

EESTI VAHEKEELE KORPUSE TÖÖRÜHMA IV SÜGISSEMINAR TEESID

On the meaning of 'disability' - related words in newspaper texts and in the attitudes of language users

Heidi Martimo
University of Oulu
heidimartimo@gmail.com

Background and significance

The latest effort to affect attitudes towards disability and persons with disabilities culminated in the United Nations Convention on the Rights of Persons with Disabilities adopted by the General Assembly in 2006. Besides prevention of discrimination, a general attitudinal change is an essential precondition for the implementation of the goals of the convention. To accomplish such a change, the member states have committed themselves to reducing prejudice and stereotypical thinking and to promoting knowledge about disability and how people with disabilities contribute to societal diversity.

According to the Finnish Government Report on Disability Policy (2006) the attitudinal climate for people with disabilities has become more positive in Finland, but they are still seen as 'objects' and as persons who predominantly prefer insider contacts in peer groups. Stereotypical and homogenic conceptions of persons with disabilities prevail and they are not always treated as individuals in their own right.

In previous linguistic research, the subject of disability was touched on in Paso's (2007) study on work-capacity discourse, 'Työkyky merkitsee', because the topics of disability, and work and functional capacity are inherently intertwined. In addition, Tiirilä's (2007) text linguistic study 'Tekstit viraston työssä' focuses on transport service documents required by the Disability Service Act. Moreover, Reinikainen (2007) investigated meanings assigned to disability and disabled life in various texts while Martin (2003) studied the concept of disability in textbooks using, among other methods, critical discourse analysis.

Aim

In this study, I aim to explore disability-related interpretations, expectations, values and attitudes as they are manifested in newspaper texts and in the responses of the study subjects to the inquiry to be conducted. The lexemes denoting disability seem to carry images and meanings suggesting helplessness, pitifulness and inadequacy. I observe the linguistic essence of disability as it is reflected in published newspaper and internet texts and in the attitudes of the study subjects, thus seeking to throw light to the connection between attitudes and language.

I focus on the following research questions:

- Which words, concepts and collocations are used when disability is discussed?
- Does the language used reflect attitudes towards disability and persons with disabilities and, if so, how are these attitudes manifested in language? Does the language used affect attitudes?
- What are the prevalent attitudes towards disability? What are these attitudes based on?

Materials and Methods

My study materials consist of the 1990s newspaper file of the Finnish Language Text Bank of CSC-IT Centre for Science Ltd, the articles on disability and persons with disabilities published in the newspaper Kaleva during the first decade of the 21st century, and the responses obtained to an attitudinal inquiry. To study the Finnish corpora of the Text Bank, I use the WWW-Lemmie 2.0 corpus tool with which concordances and word clouds can be collected and frequency tables made. I will then conduct an inquiry on the attitudes and conceptions relating to disability and persons with disabilities using the Webropol survey tool. Webropol is an Internet-based solution, by means of which inquiries can be implemented from the drafting of the questionnaire to the reporting of the results. My purpose is to find out whether the results of the attitudinal survey support or vary from the conclusions suggested by the results of the newspaper text corpus analysis.

I investigate the semantic prosody and semantic preference of disability-related words. Semantic prosody is conventionally described by reference to positive or negative meaning. Semantic preference is defined as the relation between the meaning of a lexeme and that of a collocate frequently appearing in the vicinity of the lexeme, i.e., as a lexical-semantic co-occurrence. I explore contextual meanings by means of the methods of corpus linguistics. According to the principles of corpus linguistics, meanings arise from usage but they are invisible. (Sinclair 1991, Jantunen 2004) Accordingly, meanings can only be studied by studying words within large quantities of texts. Certain collocates appear with certain lexemes repeatedly. They are chosen for and function within a particular context and their meaning is determined by the context.

