

**Tartu Ülikool**

**Filosoofiateaduskond**

**Germaani, romaani ja slaavi filoloogia instituut**

**ÜKSKEELSE SPETSIALISEERITUD KORPUSE  
RAKENDUSVÕIMALUSTEST KIRJALIKUS TÕLKES  
RADIOLOOGILISTE KIRJELDUSTE NÄITEL**

**Magistritöö**

**Eola Valdre**

**Juhendajad:**

**Prof Krista Vogelberg**

**Dr Peeter Ross**

**Tartu 2010**

*„Haiguste tundmine on esimene ja ülem asi, selleta arst vigadust mitte ei või parandada.  
Juhtub – kui meie paigus sagedaste, et tohter haigeid oma silmaga ei näha, veel vähem tema  
viga läbi katsuma saab, siis ei jää muud teejuhti, kui paljas sõnakandja jutt.“*

*Fr. R. Kreutzwald (Kreutzwald 1900: 12)*

# SISUKORD

1	SISSEJUHATUS.....	1
1.1	Töö teema.....	1
1.2	Teemavaliku põhjendus.....	1
1.3	Töö eesmärk.....	3
1.4	Aktuaalsus, tähtsus, uudsus ja rakendusvõimalused.....	4
2	TAUST.....	6
2.1	Eesti meditsiinkeel.....	6
2.1.1	Meditsiinkeel Euroopas läbi aegade.....	6
2.1.2	Inglise meditsiinkeele arengulugu.....	9
2.1.3	Eesti meditsiinkeele teke ja areng.....	15
2.2	Kommunikatsioon meditsiinis, meditsiinkeel ja e-tervise ajastu.....	18
2.2.1	Terviseandmed, meditsiinkeel ja e-tervis.....	20
2.2.2	Radioloogi vastus: meditsiiniline narratiiv ja struktureeritud tekst.....	21
2.3	Lingvistika, meditsiin ja tõlkimine.....	28
2.3.1	Keeletehnoloogiaressursid.....	28
2.3.1.1	Tekstikogu ja korpus.....	29
2.3.1.2	Tõlketeadus, tõlkimine ja korpused.....	31
2.3.1.3	Meditsiinkeele korpused.....	33
3	KASUTATUD MATERJALID JA VAHENDID.....	36
3.1	Uurimisobjekt.....	36
3.2	Inglisekeelsetes lülisamba MRT-uuringute kirjeldustes kasutatavad laused.....	37
3.3	Info- ja keeletehnoloogiavahendid.....	37
4	MEETODID.....	38
4.1	Korpuse korrastamine: lülisamba MRT-uuringute tööfaili teisendused enne keelelist analüüsi 38	
4.1.1	Lülisamba MRT-uuringute kirjelduste struktuuri ühtlustamine.....	38
4.1.2	Lülisamba MRT-uuringute kirjelduste vabateksti keeleline korrastamine.....	40
4.1.3	Korpuse liigendus ja märgendus.....	42
4.2	Lülisamba MRT-uuringu kirjelduse kliiniline struktuur.....	43
5	TULEMUSED.....	44
5.1	Lülisamba MRT-uuringute keeleline analüüs.....	44

