ltd.

Selinker, Larry 1972. Interlanguage. – *International Review of Applied Linguistics* 10, 209–231.

Tiainen, Outi 1997. Suomen eksistentiaalilause – päättymätön tarina. – *Virittäjä*, 563–571.

Tognini-Bonelli, Elena 2001. *Corpus Linguistics at Work* (Studies in Corpus Linguistics 6). Amsterdam: John Benjamins Publ. Co.

Vähämäki, Börje 1987. On defining basic sentences in Finnish. – *Fennistica festiva in honorem Göran Karlsson* / Toim. Mauno Koski, Eeva Lähdemäki, Kaisa Häkkinen. Turku: Åbo Akademis förlag, 174–188.

Aineslähteet

LAS2 = Edistyneiden suomenoppijoiden korpus. Suomen kielen oppiaine, kieli- ja käännöstieteiden laitos, Turun yliopisto. Turku. Aineistona on korpukseen vuonna 2007 tallennetut tenttivastaukset.

- LaX = Lauseopin X-arkisto. Kotimaisten kielten tutkimuskeskus – Turun yliopiston kieli- ja käännöstieteiden laitos, Lauseopin arkisto. Turku
- LaX-1 = Aineisto: Yleiskieli/Tietokirjallisuuden kieli/Tietosanakirjat
XPath: //cl[w[@fun='npsubj'] and w[@fun='pred']]
[Viitattu 02.07.2010.]
- LaX-2 = Aineisto: Yleiskieli/Tietokirjallisuuden kieli/Tietosanakirjat
XPath: //cl[@type='affdecl' and w[@fun='npsubj'] and following-sibling::w[@fun='pred']]
[Viitattu 02.07.2010.]
- LaX-3 = Aineisto: Yleiskieli/Tietokirjallisuuden kieli/Tietosanakirjat
XPath: //cl[@type='negdecl' and w[@fun='npsubj'] and following-sibling::w[@fun='pred']]
[Viitattu 02.07.2010.]
- LaX-4 = Aineisto: Yleiskieli/Tietokirjallisuuden kieli/Tietosanakirjat
XPath: //cl[@type='affdecl' and w[@fun='pred' and following-sibling::w[@fun='npsubj']]]
[Viitattu 02.07.2010.]
- LaX-5 = Aineisto: Yleiskieli/Tietokirjallisuuden kieli/Tietosanakirjat
XPath: //cl[@type='negdecl' and w[@fun='pred' and following-sibling::w[@fun='npsubj']]]
[Viitattu 02.07.2010.]
- VA = tekijän hallussa oleva vertailuaineisto

What Can a Syntactically Encoded Corpus Reveal about Word Order?

Predicate and Subject in the Existential Sentences of Advanced Learners of Finnish

Ilmari Ivaska

Summary

Existential sentence is a clause type that is included in all clause-type classifications of Finnish. A prototypical existential sentence differs from so-called normal clauses of Finnish in the word order, case marking of the subject constituent and the lack of congruence between subject and verb. This paper discusses the word order in the existential sentences of advanced learners of Finnish. The study is part of the Advanced Finnish Learners' Corpus -project (LAS2) at the University of Turku.

The results show that the word order is VS in 90 % of the existential sentences and SV in 10 % of the existential sentences. The division is roughly the same in the comparison material. The analysis of the co-occurrences reveals that SV word order, the non-prototypical variant in the existential sentences, is more probable in the negative sentences both in the study material and in the comparison material. This lexicogrammatical co-occurrence works only in the context of the existential sentences as, in general, the negation does not impact on the word order of Finnish.

However, while something invariably motivated the SV word order in the comparison material, those motivators could not always be pinpointed in the study material. This might

suggest that given the analogy with the target language, negation may have become a motivator of word order inversion in existential sentences in the inter-languages of the informants.

The results show that analogical learning appears to be of great significance in the grammatical system of the inter-language. This encourages us to research the quantitatively probable co-occurrences in the learner language also from the syntactical point of view. A more thorough study on the tendencies pointed out in this article would hopefully tell us more about the grammatical and lexico-grammatical collocations.

Keywords: Finnish as a Second Language, Corpus Linguistics, Word Order, Negation, Syntax

Esittely

FM Ilmari Ivaska on työskennellyt Edistyneiden suomenoppijoiden korpus -tutkimushankkeessa sen alusta (2007) alkaen. Ivaska kirjoitti pro gradu -tutkielmansa edistyneiden suomenoppijoiden eksistentiaalisista lauseista. Tällä hetkellä hän tekee korpusvetoisena tutkimuksena toteutettavaa, edistyneiden suomenoppijoiden lausetyyppejä käsittelevää väitöskirjaa.

ilmari.ivaska@utu.fi

EESTI KEELE SÕNAJÄRJE VEALEIDJA PROTOTÜÜBI ARENDAMINE

Erika Matsak, Pille Eslon, Jaagup Kippar

Ülevaade

Eesti vahekeele korpuse (EVKK) materjalide põhjal on kindlaks tehtud, et eesti keele õppijale valmistab kõige sagedamini raskusi sõnajärg, nt **Majandus kiiresti arenes*. Artiklis kirjeldatakse eesti keele sõnajärjevigade tuvastamise võimalusi, meetodit ja tulemusi. Kõigepealt antakse lühiülevaade olemasolevate automaatsete vealeidjate tööpõhimõtetest ning kirjeldatakse Tallinna Ülikoolis VAKO-projekti raames välja töötatud sõnajärjevealeidja prototüüpi¹. Sõnajärjevealeidja on statistikapõhine programm, mille tööpõhimõte sarnaneb teatud määral n-grammidega. Programmi aluseks on võetud lause sõnajärje seisukohalt üheksa olulise lauseliikme korrektsed sõnajärjemallid, mille regulaarne ilmumine keelekasutuses moodustab sõnajärjemustreid. Programmi efektiivsuse tõstmiseks on sõnajärjemustrid paigutatud andmepuusse, kust prototüüp otsib kõigepealt õige algusmärgendiga puu ning seejärel sagedasema

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

sõnajärjemustri. Prototüübi tööd on testitud *Eesti vahekeele korpuse* tekstidel. Artiklis tuuakse välja sõnajärjevigade tuvastamisel tekkinud probleemid ja pakutakse võimalusi vealeidja efektiivsuse suurendamiseks. Sõnajärvealeidja prototüübi näol on tegemist universaalsusele pretendeeriva programmiga, sest kasutatud analüüsimeetodit ja sõnajärvealeidja prototüübi algoritmi saab rakendada ka teiste keelte sõnajärje uurimiseks.

Võtmesõnad: morfosüntaks, automaatne veatuvastus, sõnajärjevead

1. Vealeidjad ja meetodid

Keeletehnoloogias kasutatakse vigade tuvastamiseks erinevaid meetodeid. Üks selline tugineb spetsiaalsele sõnastikule ehk leksikonile, milles teksti iga sõnet võrreldakse õige sõnavormiga (vt nt Damerau 1964): kui sõnastikust vastet ei leita, loetakse sõnavorm valeks. Niisuguste programmidega on tavaliselt võimalik poolautomaatselt (mõningate sõnade puhul ka täisautomaatselt) parandada vigu, kusjuures aluseks võetakse algoritmid, mis leiavad lähima õige sõna.

Teine meetod põhineb n-grammidel (vt nt Beesley 1988): võrreldakse osasõna ehk näiteks kahte või kolme järjest paiknevat tähte. Ka sel juhul on õigete sõnade loendist genereeritud mallid ning on teada, millised osasõnad millise sagedusega esinevad. Väiksema sagedusega osasõnad on tihti vea tunnuseks. Lisaks on nii sõnade kui osasõnade tasemel võimalik moodustada sagedasemate vigade nimistud (vt nt Pedler 2007). Selliste loendite saamiseks on sõnad korpustes märgendatud: igale vigasele sõnale vastab õige.

Sõnajärje kontrollimiseks erinevates keeltes on n-gramme ja statistilisi meetodeid kasutatud koos (vt nt Athanaselis, Baka-

midis, Dologlou 2006). Sellisel juhul on grammiks terve sõna. Levinud on ka lingvistiline reeglipõhine lähenemine (vt nt Arppe 2000).

Eesti keele sõnajärge on kirjeldatud meetodeid rakendades uuritud suhteliselt vähe, kuna eesti lause sõnajärjestruktuuri analüüsiks vajalikke infotehnoloogilisi tugisüsteeme pole seni loodud. Järgnevalt tutvustame *Eesti vahekeele korpuse* sõnajärjevealeidja prototüübi tööpõhimõtteid.

2. Sõnajärjevealeidja prototüübi alus

Tallinna Ülikoolis on VAKO-projekti raames sõnajärje kontrollimiseks loodud prototüüp, mille aluseks on eesti kirjakeele korrektsed sõnajärjemallid. Nende väljaselgitamiseks uuriti sõnade süntaktilisi rolle lauses, misjärel töötati välja sõnajärje analüüsimeetod, mis sarnaneb teatud määral n-grammidega ning on kohandatud eesti keelele (vt Matsak, Metslang, Kippar 2010). Võimalikuks sai see tänu Kaili Müürisepa loodud parserile², mis on implementeeritud *Eesti vahekeele korpuse*. Parser eristab 27 erineva süntaktilise funktsiooni märgendit³, kuid lausete sõnajärje analüüs näitas, et suurem osa neist ei mängi sõnajärjes kuigi olulist rolli. Lause sõnajärje seisukohalt on olulised verbi ja lause põhja üheksa märgendit (vt Metslang, Matsak 2010):

1. verbi märgendid
 - @FMV – finiiitne verb
 - @IMV – infiniitne verb
 - @FCV – *olema* liitaegades ning modaalverbid ahelverbides, finiiitne vorm

² Vt: EstCG Parser 1.0a, vt <http://www.cs.ut.ee/~kaili/parser/>, 25.07.2010.

³ Vt: <http://www.cs.ut.ee/~kaili/parser/demo/synttags.html>, 25.07.2010.

@ICV – *olema* liitaegades ning modaalverbid ahelverbides, infiniitne vorm

@NEG – verbi eitus

2. lause põhja märgendid

@SUBJ – alus ehk subjekt

@OBJ – sihitis ehk objekt

@PRD – öeldistäide ehk predikatiiv

@ADVL – määrus ehk adverbiaal, sh fraasiadverbiaal.

Sõnajärvealeidja prototüübi loomisel on iga esimese osalause ja lihtlause piires otsitud lause põhja märgendite regulaarselt esinevaid järjendeid, mis moodustavad erineva sagedusega ilmnevaid sõnajärjemustreid. Nii näiteks annab eesti keele parser lausele (1) järgmise morfosüntakstilise väljundi:

(1) *Loomulikult ei tohi tunnistada, et sul endal ka need olemas on.*

"<s>"

"<Loomulikult>"

"loomulikult" L0 D cap @ADVL #1->3

"<ei>"

"ei" L0 V aux neg cap @NEG #2->3

"<tohi>"

"tohti" L0 V main indic pres ps neg cap <FinV> <Intr>

@FMV #3->3

"<tunnistada>"

"tunnista" Lda V main inf cap <NGP-P> @IMV #4->4

"<>"

"," Z Com CLB #5->5

"<et>"

"et" L0 J crd cap @J #6->12

"et" L0 J sub cap @J #6->12

"<sul>"

"sina" L1 P pers ps2 sg ad cap @NN> @ADVL #7->7

"<endal>"
 "ise" L1 P pos det refl sg ad cap @<NN #8->8
 "<ka>"
 "ka" L0 D cap @ADVL #9->12
 "<need>"
 "see" Ld P dem pl nom cap @SUBJ #10->12
 "<olemas>"
 "ole" Lmas V main sup ps in cap <Intr> @ADVL #11->12
 "<on>"
 "ole" L0 V main indic pres ps3 pl ps af cap <FinV>
 <Intr> @FMV #12->12
 "<.>"
 "." Z Fst #13->13
 "</s>"

Pärast sõnajärje seisukohalt ebaoluliste märgendite ning sõnade (nt *loomulikult, arvatavasti, tõenäoliselt, niisiis* jt) eemaldamist jätkab programm analüüsi esimese osalause piirini CLB. Toodud näites jääb alles kolme olulise süntakstilise märgendi järjend ['@NEG', '@FMV', '@IMV']. Lihtlause *Internetis (@ADVL) on (@FMV) võimalik (@PRD) kasutada (@SUBJ) mitmeid (@NN>) teenuseid (@OBJ)* sõnajärjemall on ['@ADVL', '@FMV', '@PRD', '@SUBJ', '@OBJ']. Mall ei arvesta sõnavormi *mitmeid*, sest nimi-sõnalise eestäiendina @NN> ei kuulu see sõnajärje seisukohalt oluliste märgendite hulka. Samalaadseid probleeme tekitavad sõnajärje analüüsis adverbiaali või fraasiadverbiaali rollis kasutatud sõnad. Järelikult tuleb eelnevalt hinnata, kui olulist osa mängib konkreetne sõna(vorm) lause sõnajärjes. Selleks tuleb iga sõnakasutus lauses üle kontrollida ja kindlaks teha vajalike/mittevajalike märgendite ja ebaoluliste sõnade loendid.

Lisaks tuleb eelnevalt kokku leppida ka selles, milliseid lauseid vealeidja prototüüp vaatleb ning milliseid mitte. Kuigi eesti keele sõnajärje aluseks on süntaktiliste põhimõtete asemel

pigem infostruktuurilised printsiibid (Lindström 2005: 173; EKG II: 13) ja eesti keele põhisõnajärjeks on üldiselt peetud SVX-järge (Vilkuna 1998, Koptjevskaja-Tamm, Wälchli 2001: 705) või statistiliselt pea sama sagedasti esinevat XVS-järge (Tael 1988), võib eesti keele kõrvallauses, eituslauses, küsisõnadega algavates küsilauses ning hüüdlausetes öeldisverb paikneda ka lause lõpus (V3-järg). Seetõttu on prototüüp programmeeritud välja sortima kõik laused, mis algavad sõnadega *kui, kuna*, mistahes küsisõnade või nende käändevormidega. Samuti jäetakse välja umbisikulises kõneviisis olevad laused või osalaused. Hüüdlause tunnuseks loeb prototüüp seni veel lauselõpulist hüüumärki.

Kuna õppijakeele puhul esineb süntaktilist valeanalüüsi, mis on tingitud õigekirjavigadest, siis on vaatluse alt välja jäetud ka kõik (osa)laused, milles esineb mõni õigekirjaviga. Nii ei analüüsi prototüüp lauseid nagu **Ema teda väga armstas, vanaema veel rohkem*, sest süntaksianalüsaator määrab vigaselt kirjutatud predikaadi *armstas* (= *armastas*) adverbiaaliks ning sellest tulenevalt leiaks prototüüp, et õppijakeele lausest puudub öeldis. Niisuguste lausete sõnajärge saab hakata kontrollima siis, kui *Eesti vahekeele korpusse* on implementeeritud lemmatiseerija-oletaja, mis seob vigaselt kirjutatud või moodustatud vormi korrektsega ja leiab üles lemma⁴.

3. Sõnajärjemustrite otsimine

Korrektsete sõnajärjemallide ja -mustrite leidmiseks kasutati esialgu ilukirjandustekstide lausetest koosnevat 3000 sõnelist pilootkorpus. Valimit analüüsiti EstCG Parser 1.0a abil, parseri

⁴ Eesti õppijakeele lemmatiseerija-oletaja on VAKO-projekti raames välja töötanud Kairit Sirts.

morfosüntaktiline väljund kopeeriti VBA skriptiga programmeeritud spetsiaalsesse aknasse Excelis (vt joonis 1). Järgnevalt korrastati väljund automaatselt: programm paigutas kogu morfosüntaktilise info tabelisse ja otsis iga lause algusosas sõnajärje seisukohalt oluliste märgendite järjestusi. Automaatselt leitud süntaktiliste märgendite järjendid kontrolliti käsitsi üle, korrektseks tunnistati 242 sõnajärjemalli.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		Süntaktitekst										
3		Vea tüüp										
4	1	Tekst korrigeerimine	Süntaksianalüsaatori väljund	Vea tüüp	Sõna vorm	Morfosüntaktiline analüüs	1	1.1	1.1.1	1.1.2	1.2	1.2.1
5	2	koost	koost // _S_ com sig et #cap // **CLB @ADVL		koost	// _S_ com sig et #cap // **CLB @ADVL						
6	3	mina	mina# // _P_ pers p3 s3 nom // @SUBI	7.1	mina#	// _P_ pers p3 s3 nom // @SUBI						
7	4	eni	eni# // _V_ aux neg // @NEG		eni#	// _V_ aux neg // @NEG						
8	5	siis	siis# // _V_ mod inde pres p3 neg #invv kon // @FCV		siis#	// _V_ mod inde pres p3 neg #invv kon // @FCV						
9	6	misidagi	misidagi# // _P_ indef sg part // @ADVL		misidagi	// _P_ indef sg part // @ADVL						
10	7	parti	parti# // _D_ // @ADVL		parti#	// _D_ // @ADVL						
11	8	halla	halla# // _A_ pers sg part // @AN		halla#	// _A_ pers sg part // @AN						
12	9	utseid	utseid# // _S_ com sg nom #? // @PRD		utseid#	// _S_ com sg nom #? // @PRD						
13	10	S	// _S_ #st //		.	// _S_ #st //						
14	11	S			.							
15	12	S			.							
16	13	S			.							
17	14	S			.							
18	15	S			.							
19	16	S			.							
20	17	S			.							
21	18	S			.							
22	19	S			.							
23	20	S			.							
24	21	S			.							
25	22	S			.							
26	23	S			.							
27	24	S			.							
28	25	S			.							
29	26	S			.							
30	27	S			.							
31	28	S			.							
32	29	S			.							
33	30	S			.							
34	31	S			.							
35	32	S			.							
36	33	S			.							
37	34	S			.							
38	35	S			.							
39	36	S			.							
40	37	S			.							
41	38	S			.							
42	39	S			.							
43	40	S			.							
44	41	S			.							
45	42	S			.							
46	43	S			.							
47	44	S			.							
48	45	S			.							
49	46	S			.							
50	47	S			.							
51	48	S			.							
52	49	S			.							
53	50	S			.							
54	51	S			.							
55	52	S			.							
56	53	S			.							
57	54	S			.							
58	55	S			.							
59	56	S			.							
60	57	S			.							
61	58	S			.							
62	59	S			.							
63	60	S			.							
64	61	S			.							
65	62	S			.							
66	63	S			.							
67	64	S			.							
68	65	S			.							
69	66	S			.							
70	67	S			.							
71	68	S			.							
72	69	S			.							
73	70	S			.							
74	71	S			.							
75	72	S			.							
76	73	S			.							
77	74	S			.							
78	75	S			.							
79	76	S			.							
80	77	S			.							
81	78	S			.							
82	79	S			.							
83	80	S			.							
84	81	S			.							
85	82	S			.							
86	83	S			.							
87	84	S			.							
88	85	S			.							
89	86	S			.							
90	87	S			.							
91	88	S			.							
92	89	S			.							
93	90	S			.							
94	91	S			.							
95	92	S			.							
96	93	S			.							
97	94	S			.							
98	95	S			.							
99	96	S			.							
100	97	S			.							

Joonis 1. Abivahend sõnajärjemallide otsimiseks

Kuna Exceli töökiirus on väike, siis sobib see abivahend idee testimiseks, mitte korpusedu automaatanalüüsiks. Seetõttu implementeeriti algoritmid Eesti vahekeele korpuse, kus lisaks korpuse tekstidele oleks võimalik analüüsida mistahes tekste ja nende keelelist korrektsust.

Prototüüpi testiti ilukirjandustekstidest võetud 20000 lausel. Valimi automaatanalüüsi tulemusena leiti vajalike süntaktiliste märgendite järjestus esimeses osalauses ja lihtlauses, määrati kindlaks ühesuguste järjendite esinemissagedus ning toodi välja regulaarselt kasutatud sõnajärjemustrid. Kinnitust leidis 600 korrektset sõnajärjemalli. Kuna uusi tekste produtseeritakse pidevalt juurde ja korpuse maht kasvabkiiresti, siis tuleb usaldusväärsete analüüsitulemuste saavutamiseks tagada programmi efektiivsus ja ka piisav töökiirus (vt Matsak, Metslang, Kippar 2010). Üheks võimaluseks on paigutada süntaktiliste

märgendite järjendid sõnajärjemustritesse ja ühesuguse algusmärgendiga mustrid andmepuudesse. Programm leiab kõigepealt üles vastava algusmärgendiga puu ja otsib puu ladvast allapoole liikudes sagedasemaid ja regulaarselt ilmnevaid sõnajärjemustreid. Kuna suure tõenäosusega moodustavad mustreid just sagedased sõnajärjemallid, siis tõhustab andmepuude kasutamine programmi tööd üsna oluliselt. Teisalt võimaldab see kasvatada andmepuude panka, lisades näiteks sõnajärjemustrite stiilivariante jm.

4. Andmepuud: süntaktilised märgendid ja sõnajärjemustrid

Andmepuud genereeriti rohkem kui 10000 lause alusel, kuna algsest 20000 lauset sisaldavast valimist ei analüüsi prototüüp 8590 lauset. Sõnajärjepuid kirjeldatakse vastavalt sagedusele. Kui sõnajärjevealeidja prototüüp genereerib puid ning tuvastab õige/vigase sõnajärje süntaktiliste märgendijärjendite ja nende osade kokkulangevuse alusel ehk paradigmaatilisel, siis eesti keele (osa)lausete regulaarselt korduvad (tüüpilised, keeleomased) sõnajärjemallid ilmnevad märgendi sageduse järgi süntagmaatilisel. Nende interpreteerimiseks tuleb iga puu haru vaadelda lineaarselt ning märgendite sagedusest lähtudes. Järgnevalt kirjeldame lühidalt eesti keele sõnajärjepuude andmepanka ja keeleomaseid korrektseid sõnajärjemalle, mis on välja toodud algusmärgendi sageduse põhjal. Allpool ei too me iga sõnajärjepuu detailseid kirjeldusi, vaid peatume sagedasematel eesti keelele omastel sõnajärjemustritel. Sõnajärjepuude ülevaade lõpeb üldistusega, milleks on levinumate sõnajärjemustrite andmepuu. Selle puu genereerimisel on nähtavaks jäetud vaid need märgendid, mida ligi 10000-lauselises valimis on esinenud vähemalt neljakümnes lauses.