References

- Finnish Government Report on Disability Policy 2006
- Jantunen, Jarmo Harri 2004: Synonymia ja käännessuomi. Korpusnäkökulma samamerkityksisyyden kontekstuaalisuuteen ja käännesskielen leksikaalisiin erityispiirteisiin. Joensuun yliopiston humanistisia julkaisuja n:o 35. Joensuu.
- Martin, Deb 2003: Dis-ability in composition textbooks: A rhetoric of difference. Thesis PhD. Texas Woman's University.
- Paso, Eija 2007: Työkyky merkitsee. Työkyvyn merkityksen ja maailman rakentuminen lehtiteksteissä. Acta universitatis Ouluensis B Humaniora 77. Oulu.
- Reinikainen, Marjo-Riitta 2007: Vammaisuuden sukupuolittuneet ja sortavat diskurssit. Yhteiskunnallis-diskursiivinen näkökulma vammaisuuteen. Jyväskylä studies in education, psychology and social research 304. Jyväskylä.

Sinclair 1991: Corpus, Concordance, Collocation. Oxford University Press.
Tiililä, Ulla 2007: Tekstit viraston työssä. Tutkimus etuuspäätösten kielestä ja konteksteista. Helsingin yliopisto, Helsinki.
United Nations, General Assembly 2006: Convention on the Rights of Persons with Disabilities, including a draft Optional Protocol. A/AC.256/2006/4 and Add.1.

***Ihminen* oppijansuomessa. Korpusanalyysi frekventistä fraseologisesta yksiköstä**

Jarmo Harri Jantunen
Oulun yliopisto
jarmo.jantunen@oulu.fi

Esityksessäni tarkastelen oppijankielen fraseologiaa yksiköitä esimerkkinäni frekventti suomen kielen *ihminen*-nomini. Tavoitteenani on selvittää, millä tavalla oppijankielen fraseologiset yksiköt voivat erota natiivikielen yksiköistä. Huomion kohteena ovat *ihminen*-nominin taivutus (paradigmaattinen morfologinen preferenssi, Jantunen 2004), nominin kollokointi lähimpien kollokaattien kanssa (2L–2R) ja ns. semanttinen preferenssi, eli sanan yhteisesiintyminen tiettyjen kontekstissa ilmenevien merkityspiirteiden kanssa. Tarkastelen myös taitotason ja oppijan äidinkielen vaikutusta edellä mainittuihin seikkoihin. Lähestymistapa aineistoon on korpusvetoinen.

Aineistonani on *Kansainvälinen oppijansuomen korpus* (ICLFI), jonka laajuus on noin 500 000 sanetta. Aineisto on koottu Oulun yliopistossa usean yliopiston yhteistyöhankkeena. Se sisältää ulkomailla opiskelevien suomenoppijoiden kirjoittamia fiktiivisiä tekstejä ja asiatekstejä. Verrannollisena korpuksena käytän *Natiivisuomen korpusta*, joka on luotu Joensuun yliopistossa kootusta *Käännössuomen korpuksesta* (Mauranen 2000).

Tämä oppijansuomen *ihmistä* koskeva analyysi on rinnakkaistutkimus EVKK-korpushankkeen johtajan Pille Eslonin oppijanvieron *inimene*-nominin analyysille.

Genitiivi kasutusest eesti õppijakeeles sõna *inimene* näitel

Pille Eslon
Tallinna Ülikool
peslon@tlu.ee

1. Materjali allikas – Eesti vahekeele korpus (EVKK)¹, Tartu Ülikooli eesti kirjakeele korpuse tasakaalustatud ilu- ja teadustekstid (TYKK)², EKI tekstikorpuse praegusaja *on-line* ajakirjandustekstid (EKI korpus)³.

¹ <http://evkk.tlu.ee> (2.10.2009).

² <http://www.cl.ut.ee/korpused/grammatikakorpus/failid/> (2.10.2009).

³ <http://www.eki.ee/corpus/> (3.09.2008).