5.1.1	Tekstides esinevate sõnade sagedus võrrelduna eesti kirjakeele korpuse sõnasagedustega	44
5.1.2	Tekstides esinevad lühendid.....	46
5.1.3	Süntaksi eripära tekstides .....	49
5.1.4	Samatähenduslikud laused .....	52
5.2	Lülisamba MRT-uuringute tekstide terminoloogia ja kliiniliste andmete struktuur .....	54
6	ARUTELU .....	61
6.1	Miks on vaja tõlkida autentseid terviseandmeid?.....	61
6.2	Kui pädev peab autentse meditsiiniteksti tõlkija olema meditsiinis?.....	62
6.3	Milliseid probleeme esineb autentsetes meditsiinitekstis?.....	64
6.3.1	Tekstide suur variatsioon – kuidas standardida?.....	64
6.3.1.1	Kirjavead .....	65
6.3.1.2	Lühendid.....	66
6.3.1.3	Nimisõnafraasid täislausete asemel.....	67
6.3.1.4	Eituse kasutamine.....	69
6.3.1.5	Samatähenduslikud laused .....	70
6.4	Milleks on hea struktureeritud tekst? .....	71
6.5	Miks on masintõlge radioloogiliste kirjelduste tõlkimiseks mõistlikum lahendus kui tõlkija palkamine? .....	72
7	KOKKUVÕTE.....	73
8	KASUTATUD KIRJANDUS .....	75
9	LISAD .....	79
	Lisa 1. Inglisekeelsetes lülisamba MRT-uuringute kirjeldustes kasutatavad laused.....	79
	Lisa 2. Mõttekriipsu kasutamine lülisamba MRT-uuringute kirjeldustes järeluste esitamiseks .....	82
10	SUMMARY .....	87

## LÜHENDID JA MÄÄRATLUSED

Käesolevas töös on lühendeid ja võtmesõnu kasutatud järgmistes tähendustes.

<b>autentne meditsiinitekst</b>	autentses tööolukorras loodud terviseandmeid sisaldav meditsiinidokumendi toimetamata tekstiosa (nii struktureeritud kui ka vabatekst), mis on mõeldud teabevahetuseks meditsiinispetsialistide vahel
<b>e-tervis</b>	elektroonseid terviseandmeid haldavate andmebaaside võrgustik
<b>korpus</b>	teatud kriteeriumide järgi valitud ja käsitletud tekstikogu
<b>lingvistika</b>	keeleteadus
<b>lähtekeel</b>	keel, millest tõlgitakse
<b>lähtetekst</b>	tekst, mida lähtekeelest tõlgitakse
<b>masintõlge</b>	arvutipõhine tõlge
<b>magnetresonantstomograafia</b>	tuumamagnetresonantsil põhinev radioloogiline uurimismeetod
<b>meditsiinkeel</b>	kogu meditsiini valdkonna spetsialiseeritud sõnavara, keelekasutus ning kirjalik teabevahetus
<b>MRT-uuring</b>	magnetresonantstomograafia uuring
<b>NLP-vahendid</b>	loomuliku keele töötlemise vahendid
<b>radioloogia</b>	meditsiinilistel visualiseerimismeetoditel ehk pildidiagnostikal põhinev meditsiiniharu
<b>radioloog</b>	radioloogias spetsialiseerunud arst
<b>radioloogiline kirjeldus</b>	radioloogilise kujutise kirjalik kirjeldus, mille radioloog annab vastuseks raviarstile
<b>sihtkeel</b>	keel, millesse tõlgitakse
<b>sihttekst</b>	sihtkeelde tõlgitud tekst
(meditsiinidokumendi) <b>struktureeritud tekst</b>	teatud liigenduse ja/või ontoloogiatega piiratud tekstiloome tulemusel tekkinud kirjalik autentne meditsiinitekst
<b>terviseandmed</b>	patsiendi tervislikku seisundit kajastavad andmed (arvnäitajad, piltkujutised, tekstid jms), mis on koondatud kas paber kandjal olevatesse (haiguslugu, ambulatoorne kaart jne) või elektroonsetesse meditsiinidokumentidesse (e-terviselugu)
(meditsiinidokumendi) <b>vabatekst</b>	struktuuriga ja/või ontoloogiatega piiramata tekstiloome tulemusel kirjutatud autentne meditsiinitekst

# 1 SISSEJUHATUS

## 1.1 Töö teema

Käesolevas magistritöös käsitletakse eesti meditsiinikeele korpusepõhise uurimise võimalusi ning uuritakse autentse meditsiinikeele ükskeelse korpuse potentsiaalseid rakendusvõimalusi masintõlke lahenduste arendamisel. Töös on kasutatud autentseid isikustamata vabatekstilisi meditsiinitekste, täpsemalt lüüsisamba magnetresonantstomograafia uuringute (MRT-uuringute) kirjeldusi. Tegu on empiirilise korpusepõhise uurimusega.