4.1. Verbi eitus

Verbi eitust @NEG sisaldavaid lauseid oli valimis kõige vähem – vaid 6. Seetõttu on nende alusel kõige lihtsam näidata andmepuu etappide kaupa ehitamist ja sõnajärjemustrite järk-järgulist genereerimist. Kõigepealt toodi välja (osa)lauseid iseloomustavad sõnajärjemallid (vt näitelauseid 2–7) ning seejärel võrreldi neid üksteisega.

(2) *Loomulikult ei (@NEG) tohi (@FMV) tunnistada (@IMV) , et sul endal ka need olemas on .*

['@NEG', '@FMV', '@IMV']

(3) *Seega ei (@NEG) saa (@FMV) me (@SUB) oma läkitust (@OBJ) kodeerida (@IMV) nii , et see oleks mõisteta .*

['@NEG', '@FMV', '@SUBJ', '@OBJ', '@IMV']

(4) *Siis ei (@NEG) lausu (@FMV) ma (@SUBJ) enam ühtki sõna (@OBJ) ega mõtle enam millelegi , kõik vajub enneaimamatusse õndsusemerre .*

['@NEG', '@FMV', '@SUBJ', '@OBJ']

(5) *Ei (@NEG) ole (@FMV) enam millest (@ADVL) rääkida (@SUBJ) , ta tahab maale saada , ehk ta seda küll ei ütle .*

['@NEG', '@FMV', '@ADVL', '@SUBJ']

(6) *Ei (@NEG) ole (@FMV) jahu (@SUBJ) põskedel (@ADVL) ja huuled on loomulikult värsked .*

['@NEG', '@FMV', '@SUBJ', '@ADVL']

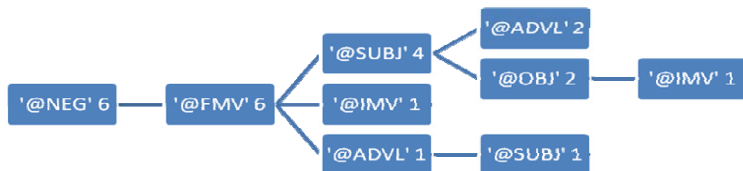
(7) *Siiski ei (@NEG) ole (@FMV) ma (@SUBJ) teie peale (@ADVL) väga tige , et te mu üles ajasite .*

['@NEG', '@FMV', '@SUBJ', '@ADVL']

Selgus, et 6. ja 7. näitelause sõnajärjemall ['@NEG', '@FMV', '@SUBJ', '@ADVL'] langeb kokku. Seejärel otsiti süntaktiliste märgendite järjendites osalisi kokkulangevusi, mis võimaldab esile tuua eitusaluse sõnajärjemustri hargnevusi. Näiteks:

1 lause	'@NEG'	'@FMV'	'@IMV'		
1 lause	'@NEG'	'@FMV'	'@SUBJ'	'@OBJ'	'@IMV'
1 lause	'@NEG'	'@FMV'	'@SUBJ'	'@OBJ'	
1 lause	'@NEG'	'@FMV'	'@ADVL'	'@SUBJ'	
2 lauset	'@NEG'	'@FMV'	'@SUBJ'	'@ADVL'	

Kuna algusmärgendite @NEG ja @FMV järgnevus kattub kõikides analüüsitud lausetes ja kolmandal positsioonil esineb teistest sagedamini märgend @SUBJ, siis ei pea andmepuus samu märgendeid kordama. Nende järgnevuste mugavamaks leidmiseks sortiti vastavat haru puus ülespoole (vt joonis 2).

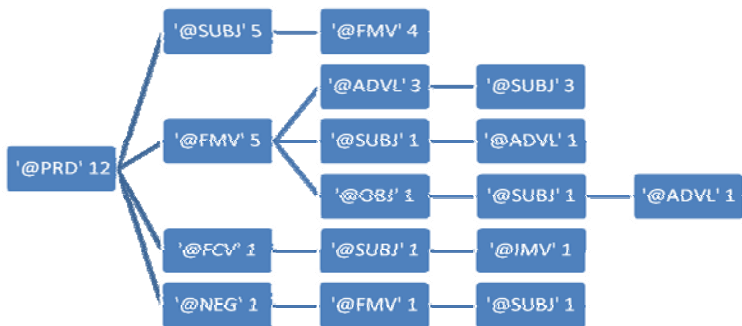


Joonis 2. Verbi eitusega algavate lausete sõnajärjestused

Joonise 2 alusel võib väita, et verbi eitusega algavatel (osa)lausetel on kaks sagedamini kasutatavat sõnajärjestust: ['@NEG', '@FMV', '@SUBJ', '@ADVL'] ja ['@NEG', '@FMV', '@SUBJ', '@OBJ'].

4.2. Predikatiiv

Predikatiiviga algavaid sõnajärjestemalle leiti ilukirjandustekstide valimist 12, seega rohkem kui verbi eitusega algavaid sõnajärjestemalle. Keerulisemad on ka puu hargnevused, sõnajärjestemustreid on rohkem ja nad varieeruvad sagedamini, vt joonis 3).



Joonis 3. Predikatiiviga algavate lausete sõnajärjemustrid

Levinumad predikatiiviga algavad sõnajärjemustrid on ['@PRD', '@SUBJ', '@FMV'] – vt 8. näidet ja ['@PRD', '@FMV', '@ADVL', '@SUBJ'] – vt 9. näidet.

(8) *Ja saatanlik (@PRD) nagu ta (@SUBJ) oli (@FMV), ei suutnud ta jätta puusi diskreetselt nõksutamata .*

['@PRD', '@SUBJ', '@FMV']

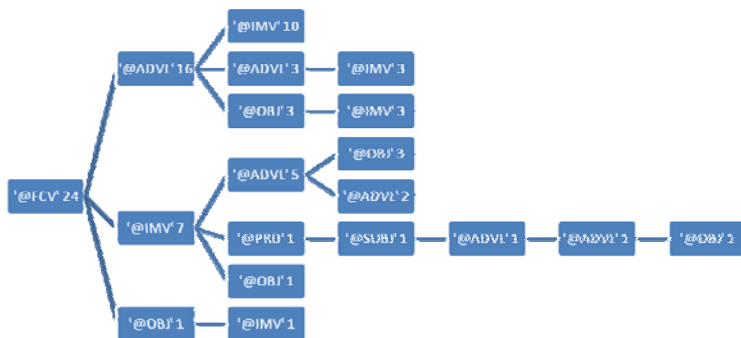
(9) *Raske (@PRD) on (@FMV) end tagasi (@ADVL) hoida (@SUBJ), kui ümberringi kõik purskab ja pritsib .*

['@PRD', '@FMV', '@ADVL', '@SUBJ']

4.3. Verb *olema* liitaegades ning modaalverbid ahelverbides (finiitvormid)

Verbi *olema* finiidse vormiga @FCV algavaid sõnajärjemalle oli valimis 24. Kõige regulaarsemalt kasutati mustrit, kus teisel positsioonil oli adverbiaal @ADVL ning poole võrra vähem

mustrit, kus verbi *olema* finiitsele vormile järgnes infiniitne verb @IMV, üksikjuhtumil objekt @OBJ (vt joonis 4).



Joonis 4. Sõnajärjemustrid, mille alguses on verbi *olema* või modaalverbide finiitne vorm

Enimkasutatud sõnajärjemustrites on liitaja vormi komponendid (*olema* finiitne vorm ja kesksõna) tavaliselt lahutatud teisel positsioonil asuva adverbialiga, vt 10. näidet.

(10) *Olin (@FCV) tihti (@ADVL) tundnud (@IMV), et see suvi oli mind välja valinud, kutsunud täitma prohvettlikku ülesannet.*
 ['@FCV', '@ADVL', '@IMV']

Järgnevad sõnajärjemustrid, milles kolmandal positsioonil on kas teine adverbialne lauselaiend (vt 11. näidet) või objekt (vt joonis 4).

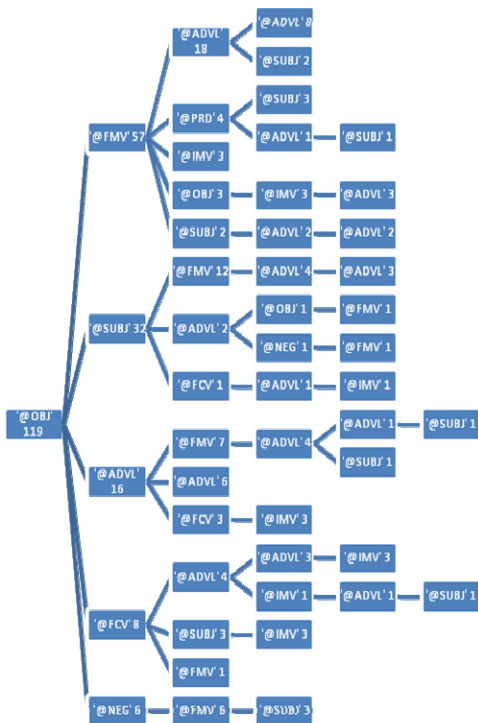
(11) *Olin (@FCV) juba (@ADVL) põlvini ürglimasse (@ADVL) vajunud (@IMV) ja vankusin .*
 ['@FCV', '@ADVL', '@ADVL', '@IMV']

Harvem tuleb ette niisuguseid sõnajärjemustreid, milles liitaja vormi on kõrvuti: ['@FCV', '@IMV', '@ADVL', '@OBJ'] – (vt 12. näidet) ja ['@FCV', '@IMV', '@ADVL', '@ADVL'] (vt joonis 4).

- (12) *Olen (@FCV) lasknud (@IMV) enesele (@ADVL) uue suve-ülikonna (@OBJ) valmistada, mis mu lühikest ja lihavat keha näib nägusamaks tegevat.*
['@FCV', '@IMV', '@ADVL', '@OBJ']

4.4. Objekt

Objektiga algavaid sõnajärjemalle oli valimis 119. Seetõttu osutus andmepuu ülesehitus eelmiste puude ülesehitusest keerukamaks ja märgendijärjendite varieeruvus mitmekesisemaks (vt joonis 5). Objektile järgneb kõige sagedamini finiiitne verbivorm @FMV (57 korda), harvem subjekt @SUBJ (32) ja adverbiaal @ADVL (16), harva verb *olema* liitaegades või modaalverbid ahelverbides @FCV (8) ning verbi eitus @NEG (6). Objektiga algava sõnajärjepuu harus, kus teisel positsioonil on verbi finiiitne vorm ['@OBJ', '@FMV'], ilmneb kolmandal positsioonil kõige järjekindlamalt ja seega suurema tõenäosusega adverbiaal @ADVL (18 korda) – vt 13. näidet. Võimalikud on ka predikaatiiv (vt 14. näidet), verbi infiniitne vorm (vt 15. näidet) ning subjekt. Kui objektiga algava sõnajärjemustri teine komponent on subjekt, siis on üsna tõenäone, et kolmandal positsioonil seisab verbi finiiitne vorm, neljandal ja viiendal adverbiaal jne (vt joonis 5). Järelikult, mida sagedasemad on keelekasutuses ühesuguste algusharudega puud, seda selgemalt tulevad esile regulaarselt kasutatavad sõnajärjemustrid. Analoogselt saab interpreteerida kõiki järgnevaid sõnajärjepuid.



Joonis 5. Objektiga algavate lausete sõnajärjemustrid

(13) *Nad (@OBJ) tunneb (@FMV) siin (@ADVL) kohe ära.*

['@OBJ', '@FMV', '@ADVL']

(14) *Luuletust (@OBJ) oli (@FMV) lihtne (@PRD) lugeda (@SUBJ) ja see jäi ruttu meelde .*

['@OBJ', '@FMV', '@PRD', '@SUBJ']

(15) *Inimesi (@OBJ) peab (@FMV) koondama (@IMV) ning esimesed kandidaadid on ütlema tagi Internetile töötajad .*

['@OBJ', '@FMV', '@IMV']

4.5. Finiitne verbivorm

Verbi finiitse vormiga @FMV algavaid sõnajärjemalle on valimis 401 (vt joonis 6). Sõnajärjemustreid moodustavad märgendite järjendid, milles teisel positsioonil on tavaliselt adverbiaal @ADVL (238 korda), harvem objekt @OBJ (70) või subjekt @SUBJ (70) ja predikatiiv @ PRD (18), üksikjuhtumitel ka infiniitne verb @IMV (3) ja verbi finiitne vorm @FMV (1).

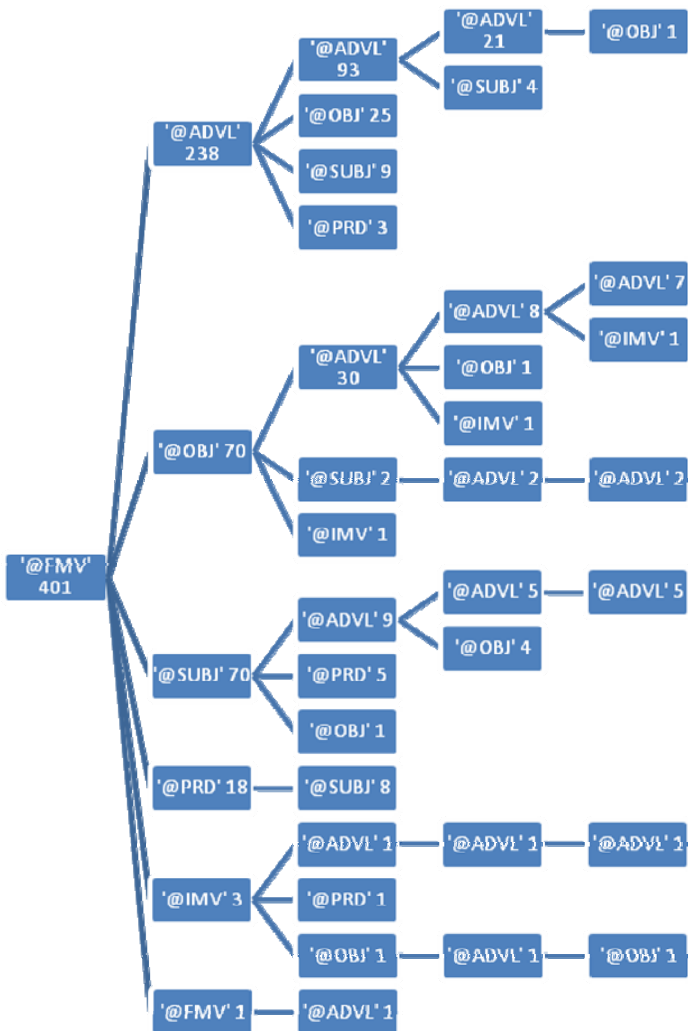
Verbi finiitse vormiga algavate sõnajärjemustrite andmepuu hargnevused näitavad ühte kindlat sõnajärjeseaduspära: kui verbi finiitsele vormile järgneb adverbiaal ['@FMV', '@ADVL'], siis on kolmandal ja neljandal positsioonil suure tõenäosusega teine ning kolmaski adverbiaal (vt 16. näidet). Samalaadne märgendite järjestus hakkab silma ka siis, kui verbi finiitsele vormile järgneb objekt või subjekt: kolmandal ja neljandal positsioonil on sel juhul tavaliselt adverbiaal (vt 17. näidet); üksikjuhul kordub see seaduspärasus ka teisel positsioonil oleva infiniitse verbiga (vt tabel 6).

(16) *Jään (@FMV) seejärel (@ADVL) kohe (@ADVL) magama.*

['@FMV', '@ADVL', '@ADVL']

(17) *Pane (@FMV) silmad (@OBJ) kinni (@ADVL) või ma tapan su ära*

['@FMV', '@OBJ', '@ADVL']



Joonis 6. Verbi finiitse vormiga algavad sõnajärjemustrid

4.6. Adverbiaal, sh fraasiadverbiaal

Adverbiaaliga algavaid sõnajärjemalle oli valimis kokku 857. Kuna sõnajärjemustrid muutuvad järjest keerulisemaks, kirjeldame adverbiaaliga algavat andmepuud alaliikide kaupa, mis on reastatud sõnajärjemustrite kahe algusmürgendi sageduse alusel:

- 1) adverbiaal + finiiitne verb ['@ADVL', '@FMV'] – 465 korda
- 2) adverbiaal + subjekt ['@ADVL', '@SUBJ'] – 128 korda
- 3) adverbiaal + adverbiaal ['@ADVL', '@ADVL'] – 122 korda
- 4) adverbiaal + verb *olema* liitaegades ning modaalverbid ahelverbides (fiiiitne vorm) ['@ADVL', '@FCV'] – 73 korda, adverbiaal + verbi eitus ['@ADVL', '@NEG'] – 31 korda, adverbiaal + infiiiitne verb ['@ADVL', '@IMV'] – 14 korda ja adverbiaal + objekt ['@ADVL', '@OBJ']) – 14 korda.

Levinum adverbiaaliga algav sõnajärjemuster on ['@ADVL', '@FMV', '@SUBJ'], kus teisel positsioonil on verbi finiiitne vorm ja kolmandal subjekt (vt 18. näidet). Neljandal positsioonil järgneb subjektile tavaliselt adverbiaal, harvem objekt või predikaatiiv (vt joonis 7).

(18) *Esimest korda (@ADVL) satub (@FMV) ta (@SUBJ) silme ette.*
['@ADVL', '@FMV', '@SUBJ']

Veidi harvem on sõnajärjemuster, kus kolmandal, neljandal ja viiendal positsioonil on adverbiaal ['@ADVL', '@FMV', '@ADVL', '@ADVL'] (vt 19. näidet ja joonis 7).

(19) *Oma ukse lävel (@ADVL) seisab (@FMV) siin (@ADVL) valge põllega (@ADVL) lihakaupleja (@SUBJ) ja pagari aknal on pinu sihvakaid saiu , pikki ja paksusid kui kasehalud .*
['@ADVL', '@FMV', '@ADVL', '@ADVL', '@SUBJ']

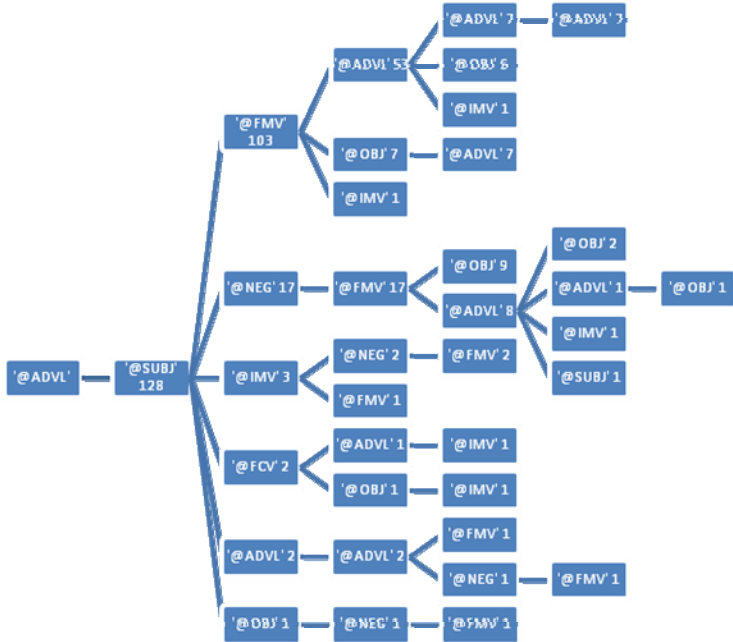
Ka selles sõnajärjemustris, kus adverbiaalile ja verbi finiiitsele vormile järgneb objekt, seisab adverbiaal tavaliselt nii neljandal kui ka viiendal positsioonil (vt joonis 7).



Joonis 7. Adverbiaaliga algavad sõnajärjemustrid (alaliik 1)

Sageduselt teine adverbiaaliga algava andmepuu alaliik on ['@ADVL', '@SUBJ'] – 128 kasutusnäidet (vt joonis 8). Tavaliselt järgneb subjektile verbi finiiitne vorm @FMV – 103 korda, tunduvalt harvemini verbi eitus @NEG (17 korda), väga harva või üksikjuhtumitel infiniitne verb, verb *olema* liitaegades ning modaalverbid ahelverbides, adverbiaal ja objekt. Sõnajärjemustri ['@ADVL', '@SUBJ', '@FMV'] neljas märgend on suure tõenäosusega teine adverbiaal (vt 20. näidet ja joonis 8). Kui

adverbiaalile ja subjektile järgneb verbi eitus, siis on võimalik, et järgmisena lisandub objekt (vt 21. näidet) või teine adverbiaal (vt joonis 8).

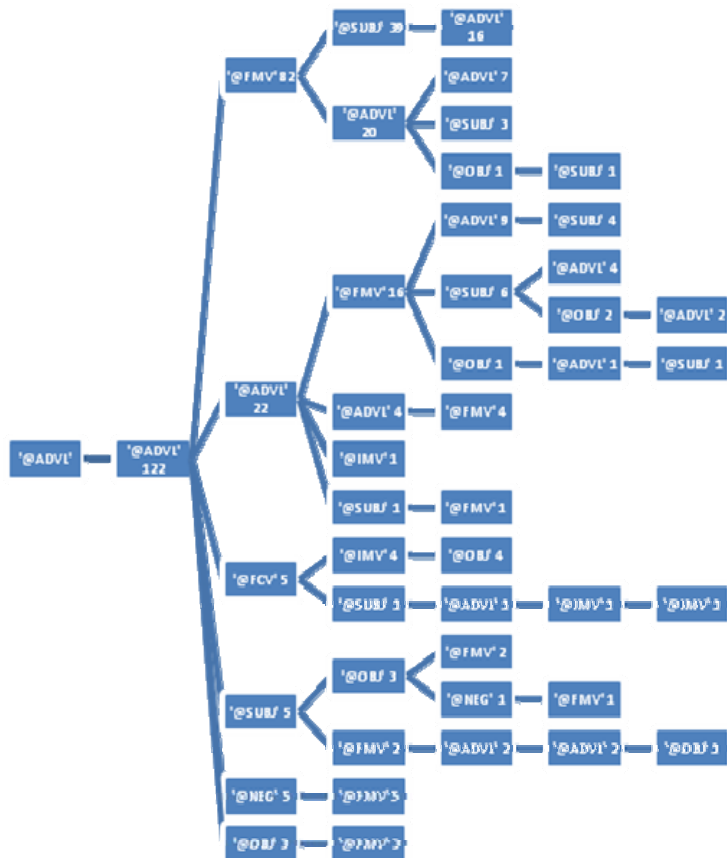


Joonis 8. Adverbiaaliga algavad sõnajärjestused (alaliik 2)

(20) Siis (@ADVL) nad (@SUBJ) tulevad (@FMV) koos (@ADVL) alla (@ADVL) ja ...