2. Analüüsivahend – statistikapõhine programm WST 5,0⁴: annab erinevate valimite sõnavormide sageduse, uuritava sõna vormide kasutusnäited e konkordantsid, aitab tuua esile sõnavormide leksikaalsed ja leksikaalgrammatilised kasutusmustrid.
3. Analüüsiüksused: sõna *inimene* ainsuse ja mitmuse genitiivi sisaldavad leksikaalsed kooslused ehk kollokatsioonid ja kollokaadid ühelt poolt ning sagedased leksikaalgrammatilised konstruktsioonid teisalt⁵.
4. Millest huvi genitiivi vastu? Analüüsinud WST 5,0 abil kolme korpusainest – peamiselt vene emakeelega inimeste eesti õppijakeelt (EVKK), üle kolmekümne-neljakümne aasta taguseid ilu- ja teadustekste (Tartu Ülikooli eesti kirjakeele korpuse tasakaalustatud osa) ning EKI tekstikorpuse *on-line* ajakirjandustekste (EKI korpus) – selgusid samalaadsused ja erijooned, mis iseloomustavad sõna *inimene* käändevormide paradigmat eelmise sajandi teisest poolest tänaseni. Grammatiliste käänete osas on õppija- ja kolmekümne-neljakümne aasta taguse eesti kirjakeele vahel selge erinevus genitiivi ja partitiivi vormide kasutuses. Kirjakeeles on sagedasemad nominatiivi ja genitiivi vormid, samas kui oluliselt vähem on kasutatud partitiivi. Seevastu eesti õppijakeeles on genitiiv ja partitiiv kohad vahetanud: partitiivi on kasutatud tunduvalt sagedamini kui genitiivi. Kui piirduda nimetatud kahe korpusainese võrdlemisega, siis võiks seda nähtust pidada õppijakeelele iseloomulikuks jooneks, mida on tavaliselt püütud seletada lähtekeele mõju, õpetamis- või õpistrateegiatega. Samas analüüsides EKI korpuse ajakirjandustekstide alusel praegusaja keelekasutust selgus, et grammatiliste käänete ainsuse ja mitmuse vormide kasutamine on samalaadne eesti õppijakeele käändekasutusega⁶. Nagu kolmekümne-neljakümne aasta taguses eesti kirjakeeles, nii on ka praegusaja kirjakeeles valdav nominatiivi vormide kasutus, sageduselt teisel kohal on aga partitiivi vormid ning seejärel genitiivi vormid, mida on kasutatud tunduvalt vähem kui aastatetaguses eesti kirjakeeles. Mõistetavalt tekib küsimus, kuidas seda seletada. Kas eesti keele käändekasutuses on toimumas muutused ning millega neid võiks põhjendada? Mis on lahti eesti genitiiviga ja kas tegelikult üldse ongi midagi lahti? Sellele küsimusele püütakse leida vastust, uurides grammatiliste käänete leksikaalset ja leksikaalgrammatilist distributsiooni diakroonselt ning erinevate registrit alusel.

Sõnajärjevigade automaatse tuvastamise probleeme

⁴ <http://www.lexically.net/wordsmith/> (3.10.2009).

⁵ Vt Biber, Douglas & Conrad, Susan & Reppen, Randi 1998. Corpus linguistics: Investigating language structure and use. Cambridge University Press; Barnbrook, Geoff 1998. Language and computers: A practical introduction to the computer analysis of language. Edinburgh University Press.

⁶ Vt Eslon, Pille & Matsak, Erika 2009. Eesti keele kasutusvariandid: korpusest tulenev käändevormide võrdlev analüüs. Toim. Helle Metslang, Margit Langemets, Maria-Maren Sepper, Reili Argus. – Eesti Rakenduslingvistika Ühingu aastaraamat 5 (79 - 110). Tallinn: Eesti Keele Sihtasutus.

Erika Matsak

Tallinna Ülikool

matsak@tlu.ee

Ettekandes antakse ülevaade tööst sõnajärje, eesti õppijakeele sagedasema vealiigi automaatsel tuvastamisel seoses Eesti vahekeele korpuse (EVKK) vealeidja prototüübi loomisega. Töö on seotud riikliku programmi Eesti Keele Keeletehnoloogiline tugi (2006-2010) projektiga VAKO - Eesti vahekeele korpuse keeletarkvara ja keeletehnoloogilise ressursi arendamine (2008-2010).