Töö on suunatud uute keeleressursside loomise vajaduse uurimisele: vaatenurgaks on nende ja juba olemasolevate keele- ja infotehnoloogiaressursside rakendusvõimalused eesti-inglise-eesti suunaliste autentsete meditsiinitekstide tõlkimisel. Kuna töö keskendub ükskeelse spetsialiseeritud korpuse võimalustele, siis ei välista see töös esitatud tulemuste rakendatavust ka teistel keelesuundadel ning eesmärkidel (nt andmekäives). Sihtrühmad, kellele käesolev töö mõned aspektid võiksid huvi pakkuda, on arstid (eelkõige radioloogid), lingvistid ja tõlkijad.

## 1.2 Teemavaliku põhjendus

Meditsiinikeel ei ole olnud Eestis intensiivse uurimise objektiks, samas kasvab meditsiini üldise arengu foonil pidevalt vajadus nii heade omakeelsete tekstide kui ka asjalike tõlgete järele. Üldse ei ole seni uuritud autentset, s.o töösituatsioonis kasutatavat, keelt. Viimastel aastakümnetel on eesti meditsiinikeelt tervikuna oluliselt mõjutanud kolm tegurit:

- inglise keele kerkimine meditsiini ja teaduse *lingua francaks* (Berghammer 2008; Wulff 2004);
- vajadus talletada ja edastada terviseandmeid elektroonsel kujul (KOM (2004) 356);

- tervishoiuteenuse muutumine piiriüleseks, seda nii veebikeskkonna loodud võimaluste kui ka inimeste vaba liikumise tõttu (Ross 2010).

Tõenäoliselt on inglise keele levik mõjutanud ka eesti meditsiinkeelt ning seda eriti uutes meditsiinivaldkondades, mida sageli ongi arstitudengitele või arstidele õpetatud kas inglise keeles või inglise keele vahendusel. Seda mõju tuleb kuigivõrd arvestada (või vähemalt teada) ka tõlkimisel.

Üheks neist uutest meditsiinivaldkondadest Eestis on MRT-uuringud, mis jõudsid Eestisse kliinilise uuringumeetodina umbes samaaegselt Nõukogude Liidu lakkamisega. Venekeelseid asjakohaseid õpikuid uuringumeetodi juurutamisel kasutada ei olnud ning eestikeelseid õppematerjale polnud ka veel tehtud. Paljud radioloogid said algusaastatel põhjalikuma väljaõppe MRT-uuringutest inglise keele vahendusel välismaal.

Andmete elektroonne salvestamine on võimaldanud integreerida olemuslikult erinevat, nii patsiendi- kui ka tervishoiuteenuse osutaja keskset teavet ühtses andmebaasis, kuid analüüsitakse sellest teabest vaid murdosa. Osaliselt on see tingitud terviseandmete kirjelduse kvaliteedist. Andmeid saab elektroonses andmebaasis esitada struktureerituna või vabatekstina. Struktureeritud andmetest on palju lihtsam otsitavat leida kui vabatekstist, kuid autentset meditsiiniteksteid on peamiselt vabatekstilised.