['@ADVL', '@SUBJ', '@FMV', '@ADVL', '@ADVL']

(21) *Siin (@ADVL) nad (@SUBJ) ei (@NEG) kalla (@FMV) õlut (@OBJ) maha (@ADVL), ei karju, ei tuigu.*
 ['@ADVL', '@SUBJ', '@NEG', '@FMV', '@OBJ', '@ADVL']



Joonis 9. Adverbiaaliga algavad sõnajärjemustrid (alaliik 3)

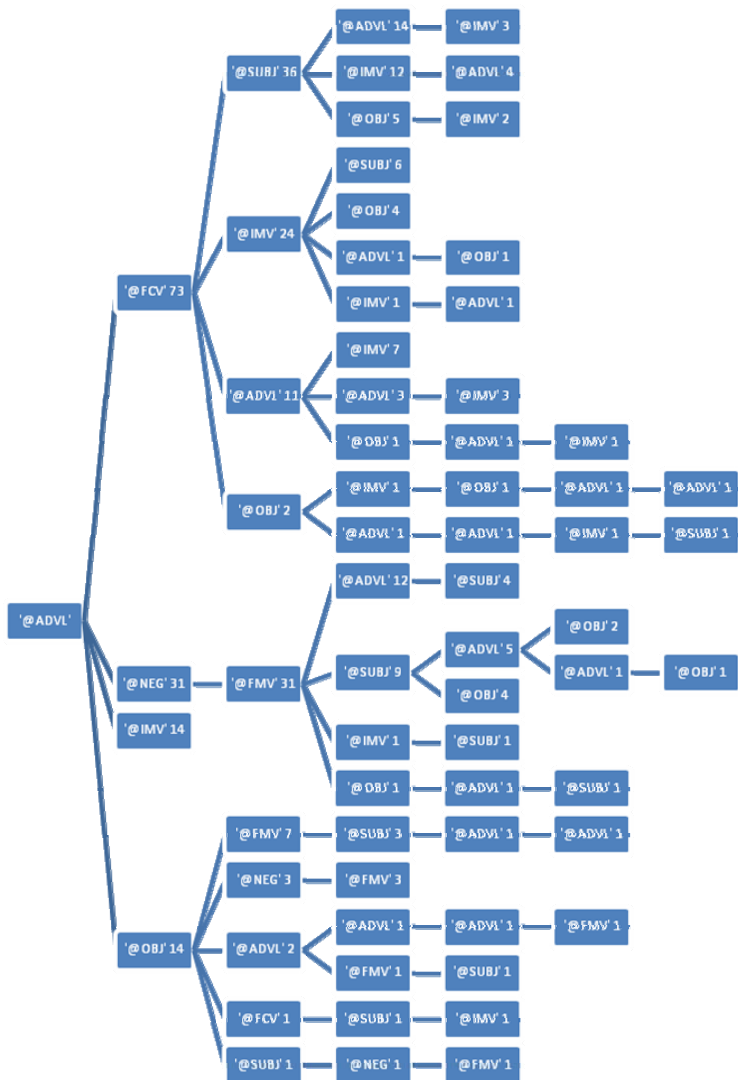
Kolmanda sõnajärjemustrite alaliigi moodustavad sõnajärjemallid, mille alguses on kaks adverbiaali ['@ADVL', '@ADVL'] (122 korda, vt joonis 9). Kõige sagedamini järgneb teisel posit-

sioonil olevale adverbiaalile verbi finiiitne vorm @FMV (82 korda), harvem kolmas adverbiaal @ADVL (22 korda), väga harva verb *olema* liitaegades ning modaalverbid ahelverbides, subjekt, verbi eitus ja objekt (vt joonis 9). Sõnajärjemuster ['@ADVL', '@ADVL', '@FMV'] jätkub tavaliselt kas subjekti (vt 22. näidet) ja adverbiaaliga või siis kolmanda ja sellele järgneva neljanda adverbiaaliga (vt 23. näidet).

(22) *Sealsamas (@ADVL) kõrval (@ADVL) oli (@FMV) viljapõld (@SUBJ) , moone täis, ja põlluservas kitkus üks naine rohtu , hall seelik üles kääritud , see võis olla tuhande aasta eest .*
['@ADVL', '@ADVL', '@FMV', '@SUBJ']

(23) *Aga teel (@ADVL) kohvikusse (@ADVL) mõtled (@FMV) ikkagi (@ADVL) ümber (@ADVL) , pistad talle viiemargase pihku ja lased jalga .*
['@ADVL', '@ADVL', '@FMV', '@ADVL', '@ADVL']

Adverbiaaliga algava andmepuu neljanda alaliigi sõnajärjemustrid (vt joonis 10) moodustuvad nende sõnajärjemallide alusel, milles teisel positsioonil on kas verbi *olema* liitajad ja modaalverbid ahelverbides ['@ADVL', '@FCV'] (73 korda), verbi eitus ['@ADVL', '@NEG'] (31 korda), infiniitne verb ['@ADVL', '@IMV'] (14 korda) või objekt ['@ADVL', '@OBJ'] (14 korda). Neljanda alaliigi sagedasemas sõnajärjemustris ['@ADVL', '@FCV'] on kolmandal positsioonil enamasti subjekt (36 korda), veidi harvem infiniitne verb (24 korda), teine adverbiaal (11 korda), üksikjuhtumitel objekt. Tavaliselt jätkab sõnajärjemustrit ['@ADVL', '@FCV', '@SUBJ'] kas teine adverbiaal (14 korda, vt 24. näidet) või infiniitne verb (12 korda, vt 25. näidet). Sõnajärjemustri ['@ADVL', '@FCV', '@IMV'] järgmine märgend on (@SUBJ (subjekt) või @OBJ (objekti) jne.



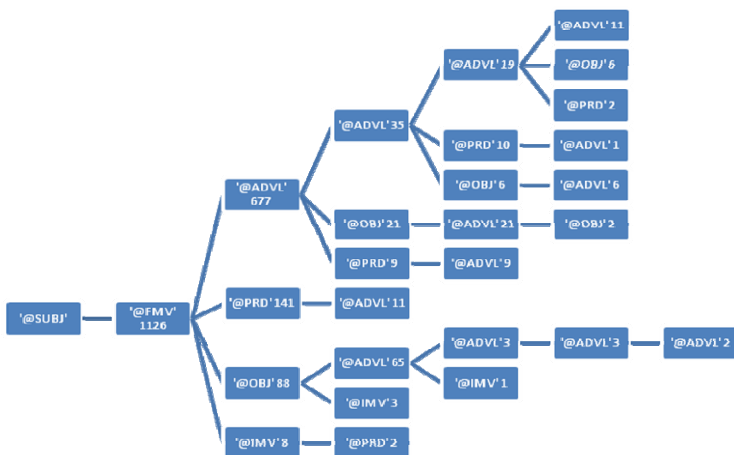
Joonis 10. Adverbiaaliga algavad sõnajärjestused (alaliik 4)

(24) *Seepärast (@ADVL) oli (@FCV) ta (@SUBJ) uksele (@ADVL) kirja (@OBJ) kinnitanud (@IMV), et tuleb varsti tagasi ja palub toas oodata.*
 ['@ADVL', '@FCV', '@SUBJ', '@ADVL', '@OBJ', '@IMV']

(25) *Minu mängu (@ADVL) oli (@FCV) sünenenud (@IMV) pikem paus (@SUBJ) – polnud raha* ['@ADVL', '@FCV', '@IMV', '@SUBJ']

4.7. Subjekt

Subjektiga algav andmepuu on genereeritud kõige sagedamini esinenud sõnajärjemallide alusel. Regulaarsemalt kasutatud sõnajärjemustrite alguses järgneb subjektile tavaliselt verbi finiiitne vorm (1236 korda, vt joonis 11). Seni vaadeldud andmepuudest erineb see puu mitme tunnuse poolest. Prototüübi genereeritud sõnajärjemallide üldise paradigma alusel on näha,



Joonis 11. Subjektiga algav sõnajärjemuster ['@SUBJ', '@FMV']

et subjektiga algav sõnajärjepuu pole struktuurilt ega variantide rohkuselt eelnevalt kirjeldatud puudest keerulisem. Lisaks domineerivale sõnajärjele ['@SUBJ', '@FMV'] kasutatakse sageli veel vaid kahte subjektiga algavat mustrit, milles teisel positsioonil on verb *olema* liitaegades, modaalverb ahelverbides (110 korda, vt joonis 12) või verbi eitus (110 korda, vt joonis 13). Oluliselt vähem esineb sõnajärge, mille puhul subjektile järgneb adverbiaal (46 korda, vt joonis 14). Seega sisaldub kõige sagedasema algusmärgendiga andmepuudes vähem sõnajärjemustreid kui eelnevalt kirjeldatud puudes, kuid need katavad suurema jao valimi osalausetest ja lihtlausetest.

Algusmärgendiga ['@SUBJ', '@FMV'] seisab pooltel juhtudel kolmandal positsioonil adverbiaal (677 korda 1126 kasutusest) ning sellele võib järgneda veel kolm adverbiaali (vt 26. näidet). Harvem esineb muster, kus kolmandal positsioonil olevale adverbiaalile järgneb objekt (vt 27. näidet).

(26) *Mu mõrvoja (@SUBJ) istub (@FMV) salongi laua ääres (@ADVL).*
 ['@SUBJ', '@FMV', '@ADVL']

(27) *See (@SUBJ) tekitas (@FMV) alati (@ADVL) hirmu (@OBJ), kui korraste ei olnud enam midagi näha läbi, vaid ainult sissepoole. Kuid asja muutis hullemaks see, et nüüd oli minu akna taga mitte tühi sein nagu lapsepõlves, vaid rõdu.*
 ['@SUBJ', '@FMV', '@ADVL', '@OBJ']

Algusmärgenditele ['@SUBJ', '@FMV'] järgneb veelgi harvem predikatiiv (141 korda) (vt 28. näidet) või objekt (88 korda).

(28) *Ma (@SUBJ) olen (@FMV) kindel (@PRD) selles (@ADVL), et ta mind hakkab armastama.*
 ['@SUBJ', '@FMV', '@PRD', '@ADVL']

Sõnajärjepuud, mille algusmärgendid on ['@SUBJ', '@FCV'], kasutatakse teistest subjektiga algavatest puudest tunduvalt

harvem (vt joonis 12). Kolmandal positsioonil on enamasti infiniitne verbivorm (64 korda, vt 29. näidet) või adverbiaal (48 korda, vt 30. näidet).



Joonis 12. Subjektiga algava lause sõnajärjemuster ['@SUBJ', '@FCV']

(29) *Pilved (@SUBJ) olid (@FCV) tulnud (@IMV) päikesele nii lähedale (@ADVL), et lausa nühkisid teda, kriimustades tema kollast lakki.*

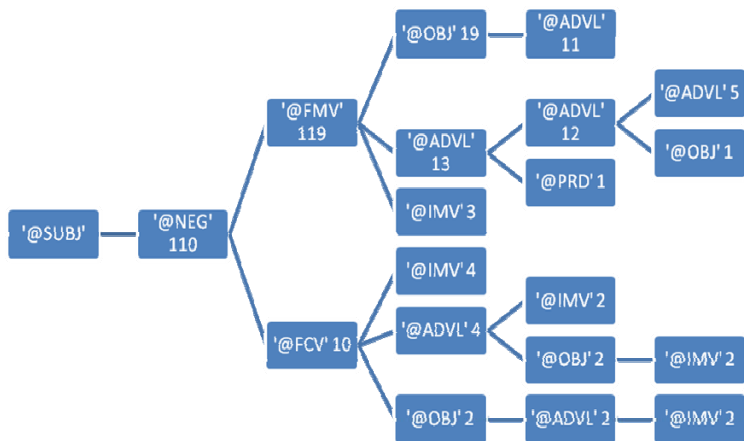
['@SUBJ', '@FCV', '@IMV', '@ADVL']

(30) *Aga tema (@SUBJ) oli (@FCV) trepi alla (@ADVL) läinud (@IMV), kus turistid seisavad ja jõllitavad.*

['@SUBJ', '@FCV', '@ADVL', '@IMV']

Subjektiga algavas verbi eitust sisaldavas sõnajärjepuus ['@SUBJ', '@NEG'] on kolmandal positsioonil valdavalt verbi finiidne vorm @FMV (119 korda, vt 31. näidet), millele järgneb kas objekt (19 korda) või adverbiaal (13 korda).

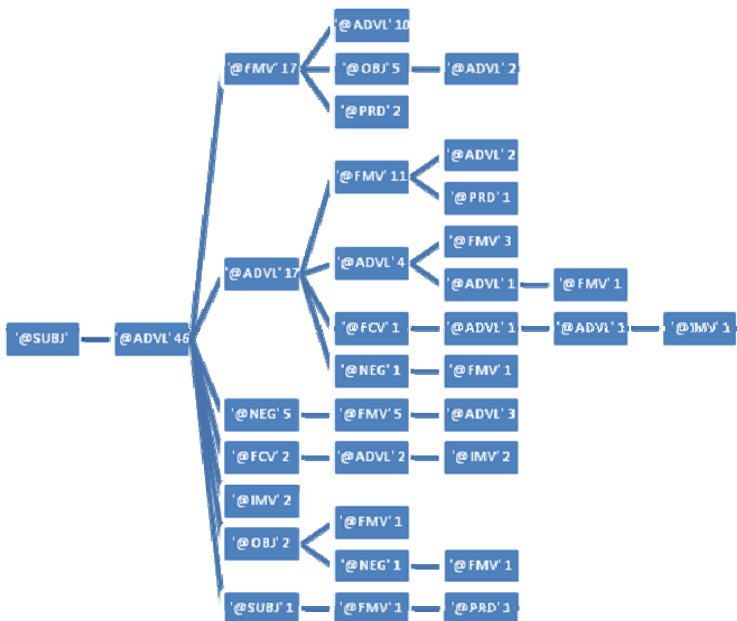
(31) *Ma (@SUBJ) ei (@NEG) tea (@FMV) , kas ma tahangi seda enam kirjutada , sulle .* ['@SUBJ', '@NEG', '@FMV']



Joonis 13. Subjektiga algav sõnajärjemuster ['@SUBJ', '@NEG']

Subjektiga algavat sõnajärjepuud ['@SUBJ', '@ADVL'], kus teisel positsioonil on adverbiaal, kasutatakse eelmistega võrreldes harva (46 korda). Kolmandal positsioonil on enamasti verbi finiiitne vorm (17 korda, vt 32. näidet) või adverbiaal (17 korda, vt joonis 14).

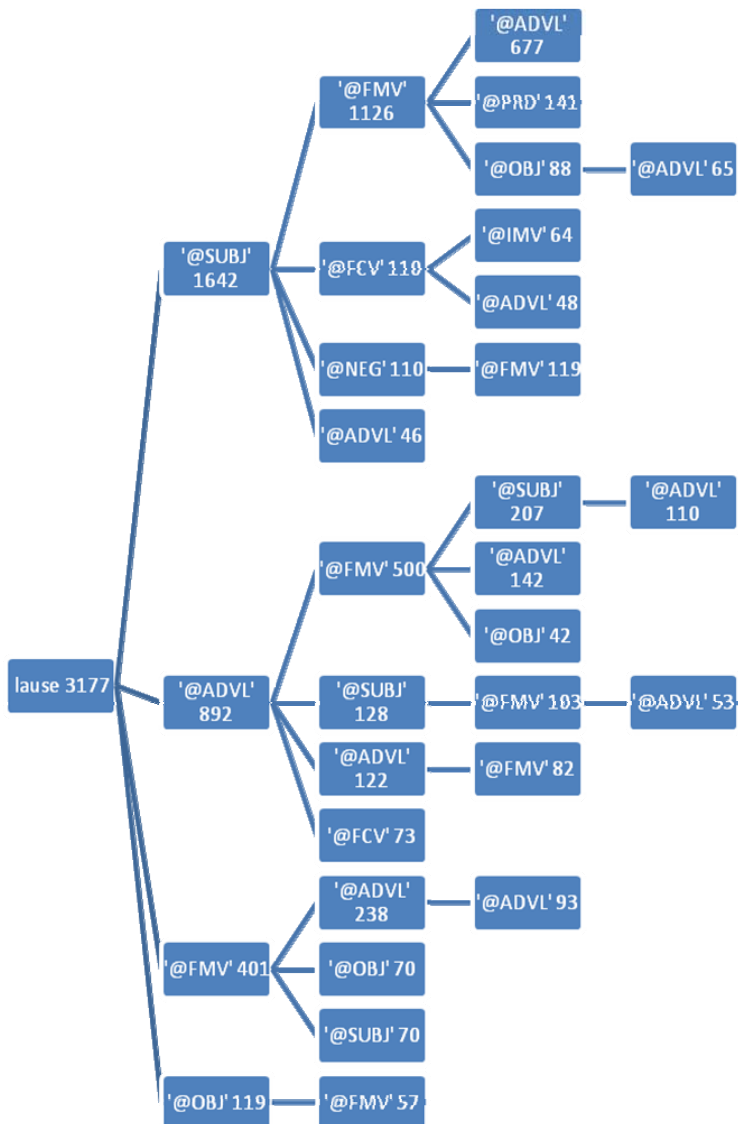
(32) *Aurutoru (@SUBJ) aina (@ADVL) puhub (@FMV) kaebavalt (@ADVL) ja kaugemalt udu seest vastavad teised laevad koledalt ja hädaohtu aimates nagu linnud , kes üksteist kiskja eest hoiatavad , kes neid kuskil luurab .*
 ['@SUBJ', '@ADVL', '@FMV', '@ADVL']



Joonis 14. Subjektiga algav sõnajärjemuster ['@SUBJ', '@ADVL']

4.8. Sõnajärjepuudest kokkuvõtvalt

Vaatamata sellele, et valimi alusel leitud korrektsetest sõnajärjemallidest genereeriti sõnajärjepuud, milles leidub rohkem või vähem regulaarseid sõnajärjemustreid, oleks mõttekas minna veel astme võrra kõrgemale ja tuua välja ka üks üldine levinumate sõnajärgede andmepuu (vt joonis 15). Selleks on programmi jaoks jätud nähtavaks vaid need märgendid, mida analüüsitud valimis on esinenud vähemalt nelikümmend korda (kokku 3177 juhtumit). Selle piirangu alusel on valimis kõige levinum sõnajärjemuster ['@SUBJ', '@FMV', '@ADVL']: *Ta (@SUBJ) ärkab (@FMV) hommikul (@ADVL)*, mis esines kokku 677 korral 3177-st ehk ligikaudu igal viiendal juhul.



Joonis 15. Valimi levinumad sõnajarjestused

5. Sõnajärjevealeidja prototüüp

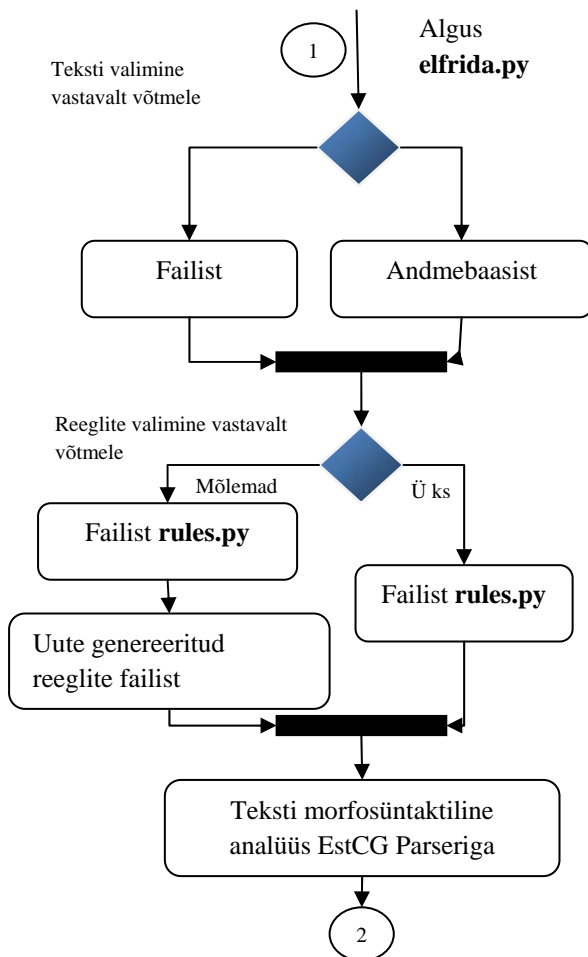
Sõnajärjevealeidja prototüübi programmeerimisel võeti aluseks leitud õigete sõnajärjemallide kogum ehk paradigma. Realiseerimiseks on kasutatud Zope andmebaasi ning programmeerimiskeelt Python.

Kõigepealt tehakse kindlaks, kas lause on vaadeldav valitud sõnajärjemallide all või kuulub selliste lausete hulka, millega meie arendatav prototüüp praegu veel ei tegele. Sõna tasandil hoitakse selliseid lisaandmeid nagu sõna lemma, morfoloogiline kirje, süntaktilised märgendid, analüsaatori töö korrektust hinnanud lingvistid parandused.