Sõnajärje uurimiseks on kasutatud eesti keele süntaksianalüsaatorit koos ühestajaga, mis võimaldab morfosüntaktiliste märgendite abil lauseid formaliseerida. Vigase sõnajärje automaatset tuvastamist on alustatud õigete sõnajärjereeglite eraldamisest. Selleks on kasutatud Tartu Ülikooli eesti kirjakeele korpuse ilukirjandustekste. Vigase sõnajärje tuvastamiseks on EVKK veebikeskkonnas loodud eraldi moodul, mis võimaldab analüüsida õppijakeelt, s.h eesti õppijakeele sõnajärge. Ettekanne on pühendatud õigete ja valede sõnajärjereeglite tuvastamisvõimalustele eesti õppijakeeles ning sellega seotud probleemidele.

Eksistentiaalisten lauseiden sanajärjestys edistyneiden suomenoppijoiden kirjoitelmissa – korpuspohjainen tarkastelu

Ilmari Ivaska
Turun yliopisto
itiv@utu.fi

Turun yliopiston Lauseopin arkistoon kehitteillä oleva Edistyneiden suomenoppijoiden korpus (LAS2) on edistyneiden kielenoppijoiden kirjoitelmista koostuva elektroninen aineisto. Korpus koodataan niin morfologisesti kuin syntaktisestikin, minkä lisäksi se hakusanoitetaan, rakenteistetaan lauseittain ja lauseisiin koodataan niiden modaalinen lausetyyppi. Tällainen koodaus mahdollistaa tarkastelun, joka keskittyy rakenteellisiin elementteihin.

Suomen kielen eksistentiaalisessa lauseessa on useimmiten suomelle muuten epätyypillinen VS-sanajärjestys. Tarkasteltaessa niitä eksistentiaalisia lauseita, joissa on lausetyypille odotuksenvastainen SV-sanajärjestys, on käynyt ilmi muutamia potentiaalisia myötäesiintymiä. Näistä myötäesiintymistä selvin on lauseen negaatio, jota voidaan tarkastella sekä kieliopillisten suhteiden välisenä kolligaationa että leksikaalis-kieliopillisten suhteiden välisenä kolligaationa.

Lauseittain rakenteistettu, syntaktisesti ja morfologisesti koodattu korpus mahdollistaa tällaisten rakenteellisesti määriteltävissä olevien kokonaisuuksien esiintymisympäristöjen ja ehtojen kvantitatiivisen tarkastelun.

Negaatiossa ei ole mitään sellaista, mikä itsessään laukaisisi subjektin ja finiittiverbin välisen inversion. Lauseen kielteisyys on kuitenkin suhteellisesti todennäköisempää sellaisissa eksistentiaalisissa lauseissa, joiden sanajärjestys on eksistentiaaliselle lauseelle epäprototyyppinen SV. Tämä myötäesiintymä ilmenee sekä suomenoppijoiden kirjoitelmista että ensikielenään suomea puhuvien vertailuaineistosta. Vertailuaineiston epäprototyyppisissä tapauksissa SV-sanajärjestykselle on kuitenkin kaikissa tapauksissa osoitettavissa jokin motivoija. Edistyneiden suomenoppijoiden aineistossa yli 40 prosentissa tapauksista epäprototyyppiselle sanajärjestykselle ei ole löydettävissä kontekstista ainakaan mitään kohdekielen mukaista motivoijaa. Tällaisia lauseita on aineiston yhdeksästä informantista seitsemällä.

Tulosta voidaan tulkita siten, että kielenoppijoiden välikieleen on saattanut muodostua kohdekieleen perustuva mutta sen kieliopin vastainen analogia. Näin välikieleen on syntynyt eksistentiaalisen lauseen sanajärjestyksestä koskeva valintapreferenssi.

Tulokset on saatu analysoimalla korpusmateriaalia manuaalisesti ennen sen koodaamista. Kaikki ilmiön muuttujat pystytään kuitenkin määrittämään koodatusta korpuksesta niin, että tulokset suodattuvat samanlaisiksi.