Seoses Eesti kuulumisega Euroopa Liitu on suurenenud ja suureneb veelgi piiriülene suhtlus ning inimeste mobiilsus. Lisaks areneb hoogsalt ka piiriülene koostöö, sh ka tervishoiuteenuse pakkujate vahelised allhanked või konsultatsioonid (Ross 2010). Seetõttu on muutunud aktuaalseks ka autentsete meditsiinitekstide tõlkimine. Radioloogia on piiriülese koostöö hea näide, sest tänapäevane arvutitehnoloogia võimaldab radioloogil kirjeldada pilte nende

tegemiskohast eemal viibides. Seega ei pea radioloog Inglismaal tehtud pildi kirjeldamiseks olema Inglismaal ega Eestis tehtud pildi korral Eestis. Radioloog tõlgina mõistab „tõlkida“ kujutise keelt meile arusaadavateks tähemärkideks, kuid see ei tähenda, et iga radioloog suudaks kirjeldada kujutist mingis võõrkeeles. Samuti ei ole võimalik loota, et tõlkimisega saaks hakkama meditsiinilise erihariduseta tõlkija: autentne meditsiinitekst on palju spetsiifilisem, sisaldab ebakonventsionaalseid lühendeid, sõnastikes kirjeldamata püsiühendeid ja termineid, lausestus on täislausete asemel esindatud peaaesjalikult nimisõnafraasidega ning tekst ise on toimetamata, sisaldades rohkeid näpuvigu. Lisaks on kogu materjali hulk tegelikkuses nii suur, et see käiks vähestele meditsiinis pädevatele tõlkijatele lihtsalt üle jõu, rääkimata tõlkimise aja- ja kulumahukusest.

Seetõttu on oluline mõista, kuidas oleks võimalik autentset meditsiiniteksti käsitleda nii, et seda suudaks lugeda ja tõlkida arvuti. Selleks et andmete loomine ja kasutamine oleksid efektiivsed, peaksid nii andmete esitus kui ka nende tõlkimine olema võimalikult standarditud.

### **1.3 Töö eesmärk**

Töö eesmärgiks on luua väike ükskeelne spetsialiseeritud, s.o autentsetest meditsiinitekstidest koosnev pilootkorpus, mille näitel uurida asjaolusid, mis võiksid osutada probleemiks tekstide standarditud käsitlemisel ja tõlkimisel. Meditsiinitekstide näiteks on valitud lülisamba MRT-uuringute radioloogilised kirjeldused.

Töö käigus tuleb lahendada järgmised ülesanded:

- korrastada, süstematiseerida ja analüüsida lülisamba MRT-uuringute vabatekstilised kirjeldused;
- analüüsida andmete esituse struktuuri;



- analüüsida lingvistiliselt autentsete meditsiinitekstide eristumist üldkeelest;
- uurida, kuidas mõjutab autentse meditsiiniteksti keeleline eripära selle tõlkimist ja tõlkerakenduste loomist.

#### **1.4 Aktuaalsus, tähtsus, uudsus ja rakendusvõimalused**

Meditsiini, arvutilingvistikat ja tõlkimist ei ole Eestis seni ühendada püütud ning ka maailmas tervikuna on see uus suund, mis on hakanud arenema eelkõige seoses terviseandmete elektroonse registreerimise ja analüüsiga. Samas on kõik kolm valdkonda eraldi võetuna Eestis väga arvestatava arengutasemega, mis peaks asjakohaste projektide olemasolul võimaldama väga efektiivseid lahendusi.

Tuleb rõhutada, et autentne meditsiinikeel ei ole sama, mis arstiteaduslikus ajakirjas, meditsiiniprojektides või teadusuuringutes kasutatav keel. „Eesti Arsti“ põhjal koostatud korpus (<http://www.cl.ut.ee/korpused/segakorpus/eestiartst/>) ei hõlma autentsetes tööolukorras kasutatavat keelt.

Käesolevas töös on esitatud väikesemahuline prototüüpne näitematerjal, millega soovin vastata küsimusele, kas ja kuidas on võimalik juba olemasolevaid (ja pidevalt arenevaid) keele- ja infotehnoloogiaressursse meditsiinitõlke otstarbel rakendada ning uusi keeleressursse haigla infosüsteemis olevatest isikustamata tekstiandmetest luua. Olemasolevate ressursside kasutamine ning meditsiinitekstide masintõlge on olemuslikult ressursi- ning kulusäästlik lähenemine. Automaatne tõlkelahendus toetaks kaudselt patsientide vaba liikumist, vähendades oluliselt piiranguid terviserikkega patsientide reisimisel võõrkeelsesse keskkonda. Olen käesoleva tööga püüdnud juhtida tähelepanu sellele, et **juba praegu on haiglate andmebaasides olemas suur hulk kirjeldavat meditsiinilist vabateksti, mida on võimalik**