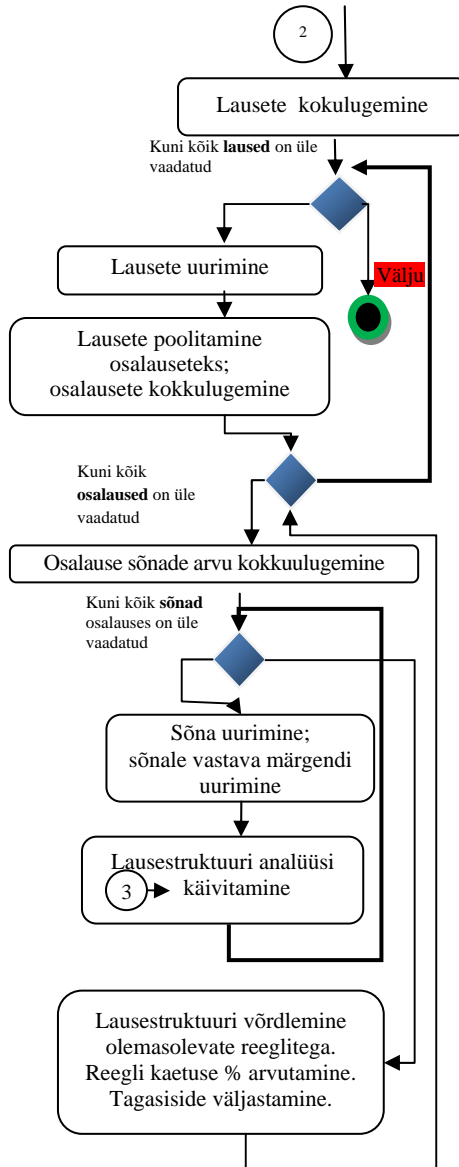
Vealeidja prototüübi põhiliseks skriptiks on *elfrida.py* (vt joonis 16). Teksti võib skripti sisestada kahel viisil: eraldi ette antud failist⁵ või andmebaasist, kus paiknevad *Eesti vahekeele korpuse* õppijakeele laused. Olemasolevad korrektse sõnajärjemallid paiknevad failis *rules.py*. Samas on eraldi fail, kus hoitakse uusi sõnajärjemalle. Kui tegu on vigadeta tekstiga, siis saab selle alusel genereerida uusi korrektseid sõnajärjemalle. Kui tegu on veamärgendusega õppijakeele tekstiga⁶, siis on võimalik tuvastada tüüpilisi sõnajärjevigu. Uued suure sagedusega esile tulevad sõnajärjemallid on otstarbekas lisada mallide paradigmasse. Et vealeidja oleks efektiivsem, selleks on võimalik rakendada kahepoolset kontrolli: ühelt poolt tehakse

⁵ Mõne teksti analüüs võib käia mitte läbi andmebaasi, vaid läbi faili, kus paiknevad just hetkel olulised laused, mille abil saadakse kiirvastuseid lause struktuuri ning märgendite kohta. Selline eraldi failist lugemine on oluline ka prototüübi testimiseks.

⁶ Eesti vahekeele korpuses on võimalik vigu märgendada ka käsitsi. Lingvistid vaatavad õppijakeele tekstid üle ning lisavad korpuse veataksnoomia alusel vastava veamärgendi. Korpuses on kõik vead liigendatud ülem- ning alamrühmadesse, vea tüübi saab valida taksonoomiast.



Joonis 16. Sõnajärje vealeidja prototüübi algoritmi



Joonis 17. Sõnajärje vealeidja prototüübi algoritm (järg)

kindlaks, kas analüüsi tulemusel korrektseks tunnistatud mall sobib konkreetse (osa)lause analüüsimiseks; teisalt vaadatakse, kas lause struktuur kuulub tüüpiliste vigaste struktuuride alla või mitte. Skriptis *elfrida.py* on arvestatud ka mõne väga sageli esineva sõnajärveega nagu eksimine V2-reegli vastu. Selle reegli kohaselt peab predikaat (@FMV või @FCV) paiknema vajalike märgendite järjestuses mitte kaugemal kui kolmandal positsioonil. Teisisõnu, kui eemaldada lausest kõik sõnad, mis ei mõjuta sõnajärge, siis ülejäänud sõnade seas ei tohi öeldis paikneda kolmandast positsioonist kaugemal.

Kui on selge, kuidas teksti sõnajärg analüsaatorisse sisestatakse ning kas on vaja kasutada uute genereeritud reeglite faili, siis suunatakse tekst süntaktilisele analüüsile.

Skript *elfrida.py* suhtleb EstCG 1,0 parseriga automaatsete päringute kaudu. Parserist saadud morfosüntaktiliselt analüüsitud tekst korrastatakse, lause ning morfosüntaktiline info pannakse eraldi muutujatesse.

Järgnevalt loetakse laused kokku ning jagatakse osalauseteks, kasutades selleks osalause piiri märgendit. Iga sõna ja sellele vastavat märgendit analüüsitakse eraldi ning võrreldakse korrektseks tunnistatud sõnajärjemallidega (vt joonis 17).

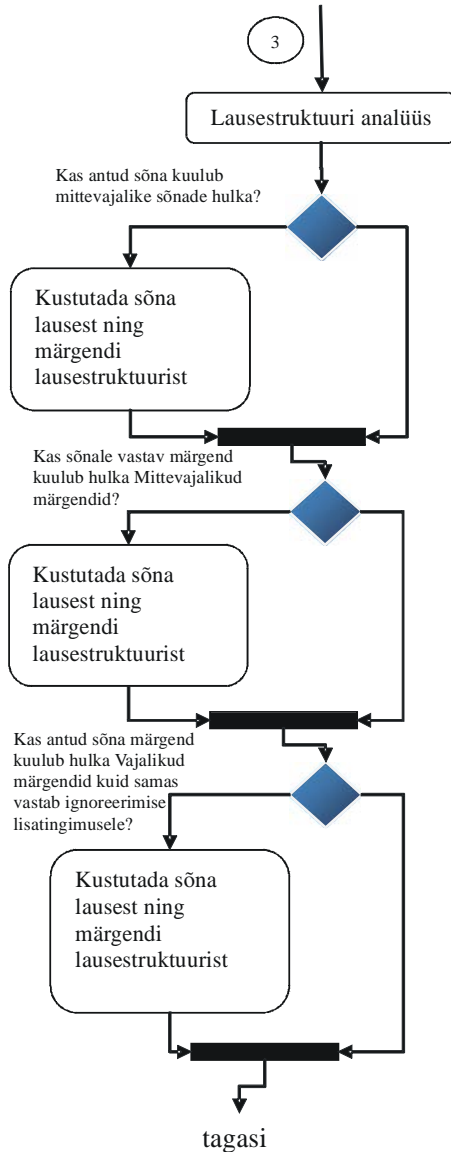
Põhimõtteliselt ei ole see algoritm piiratud esimese osalause töötlemisega, programmi ülesehitus võimaldab analüüsida ka teisi laoseosi.

Osalause leidmisel on suurimaks probleemiks õppijakeeles tehtud grammatilised ja interpunktuatsiooni vead. Kui lauses on koma puudu või sõna valesti kirjutatud, on morfoloogiliselt valesti analüüsitud ka sellised sõnaliigid nagu side- ja tegusõna

ning osalausepiiri leidmine võib nurjuda (vt Müürisep, Puolakainen 2007). Ka valesti kirjutatud sõna morfosüntaktiline analüüs võib vastuseks anda valed märgendid, mille

võrdlemine korrektsetega võib põhjustada vale tagasisidet ka lauseehituse kohta. Väljundi uurimisel tagastatakse Eesti vahekeele korpusele integreeritud väike lisaskript – info selle kohta, kas sõna on kirjutatud õigesti või mitte, mis annab võimaluse automaatselt välja selekteerida ainult need osalused, mis on kirjutatud grammatiliselt korrektselt ning määrata õigesti osalausepiir. Nagu eespool seletatud, ei vaatle sõnajärjevalitud prototüüp kirjavigu sisaldavaid lauseid.

Prototüübi järgmine samm on lausestruktuuri analüüs (vt joonis 18). Vaatluse all on üksteisele järgnevad sõnad ning nende süntaktilised märgendid. Analüüsi alustatakse esimesest sõnast ja vaadatakse, millisesse hulka see kuulub. Siin võib olla kaks võimalust, mida on eespool näidete varal demonstreeritud. Kui sõna kuulub sõnajärje seisukohalt ebaoluliste hulka, siis sõna ja sellele vastav märgend kustutatakse. Teisel juhul uuritakse sõnale vastavat märgendit ja kui selgub, et märgend kuulub mittevajalike märgendite hulka, siis nii märgend kui ka märgendile vastav sõna kustutatakse. Samas võib esineda keerulisemaid olukordi, kus märgend kuulub vajalike märgendite hulka, ent tuleb eemaldada positsiooni tõttu lauses (vt eespool).



Joonis 18. Sõnajärje vealeidja prototüübi algoritm (järg)

Kui kõik sõnad lauses ja neile vastavad märgendid on kirjeldatud moel analüüsitud, siis on saadud vajalike märgendite järjendeid võimalik võrrelda olemasolevate korrektseks tunnistatud sõnajärjemallidega. Samas võib tekkida olukord, kui (osa)lauses on sõnu, mida olemasolevad korrektset sõnajärjemallid ei kata. Sel juhul tuleb välja arvutada kaetuse koefitsient. Näiteks osalausele *Kõnelemine oli lahjaks läinud* vastav sõnajärjemall on ['@SUBJ', '@FCV', '@ADVL', '@IMV'] ning kaetuse koefitsient on 1 ehk 100%. Osalause *Ta näis nüüd lambi laualt võtvat* sõnajärjemall on andmebaasis olemas, kuid üles kirjutatud lühemana: ['@OBJ', '@FMV', '@ADVL']. Sellest johtuvalt on ka kaetuse koefitsient 0,5 ehk 50% ning lauseosa *lambi laualt võtvat* ei ole struktuuri õigsuse kohta tagasisidet saanud.

Lauseid, mille puhul kaetuse koefitsient on alla 1, on võimalik välja filtreerida, et lingvist saaks jätkata nende uurimist ja leida optimaalseid formaliseermisvõimalusi.

6. Prototüübi testimistulemused

Prototüübi testimiseks kasutati *Eesti vahekeele korpuse* B-taseme tekste, mille hulgast võeti juhuvaliku alusel 5880 lauset. Nagu eespool selgitatud, lasti prototüübil välja sorteerida laused, mis algavad sõnaga *kui, kuna* või ükskõik millise ja millises käändes küsisõnaga. Sõnajärjevigu ei otsitud ka õigekirjavigu sisaldavatest, hüüumärgiga lõppevatest või umbisikulises kõneviisis olevatest lausetest.

Kõiki ülejäänud lauseid kontrollis prototüüp järgmiste vigade osas: 1) märgend @FMV paikneb kaugemal kui teisel positsioonil; 2) enne osalausepiiri (CLB) ei ole märgendit @FMV ega @FCV; 3) märgend @PRD ei ole lauses viimasel positsioonil; 4) märgend @IMV ei ole lauses viimasel positsioonil.

Kui prototüüp ei leidnud ühtegi loetletud veakirjeldust, siis võrreldi lausete sõnajärge korrektsete sõnajärjemustritega, mida oli kokku 600. Prototüüp luges laused õigeks, kui lause oli kaetud sobiva õige sõnajärjemalliga.

Prototüübi töö efektiivsuse hindamiseks valiti 5880 lausest omakorda 300 juhuslikku lauset, kusjuures jälgiti, et suhtarvuliselt oleks väljajätavate, õigeks hinnatud ja veakahtlusega lausete hulk sama.

Prototüübi hinnang lause sõnajärje korrektsuse kohta loeti õigeks, kui see langes lingvisti hinnanguga kokku⁷. Erand tehti lausete puhul, mis sisaldasid mõnda õppijakeele viga (nt valesti kasutatud kirjavahemärk või kirjavahemärgi puudumine tingib osalause piiri vale analüüsi: **Laulupidu see on väga tähtis üüritus eestlasteks, sest seal ...*). Ka need laused, mis sisaldasid pärisnime ja mida Filosoofi eesti keele speller ei tundnud (nt *Kõige meelsamini vaatab Zaura vanu must-valgeid filme*), ei läinud prototüübi vealeidja töö efektiivsuse üle otsutamisel arvesse. Pärinimede morfoloogilise analüüsi probleemiga seoses on Heiki-Jaan Kaalep ja Tarmo Vaino (2000) ühe lihtsaimalt rakendatava lahendusena välja pakkunud tüpograafiliste konventsioonide kasutamise: pärisnimed algavad suurtähega; nimede äratundmise teeb lihtsamaks asjaolu, et sõnastikust puuduvad sõnad kuuluvad teatud väikesesse arvu muuttüüpidesse.

Kokkuvõttes pidas prototüüp juhuvaliku alusel saadud 300 lausest vigaseks 143, korrektseks 72 ja väljajätmisele kvalifitseeruvaks 85 lauset. Lingvisti hinnangu alusel olid samad näitajad vastavalt 146, 75 ja 79. Seega langesid vaadeldud lausete puhul prototüübi töö ja lingvisti hinnangud kokku 87,82% ulatuses.

⁷ Prototüübi tööd kontrollis lingvistika magistrant Hanna Sinijärv

Kümnel korral leidis lingvist, et lause on korrektne, kuid prototüüp pidas seda vigaseks. Nendest juhtudest kaheksal korral leidis prototüüp, et kuna öeldis paikneb lauses kaugemal kui kolmandal positsioonil (nt *Esialgu tal see ebaõnnestub, sest ...*), siis on see sõnajärg vigane. Kahel ülejäänul korral ei olnud predikatiiv või infiniitne verb lauses viimasel positsioonil (*Ma ise olen **olnud** Inglismaal, Soomes, Venemaal, Ukrainas, Lätis, Leedus*) ning ka selle sõnajärje tunnistas prototüüp ebakorrektsesks. Põhjus on selles, et prototüüpi pole veel treenitud neid sõnajärjemalle korrektsetena vaatlema.

Kontrollitud lausetest olid 18 niisugused, mille puhul leidis programm, et lause on korrektne, kuid lingvist luges sõnajärje vigaseks. Enamasti põhjustas laharvamuse adverbiaali positsioon, nt **See (@SUBJ) on (@FMV) armastusest (@ADVL) film (@PRD)*, vrd *See on armastusfilm* või *See film on armastusest*. Et prototüüp semantikaga arvestada ei oska ja sama märgendijärjend võiks olla ka täiesti korrektne (nt *Mari on hommikuti ilus*), siis on taoliste juhtumite kvalifitseerimine prototüübi edasisel arendamisel suur proovikivi.

Kuigi Eesti vahekeele korpuse vealeidja, sh sõnajärjevealeidja arendamisel on ees palju tööd, saab prototüüpi pidada küllaltki tõhusaks, mis annab soodsa stardipositsiooni edaspidiseks. Seda enam, et paljude praeguste kitsaskohtade lahendused on juba töös. Näiteks suudab automaatne sõnajärjevealeidja koos VAKO-projekti raames loodud õppijakeele lemmatiseerija-oletajaga tulevikus tõenäoliselt analüüsida ka neid lauseid, kus õppija on sõna valesti kirjutanud või eksinud vormimoodustusreeglite vastu.

7. Kokkuvõte

VAKO-projekti *Eesti vahekeele korpuse keeletarkvara ja keeletehnoloogilise ressursi arendamine (2008–2010)* raames on uuritud eesti keele sõnajärjemalle ja -mustreid, genereeritud sõnajärje andmepuud ning loodud sõnajärje vealeidja prototüüp, mis võimaldab kontrollida (osa)lause süntaktilist struktuuri. Prototüübi graafiline liides on valmimisjärgus. Edasine eesmärk on välja arendada keeleõppijale mõeldud tugisüsteem, mille abil saab nii verifitseerida lauseehituse ja sõnajärje korrektsust kui anda soovitusi õige sõnajärjemalli valimiseks ja kasutamiseks. Sellised lahendused on mõnede teiste keelte puhul juba osaliselt realiseeritud, nt inglise keele speller pakub kasutajale õige sõnajärje variandi. Senised VAKO-projekti raames saadud tulemused on olulised ka eesti keele spelleri edasiseks täiustamiseks.

Eesti keele sõnajärjevealeidja prototüüpi saab kasutada uute sõnajärjemustrite otsimiseks ning nende sarnasuse ja kasutuserinevuste võrdlemiseks näiteks sama autori eri ajal kirjutatud tekstides, erinevates allkeeltes ja žanrites, erinevate korpustainete analüüsis. Artiklis kirjeldatud meetod lubab võrrelda eesti keele erinevaid kasutusvariante (nt õppijakeelt ja kirja-keelt), erinevate keeleoskustasemetega (nt A2–B1–B2–C1) sõnajärjemustreid ja morfosüntaksit, individuaalse keelekasutuse eripära ning välja tuua ühe või teise sõnajärjemustri eelistamisega kaasnevad morfosüntaktilised piirangud. Seega on sõnajärje empiirilise uurimise käigus saadud andmepuud olulised mitte ainult keeletehnoloogilistes rakendustes, vaid ka keeleteaduses ja eesti keele õppes. Kuna tegu on universaalsusele pretendeeriva statistikal põhineva programmiga, siis saab kirjeldatud meetodit ja prototüübi algoritmi kasutada ka teiste keelte sõnajärje uurimisel. Tingimuseks on nõue, et uuri-

tava keele lause süntaktiline struktuur lubaks jaotust vajalike ja mittevajalike märgendite vahel.

Eesti keele uurimise seisukohast avab käesolev sõnajärjeuuring selle keerulise nähtuse piire, luues uusi perspektiive sõnajärjega seotud probleemide lahendamiseks.

Kirjandus

Arppe, Antti 2000. Developing a Grammar Checker for Swedish. – Proceedings from the 12th Nordiske datalingvistikkdager, Trondheim, December 9-10, 1999 / Ed. by Torbjorn Nordgard. Department of Linguistics, Norwegian University of Science and Technology (NTNU). Trondheim: University of Trondheim, 13–27. <http://www.ling.helsinki.fi/~aarppe/Publications/Nodalida-99.pdf>, 28.08.2010.

Athanaselis, Theologos, Stelios Bakamidis, Ioannis Dologlou 2006. A Fast Algorithm for Words Reordering Based on Language Mode. – 16th International Conference, Athens, Greece, September 10–14, 2006. Proceedings, Part II, 943–951. <http://www.springerlink.com/content/q646768285871122/fulltext.pdf>, 02.01.2010.

Beesley, Kenneth R. 1988. Language identifier: a computer program for automatic natural-language of on-line text. – Language at crossroads: Proceedings of the 29th Annual Conference of the American Translators Association, Oct 12–16, 47–54.

Damerau, Fred 1964. A Technique for Computer Detection and Correction of Spelling Errors. – Communications of the ACM 7(3), 171–176.

EKG II = Erelt, Mati; Kasik, Reet; Metslang, Helle; Rajandi, Henno; Ross, Kristiina; Saari, Henn; Tael, Kaja; Vare, Silvi 1993. Eesti keele grammatika II. Süntaks. Lisa: Kiri. Eesti Teaduste Akadeemia Keele ja Kirjanduse Instituut. Tallinn.

Kaalep, Heiki-Jaan, Tarmo Vaino 2000. Teksti täielik morfoloogiline analüüs lingvisti töövahendite komplektis. – Arvutuslingvistikalt inimesele. Tartu: Tartu Ülikooli kirjastus, 87–99.

Koptjevskaja-Tamm, Maria, Bernhard Wälchli 2001. The Circum-Baltic languages. An areal-typological approach. – Circum-Baltic languages / Ed. by Östen Dahl, Maria Koptjevskaja-Tamm. Vol. 2. Amsterdam, Philadelphia: John Benjamins, 615–750.

Lindström, Liina 2005. Finiitverbi asend lauses. Dissertationes philologiae estonicae Universitatis Tartuensis 16. Tartu: Tartu Ülikooli kirjastus.

Matsak, Erika, Helena Metslang, Jaagup Kippar 2010. The prototype of word order assessment at the Estonian Interlanguage Corpus. – The 2010 International Conference on Artificial Intelligence ICAI 2010. Las Vegas, Nevada, USA (July 12–15, 2010). Vol. II. Las-Vegas: CSREA Press, 870–875.

Metslang, Helena, Erika Matsak 2010. Kesksete lausekomponentide järjestus õppijakeeles: arvutianalüüsi katse. – Eesti Rakenduslingvistika Ühingu aastaraamat 6 / Toim. Helle Metslang, Margit Langemets, Maria-Maren Sepper. Tallinn: Eesti Keele Sihtasutus, 175–193.

Müürsep, Kaili, Tiina Puolakainen 2007. Eesti keele formaalne grammatika: mudelist rakenduseni. Eesti Matemaatika Seltsi aastaraamat 2002–2003 / Toim. Andi Kivinukk, Gert Tamberg, Jan Willemson. <http://www.matemaatika.eu/printpdf/ar2002s>, 28.08.2010.

Müürsep, Kaili, Tiina Puolakainen, Kadri Muischnek, Mare Koit, Tiit Roosmaa, Heli Uiho 2003. A New Language for Constraint Grammar: Estonian. – International Conference "Recent Advances in Natural Language Processing". Proceedings. Borovets, Bulgaria, 10–12 September 2003, 304–310. <http://math.ut.ee/~kaili/papers/ranlp03.pdf>, 28.08.2010.

Müürsep, Kaili 2001. Parsing Estonian with Constraint Grammar. – Online proceedings of NODALIDA'01. Uppsala. <http://stp.ling.uu.se/nodalida01/pdf/myyrisep.pdf>, 28.08.2010.

Pedler, Jennifer 2007. Computer Correction of Real-word Spelling Errors in Dyslexic Text. PhD Thesis. Department of Computer Science and Information Systems. Birkbeck: University of London.

Tael, Kaja 1988. Sõnajärjemallid eesti keeles (võrrelduna soome keelega). Preprint KKI-56. Tallinn.

Vilkuna, Maria 1998. Word order in European Uralic. – Constituent Order in the Languages of Europe. Empirical Language Typology. EURO-TYP 20–1 / Ed. by Anna Siewerska. Berlin, New York: Mouton de Gruyter, 173–233.