**Korpuse morfoloogiline tükeldamine
Silvi Vare sõnaperede alusel**

Kairit Sirts

Tallinna Tehnikaülikool
kairit.sirts@hotmail.ee

Eesti Keele Instituudi juures on valmimas Silvi Vare poolt Sõnaperede sõnaraamat, mis defineerib ära seal esinevate sõnade jagamise morfeemideks sõnatuletuse alusel. Pakume välja meetodi, kuidas tükeldada sõnu morfeemideks Silvi Vare sõnaperede sõnastiku alusel. Tükeldamise aluseks on tekst, mis on olemasolevate morfoloogiliste analüüsi vahenditega jagatud tüve- ja lõpumorfeemideks ning Silvi Vare sõnaperede sõnastik. Meetodi sisuks on algoritm, mis koosneb kolmeteistkümnest reeglist. Kõik sõnad tekstis jagatakse kolmeteistkümnesse mittekattuvasse hulka ning iga hulga peal rakendatakse ühte reeglit. Kuna hulgad ei kattu, siis on reeglite rakendamise järjekord vaba. Reeglite aluseks ei ole võetud lingvistilised kategooriad, vaid need baseeruvad puhtalt mustrite sobitamisele.

Verbide lemmatiseerimine ja märgendamine eesti vahekeele korpuses

Anni Muru
Tallinna Ülikool
anni.muru@tlu.ee

Töö haakub riikliku programmi Eesti Keele Keeletehnoloogiline tugi (2006-2010) projektiga VAKO – Eesti vahekeele korpuse keeletarkvara ja keeletehnoloogilise ressursi arendamine (2008-2010), mille eesmärk Eesti vahekeele korpuse (EVKK) tekstide käsitsimärgendamiselt üle minna automaatsele.

Eesmärk on lemmatiseerida ja märgendada kõik EVKKs sisalduvad verbivormid. Esmalt taandatakse need algvormi kujule. Sõna leksikaalse tähenduse määramiseks tekstis kasutatakse sõna lähikonteksti.

EVKK normipäraste vormide lemmatiseerimisel ja märgendamisel saab toetuda eesti kirjakeele märgendamise kogemusele (Kaalep jt 2000; Müürisep 2000; Müürisep jt 2001; Muischnek & Vider 2005), kuid grammatikareeglitele mittevastavate vormidega tegeledes lähtutakse ühelt poolt EVKK veaklassifikatsioonist ja teisalt mittegrammatiliste vormide sidumisest normipärastega.

Ettekandes räägitakse lemmatiseerimis- ja märgendamisprotsessist, antakse lühiülevaade sellest, kuidas eesti kirjakeeles verbivorme märgendatakse ja mis on suurimad probleemid. Amuti on juttu sellest, kui kaugele on jõutud EVKK verbivormide lemmatiseerimisel ja märgendamisel.

Kaalep jt 2000 = Heiki-Jaan Kaalep, Kadri Muischnek, Kaili Müürisep, Andriela Rääbis, Külli Habicht. 2000. Kas tegelik tekst allub eesti keele morfoloogilistele kirjeldustele? Eesti kirjakeele testkorpuse morfosüntaktilise märgendamise kogemusest. – Keel ja Kirjandus 9, lk 623-633

Muischnek, Kadri, Vider, Kadri. 2005. Sõnaliigituse kitsaskohad eesti keele arvutianalüüsis. Eesti Rakenduslingvistika ühingu aastaraamat 1(2004). Eesti Keele Sihtasutus, Tallinn, lk 99-114

- Müürisep jt 2001 = Kaili Müürisep, Tiina Puolakainen, Heli Uibo, Mare Koit, Tiit Roosmaa, Kadri Muischnek. 2001. Eesti keele formaalne grammatika. Tartu: Tartu Ülikooli Kirjastus <http://math.ut.ee/~kaili/Loengud/Mudelid08/ems02.pdf> (14.09.2009)
- Müürisep, Kaili. 2000. Eesti keele arvutigrammatika: süntaks. Dissertationes Mathematicae Universitatis Tartuensis 22. Tartu <http://math.ut.ee/~kaili/thesis> (06.09.2009)

A corpus study on Finnish learner language: The use of the partitive case

Marianne Spoelman
University of Oulu
marianne_spoelman@hotmail.com

Finnish belongs to the Finno-Ugric language family and is particularly well known for its rich and complex morphology. Consisting of fifteen cases, the Finnish case system also comprises a partitive case, which is a typical case characterizing the Balto-Finnic languages. The partitive has often been found to be problematic for foreign learners of Finnish, because its use involves polarity, aspect, definiteness as well as divisibility of the subject, object or referent.