**lisaks tervisega seotud andmekaevele kasutada ka rakenduslikeks keeleuuringuteks, sh masintõlke rakenduste loomiseks tulevikus.** Masintõlke võimalus hõlbustaks oluliselt ja tõenäoliselt ka standardiks piiriülest meditsiiniinfo vahetamist. Olemasolevad keeletehnoloogiavahendid ja autentsed meditsiinitekstid võimaldavad eestikeelse meditsiinikeele sõnavara ja süntaksi uurimist, taksonoomiate, tõlkeotstarbeliste tarkvaralahenduste ja (arvuti)leksikonide loomist.

## 2 TAUST

### 2.1 Eesti meditsiinkeel

#### 2.1.1 Meditsiinkeel Euroopas läbi aegade

Meditsiinkeele ajalooline areng hõlmab pea iga Euroopa keele korral vähemalt nelja arvestatavat perioodi, mis on meditsiinkeelt Euroopas mõjutanud: kreeka keele, ladina keele, riiklike keelte ja inglise **meditsiini**keele ajastud. Vanimad teadaolevad meditsiinitekstid on 5.–4. sajandil e.m.a kirjutatud Hippokrates. Pikka aega oli ka Rooma impeeriumi meditsiinkeeleks kreeka keel. Esimese sajandi alguses, kui kreeka keel oli ikka veel meditsiinkeeleks Rooma impeeriumis, kirjutas Aulus Cornelius Celsus kreekakeelsetel allikatel põhineva „*De Medicina*“. Teose erakordselt kauni ladina keele pärast hakati autorit nimetama ka *Cicero medicorum*iks. Kuid Celsuse suurimaks mureks oli, et **enamikul kreeka meditsiiniterminitest puudusid ladina vasted**. Tema lahendus probleemile ongi lingvistika seisukohalt huvipakkuv. Esiteks võttis ta mõned sõnad kreeka keelest otse üle, säilitades isegi nende grammatilised lõpud ning kirjutas need lihtsalt kreekakeelsete tsitaatsõnadena kreeka keeles oma teksti. Seejärel muutis ta sõnad ladinapäraseks, asendades kreekapäraseid sõnalõpud ladinapärasega. Kolmandaks kasutas ta ka tõlkevasteid (Wulff 2004). **Ladina keel muutus aga peagi mitte ainult meditsiinist kirjutamise keeleks, vaid ka arstiteadlaste omavahelise kommunikatsiooni keeleks.**

Enamik paberandjal olevast meditsiinimõttest keskaegses Euroopas oli pärit kreeka meditsiinist. Originaalkirjutisi oli vähe. Klassikaliste tekstide taustaks olid ka nii suured autoriteedid, et nende seisukohtade vaidlustamine oli tõenäoliselt lihtsalt mõeldamatu. Keskajal lisandus kreeka ja ladina keelele veel araabia keel, sest mitmed kreekakeelsed tekstid tõlgiti araabia keelde. Araabiamaa arstiteadlased andsid samuti oma panuse meditsiini ning

nii kandus osa termineid meieni ka araabia keelest. Kreekakeelsed allikad jõudsid Euroopa teadlasteni kas ladina või araabia keele kaudu. Just tänu Araabia arstiteadlastele säilisid paljud Kreeka meditsiinitekstidest, mis oleksid pöördumatult kadunud varasel keskajal.

**Araabiakeelsetena säilinud tekstid taaselustasid X sajandi paiku ladinakeelsete meditsiinitekstide traditsiooni ning Toledo Hispaanias ja Salerno Itaalias said tuntuks meditsiini ja meditsiinitekstide tõlkekeskustena** (Mäkinen 2006: 2).