The development of the prototype for an automatic word order error detector for Estonian

Erika Matsak, Pille Eslon, Jaagup Kippar

Summary

The article presents the possibilities for recognizing word order errors in Estonian, the methods used and the current results. The article concentrates on the prototype for an automatic word order error detector for Estonian developed in Tallinn University. The statistic-based program works on a method that is similar to n-grams and the rules used are the patterns formed with 9 compulsory parts of a sentence. The set of correct word order patterns were found from the fiction sub-corpus of Tartu University's Corpus of Written Estonian.

For the statistically reliable results and the utmost efficiency and speed of the program, the rules were placed in a tree structure. The prototype starts the searches by finding a proper initial tag and continues to find a correct compatible pattern that has the highest frequency rate.

At current stage the work is focused on detecting the right/wrong position of the finite/infinite verb and the predicative (since most commonly Estonian is known as a verb second language). Prototype's efficiency was tested on Estonian learner language corpus texts. In the test described in the article 5880 sentences were analyzed with the error analyzer and 300 sentences of the output were assessed. The prototype estimated the correctness of the word order properly in 87.82% of the cases.

Although there are a number of problems that still need to be solved including the misspelled or unknown words (i.e. proper nouns) and erringly unmarked clausal border, the method and the algorithm of the prototype for an automatic word order error detector for Estonian could also be used on other languages' word order studies as well.

The article is summarized with the survey of the problems occurred on word order detection and the possible ways to make the detector more efficient.

Keywords: morphosyntax, automatic error detection, word order errors

Autorid

Erika Matsak, Tallinna Ülikooli informaatika instituudi dotsent
matsak@tlu.ee

Pille Eslon, Tallinna Ülikooli eesti keele ja kultuuri instituudi
vanemteadur, dotsent
peslon@tlu.ee

Jaagup Kippar, Tallinna Ülikooli informaatika instituudi lektor
jaagup.kippar@tlu.ee

Menen seuraavaan kurssiin

**KORPUSPOHJAINEN TUTKIMUS
RUOTSINKIELISTEN
SUOMENOPPIJOIDEN ILLATIIVIN
JA ALLATIIVIN KÄYTÖSTÄ
KIRJALLISESSA TUOTANNOSSA**

Tuija Määttä

Abstrakti

Artikkelissa tarkastellaan ruotsinkielisten suomenoppijoiden illatiivin ja allatiivin käyttöä. Tutkimus on korpuspohjainen, Tutkimusaineistona on käytetty Kansainvälisen oppijansuomen korpukseen (ICLFI, International Corpus of Learner Finnish) sisältyvää Ruotsin osakorpusta, joka tutkimuksen tekovaiheessa sisälsi 43 496 sanetta. Aineisto on käsitelty WordSmith Tools -ohjelmaan sisältyvällä Concordance-ohjelmalla. Suurimassa osassa illatiivissa ja allatiivissa taivutetuista sanoista sijapääätteen valinta on norminmukainen. Aineistossa on kuitenkin tapauksia, joissa illatiivia on käytetty allatiivin sijasta ja myös allatiivia illatiivin sijasta. Artikkelissa pohditaan tällaiseen norminvastaiseen käyttöön johtaneita mahdollisia syitä. Esille otetaan mm. lähdekielen vaikutus, maailmankuvan hahmottaminen ja käytettyjen oppimateriaalien, lähinnä sanaston, vaikutus illatiivin ja allatiivin valintaan. Tässä tutkimusvaiheessa on kuitenkin liian varhaista esittää mitään

varmoja tuloksia norminvastaiseen sijavalintaan vaikuttaneista tekijöistä.

Avainsanat: oppijansuomi, korpuspohjainen tutkimus, illatiivi, allatiivi

1. Johdanto

Suomen kieliä on nykyään monenlaisia. Yksi variantti muiden joukossa on *oppijansuomi*. Oppijansuomella tarkoitetaan sellaista suomen kielen muotoa, jota suomen kieltä opiskelevat eri äidinkielliset opiskelijat ja oppijat käyttävät. Oppijansuomen käyttäjien määrä kasvaa samoin kuin muidenkin oppijankielivarianttien käyttäjien. Suomea on jo kauan voinut opiskella lukuisissa opetuspisteissä kautta maailman, nykyisin yli 100 opetuspisteessä. Syitä, jotka vauhdittanevat kiinnostusta suomen kielen opintoja kohtaan Suomen rajojen ulkopuolella, ovat vilkas kansainvälistyminen ja monin tavoin parantuneet liikkumis- ja kontaktimahdollisuudet. Suomessa syy oppijansuomen käyttäjien ja käytön lisääntymiseen on kasvanut maahanmuutto.

Suomessa oppijansuomi tutkimuskohteena on aika nuori tutkimusala. Eri näkökulmista ja eri tutkimusteorioita käyttäen oppijansuomea on tutkittu vajaa parisenkymmentä vuotta. Aiheesta on julkaistu artikkeleita, antologioita ja eritasoisia opinnäytetöitä useissa yliopistoissa¹. Aineistot näissä tutkimuksissa eivät aina ole olleet kovin laajoja, ja tutkimusmenetelminä on useimmiten käytetty manuaalisia menetelmiä.

¹ Ks. esim. https://www.jyu.fi/hum/laitokset/kielet/ppiaineet_ks/fennicum/opiskelu/s2/s2, 15.09.2010.

Joitakin laajempiakin tutkimuksia on julkaistu. Mainittakoon niistä esim. Yrjö Lauranton, joka on tarkastellut espanjankielisten suomenoppijoiden paikallissijojen käyttöä ja merkityksiä. Tutkimuksessa esitetään myös näkemyksiä paikallissijojen opettamisesta ja oppimisjärjestyksestä (ks. Lauranto 1997). Kirsti Siitonen on tarkastellut edistyneiden suomenoppijoiden kielimuotoa. Keskiössä ovat *U*-johdoksiset verbit ja se, kuinka suomenoppijat tuntevat näitä verbejä ja hallitsevat niiden käyttöä (ks. Siitonen 1999).

2. Ruotsinkielisten suomenoppijoiden illatiivin ja allatiivin käyttö tutkimuskohteena

Ruotsinkielisillä opiskelijoilla, jotka opiskelevat suomea vieraana kielenä alkeistason kurssilla, on joukko kieliopillisia asioita, joiden omaksumiseen ja jotakuinkin aktiiviseen käyttämiseen kuluu aikaa. Yksi tällainen asia on suomen kielen paikallissijat. Kestää suhteellisen kauan, ennen kuin opiskelijoille selkenee esimerkiksi sisä- ja ulkopaikallissijojen peruserot ja -käyttötavat.

Ruotsinkieliset suomenoppijat sekoittavat joskus illatiivin ja allatiivin käytön erilaisissa teksti- ja kommunikointiyhteyksissä. Tässä ensimmäisessä tutkimusvaiheessa kartoitan mainittujen sijojen käyttöä ja pohdin mahdollisia syitä, joista tällainen sekakäyttö saattaisi johtua.

Tutkimukseni tässä vaiheessa en pysty vielä esittämään mitään konkreettisia tuloksia niistä syistä, jotka voisivat vaikuttaa illatiivin ja allatiivin sekoittuneeseen käyttöön. Nostan kuitenkin esille seuraavat kysymykset:

- 1) Onko kohdekielen, suomen, ja lähdekielen, ruotsin, rakenteellisilla eroilla tekemistä sijavalinnan kanssa?
- 2) Vaikuttaako yleinen maailmankuvan hahmottaminen sijavalintaan?
- 3) Onko opetuksessa käytetyillä oppimateriaaleilla vaikutusta illatiivin ja allatiivin käyttötapoihin oppijankielessä?

2.1 Tutkimusaineistosta

Nykyaikainen tekniikka on mahdollistanut ja myös nopeuttanut sähköisten tutkimusaineistojen, korpusten, keruuta ja käsittelyä. Tässä artikkelissa esittelemäni tutkimus pohjautuu tutkimushankkeessa *Korpustutkimus oppijankielen kielikohtaisista ja universaaleista ominaisuuksista* (Corpus study on language-specific and universal features in learner language) koottavaan *Kansainvälisen oppijansuomen korpuksen* (International Corpus of Learner Finnish, ICLFI) Ruotsin osakorpukseen, jonka sisältämät tekstit on kerätty Uumajan yliopistossa.

Tehdessäni tämän tutkimuksen syksyllä 2009 Ruotsin osakorpus sisälsi 43 496 sanetta. Kyseinen osakorpus on kerätty vuosien 2006–2009 välisenä aikana. Suomenoppijat ovat äidinkieltään ruotsinkielisiä, ja he opiskelevat suomea vieraana kielenä suomen kielen vasta-alkajien alkeistason kursseilla. Kurssi etenee puolivauhdilla ja kestää täten koko lukuvuoden. Kurssi sisältää kielioppia, suullista ja kirjallista kielivalmiutta ja tekstinyymmärtämistä. Viikon aikana pidetään kuusi oppituntia.

Kirjallisen kielivalmiuden osioon kuuluu pakollisena kirjoitelmien tuottoa. Lukuvuoden aikana kirjoitetaan keskimäärin 10 tuotosta, joiden pituus on n. 30 lausetta. Tehtävät suoritetaan

itsenäisesti kotona ja apuneuvot ovat sallittuja. Näiden tekstien lisäksi korpuksessa on mukana erityyppisiä tekstejä, jotka on kirjoitettu kirjallisen kielitaidon kurssin loppu-
tenteissä ilman apuneuvoja, ja kirjoitusaika on ollut rajattu.

Kirjoitelmien aiheet ovat opettajan antamia, ja ne pohjautuvat suurimmaksi osaksi oppikirjan lukuteksteihin. Oppikirjana on mainittujen vuosien aikana ollut Anna-Liisa Lepämaan ja Leena Silfverbergin kirjoittama *Suomen kielen alkeisoppikirja* (Lepämaa, Silfverberg 2009). Oppikirja on käytössä kurssin kaikissa osioissa, ja lisäksi on käytetty opettajan laatimia deskriptiivisiä kielioopin opintomonisteita (Määttä 2006; 2007), sanastoja ja muuta apumateriaalia.

ICLFI-korpuksen kaikki tekstit on koodattu mm. eri tekstityyppeihin. Ruotsin osakorpuksen teksteistä suurin osa on kertomuksia. Mukana on myös kirjeitä, kortteja, sarjakuva-
tekstejä, päiväkirjoja, oppikirjan tekstin jatkamistekstejä ja uutisia. Tässä tutkimuksessani en ole ottanut huomioon näitä eri tekstityyppejä, vaan olen käsitellyt aineiston yhtenä massana.

ICLFI-korpuksen sisältämät tekstit ovat eri taitotasoilta. Tasojen mittarina on käytetty yleiseurooppalaisen viitekehyksen (*Common European Framework of Reference for Languages: Learning, Teaching, Assessment, CEFR*) taitotasoasteikkoa, jonka mukaan taitotasot ovat seuraavat: alkeistaso A1-A2, keskitaso B1-B2 ja edistyneet C1-C2. Koko korpusta ei ole vielä arvioitu viitekehyksen mukaan, vaan nykyisellään suomen kielen kurssien opettajat itse arvioivat opiskelijoidensa tekstien tasot. Uumajan osakorpuksesta käyttämiäni tekstien taitotaso on A2.

2.2 Tutkimusmenetelmästä

Korpuksen analysoinnissa olen käyttänyt *WordSmith Tools* -ohjelmaan sisältyvää *Concordance*-ohjelmaa. Ohjelmaa käytin niin, että sain listatuksi kaikki korpuksessa esiintyvät sanat, jotka päättyvät illatiivin *Vn-*, *hVn-* ja *sVVn-*sijapäätteisiin ja allatiivin *lle-*sijapäätteeseen. Suurin osa näissä kahdessa sijassa taivutetuista sanoista on yksikössä. Korpuksessa on joitakin monikon illatiivin sijapäätteisiä yksittäisiä sanoja ja muutamia fraaseja kuten *mennä naimisiin*, *mennä kihloihin*.

Artikkelin kaikki numeroidut esimerkit ovat peräisin Ruotsin osakorpuksesta. Näyte lauseista (esim. 1–12), joissa on *WordSmith Tools* -ohjelmalla haettuja sanoja taivutettuna allatiivimuodossa (kursivoinnit kirjoittajan):

- (1) tarkoitan, että joskus kun menet johtaja *Närhelle*,
- (2) että mina aion tehdä harakka-tatuointi *käsivarelle*.
- (3) Timo päti että oli hyvä aika soittaa *Kaisalle*.
- (4) tilaisin heti matkan *New Yorkille*.
- (5) Menisin pitkälle *matkalle*.
- (6) Illalla menimme setä *Håkanille* ja katsoimme heidän pienen
- (7) hakea rahaa niiten kirjoitus- tai *taideprojektille*.
- (8) Menen sitten juomaan kahvia *poliisiasemaalle*.
- (9) Työn jälkeen tavasin *Cecilialle* ja lainasin hänelle kolmen
- (10) Onni *Karvoiselle*, niin Osku on aika tyhmä koira
- (11) Minä haluaisin lahjoittaa rahaa köyhälle *ihmiselle*
- (12) mutta hän haluisi ottaa hevosen mukaan *päiväkodille*

Saatuani esille kaikki illatiivissa ja allatiivissa taivutetut sanat koostin korpuksesta sellaiset tapaukset, joissa ei ollut käytetty

yleiskielen normien vaatimia oikeita sijamuotoja. Korpuksessa on monenlaisia esimerkkejä sekä ulko- että sisäpaikallissijojen käytöstä, mutta tarkastelen tässä artikkelisani vain sellaisia esiintymiä, joissa illatiivin ja allatiivin käyttö on sekoittunut, esimerkiksi seuraavalla tavalla:

a) illatiivi allatiivin sijasta

(13) -- tiskasin yksi lasi tippui *lattiaan* -- (lattialle)

b) allatiivi illatiivin sijasta

(14) -- ei ikinä voi tietää, kun *autolle* tulee rengasrikko -- (autoon)

Sisä- ja ulkopaikallissijoilla on suomen kielessä useita eri funktioita. Tässä tutkimuksessa suurin osa esiintymistä on sellaisia tapauksia, joissa illatiivi- ja allatiivimuotoiset sanat merkitsevät jonkinlaista konkreettista paikkaa, johon/jolle esimerkiksi mennään, lähdetään, ajetaan, tullaan. Tässä tutkimusvaiheessa en ole ottanut huomioon illatiivi- ja allatiivitausten eri funktiota enkä myöskään erotellut aineistosta rektiotapauksia. (Ks. Lauranto 1997: 178–180, Siivelt 2009: 122–126.)

ICLFI-korpuksen annotointi on vasta aluillaan. Tämä merkitsee sitä, että tutkijan ei ole vielä mahdollista tietokoneohjelmia käyttäen saada teksteistä tietoa esimerkiksi morfologisista tai syntaktisista ilmiöistä.

Keaty Siivelt on testannut tekemässään pilottitutkimuksessa vironkielisten suomenoppijoiden sisäpaikallissijojen esiintymisfrekvensseistä ja täsmällisyydestä ICLFI-korpuksen kuuluvaa materiaalia, joka on annotoitu Connexor fi-fdg -jäsennysohjelmalla. Siiveltin materiaalin annotointi sisältää vain morfologisen koodauksen (Siivelt 2010: 3).

Siivelt toteaa, että materiaalista ei voinut saada esiin mm. sellaisia tapauksia, joissa suomenoppija oli käyttänyt väärää sijaa: *Lähden yliopistossa, käyn suihkuun* (Siivelt 2010: 8). Tällaiset tapaukset kuuluvat morfosyntaksin alueelle, mikä merkitsee, että aineiston olisi pitänyt olla koodattu myös syntaktisesti. Esimerkiksi Ilmari Ivaskan ja Kirsti Siitosen mukaan annotoinnissa on kolme koodausvaihtetta: morfologinen, syntaktinen ja semanttinen ja/tai virhekkoodaus (Ivaska, Siitonen 2009: 54–71). Oppijankielen projektin tavoitteena on koodata ICLFI-korpus kokonaisuudessaan kaikkia näitä kolmea koodausta käyttäen.

2.3 Paikallissijojen opettamisesta

Uumajan yliopiston suomen kielen alkeiskurssilla käytettävässä oppikirjassa (ks. Lepämaa, Silfverberg 2009) paikallissijat esitellään kolmannessa kappaleessa. Sijat esitellään ensin kuvasarjan avulla ja sitten annetaan eri nimityyppien joukosta valmiiksi taivutettuja paikallissijamuotoja. Tässä yhteydessä opiskelijat tutustutetaan myös kyseisten nimityyppien vartaloiden käsitteeseen. Nimityyppien joukkoon on ujutettu kaksi sanaa, *katu* ja *pöytä* (esiintyvät myös tekstissä), joiden taivutetuissa muodoissa on astevaihtelu. Yleensä opiskelijat reagoivat näiden sanojen muotoihin ja haluavat tietää, mitä sanoissa on tapahtunut. Astevaihtelun perustapaukset käsitelläänkin sitten oppikirjan neljännessä kappaleessa.

Paikallissijojen esittelyn jälkeen oppikirjassa on teksti, jossa on käytetty paikallissijoja. Melkein kaikki esiintymät ovat paikan funktiossa: mennään johonkin/jollekin, ollaan jossakin/jollakin,

tullaan jostakin/joltakin. Ainoastaan adessiivin käytöstä mainitaan erikseen sen instrumentaalinen käyttö (*Millä sinä tulet työhön? Tulen junalla.*).

Lauranto on pohtinut mm. paikallissijojen oppimista ja opettamista. Hän ottaa esille esimerkiksi paikallissijojen opettamisen ajankohdan ja asettaa kysymyksen: ”Voiko paikallissijoja alkaa käyttää alkeisopetuksessa ennen vartalon, vartalomuutosten ja vokaaliharmonian analyysiä?” (Lauranto 1997: 191). Hänen vastauksensa on myönteinen, ja perusteluna annetaan mm. opiskelijoiden kyky havainnoida asioita, joita ei ole vielä tarkasti analysoitu. Tällä tarkoitetaan esimerkiksi vartaloissa tapahtuvia muutoksia (astevaihtelu) ja sijapäätteiden vokaalivaihteluja (vokaaliharmonia). Näin menetellessä kysyiset kielen ilmiöt ovat jo tuttuja siinä vaiheessa, kun astevaihtelua ja vokaaliharmoniaa tarkastellaan tarkemman analysoinnin kautta.

Paikallissijojen opetusjärjestyksestä Lauranto esittää kuusi-kohtaisen järjestykskaavion, jota hän on soveltanut julkaisemassaan oppikirjassa (Lauranto 1995). Opetus etenee seuraavasti: 1. sporadiset esiintymät, jotka esiintyvät oppikirjassa ja opettajan puheessa, 2. aktiivinen/passiivinen käyttö, 3. esittely ja alustava analyysi, jolloin sijat esitellään luonnollisissa jokapäiväisissä kommunikointiyhteyksissä, 4. aktiivinen käyttö, 5. varsinainen analyysi, paikallissijojen muodostukseen liittyvät asiat, 6. aktiivinen käyttö (Lauranto 1997: 193).

Lepäsmaan ja Silfverbergin oppikirjassa (2009) on sekä ensimmäisessä että toisessa kappaleessa joitakin paikallissijoissa taivutettuja sanoja ja fraaseja. Molemmissa kappaleissa on paljon dialogeja, joissa on kysymyksiä ja vastauksia. Näiden kautta opetellaan eri kysymyssanoilla alkavia kysy-

myksiä ja verbin persoonataivutusta. Paikallissijoissa esiintyvistä kysymyssanoista ovat mukana vain *missä* ja *mihin* (*aikaan*). Kysymyssanojen lisäksi dialogeissa esiintyy deiktisistä adverbeista muodot *täällä*, *tuolla* ja *siellä*.

Oppikirjan toisessa kappaleessa esitellään mm. lukusanat, kellonajan ja vuorokaudenaikojen ilmaiseminen. Dialogeissa vuorokaudenajoista on mukana adessiivimuotoiset ilmaukset *illalla* ja *aamulla*. Kaikkien mainittujen yksittäisten sanojen lisäksi opetetaan fraasit *käydä suihkussa*, *olla työssä*, *olla ajoissa/myöhässä* ja *olla oikeassa/väärässä*, mutta tässä opetusvaiheessa taivutettuihin muotoihin ei puututa.

Ruotsinkieliset suomenoppijat ovat siis saaneet tutustua etukäteen joihinkin paikallissijoissa taivutettuihin sanoihin ja fraaseihin ennen niiden varsinaista opettamista ja oppimista. Näin ollen he tunnistavat helpommin jotkin uudet sanat ja taivutusmuodot.

3. Lähde- ja kohdekielen paikallisuuden ilmaisemisen erilaisuus paikallissijan valintaan vaikuttavana tekijänä

Yksi suomen kielelle tyypillinen piirre on sen paikallissijasysteemi, joka on kolmijakoinen. Systeemiin kuuluu tulosijat (illatiivi, allatiivi), osijat (inessiivi, adessiivi) ja erosijat (elatiivi, ablatiivi). Vastaavanlaista paikkaa ilmaisevaa järjestelmää ei ole esimerkiksi indoeurooppalaisissa kielissä kuten ruotsissa (Hakulinen 1979: 522). Kun vielä otetaan huomioon se, että suomenoppijan on kyettävä ajattelemaan, tapahtuuko jokin toiminta jonkin tilan sisäpuolella (sisäpaikallissijat illa-

tiivi, inessiivi, elatiivi) vai ulkopuolella (ulkopaikallissijat allatiivi, adessiivi, ablatiivi), voi ymmärtää, että oikean sijan valinta voi olla hyvin vaikeaa.

Suuraavista esimerkeistä (ks. 15–17) käy hyvin ilmi se, että suomenoppijan on vaikea pitää tilakäsitys mielessään valitessaan sijamuotoa.

(15) Nukuimme *lentokonella*. (lentokoneessa)

(16) Joskus menemme *pubille*. (pubiin)

(17) Vien lasin *tiskipöytään*. (tiskipöydälle)

ISK 2004 definioi, että konkreettisesti paikallisessa käytössään illatiivi, inessiivi, elatiivi ilmaisevat sijaintia rajatun tilan sisäpuolella, reunojen tai epämääraistenkin rajojen ympäröimänä, sekä siirtymistä sellaisten rajojen sisältä sieltä pois. Ulkopaikallissijat, allatiivi, adessiivi ja ablatiivi, ovat puolestaan ISK:n mukaan sellaisia jotka konkreettisesti paikallisessa käytössään ilmaisevat sijaintia entiteetin ulko- tai yläpinnalla tai siirtymistä sellaiselle tai sellaiselta pois.