This corpus study investigates the use of the partitive case in Finnish learner language, focusing on Estonian, German and Dutch learners of Finnish. The source languages (L1s) have been chosen based on their genetic and typological distance to Finnish, since linguistic phenomena are assumed to be more difficult to learn, when not existing in a learner's L1.

The purpose of the study is to provide valuable insight into the nature of learner language and the role of L1 influence. The study is part of the project *Corpus study on language-specific and universal features in learner language*, in which the *International Corpus of Learner Finnish (ICLFI)* is compiled from 2007 onward.

In this talk, a frequency pilot study on the use of the partitive case in Finnish learner language will be discussed. This pilot study has been based on the use of partitive objects, predicatives and subjects in the Estonian, German and Dutch subcorpora of the ICLFI. Frequencies observed from the learner corpora were compared with each other and with a native Finnish reference corpus. In addition, some preliminary outcomes of an error analysis focusing on partitive overuse errors (i.e. partitives used instead of other cases) and partitive underuse errors (i.e. other cases used instead of partitives) will be discussed.

**Korpuspõhine uurimus eestikeelsete soome keele õppijate
sisekohakäänete kasutusest**

**Korpuspohjainen tutkimus vironkielisten suomenoppijoiden
sisäpaikallissijojen käytöstä**

Keaty Siivelt
Tallinna Ülikool
ksiivelt@tlu.ee

Ettekandes vaadeldakse eestikeelsete soome keele õppijate sisekohakäänete kasutust Soome õppijakeelekorpusel (*International Corpus of Learner Finnish - ICLFI*). Õppijakeelel põhinev uurimus võimaldab vaadelda nii lähtekeele mõju, õppijakeelele omaseid jooni kui ka õppematerjalide kohandamist õppija vajadustele (Eslon 2007: 98). Ettekandes analüüsitakse kuidas eestikeelsed soome keele õppijad kasutavad sisekohakäändeid ja milline on võimalik lähtekeele mõju.

Uurimus põhineb eesti ja soome sisekohakäänete kontrastiivsel analüüsil, mille tulemusi on võimaliku lähtekeele mõju selgitamiseks võrreldud sisekohakäänete korpuspõhise analüüsi tulemustega. Korpusel põhinev uurimus tagab tänu suurele ainesikule, tulemuste parema üldistatavuse. Uurimuse tulemustest selgus, et eestikeelsetele soome keele õppijatele tekitavad enim probleeme eesti ja soome morfoloogia ja fonoloogia, mitte aga morfosüntaksi erinevused. Tulemustest selgus, et lähtekeele mõju seisukohast on olulised kõik lähte- ja sihtkeele keeletasemete vahelised sarnasused ja erinevused. Sarnasused keelte erinevatel keeletasanditel võivad tugevdada võimalikku positiivset lähtekeele mõju.

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

Korpuspohjainen tutkimus ruotsinkielisten suomenoppijoiden illatiivin ja allatiivin käytöstä kirjallisessa tuotannossa

Tuija Määttä
Umeå universitet
tuija.maatta@finska.umu.se

Ruotsinkielisillä opiskelijoilla, jotka opiskelevat suomea vieraana kielenä alkeistason kurssilla, on joukko kielipillisiä asioita, joita heidän on hankala omaksua ja käyttää. Pitkäaikaisen opetusurani aikana olen huomannut, että yksi tällainen seikka on suomen paikallissijat. Kestää kauan, ennen kuin heille selkenee esimerkiksi sisä- ja ulkopaikallissijojen peruserot ja -käyttö.