Paikallisuus on täten luonteeltaan kaksijakoinen. Suomenoppijan on kyettävä ajattelemaan valitessaan sisä- ja ulkopaikallissijojen välillä, missä tilassa jokin toiminta tapahtuu.

Taulukko 1. Toiminnan paikallisuus

Toiminta tapahtuu jonkin (rajatun paikan) sisällä	Toiminta tapahtuu paikan päällä (pinta) tai jonkin paikan ulkopuolella
Illatiivi <i>Kävelen kauppa/an</i>	Allatiivi <i>Menen tori/lle</i>
Inessiivi <i>Olen sauna/ssa</i>	Adessiivi <i>Seison piha/lla</i>
Elatiivi <i>Tulen koulu/sta</i>	Ablatiivi <i>Palasin asema/lta</i>

Paikallissijan valinta on vaikeaa, koska samanaikaisesti suomenoppijan on pidettävä mielessä paikallisuuden kaksijakoisuus ja tämän lisäksi myös hahmotettava liikkeiden eri suunnat:

a) toiminta paikan sisäpuolella

(18) Timo meni *baarille* ja osti kaksi olutta niille. (baariin)

b) toiminta paikan pinnalla

(19) Ja lapset pelaavat peli *tontissa*. (tontilla)

Rune Ingo kiinnittää huomiota suomen kielen dynaamiseen maailmankuvaan. Hän toteaa, että ”dynaaminen maailmankuva näkyy hyvin todellisen, ajatellun tai ajateltavissa olevan **liikkeen** tarkassa ilmaisemisessa”. Hänen mukaansa liikkeen suunta ratkaisee paikallissijan valinnan: ilmaistaessa jonkin toiminnan liikettä johonkin käytetään tulosijoja (tullaan johonkin/jollekin). Jos taas toiminnan laatu on staattista tai pysyy jossakin tietyssä paikassa, sijaksi valitaan olosija (tehdään jotakin tai ollaan jossakin/jollakin). Toiminnan liikkeen suunnan ollessa pois päin jostakin on käytettävä erosijoja (tullaan (pois) jostakin/joltakin) (Ingo 2000: 271–272).

Taulukko 2. Liikkeen suunta

Liikkeen suunta	Sisäpuolella	Ulkopuolella
Tulosija	Illatiivi	Allatiivi
Olosija	Inessiivi	Adessiivi
Erosija	Elatiivi	Ablatiivi
↓	↓	↓
<i>Tyttö menee</i>	<i>talo/on</i>	<i>ranna/lle</i>
<i>Tyttö on</i>	<i>talo/ssa</i>	<i>ranna/lla</i>
<i>Tyttö tulee</i>	<i>talo/sta</i>	<i>ranna/lta</i>

Suomen kolmijakoinen paikallissijasysteemi on vieras ruotsinkielisille siksikin, että ruotsissa käytettävät jotkin prepositiot ovat käytössä ilmaistaessa sekä ero- että tulosijoja. Ruotsissa ei siis huomioida paikallisuuden laatua, tapahtuuko jokin toiminta jonkin tilan sisä- vai ulkopuolella. Myös Eeva Lähdemäki, joka on tutkinut kontrastiivista analyysiä ja virheanalyysiä käyttäen suomenruotsalaisten ns. pienessä kielikohteessa ja Suomen ruotsinkielisissä kouluissa kirjoitettuja suomen kielen A-tason ylioppilasaineita, on todennut, että sijajärjestelmän kolmijakoisuus on vaikea ruotsinkielisille (Lähdemäki 1995: 91).

Ruotsin substantiivit esiintyvät perusmuodossa (vastaa suomen nominatiivia) ja genetiivissä (ks. Teleman *et al* 1999: 112). Olof Thorellin mukaan sija ei vaikuta substantiivin omaan merkitykseen vaan ainoastaan sen syntaktisiin funktioihin. Mainittakoon, että joillakin pronomineilla on kuitenkin objektin funktiossa oma muoto (ks. Thorell 1977: 45, 78). Lähdemäen mukaan nykyruotsin frekvenssisanakirjassa (Allén 1980) on 69 prepositiota, joista ”vajaalle 20:lle saattaa löytää vastineen suomen sijamuodoista” (Lähdemäki 1995: 91). Ulla Laurén, joka on tarkastellut kaksikielisten oppilaiden kielivirheitä ja interferenssiä, sanoo vastaavuuksien prepositioiden ja sijamuotojen välillä olevan mutkikkaita. Esimerkkinä hän mainitsee, että yhdellä ja samalla sijamuodolla voi ruotsissa olla viisikin eri prepositiota, ja kääntäen, että yhtä prepositiota kohti voi suomessa olla kymmenisen vastinetta (Laurén 1991: 191).

Seuraava taulukko näyttää, mitä prepositioita ruotsissa käytetään vastaamassa suomen sisä- ja ulkopaikallissijoja.

Taulukko 3. Suomessa paikallissija – ruotsissa prepositio

Paikallissija	Prepositio
Sisäpaikallissijat	
illatiivi	till, in i
inessiivi	i, inne i
elatiivi	från, (ut) ur
Ulkopaikallissijat	
allatiivi	till, mot
adessiivi	på
ablatiivi	från

Taulukosta näkee, että ruotsin prepositiojärjestelmässä on sellaisia prepositioita (*till, från*), joita käytetään vastaamaan suomessa sekä sisä- että ulkopaikallissijoja. Ingo toteaaakin, että mm. ruotsissa ei huomioida paikallisuuden laatua, ja näin ollen tulosijojen (illatiivi, allatiivi) ja erosijojen (elatiivi, ablatiivi) vastineena esiintyvät samat prepositiot (Ingo 2000: 272). Tämäkin eroavaisuus kielten välillä on seikka, joka aiheuttaa ruotsinkielisille suomenoppijoille vaikeuksia ajatella paikallisuutta ”suomalaisittain” ja päätyä oikean paikallissijan valintaan (ks. esim. 20–21).

(20) Ylihuomenna matkustamme *Sveitsille*. (Sveitsiin)

’I övermorgon reser vi *till* Schweiz.’

(21) Kesällä norjalaiset tulevat *Bodenille*. (Bodeniin)

’På sommaren kommer norrmännen *till* Boden.’

Toinen aspekti, joka on tyypillinen suomen kielen maailmankuvan hahmottamisessa, on **tilan** käsittäminen. Tässäkin suhteessa suomi ja ruotsi eroavat toisistaan siten, että suomessa ajatellaan useimmiten sisäpaikallissijasysteemin kautta (ks. esim. 22–23). Vastaavissa tapauksissa ruotsissa käytetään prepositioita *på* ja *vid* (Ingo 2000: 277).

(22) Se on parempi nyt kun kun lunta on *maalla*. (maassa)
'Det är bättre nu när det är snö *på* marken.'

(23) *Samalla lentokoneella* istui Peter Forsbergin isä, Kent.
(samassa lentokoneessa)
'*På samma flygplan* satt Peter Forsbergs pappa Kent.'

4. Ruotsinkielisten suomenoppijoiden illatiivin ja allatiivin käytöstä

4.1 Illatiivin ja allatiivin esiintymistä

Tässä tutkimuksessa käytetyn korpuksen koko on 43 496 sanetta ja 4 865 lausetta. Seuraavassa taulukossa esitellään illatiivin ja allatiivin esiintymämäärät. Luvut sisältävät sekä norminmukaiset että -vastaiset esiintymät.

Taulukko 4. Illatiivin ja allatiivin esiintymämäärät

Illatiivi	
Vn	1021
hVn	111
sVVn	52
Yhteensä	1184 (69 %)
Allatiivi	
Ile	521
Yhteensä	521 (31 %)
Illatiivi ja allatiivi yhteensä	1705

Suomenoppijat ovat käyttäneet kirjallisissa tuotoksissaan illatiivia huomattavasti enemmän kuin allatiivia. Syynä saattaa yksinkertaisesti olla se, että suomen kielessä illatiivi on

yleisempi sija kuin allatiivi. Sijamuotojen frekvensseistä on tehty lukuisia tutkimuksia mm. murteista ja yleiskielestä. Tuloksista käy ilmi, että useimmiten illatiivi on yleisin sijamuoto (esim. Leino 1993: 176). Anne Nesser on tutkinut sisä- ja ulkopaikallissijojen käyttöä ruotsinsuomalaisten oppilaiden kirjoittamissa aineissa. Myös hänen tuoksensa näyttää illatiivin käytön ylivoimaisuuden. Allatiivi sen sijaan näyttää olevan frekventein ulkopaikallissijoista, ja se sijoittuu yleensä heti sisäpaikallissijojen jälkeen (Nesser 1986: 144–145). (Vrt. Lehtinen 1987 taulukko 5.)

Tarkasteltaessa illatiivin eri sijapäätteiden yleisyyttä taulukko 4 näyttää, että päätte Vn on frekventein. Sen osuus on peräti 86 % kaikista illatiivitapauksista. Päätteen hVn osuus on hieman yli 9 % ja päätteen $sVVn$ 4 %. Syy näinkin vaihteleviin käyttöeroihin on varmasti oppimateriaaleissa käytetyssä sanastossa. Suurin osa niissä esiintyvistä sanoista on sellaista nominityyppiä, jonka taivutusvartalo päättyy yhteen vokaliin, esimerkiksi, *kala-*, *tuoli-*, *auto-*, *koulu-*, *hylly-*, *päivä-*, *hölmö-*. Nähtävästi juuri se seikka, että tällaiset nominityypit toistuvat tarpeeksi usein, ja se, että illatiivimuodon muodostamiseen saa apua sanan vartalovokaalista, helpottavat oppijaa muodostamaan oikeita illatiivimuotoja helpoiten.

Yksitavuisia sanoja oppimateriaaleissa on harvakseltaan. Lepämaan ja Silfverbergin oppikirjan (2009) sanastossa on mukana vain seuraavat 14 sanaa (pronominit pois laskien): *jää*, *kuu*, *maa*, *muu*, *puu*, *pää*, *suo*, *suu*, *sää*, *tie*, *työ*, *voi*, *vyö* ja *yö*. Päätteen hVn käytön määrään oppijankielen korpuksessa on vaikuttanut sanan *työ* frekventti käyttö (46 tapausta). Sana esiintyy sekä yksikössä että monikossa, *mennä työhön/töihin*. Lisäksi korpuksessa on mukana jokseenkin paljon eri pronomien illatiivimuotoja, esimerkiksi *siihen*, *tähän* ja *mihin*.

Vähiten korpuksessa on *sVVn*-päätteisiä illatiivimuotoja. Luultavasti tämänkin illatiivipäätteen suhteellisen vähäinen käyttö johtuu yksinkertaisesti oppimateriaalien sanastosta ja myös siitä, että oppikirjassa esitellään ainoastaan neljä nimityyppiä (*e-*, *As-*, *is-* ja *nUlt*-nimityypit), joissa käytetään *sVVn*-illatiivipäätettä. Lisäksi sanastossa on mainittu vain muutama paikannimi, joiden nominatiivimuoto ja taivutusvartalo päättyvät kahteen vokaaliin (*Lontoo-*, *Espoo-*, *Porvoo-*).

Korpuksesta ja opiskelijoiden erilaisista harjoitustehtävistä on nähtävissä, että juuri *sVVn*-illatiivityypin käyttö on hankalinta, ja tuotoksen tuloksena on hyvin usein sanan genetiivimuoto. Saman havainnon on tehnyt Nesser tarkastellessaan ruotsin-suomalaisten oppilaiden illatiivin käyttöä (Nesser 1985: 60). Jostakin syystä ei omaksuta kovinkaan helposti sitä, että mainittujen nimityyppien pitkä vartalonvokaali edellyttää *sVVn*- päätteen käyttöä (ks. esim. 24–25).

(24) He menivät Pekan *huaneen*. (huoneeseen)

(25) Hotellissa menisin *huoneen* nukkumaan. (huoneeseen)

4.2 Norminvastaisesta illatiivin ja allatiivin käytöstä

4.2.1 Illatiivi allatiivin sijasta

Kuten olen maininnut, ruotsinkieliset – ja muunkin kieliset – suomenoppijat valitsevat jostain syystä joissakin tapauksissa väärän paikallissijamuodon. Käyttämässäni korpuksessa on 62 tapaus, joissa suomenoppija on käyttänyt illatiivia allatiivin sijasta (ks. esim. 26–30).

(26) Matkustaisin lentokoneella *autioihin saariin*. (autioille saarille)

(27) Menimme tunnettuun *hautausmaahan*. (hautausmaalle)

(28) Jokin päivän jälkeen menisin *safariin*. (safarille)

(29) -- viisitoista yli yksi menen *seuraavaan kurssiin*. (seuraavalle kurssille)

(30) Joten kello 8 menin bussilla *juna-asemaan*. (juna-asemalle)

Poikkeavan sijamuodon valinnan taustalla saattaa olla joissakin tapauksessa sanaston vaikutus. Tällä tarkoitan sitä, että suomen kielessä on sanoja, jotka taipuvat ainoastaan joko sisäpaikallissijoissa tai ulkopaikallissijoissa. On kuitenkin koko joukko sellaisia sanoja, jotka voivat taipua sekä sisä- että ulkopaikallissijoissa. Tällöin kuitenkin muotojen välillä on merkitysero, esimerkiksi *Muutan maasta. Muutan maalta*. Paikallissijoihin tutustuessaan suomenoppijani saavat erikseen tietoa joistakin oppimateriaaleissa usein toistuvista ulkopaikallissijoissa taivutettavista sanoista, esimerkiksi *asema, tori, katu, kurssi*.

Luultavaa on, että suomenoppija ainakin opintojensa alkuvaiheessa kytkee paikallissijat kiinteästi koskemaan vain konkreettisia paikkoja ilmaisevia sanoja. Siksipä sellaiset sanat kuin esimerkiksi *kurssi, loma, maa* ja *lounas* joutuvat jonkinlaiselle "harmaalle alueelle", ja suomenoppija on ymmällään sijavalinnan edessä.

4.2.2 Allatiivi illatiivin sijasta

Allatiivi illatiivin asemesta esiintyy korpuksessa 51 kertaa. Määrä ei ole mitenkään silmiinpistävä ajateltaessa eri sijamuotojen yleistä esiintymistiheyttä ja sitä, että ulkopaikallissijojen käyttö on vähäisempää kuin sisäpaikallissijojen.

Allatiivia illatiivin sijasta esiintyy selvää sisätilaa merkitsevien sanojen yhteydessä. Syytä tällaiseen käyttöön on vaikea selittää. Useimmissa tapauksissa ruotsissa käytetään *på*-prepositiota, joka edellyttäisi suomessa inessiivia, vaikka oikea muoto onkin illatiivi (ks. esim. 31–36).

- (31) Sitten menimme *hotellille*. (hotelliin)
- (32) Myöhemmin menimme *yökerholle*. (yökerhoon)
- (33) -- päätin lähteä *uimahallille*. (uimahalliin)
- (34) -- he menivät *elokuvateatterille*. (elokuvateatteriin)
- (35) Hän täytyi juosta *pankille*. (pankkiin)
- (36) Ja muuttaisin toinen *asunnolle*. (asuntoon)

4.2.3 Illatiivi ja allatiivi maantieteellisissä nimissä

Ruotsin osakorpuksessa on mukana sellaisia kertomuksia ja kuvauksia, joissa suomenoppijat kirjoittavat matkustamisesta tai olostaan jossakin maassa tai kaupungissa. Näissä kirjoitelmissä he ovat käyttäneet erityyppisiä paikannimiä. Alkeiskurssin opetuksessa ei käsitellä paikannimiä eikä niiden taivuttamiseen liittyvää säännöstöä, joten oppijat ovat itse ratkaisseet sijavalinnan. Ainoat paikannimet, joista opetuksessa mainitaan erikseen, ovat *Tampere* ja *Venäjä*, jotka taivutetaan ulkopaikallissijoissa. Ei siis ole mitenkään yllättävää, että paikannimissäkkin esiintyy illatiivin ja allatiivin horjuntaa. Seuraavissa esimerkeissä (ks. 37–41) on allatiivin sijasta käytetty illatiivimuotoja.

- (37) Haluaisin matkustaa *Färsaariin* oppimaan --. (Fär-saarille)
- (38) -- matkustaisin laivalla *Venejään*. (Venäjälle)

(39) Perjantai-iltana tulee lähtöä *Lapinjärveen*. (Lapinjärvelle)

(40) Ostan nyt netistä halvat liput *Kreetaan*. (Kreetalle)

(41) Ensi vuonna ehkä menimme *Posioon*. (Posiolle)

Suomenoppijoiden tuotoksissa on mukana joukko maantieteellisiä nimiä, jotka yleiskielen normin mukaan on taivutettava illatiivissa. Suomenoppijat käyttävät kuitenkin hyvin useissa tapauksissa allatiivia (ks. esim. 42–46).

(42) Sitten ajaisin *Kiinalle* isän isän kansa --. (Kiinaan)

(43) Tilaisin heti matkan *New Yorkille*. (New Yorkiin)

(44) Kesällä ajaisin *Walesille* --. (Walesiin)

(45) -- lähtisimme sitten joko *Kanadalle* tai *Etelä-amerikalle* --. (Kanadaan, Etelä-Amerikkaan)

(46) Sitten matkustamme *Geneville*. (Geneveen)

Esimerkkien (37)–(46) sijavalintoja on vaikea tulkita ainakin tässä tutkimusvaiheessa. Taustalla voi olla yksinkertaisesti se seikka, että ruotsissa näissä kaikissa tapauksissa käytetään *till*-prepositiota (*till Färöarna/Kreta*, *till Kina/Geneve*), joten suomenoppija ei saa minkäänlaista apua lähdekielestään.

Yleisestikin ottaen maantieteellisten paikannimien taivutus voi tuottaa pulmia, koska taituvussäännöstö on monimutkainen, myös kielen natiivipuhujalle. Usein säännöistäkään ei saa apua taivutuspulmiin, koska jotkin paikannimet taivutetaan paikallisen perinteen mukaan. Suomenoppijoiden poikkeavan sijamuodon valintaan vaikuttanee myös se, että maantieteellisiä paikkoja ei hahmoteta minkäänlaisiksi sisä- tai ulkopaikoiksi. Tästä johtuen suomenoppija saattaa yksinkertaisesti valita sijan arvaamalla ja ihan sattumanvaraisesti.

5. Päätteeksi

Ruotsinkielisten suomenoppijoiden kirjallisissa tuotoksissa on havaittavissa illatiivin ja allatiivin käytössä tietynlaista horjuvuutta. Suurimmassa osassa näiden kahden paikallissijan tuotoksista muodot ovat oikein rakennettuja ja ne ovat kontekstiin sopivia. Joissakin tapauksissa sijamuodot ovat sekoittuneet; on käytetty illatiivia allatiivin sijasta ja vice versa.

Mahdollisesti tällaisissa tapauksissa on kyse lähdekielen, ruotsin, vaikutus suomeen, kohdekieleen. Ruotsi ja suomi kuuluvat eri kieliperheisiin, ja tästä johtuu esimerkiksi se, että suomenkielinen ja ruotsinkielinen hahmottavat maailmankuvansa eri tavalla. Paikallisuutta ajatellessaan suomenkielinen ajattelee jonkin toiminnan tapahtumasuuntaa, liikettä. Myös paikallisuuden laatu on tärkeä aspekti: tapahtuuko jokin toiminta jonkin paikan sisä- vai ulkopuolella. Ruotsinkieliselle tällainen ajattelutapa on vieras. Tämän osoittaa esimerkiksi se, että samat prepositiot ovat käytössä ero- ja tulosijoissa olipa sitten kyseessä sisä- tai ulkotila.

Toinen mahdollinen tekijä virheiden syntymiseen saattaa olla sanastollinen. Se sanasto, joka vasta-alkajien oppimateriaaleissa esiintyy on rajallinen. Jotkin nomini- ja verbityypit voivat olla ylikorostetusti esillä, toiset taas esiintyvät harvemmin. Tärkeää on myös, että sanoja toistetaan tarpeeksi tiheään, jotta niiden erilaiset taivutetut muodot tulevat tutuiksi ja pikku hiljaa automatisoituvat suomenoppiijan mielessä ja kielessä. Näkisinkin hyvin tärkeänä panostaa oppimateriaaleja tehtäessä niissä käytettyihin sanoihin ja fraaseihin. Oppimateriaalithan ovat usein oppijan ainoita väyliä ja kosketuksia opittavaan kieleen. Oppimateriaalit eivät ole vain kieli-materiaaleja vaan käytetyn sanaston kautta oppija saa myös

muunlaista tietoa, esimerkiksi kulttuuria, tapoja, perinteitä ja arvoja heijastavaa tietoa, opittavan kielen ympäriltä.

Suomea opetettaessa vieraana kielenä ruotsinkielisille opiskelijoille – myös muun kielisille - olisi nähtävästi lisättävä tietoa ja tietoisuutta siitä, miten suomenkielinen ajatusmaailma toimii ja miten suomalainen hahmottaa maailmaa.

Kirjallisuus

Allén, Sture 1980. Nusvensk frekvensordbok. Stockholm: Ahlqvist & Wiksell.

Hakulinen, Lauri 1979. Suomen kielen rakenne ja kehitys. Neljäs korjattu ja lisätty painos. Helsinki: Otava.

Ingo, Rune 2000. Suomen kieli vieraan silmin. Vaasan yliopiston käännösteorian ja ammattikielten tutkijaryhmän julkaisut N:o 26. Vaasa: Vaasan yliopisto. Humanistinen tiedekunta.

ISK 2004 = Hakulinen, Auli, Maria Vilkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen, Irja Alho 2004. Iso suomen kielioppi (Suomalaisen Kirjallisuuden Seuran Toimituksia 950). Helsinki: SKS.