Tässä esitelmässäni keskityn ainoastaan ruotsinkielisten suomenoppijoiden allatiivin ja illatin käyttöön eri tekstityypeissä. Tutkimukseni on osa tutkimushanketta *Korpus-tutkimus oppijankielen kielikohtaisista ja universaaleista ominaisuuksista*, jota johtaa dosentti Jarmo H. Jantunen Oulun yliopistosta. Käyttämäni korpus, joka sisältää vain alkeistason opiskelijoiden kirjoittamia tekstejä, on hankkeessa kerättävän *Kansainvälisen oppijansuomen korpuksen* (ICLFI) osakorpus.

Lukuvuoden aikana suomenoppijat kirjoittavat n. 10 tekstiä, joiden pituus on n. 30 lausetta. Tekstityypeinä ovat mm. kertomus, sarjakuva, kortti, päiväkirja ja jonkin oppikirjan tekstin jatkaminen. Tehtävät suoritetaan

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

Korpuksen analysoinnissa olen käyttänyt WordSmith Tools -ohjelmaan sisältyvää Concordance-ohjelmaa.

Ruotsinkieliset suomenoppijat sekoittavat illatiivin ja allatiivin käytön erilaisissa tekstiyhteyksissä. Tutkimukseni kautta pyrin löytämään niitä syitä, joista tällainen ”sekakäyttö” johtuu. Kysymyksiä, joihin haen vastuksia, ovat mm. onko suomen ja ruotsin kielten rakenteellisilla eroilla tekemistä väärän sijavalinnan kanssa, vaikuttaako yleinen maailmankuvan hahmottaminen sijavalintaan ja onko kirjoitteilla olevassa tekstissä mahdollisesti jotakin, joka aiheuttaa paikallissijan vaihtumisen?

***ma-* ja *da-*infinitiivi kasutuse võrdlus eesti kirjakeeles ja õppijakeeles**

Anastassia Šmõreitsik
Tallinna Ülikool
nastik@tlu.ee

Ettekanne põhineb 2009. aasta kevadel Tallinna Ülikoolis kaitstud magistritööol, kus võrreldi *ma-* ja *da-*tegevusnime kasutust kahes eesti keele variandis – kirjakeeles ja õppijakeelses. Korpusainesena on kasutatud Tartu Ülikooli eesti kirjakeele korpuse tasakaalustatud osa (TÜKK)⁷ ja Eesti vahekeele korpuse (EVKK)⁸ märgendatud osa tekstipankasid. Uurimuse metodoloogia põhineb korpuspõhisel uurimissuunal.

Keelendite analüüsis on rakendatud multidimensionaalsuse printsiipi, erinevaid kvantitatiivse ja kvalitatiivse analüüsi meetodeid. Kvalitatiivses analüüsis tuginetakse funktsionaalsele kontrastiivlingvistikale, milles uuritakse *ma-* ja *da-*infinitiivi kasutusmustreid eesti kirjakeeles ja õppijakeeles.

Mõlemast korpusest on tehtud representatiivne valim, mille põhjal on välja toodud *WordSmith Toolsi* programmide abil *ma-* ja *da-*infinitiivi statistiline olulisus nendes, samuti nimetatud vormide kasutussagedus ja tüüpilised infinitiive sisaldavad konstruktsioonid. Kõiki arvandmeid on töödeldud tabelarvutusprogrammis *Excel*.

***ma-* and *da-* infinitives usage comparison in standard Estonian and Estonian interlanguage**

Anastassia Šmõreitsik
Tallinna Ülikool
nastik@tlu.ee

This presentation is based on the master thesis defended in spring 2009. The presentation aims to compare the use of *da-* and *ma-* infinitives in standard Estonian and Estonian interlanguage. I have based this on a balanced use of the Estonian corpus at Tartu University and the Estonian inter-language

⁷<http://www.cl.ut.ee/korpused/grammatikakorpus/> (06.09.2009)

⁸ <http://evkk.tlu.ee/> (06.09.2009)

corpus. I carried out a representative selection from both of these language resources (using *Excel* spreadsheets). Using selection, I pointed out the frequency of statistically important forms of *ma-* and *da-* infinitives as well as the patterns of use of the most typical constructions in different contexts (using *WordSmith Tools*). I compared the use of *ma-* and *da-* infinitives in learner language quantitatively and qualitatively with the forms used in standard language.