Ivaska, Ilmari, Kirsti Siitonen 2009. Syntaktisesti koodattu oppijan-kielen korpus: mahdollisuuksia ja kysymyksiä. – Korpusuuringute metodoloogia ja märgendamise probleemid / Toim. Pille Eslon, Katre Õim (Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 11). Tallinn: TLÜ kirjastus, 54–71.

Lauranto, Yrjö 1995. Elämän suolaa. Suomen kielen alkeita 1 ja 2. Soveltavan kielentutkimuksen keskuksen julkaisuja n:o 112. Jyväskylä: Jyväskylän yliopisto.

Lauranto, Yrjö 1997. Ensi askeleita paikallissijojen käyttöön. Espanjan-kielisten suomenoppijoiden sisä- ja paikallissijat konseptuaalisen

semantiikan näkökulmasta. Kakkoskieli 2. Helsinki: Helsingin yliopiston suomen kielen laitos.

Laurén, Ulla 1991. Språkfel och interferens hos tvåspråkiga skolelever. En studie i fri skriftlig produktion på svenska. Vaasan yliopiston julkaisuja. Tutkimuksia N:o 156. Filologia 20. Vaasa: Vaasan yliopisto.

Leino, Pentti 1993. Polysemia – kielen moniselitteisyys. Kieli 7. Helsinki: Helsingin yliopiston suomen kielen laitos.

Lehtinen, Tapani 1987. Suomen kielen historiallista muoto-oppia. Helsinki: Helsingin yliopiston suomen kielen laitos.

Lepäsmä, Anna-Liisa, Leena Silfverberg 2009. Suomen kielen alkeisoppikirja. Helsinki: Oy Finn Lectura Ab.

Lähdemäki, Eeva 1995. Mikä meni pieleen? Ruotsinkielisten virheet suomen ainekirjoituksessa. Fennistica 11. Åbo: Åbo Akademi.

Määttä, Tuija 2006. Byggstenar för din finska. Del 1. Kompendium. Umeå: Umeå universitet, Institutionen för moderna språk.

Määttä, Tuija 2007. Byggstenar för din finska. Del 2. Kompendium. Umeå: Umeå universitet, Institutionen för moderna språk.

Nesser, Anne 1986. "Se oli kurja käya koulusa". Bruket av inre och yttre lokalkasus i sverigefinska skolbarns språk. Acta Universitatis Upsaliensis. Studia Uralica et Altaica Upsaliensia 16. Uppsala: Uppsala universitet.

Siitonen, Kirsti 1999. Agenttia etsimässä. *U*-verbijohdokset edistyneen suomenoppijan ongelmana. Suomalaisen ja yleisen kielitieteen laitoksen julkaisuja 63. Turku: Turun yliopisto.

Siivelt, Keaty 2009. Korpuspohjainen tutkimus vironkielisten suomenoppijoiden sisäpaikallissijojen käytöstä. – Korpusuuringute metodologia ja märgendamise probleemid / Toim. Pille Eslon, Katre Öim (Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 11). Tallinn: TLÜ kirjastus, 115–135.

Siivelt, Keaty 2010. Vironkielisten suomenoppijoiden sisäpaikallissijojen frekvenssit ICLFI annotoidun aineiston pohjalta. Suomi toisena ja vieraana kielenä. Oppijankieli ja korpusteknologia opetuksessa. Harjoitustyö. Oulu: Oulun yliopisto. Suomi toisena ja vieraana -kielenä oppiaine.

Teleman, Ulf, Staffan Hellberg, Erik Andersson (red.) 1999. Svenska Akademiens grammatik 2. Ord. Stockholm: Norstedts förlag.

Thorell, Olof 1997. Svensk grammatik. Stockholm: Esselte Studium.

Menen seuraavaan kurssiin – I'll go into the next course

Corpus-based study of the Swedish speaking learners' usage of the illative and allative cases in text production

Tuija Määttä

Summary

This article contains the very first research results concerning the use of the interior local case illative and the exterior local case allative by Swedish speaking students learning Finnish as a foreign language at the beginners' level at university. The main goal of this research has been to construct a general picture of the use of these two local cases.

The electronic material that the research is based on is the Swedish subcorpus of the *International Corpus of Learner Finnish*, ICLFI, which contains 43 496 tokens. The corpus consists of different forms of texts: essays, cards, letters,

diaries, short stories et cetera. The corpus was analysed using a concordance programme that was included in the WordSmith Tools software package. Through the usage of different kinds of illative and allative case suffixes it was possible to find both the correct and wrong case forms.

The analysis shows that there were a total of 1184 illative forms and 521 allative forms. Among the illative forms there are 62 cases in which the correct case would have been the allative case and in 51 cases the students had used allative instead of illative. Why do the students choose a wrong local case?

This article also presents hypotheses that potentially enable a better understanding of why the illative and the allative cases have been used in a divergent way: 1) Are there some structural differences between the Finnish and Swedish languages? 2) Is it possible that people with different mother tongues picture the world differently? 3) Has the language in the teaching material that is mainly the vocabulary (the overuse/underuse) of some words or phrases, some influence on the use of the illative and allative cases?

The fact that prepositions are used to indicate locality in Swedish, and that the illative and allative are expressed by the Swedish word 'till' can cause problems when the students have to choose between the two cases.

One important thing to note when studying Finnish is that the perception of the world is dynamic. This dynamic aspect means that real, taught and conceivable movement is expressed, which means that the local cases have three different forms for existence, separation and arrival.

There are certain aspects that have to be considered when the various local cases are used, e.g. the direction of the movement and the nature of the locality. The nature of the locality is not grammatically important in the Swedish language; there is no distinction made between interior and exterior locality. Even the direction of movement is of minor grammatical importance.

Keywords: learner Finnish, corpus-based research, illative, allative

Esittely

FT Tuija Määttä toimii yliopistonlehtorina ja tutkijana Uumajan yliopiston Kieliopintojen laitoksessa (Umeå universitet, Institutionen för språkstudier). Syksystä 1995 lähtien Tuija Määttä on vastannut suomen kielen oppiaineen vasta-alkajien suomi vieraana kielenä -opetuksesta. Vuonna 2006 käynnistyi tutkimushanke *Korpustutkimus oppijankielen kieli-kohtaisista ja universaaleista ominaisuuksista*, jossa Tuija Määttä on ollut mukana hankkeen perustamisesta lähtien,

tuija.maatta@finska.umu.se

THE COPULA CONSTRUCTION IN FINNISH LEARNER LANGUAGE: ESTONIAN, GERMAN AND DUTCH LEARNERS' USE OF PARTITIVE PREDICATIVES

Marianne Spoelman

Abstract

The partitive case has often been found to be problematic for foreign learners of Finnish. This study, which is part of a larger corpus study on the use of the partitive case in Finnish learner language, investigates the use of partitive predicatives in the written Finnish of Estonian, German and Dutch learners of Finnish as a foreign language. The purpose of the study is to identify and explain the learner's major stumble blocks in the use of partitive predicatives and to address possible cases of L1 influence. Research materials (Estonian Learner Corpus 90.236, German LC 32.300 and Dutch LC 46.670 words) were selected from the *International Corpus of Learner Finnish* (ICFLI). A subset (approx. 2.7 million words) of the *Native Finnish Corpus* was used as a native speaker reference corpus. The outcomes of a combined frequency-error analysis revealed conspicuous differences between the learners from a closely related L1 background (i.e. Estonian), on the one hand and the learners from a distant L1 background (e.g. German and

Dutch), on the other. While the Estonian learners showed instances of both positive L1 influence (i.e. practically the lack of partitive overuse errors) and negative L1 influence (e.g. the transfer of Estonian copula constructions involving plural predicative), the German and Dutch learners seemed to use partitive predicatives in a far more arbitrary way, sometimes even by using the non-inflected basic form as a default. These findings provide valuable and additional insights into the phenomenon of L1 influence and can be used to draw pedagogical implications.

Keywords: Finnish learner language, the copula construction, the use of partitive predicatives

1. Introduction

The partitive is a typical case characterizing the Finnic languages (Kiparsky 1998). By origin, the partitive case was a general locative case with separative meaning (ISK 2004: §1226). However, the partitive gradually developed into a case expressing more abstract syntactical relationships, a development that has gone further in Finnish than in any of the other Finnic languages (Denison 1957). Being in Modern Finnish nowadays first and foremost an object, subject and predicative case expressing partiality, unboundedness or negative polarity, the partitive case covers a wide range of functions related to a myriad of syntactic and semantic conditions.

Researchers and the Finnish language teachers have often observed the use of the partitive case as to be problematic for foreign learners of Finnish (e.g. Denison 1957: 15; Schot-Saikku 1990). It has, however, not systematically been investigated

what it is that makes the Finnish partitive difficult, whether some partitive functions are more difficult than others and whether certain points of difficulty are common to all groups of learners or specific to certain different groups of learners. In addition, studies on the use of the partitive case have, so far, particularly been based on class room observations and more general observations, and not on learner corpora, a fairly recent development. Corpus-based research would however lead to a better understanding of the phenomenon of the learner language as the learner corpora are likely to reflect the actual performance of language learners, while the analyses based on experimental data provide, us in particular with information about learners' abstract knowledge (Nesselhauf 2004). Therefore, this study adopts a corpus-based approach in order to provide valuable insights into Finnish learner language and L1 influence.

The current study investigates the use of partitive predicatives in the written Finnish of Estonian, German and Dutch learners of Finnish as a foreign language. The learner corpus data are selected from the Estonian, German and Dutch subcorpus of the *International Corpus of Learner Finnish* (ICLFI; cf. Jantunen, Piltonen 2009). Partitive predicative constitutes the object of this study for an obvious reason. The study is part of a Ph.D. project on the use of the partitive case. The purpose of this project, which is written within the framework of the research project 'Corpus study on language-specific and universal features in learner language' is to identify and explain the learners' main stumble blocks in the use of the partitive case. Which stumble blocks are common to all groups of learners and which are specific to a certain group of learners: Does L1 influence play a role? How can these insights be used to improve the teaching of Finnish as a foreign language?

In order to identify the learners' main points of difficulty in the use of the partitive, a systematic analysis of the main functions of the partitive (i.e. object, subject and predicatives) is required. This study on partitive predicatives is thus part of a larger ensemble that will come together as time progresses. Further on, the phenomenon of L1 influence will be discussed in more detail and partitive predicatives as well as the Finnish nominative-partitive predicative alternation will be outlined, partly from a crosslinguistic perspective.

2. The influence of the first language

During the past few decades, the influence of the learner's first language (L1) on his or her second language (L2), commonly referred to as *L1 influence*, *transfer* or *crosslinguistic influence* has probably been one of the most extensively investigated phenomena in the SLA field (Jarvis 2000). Studies on L1 influence have nevertheless largely focused on Indo-European languages, particularly English, and not on structurally and typically different languages such as Finnish (Kaivapalu, Martin 2007).

L1 influence is both complex and multifaceted (Dechert, Raupach 1989). The said L1 influence can be characterized across a host of different dimensions (cf. Jarvis, Pavlenko 2010: 20) manifesting itself in various forms (e.g. overuse, errors, facilitation, preference, avoidance) (Odlin 2003) and operating in complex interaction with factors that possibly promote or inhibit the role of L1 influence (i.e. constraints on L1 influence) (Ellis 2008: 379).

One of the most widely recognized constraints on L1 influence is the relationship or degree of congruence between the learner's L1 and L2, commonly referred to as *crosslinguistic similarity* or *language distance*. Crosslinguistic similarity can be both objective and subjective. *Objective similarity* is the actual degree of congruence between languages; while *subjective similarity* corresponds to the degree of congruence the learners perceive or assume to exist (Jarvis, Pavlenko 2010: 176). Objective and subjective similarity can hypothetically fully overlap if the learner accurately perceives or assumes the objective similarities between two languages, but this appears to be relatively rare (Ringbom, Jarvis 2009). Rather, learners frequently make use of an oversimplified equivalence hypothesis, e.g. L2 structure = L1 structure (Ringbom 2007: 55).

Robert Lado (1957) emphasized the importance of actual language distance, the objective degree of similarity between languages, in his Contrastive Analysis Hypothesis (CAH):

"We assume that the student who comes in contact with a foreign language will find some features of it quite easy and others extremely difficult. Those elements that are similar to his native language will be simple for him and those elements that are different will be difficult" (Lado 1957: 2).

By assuming crosslinguistic difference to be equivalent to linguistic difficulty (Odlin 1989: 17), the CAH thus clearly overpredicts the transferability of elements, as it fails to explain when elements are transferred and when they are not, but rather assumes that objective language distance constitutes the ultimate constraint on language transfer (Jarvis, Pavlenko 2010: 1976).

In fact, subjective similarities have a more profound and direct effect on foreign second language learning than objective

similarities (Odlin 1989: 142). As foreign language learning is based on prior knowledge, learners constantly look for whatever L1-L2 similarities they can establish on the basis of their prior linguistic knowledge. At least at the beginning stages of foreign language learning, learners do not often actually perceive crosslinguistic similarities but they merely assume them to exist. Unlike objective similarities, linguistic by nature subjective similarities reflect the learning process (Ringbom 2007).

This is however not to say that objective distance is irrelevant to L1 influence: it is not the objective differences or similarities that cause L1 influence to occur but rather the L1-L2 similarities that learners perceive or assume to exist that serve as the main driving force behind the mental associations leading to instances of L1 influence (Jarvis, Pavlenko 2010). In other words, objective language distance merely provides or withholds the opportunity for the occurrence of L1 influence. Because of the presence versus lack of objective similarities, it is namely more likely that learners establish (accurate) L1-L2 similarities in case of closely related languages than in case of typologically distant languages. The similarity relations established by the learners determine, nevertheless, together with all other constraints, the likelihood of L1 influence actually occurring.

Considering these similarity relations in detail, there is no sharp or absolute boundary between similarity and difference. In the continuum of crosslinguistic similarity relations, three distinct types can be discerned: similarity, contrast and zero relations (Ringbom 2007: 5). A similarity relation means that an L2 pattern is perceived or assumed to be similar to an L1 pattern. In a contrast relation, the learner perceives an L2 pattern as in important ways differing from an L1 pattern,

though there is an underlying similarity between them (Ringbom, Jarvis 2009). A zero relation does not mean a complete lack of crosslinguistic similarity relations, but rather that the L2 patterns, at least at early stages of learning, seem to have little or no relation to the learner's L1 (Ringbom 2007: 5). The way in which objective and subjective similarities relate indicates whether L1 influence is positive or negative in nature. Negative L1 influence occurs when subjective and objective similarities conflict, while positive L1 influence occurs when subjective and objective similarities are compatible (Jarvis, Pavlenko 2010).

When analyzing the role of L1 influence in learner language, multiple comparisons need to be drawn. L1 influence, in particular positive L1 influence, can only be reliably identified on the basis of comparisons between different groups of learners, ranging from groups of learners of closely related L1 backgrounds to groups of learners of distant L1 background (cf. Jarvis 2000; Ellis 2008). Therefore, learner corpus data from Estonian, German and Dutch learners of Finnish were selected as the research materials of this study, as these three L1s vary according to their genetic and typological distance to Finnish. Estonian, like Finnish, belongs to the Finnic branch of the Finno-Ugric language family. Being very closely related, both Finnish and Estonian are languages with a rich structure of nominal and verbal inflections. The Finnish and Estonian declinational systems, consisting of fifteen versus fourteen cases, respectively, both comprise a partitive case. German and Dutch, in contrast, belong to the West-Germanic branch of the Indo-European language family. Although in German, a relatively rich case system has been preserved as compared to other Indo-European languages, the German declinational case system comprises merely four grammatical cases, and no partitive case. Considering the Dutch language, the Middle

Dutch case system has gradually disappeared. These objective differences between Finnish, Estonian, German and Dutch provide an excellent starting point for identifying possible instances of L1 influence. In what follows, the objective differences between Finnish, Estonian, German and Dutch predicatives will be outlined.

3. Partitive predicatives:

The Finnish predicative case alternation

Predicatives are part of the copula construction. In addition to the predicative, Finnish copula constructions prototypically involve a subject noun phrase and the copula verb *olla* ('to be'). A predicative can either be an adjective phrase or a noun phrase functioning as the copula complement of the sentence (ISK 2004: §944).

With the exception of a small group of non-alternating predicatives that indicate category membership or group inclusion (e.g., *Hän on suurta sukua* 'He is of a noble family'), Finnish predicatives alternate between nominative and partitive case. Alternating predicatives are also called *distributive* predicatives. Table 1 provides an overview of this predicative case alternation. As illustrated by the table, the case marking of adjective predicatives depends on the referent (i.e. the subject) and on how the referent is perceived (Vilkuna 1996: 105). The general principles are that the predicative conforms to the subject in number and expresses the divisibility of the referent: Divisible referents take a partitive and indivisible referents a nominative predicative (ISK 2004: §946; cf. Chesterman 1991 for a detailed description of the concept of divisibility). As the divisibility distinction is a semantic concept that separates singular count nouns

(indivisible) from plural count nouns and mass nouns (divisible) (Chesterman 1991: 133), the Finnish predicative case alternation is thus semantic in nature.

	Referent characteristics	Predicative characteristics	Case of the predicative	Example sentences
Sg	indivisible	adjective phrase	nominative	(1a) <i>Nainen on <u>kaunis</u>.</i> Woman(Nom.Sg) be-3Sg pretty(Nom.Sg) ‘The woman is pretty.’
	Divisible	adjective phrase	partitive	(1b) <i>Kahvi on <u>hyvää</u>.</i> Coffee(Nom) be-3Sg good-Part.Sg ‘Coffee is nice.’
Pl	divisible	adjective phrase	partitive	(1c) <i>Naiset ovat <u>kauniita</u>.</i> Woman-Nom.Pl be-3Pl pretty-Part.Pl ‘The women are pretty.’
	Indivisible	adjective phrase	nominative	(1d) <i>Nämä sakset ovat <u>terävät</u>.</i> These scissors-Nom.Pl be-3Pl sharp-Nom.Pl ‘This pair of scissors is sharp.’

Sg	indivisible	NP / indivisible	nominative	(2a) <i>Tämä on <u>uusi auto</u>.</i> this(Nom) be-3Sg new car(Nom.Sg) 'This is a new car.'
	indivisible	NP / divisible	partitive	(2b) <i>Tämä on <u>sitä hyvää kahvia</u>.</i> this(Nom) be-3Sg that good coffee-Part.Sg 'This is that nice coffee.'
	indivisible	NP / indivisible	nominative	(2c) <i>Hän on maailman <u>paras laulaja</u>.</i> he(Nom) be-3Sg world-Gen.Sg best singer(Nom.Sg) 'He is the best singer in the world.'
	divisible	NP / indivisible	nominative	(2d) <i>Kahvi on maailman <u>paras asia</u>.</i> coffee(Nom.Sg) be-3Sg world- Gen.Sg best thing(Nom.Sg) 'Coffee is the best thing in the world.'

Table 1. Overview of the Finnish predicative case alternation

More specifically, the adjective predicative assumes nominative case if the referent is (conceived of as) a count noun, thus denoting an indivisible entity (1a). Similarly, the adjective predicative takes partitive case if the referent is either a mass noun (1b) or a plural form (1c), as a mass noun designates a divisible mass or substance and a plural form a divisible

multiplicity of referents (Huumo 2007; 2009). Because they refer to an indivisible referent, marginally occurring *pluralia tantum* (1d) and plural forms that otherwise holistically denote a set of entities (e.g. *tytön silmät ~ hiukset* ‘the girl’s eyes ~ hair’) form an exception to the rule of plural predicative case-marking (Vilkuna 1996: 106).

The case of plural substantive predicatives alternates in exactly the same way as the case of plural predicative adjectives alternates (Huumo 2007). However, as illustrated in table 1, it is not only the nature of the referent but also the nature of the substantive predicative itself that possibly affects the case assignment of singular substantive predicatives (ISK 2004: §944). Regardless of the divisible or indivisible nature of the referent (i.e. the subject), a singular predicative substantive assumes partitive when the predicative itself is divisible in nature (2b) and nominative when the predicative itself is indivisible in nature (2a, 2c, 2d) (Huumo 2007). The reason for the fact that substantive predicatives, do but adjective predicatives do not influence predicative case-marking is simple: divisibility is a semantic feature that nouns but not adjectives possess.

A similar predicative case alternation is not found in other Finnic languages, as the Finnish predicative case alternation has been the result of a recent development during which the partitive predicative gradually took over the function of the nominative predicative in case of divisible referents (Sadeniemi 1950). The occurrence of partitive predicatives is, as a consequence, very limited in Estonian (Erelt 2009). As Estonian lacks a predicative case alternation similar to Finnish (Metslang 1994: 210), the Estonian distributive predicative always takes nominative case (Denison 1957: 247; cf. Ex1a-b). The occurrence of partitive predicatives restricts itself to the

category of non-alternating predicatives expressing category membership or group inclusion. Like in Finnish, these partitive predicatives nevertheless have a very limited occurrence (Erelt 2003: 97-98).

The crosslinguistic similarity relation between Finnish and Estonian predicatives can be considered a good example of the *contrast relation* introduced above. The copula construction is essentially similar in both languages, the only (but substantial) difference being that distributive predicatives show a nominative-partitive case alternation in Finnish while they assume nominative case in Estonian (cf. Ex1). Taking into consideration that Finnish and Estonian are very closely related and that the object and subject case alternations are similar in both languages, it could be expected that Estonian learners of Finnish are sometimes inclined to establish an oversimplified equivalence hypothesis (cf. Ringbom, 2007: 55), i.e. Finnish partitive predicative = Estonian partitive predicative. As for German and Dutch, those languages are typologically and genetically different from Finnish. Dutch is a morphologically poor language and the German declinational system does not contain a partitive case. As illustrated in Ex2, the predicative in both languages therefore always appears in the non-inflected basic form (which is identical to the nominative singular in German). This study aims at providing insights into learner's use of partitive predicatives and attempts to identify possible differences between groups of learners in order to anticipate pedagogical implications as to how partitive predicatives could be best taught to learners of Finnish from closely related L1 language backgrounds and distant L1 language backgrounds.