After analysing the results, I pointed out the collocational entities and different syntactic constructions typical in different contexts that are characteristic to standard Estonian and learner's Estonian.

Soome *lähteä-* ja *mennä-*verbide korpuspõhine võrdlus Suomen *lähteä-* ja *mennä-*verbien korpuspohjainen vertailu

Kristi Pällin
Tallinna Ülikool
kristi.pallin@tlu.ee

Hoolimata sellest, et eesti ja soome on lähedased sugulaskeeled on *minema* ja *mennä-*verbides toimunud ajalooline lahkuareng. Eesti keeles on *minna* : *lähen* verbid ajalooliselt komplementaarses distributsioonis, kuid soome keeles on säilinud kaks iseseisvat täielikku paradigmat: *mennä* : *menen* ja *lähteä* : *lähden*.

Ettekandes analüüsin korpusuuringu põhjal, millistes kasutuskontekstides soome *lähteä-* ja *mennä-*verbe kasutatakse.

Analüüsist selgus, et struktuurilisel tasemel on *lähteä-* ja *mennä-*verbid sünonüümid, kuid kummalgi verbil on oma ülesanne. *lähteä-*verbi kasutatakse tegevuse algust tähistavates lausetes (nt. *Juna lähtee kello 6.30*. 'rong väljub kell 6.30') ja *mennä-*verbi kasutatakse protsessi kestust tähistavates lausetes (nt. *Juna meni ohi*. 'rong sõitis mööda').

Analüüsi tulemustest tuli välja ka hulk sõnu, mis moodustavad kollokatsioone vaid ühe verbiga ja teisega seda ei kasutata.

Lähtekeele mõju uurimise ühtne metodoloogiline raamistik ja selle kasutamine õppijakeele korpusanalüüsis

Lähdekielen vaikutuksen tutkimisen yhteinen metodologinen viitekehys ja sen soveltaminen oppijankielen korpusanalyyssissa

Annekatriin Kaivapalu
Tallinna Ülikool
kaivapa@tlu.ee

Lähtekeele mõju sihtkeele omandamisele on keeleõppealase uurimistöö keskseid, enim käsitletud, kuid siiani vähim selgeid tulemusi andnud valdkondi. Suurt osa alal valitsevast seisukohtade paljususest on võimalik vältida ühtse metodoloogia abil: ühesugune lähenemisviis muudab tulemused võrreldavaks. Ettekandes tutvustatakse Scott Jarvisi (2000) lähtekeele mõju uurimise ühtset metodoloogilist raamistiku ja selle edasiarendusi ning analüüsitakse selle metodoloogia rakendusvõimalusi eesti ja soome õppijakeele korpuspõhisel uurimisel. Samuti käsitletakse kontrastiivanalüüsi ja veanalüüsi uurimismeetodina ning tutvustatakse korpusainestiku märgendamise võimalusi Euroopa keeleõppe raamdokumendi (2007) keeleoskustasemetest lähtuvalt.

Jarvis, Scott 2000. Methodological Rigor in the Study of Transfer: Identifying L1 Influence in the Interlanguage Lexicon. – Language Learning 50: 2, June 2000, 245–309.

Euroopa keeleõppe raamdokument: õppimine, õpetamine, hindamine 2007. Tartu: Haridus- ja Teadusministeerium.

Teaduskeelekorpuse loomise valikuid

Peep Nemvalts
Tallinna Ülikool
peep@tlu.ee

Olemasolevaist korpustest lähtudes mõtisklen selle üle, milline peaks tulema TLÜ teaduskeelekeskuse eesti teaduskeele korpus.

Missuguseid tekste millises mahus oleks otstarbekas koguda ja mil moel?

Kas ja mille poolest teaduskeelekorpuse töötlemisvõimalused võiksid erineda vahekeelekorpuse omadest?

Mida võiks saada teaduskeelekorpuse põhjal uurida?