(Ex1)	Finnish	Estonian
a.	<i>Kahvi on <u>hyvää</u>.</i> coffee(Nom.Sg) be-3Sg good-Part. Sg 'Coffee is nice.'	<i>Kohv on <u>hea</u>.</i> coffee (Nom.Sg) be-3Sg good(Nom.Sg) 'Coffee is nice.'
b.	<i>He ovat <u>nuoria</u>.</i> they(Nom) be-3Pl young-Part.Pl 'They are young.'	<i>Nad on <u>noored</u>.</i> they(Nom) be-3Pl young-Nom.Pl 'They are young.'
(Ex2)	German	Dutch
a.	<i>Kaffee ist <u>lecker</u>.</i> coffee(Nom.Sg) be-3Sg good(Nom.Sg) 'Coffee is nice.'	<i>Koffie is <u>lekker</u>.</i> coffee be-3Sg good 'Coffee is nice.'
b.	<i>Sie sind <u>jong</u>.</i> they(Nom) be-3Pl young(Nom.Sg) 'They are young.'	<i>Zij zijn <u>jong</u>.</i> they(Nom) be-3Pl young 'They are young.'

4. Method

A subset of the International Corpus of Learner Finnish (ICLFI) was selected as the learner data of the study. The ICLFI (cf. Jantunen, Piltonen 2009) was launched in 2007 and it comprises texts of different genres, mainly essays. Presently, the corpus contains over 500.000 words of learners of Finnish as a foreign language from 13 different language backgrounds. The corpus is made up of 13 different subcorpora, each

covering one of those language varieties. Subsets of the subcorpora of Estonian, German and Dutch learners of Finnish were selected as the materials of this study. The materials were selected according to specific criteria: only texts written by native speakers of Estonian, German and Dutch whose parents were also both native speakers of respectively Estonian, German and Dutch were selected.

A native-speaker reference corpus of over 2.7 million words was selected from the Native Finnish Corpus, compiled by Anna Mauranen as the native component of the Corpus of Translated Finnish (Mauranen 2000). The native Finnish reference corpus consists of professional writing and involves several genres, i.e. academic texts, popular non-fiction, children’s literature, fiction and biographies. Figure 1 provides an overview of the research materials.

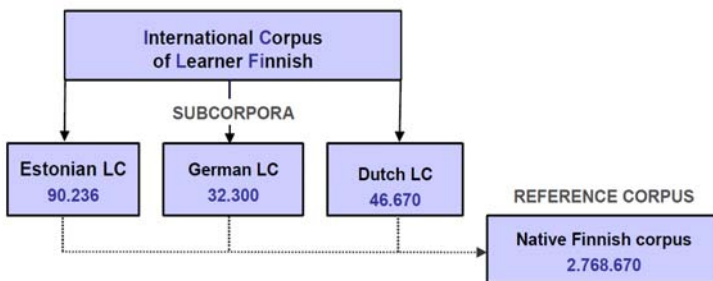


Figure 1. The materials of the study (corpus sizes in words)

Because the selected corpus data were not linguistically annotated, it was decided to run a problem-oriented annotation and tagging procedure (cf. McEnery, Wilson 2001: 69). On the basis of this procedure, the texts were annotated and error-tagged on the use of partitive noun phrases with the help of a

set of macros designed to automatically identify partitive forms and to simplify morphosyntactic annotation and error-tagging. The learner corpora were error-tagged for the occurrence of both partitive overuse and underuse errors. Partitive overuse errors are deviant usages in which the partitive has been incorrectly used as the case of the predicative. Partitive underuse errors, in contrast, correspond to those contexts in which a partitive predicative was required but another case was realized as the case of the predicative.

The absolute frequencies of partitive predicatives, partitive predicative overuse errors and partitive predicatives underuse errors were extracted from the corpora with the help of WordSmith Tools 5.0 (Scott 2008). Obligatory contexts for partitive predicatives were calculated by adding the amount of partitive underuse errors to the number of correctly used partitive predicatives. These obligatory contexts thus basically reflect the number of contexts in which a partitive predicative was required, regardless of whether or not it was indeed realized, and the frequencies of partitive predicatives indicate the number of contexts in which a partitive predicative was produced, regardless of whether they were produced correctly or incorrectly.

Statistical comparisons were drawn between the frequencies of partitive predicatives produced and between obligatory contexts for partitive predicatives observed in different learner corpora and those observed in native Finnish reference corpus. Furthermore, comparisons were drawn between the frequencies of partitive overuse and underuse errors extracted from the Estonian, German and Dutch learner corpora. For all comparisons, the α decision level was set at 0.05.

5. Results

An overview of the frequency analysis on partitive predicatives is provided in figure 2. Of the categories represented in figure 2, the category of partitive predicatives produced purely reflects the outcomes of the frequency analysis, as this category refers to the actual occurrence of partitive predicatives in the corpora, i.e. regardless of whether or not correctly used. Considering this category of partitive predicatives produced, statistical testing (Log-Likelihood chi-square) revealed that partitive predicatives were significantly more frequently used in the Dutch learner corpus than in the reference corpus (LL = 12.08; $p < .0001$) and significantly less frequently in the Estonian learner corpus than in the reference corpus (LL = 152.99; $p < .0001$). No significant differences were found between the frequencies of partitive predicatives observed in the German learner corpus and the native-speaker reference corpus (LL = 3.80).

Gradually proceeding from the mere frequency analysis toward an analysis taking into consideration partitive predicative errors, the relative frequencies of correctly used partitive predicatives and obligatory contexts are provided as well. Obligatory contexts for partitive predicatives occurred significantly more frequently in the Dutch learner corpus (LL = 45.78; $p < .0001$) and the German learner corpus (LL = 68.17; $p < .0001$) than in the native Finnish reference corpus. In addition, obligatory contexts for partitive predicatives were found to occur less frequently in the Estonian learner corpus than in the reference corpus (LL = 4.63; $p < .05$).

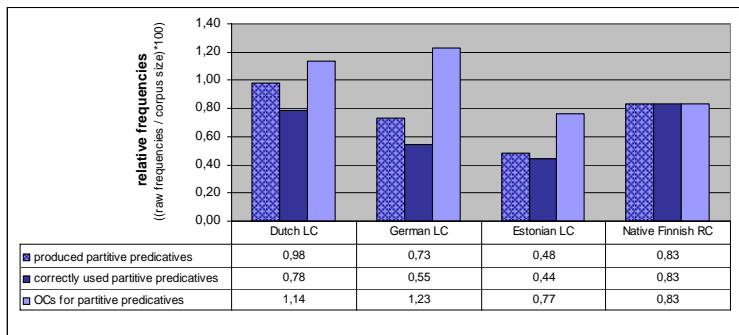


Figure 2. The use of partitive predicatives

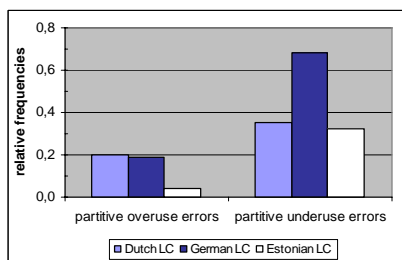


Figure 3. Partitive predicative errors

The frequencies of partitive predicative overuse and underuse errors, which are also covertly present in figure 2, are graphically represented in figure 3. From this graph, it can be inferred that partitive underuse errors were

more frequently represented in all learner corpora than partitive overuse errors.

Moreover, the Estonian learner corpus particularly shows a small number of partitive overuse errors, while a relatively large number of underuse errors was observed in the German learner corpus.

Statistical testing (Log-Likelihood chi-square) showed that partitive predicative overuse errors occurred significantly less frequently in the Estonian learner corpus than in the Dutch learner corpus (LL = 77.22; $p < .0001$) and the German learner

corpus (LL = 53.68; $p < .0001$). No significant differences were found between partitive overuse errors observed in the Dutch and the German learner corpus (LL = 0.24). Considering partitive underuse errors, the Dutch and Estonian learner corpora did not appear to differ significantly. However, partitive underuse errors occurred significantly more frequently in the German learner corpus compared to both the Dutch learner corpus (41.76; $p < .0001$) and the Estonian learner corpus (LL = 65.69; $p < .0001$).

6. Discussion and conclusions

The outcomes of the frequency analysis suggested a significant overuse of partitive predicatives in the Dutch learner corpus and a significant underuse in the Estonian learner corpus, as compared to the native-speaker reference corpus. However, obligatory contexts for partitive predicatives were found to occur significantly more frequently in both the German and Dutch learner corpora than in the reference corpus indicating a significant overrepresentation of obligatory contexts in both learner corpora. In contrast, the obligatory contexts for partitive predicatives observed in the Estonian learner corpus slightly but significantly differed from the obligatory contexts observed in the native reference corpus in the way that obligatory contexts for partitive predicatives were under-represented in the learner data.

Considering the case-marking of predicatives, the error analysis first and foremost indicated that partitive underuse errors occurred more frequently than partitive overuse errors. The German learner corpus was, however, particularly characterized by a large number of underuse errors. Partitive overuse errors were found to occur significantly less frequently

in the Estonian learner corpus than in the other learner corpora. In fact, the Estonian learners nearly failed to show any instances of partitive overuse errors, while the German and Dutch learners showed comparable error frequencies.

These outcomes and in particular the error patterns and the way in which the actually produced partitive predicatives relate to the obligatory contexts for partitive predicatives provide valuable insights into the learners' use of partitive predicatives. Unlike the German and Dutch learners of Finnish, the Estonian learners showed a significant underuse of partitive predicatives. Based on the frequency analysis only, it could be expected that the underuse was a consequence of avoidance of partitive predicatives. The error analysis, however, revealed the underuse to be due to a large number of partitive underuse errors. Connecting these findings to the fact that the Estonian learner corpus showed an almost complete lack of partitive overuse errors, this could possibly indicate that the Estonian learners of Finnish use partitive predicatives in a less arbitrary way than the other groups of learners do, as a result of the presence of similarity relations between Finnish and Estonian versus the relative lack of L1-L2 similarities for the German and Dutch learners of Finnish.

In order to further explore the above outlined hypothesis, a detailed analysis was conducted concerning partitive predicative underuse errors. The outcomes of this detailed analysis are provided in table 2.

Partitive underuse error category	Dutch LC	German LC	Estonian LC
Nominative singular instead of partitive singular	72.1 %	47.0 %.	52.0 %
Nominative plural instead of partitive plural	9.7 %	24.9 %.	42.9 %
Nominative singular instead of partitive plural	16.4 %	24.0 %.	4.8 %
Remaining underuse errors	1.8 %	4.1 %.	0.3 %

Table 2. Detailed analysis of partitive predictive underuse errors

With respect to all groups of learners, most partitive underuse errors appeared to involve the use of nominative singular instead of partitive singular. As for the use of nominative plural predicatives instead of partitive plural predicatives, this error category has relatively frequently been observed from the Estonian learner corpus (in 42.9% of all underuse errors) and substantially less frequently from the other learner corpora. In contrast, the Dutch and German learner corpora show relatively high percentages of errors in which the nominative singular (i.e. the basic or dictionary form) has been used as the case of the predicative, while a partitive plural predicative was required. Similar errors represent a marginal error category in the Estonian learner corpus. These outcomes suggest conspicuous differences between the Estonian learners on the one hand and the German and Dutch learners on the other hand and simultaneously provide additional evidence supporting the hypothesis that L1 influence plays a role in the Estonian learners' use of partitive predicatives.

The finding that underuse errors in which the nominative plural was used instead of the partitive plural were observed substantially more frequently in the Estonian learner corpus than in the other learner corpora could be explained by taking into consideration the contrast relation between Finnish and Estonian predicative case assignment. As illustrated in (Ex3), a morphosyntactically similar sentence would be grammatically correct in Estonian, indicating that at least part of the errors would be due to an incorrectly assumed L1-L2 similarity, i.e. transfer of the Estonian morphosyntactic construction.

(Ex3)

<i>Produced structure</i>	<i>Target-like structure</i>	<i>Equivalent structure in Estonian</i>
He ovat *<u>opettajat</u>. they(Nom) be-3Pl teacher-*Nom.Pl 'They are teachers.'	He ovat <u>opettajia</u>. they(Nom) be-3Pl teacher-Part.Pl 'They are teachers.'	Nad on <u>õpetajad</u>. they(Nom) be-3Pl teacher-Nom.Pl 'They are teachers.'

Moreover, underuse errors in which the nominative singular basic form was used instead of a partitive plural predicative were relatively frequently observed in the German and Dutch learner corpora but were nearly absent in the Estonian learner corpus. This difference between the German and Dutch learners on the one hand and the Estonian learners on the other, could probably partially be explained by taking into account the relative absence of similarity relations between German/Dutch and Finnish. Unlike the German and Dutch learners, the Estonian learners are familiar with the phenomenon of predicative case-marking and extensive nominal (and also verbal) inflection. Therefore, it could be the case that German and Dutch learners more often than Estonian learners of

Finnish choose the (non-inflected) basic predicative form in order to, for example, decrease their processing load, while the Estonian learners in similar cases sometimes seem to rely on their L1. It could also be argued that the German and Dutch learners are inclined to use the non-inflected predicative form because predicatives are not inflected in their L1 either (cf Ex2, pg. 5). However, the former explanation (i.e. about the lack of processing load) is far more likely when taking into account the distant relationship between Finnish and respectively German and Dutch.

From the detailed analysis of partitive underuse errors, it was also revealed that the use of the nominative singular as the case of the predicative instead of partitive singular was highly frequent in all learner corpora. This outcome might be related to the non-transparency of the rules of predicative case-marking concerning abstract referents, zero-referents and dependent clauses or infinitive constructions functioning as the referent of the predicative. This will, however, be dealt with in further investigations, in which also general implications for teaching will be tackled. Some ideas of pedagogical value would be that teaching of partitive predicatives to Estonian learners of Finnish should emphasize L1-L2 differences, while learners from distant L1 backgrounds like German and Dutch learners of Finnish would benefit from focusing on 'the big picture', i.e. the general characteristics of Finnish copula constructions as well as the notion of divisibility.

To conclude, conspicuous differences were found between the use of partitive predicatives by learners of Finnish from distant mother tongue backgrounds (i.e. German and Dutch) and learners from a very closely related mother tongue background (i.e. Estonian). As a possible consequence of a lack of L1-L2 similarities, the German and Dutch learners seemed

to use partitive predicatives in a more arbitrary way than the Estonian learners and they were sometimes even inclined to rely on the non-inflected basic form as a default predicative form. In contrast, the contrast relation between Finnish and Estonian predicative case-marking seemed to play an important role in Estonian learners' use of partitive predicatives, resulting in instances of both positive L1 influence (e.g. the almost complete lack of partitive overuse errors) and negative L1 influence (e.g. the transfer of Estonian structures involving plural predicatives). The occurrence of both positive and negative L1 influence is in line with Jarvis & Pavlenko (2010: 176-183), who claim that positive L1 influence theoretically occurs when assumed similarities are compatible with objective similarities and negative L1 influence when these two are incompatible, but that L1 influence often ends up both positive and negative at the same time given that L1-L2 correspondences are rarely or exactly the same or completely different.

References

- Chesterman, Andrew 1991. *On definiteness: A study with special reference to English and Finnish*. Cambridge: Cambridge University Press.
- Dechert, Hans W., Manfred Raupach 1989. Introduction. – *Transfer in language production* / Ed. By Hans W. Dechert, Manfred Raupach. Norwood NJ: Ablex Publishing Corporation, ix–xvii.
- Denison, Norman 1957. *The partitive in Finnish* (*Annales Academiae Scientiarum Fennicae Ser. B, Vol. 108*). Helsinki: Suomalainen Tiedeakatemia.
- Ellis, Rod 2008. *The study of Second Language Acquisition*. 2nd edition. Oxford: Oxford University Press.

- Erelt, Mati 2003. Structure of the Estonian language: Syntax. – Estonian Language / Ed. by Mati Erelt (Linguistica Uralica Supplementary Series 1). Tallinn: Estonian Academy of Science, 93–129.
- Erelt, Mati 2009. Typological overview of Estonian syntax. – Language Typology and Universals 62: 1/2, 6–28.
- Huumo, Tuomas 2007. Kvantiteetti ja aika II: nominaalinen aspekti ja suomen predikatiivin sijnvaihtelu. – Virittäjä 111, 3–23.
- Huumo, Tuomas 2009. Fictive dynamicity, nominal aspect, and the Finnish copulative construction. – Cognitive Linguistics 20, 43–70.
- ISK 2004 = Iso Suomen Kieloppi 2004 / Ed. by Auli Hakulinen jt. Helsinki: SKS.
- Jantunen, Jarmo H., Saana Piltonen 2009. Oppijansuomen ja -viron sähköiset tutkimusaineistot. – Virittäjä 113: 3, 449–458.
- Jarvis, Scott 2000. Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. – Language Learning 50, 245–309.
- Jarvis, Scott, Anna Pavlenko 2010. Crosslinguistic influence in language and cognition. New York/London: Routledge.
- Kaivapalu, Annekatrin, Maisa Martin 2007. Morphology in transition: Plural inflection of Finnish nouns by Estonian and Russian learners. – Acta Linguistica Hungarica 54, 129–156.
- Kiparsky, Paul 1998. Partitive Case and Aspect. – The Projection of Arguments: Lexical and Compositional Factors / Ed. by Miriam Butt, Wilhelm Geuder. Stanford: CSLI, 265–308.
- Lado, Robert 1957. Linguistics across cultures. Ann Arbor: Michigan University Press.
- Mauranen, Anna 2000. Strange Strings in Translated Language: A Study on Corpora. – Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects / Ed. by Maeve Olohan. Manchester: St. Jerome Publishing, 119–141.
- McEnery, Tony, Andrew Wilson 2001. Corpus Linguistics. 2nd edition. Edinburgh: Edinburgh University Press.
- Metslang, Helle 1994. Kielet ja kontrastit. – Virittäjä 98, 203–226.
- Nesselhauf, Nadja 2004. Learner corpora and their potential for language teaching. – How to use corpora in language teaching / Ed.

by John McH. Sinclair. Amsterdam/Philadelphia: John Benjamins Publ. Co.

Odlin, Terrence 1989. *Language Transfer. Cross-linguistic influence in language learning*. Cambridge, USA: Cambridge University Press.

Odlin, Terrence 2003. *Crosslinguistic influence*. – *The handbook of second language acquisition* / Ed. by Catherine J. Doughty, Michael H. Long. Oxford: Blackwell Publishing Ltd, 436–486.

Ringbom, Håkan 2007. *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters LTD.

Ringbom, Håkan, Scott Jarvis 2009. *The importance of cross-linguistic similarity in foreign language learning*. – *Handbook of language teaching* / Ed. by Catherine J. Doughty, Michael H. Long. Malden, MA: Wiley-Blackwell, 106–118.

Sadeniemi, Matti 1950. *Totaalisesta ja partiaalisesta predikatiivista*. – *Virittäjä* 54, 46–53.

Scott, Mike 2008. *WordSmith Tools version 5. Lexical Analysis Software*. Liverpool.

Schot-Saikku, Päivi 1990. *Der Partitiv und die Kasualternation: Zum Fall Partitiv in der Finnischen Syntax*. Hamburg: Buske.

Vilkuna, Maria 1996. *Suomen lauseopin perusteet*. Helsinki: EDITA.

Koopulakonstruktsioonid soome õppijakeeles: partitiivpredikatiivide kasutus eesti, saksa ja hollandi lähtekeelega õppijatel

Marianne Spoelman

Resümee

Partitiivi on peetud soome keele õppijate jaoks problemaatiliseks käändeks. Käesolev uurimus, mis on osa laiaulatuslikumast soome õppijakeele partitiiviuringust, käsitleb parti-

tiivpredikatiivi kasutust eesti, soome ja hollandi lähtekeelega õppijate kirjalikes tekstides. Eesmärk on kindlaks teha ja selektada õppijate peamisi komistuskive partitiivi kasutamisel ning analüüsida lähtekeele võimalikku mõju. Uurimisaines on valitud Oulu ülikooli soome õppijakeele korpusest: eesti lähtekeelega õppijate allkorpus sisaldab 90 236, saksa lähtekeelega õppijate allkorpus 32 300 ja hollandi lähtekeelega õppijate allkorpus 46 670 sõnet. Kontrollkorpuseks on kasutatud soome kirjakeele korpuse põhjal koostatud alamkorpust, milles on umbes 2,7 miljonit sõnet. Mitmedimensionaalse veaanalüüsi tulemusel selgusid silmnähtavad erinevused ühelt poolt eesti lähtekeelega ning teiselt saksa ja hollandi lähtekeelega õppijate partitiivpredikatiivide kasutuses. Kui tegu on lähisugulaskeeltega, näitasid analüüsi tulemused nii lähtekeele positiivset (nt partitiivi ülekasutuse puudumine) kui ka negatiivset mõju (nt eesti keele partitiivpredikaate sisaldavate koopulakonstruktsioonide ülekanne). Kui tegu on mittesugulaskeeltega, oli partitiivpredikatiivide kasutus ebasüsteemiline, mõnikord algvormi eelistav. Saadud uurimistulemused täiendavad olemasolevaid, lisades lähtekeele mõju olemuse kohta väärtuslikku teavet.

Autor

MA Marianne Spoelman, University of Oulu, Finnish as a Second and Foreign Language.

marianne.spoelman@oulu.fi

TALLINNA ÜLIKOOI EESTI KEELE JA KULTUURI INSTITUUDI TOIMETISED 2007–2010

Tallinna Ülikooli keelekorpuste optimaalsus, töötlemine ja kasutamine. Toim. Pille Eslon. (Tallinna Ülikooli eesti filoloogia osakonna toimetised 9.) Tallinn: TLÜ Kirjastus, 2007.

Õppijakeele analüüs: võimalused, probleemid, vajadused. Toim. Pille Eslon. (Tallinna Ülikooli eesti filoloogia osakonna toimetised 10.) Tallinn: TLÜ Kirjastus, 2008.

Korpusuuringute metodoloogia ja märgendamise probleemid. Toim. Pille Eslon, Katre Õim. (Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 11.) Tallinn: TLÜ Kirjastus, 2009.

Korpusuuring ja meetodid. Toim. Pille Eslon, Katre Õim. (Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 12.) Tallinn: TLÜ Kirjastus, 2010.