Renessansiajastul oli valdavaks juba ladina keel. Celsuse „*De medicina*“ ilmus trükituna 1478. aastal, see on vaid paar aastakümnet hiljem, kui trükkimine üldse leiutati. Tasapisi hakati ka riigikeeltes meditsiinist kirjutama ning nt Suurbritannias oli tõenäoliselt viimaseid ladina keeles kirjutatud meditsiinitekste 1802. aastal ilmunud William Heberdeni „*Commentarii*“. Hiljem on autorile viidatud kui *ultimus Romanorum*le (viimasele roomlastest) (Wulff 2004).

Ladina ajastule meditsiinitekstiloomes järgnes riigikeeltes kirjutamise aeg. **Kuigi Euroopa keeltesse võeti ladina keelest põhimõtteliselt üle samad terminid, tehti seda siiski keeleti erinevalt.** Näiteks germaani keeltes (saksa, hollandi ja Skandinaavia keeled) võeti terminid üle korrektsete ladina lõppudega (nt *nervus musculocutaneus* või *ulcus ventriculi*), seevastu romaani keeltes kohandati need lõpud oma keele päraseks (nt prantsuse keeles *le nerf musculo-cutané* ja *ulcère gastrique*). Kuigi inglise keel on germaani keel, on selle sõnavarast üsna suur osa romaani päritolu. Seega kalduti ka inglise keeles järgima romaani keeltele iseloomulikku adapteerimist, kui välja arvata omadussõnade paigutamine nimisõna ette (seega siis *musculocutaneous nerve* ja *gastric ulcer*). Slaavi keeltes on meditsiinitermine jaoks traditsiooniliselt kasutatud pigem tõlkevasteid (nt vene keeles *кожно-мышечный нерв* ja *язва желудка*). Tasub märkida, et kreeka keeles lubatakse tänapäeval vaid kreekakeelsete terminite kasutamist. See kehtib isegi nende terminite kohta, millele juba Celsus ladina tõlkevasted

pakkus. Muude Euroopa keelte puhul on tegu pigem üldise tendentsi kui kindla reegluga, on ju ka inglise keelde üle võetud kohandamata ladinakeelseid termineid (nt *diabetes mellitus*) ning saksa või rootsi keeles neid hoopiski kohandatud (nt südame pärgarterid saksa keeles *Coronararterien*) või tõlgitud (nt maohaavand saksa keeles *Magengeschwür*, rootsi keeles *magsår*) (Wulff 2004; Nyman 2003). **Peab siiski nentima, et ladina terminite suhteliselt hägusate reeglitega keelde ülevõtmine on tekitanud teatava ebakindluse nende ortograafia suhtes.** Rootsi keeles esinevatele sagedastele vigadele, mis tõenäoliselt tulenevad sellest, et arstid päris kindlalt enam terminite etümoloogiat ei tea, on juhtinud tähelepanu Hans Nyman, kes on püüdnud tekkivaid vigu ka rühmitada. Esimesed neist moodustavad n-ö tavalised näpukad: nt *ileum* (niudesool), mitte *ilium* (niudeluu). Teiseks puudub selge teadmine ladina keele grammatilistest vormidest ja sõnatüvest (nt *femoro-*, mitte *femuro-*: kuigi sõna on nominatiivis *femur* (reieluu, reis), võetakse tüvi genitiivist, mis on *femoris*. Kolmandal juhul puudub arusaam ladinakeelsete sõnade etümoloogiast, nt õige on *fetus* (vililane, hilisemal arengustaadiumil – loode), mitte *foetus*. Ladina morfeem *fe* tähendab õigupoolest viljakat, sigivat ning sellest on tulenevad ka nt sellised sõnad nagu *femina* (naine), *fecundus* (viljakus) (Nyman 2006). Sagedased on ka eksimused sõnade arvu ja soo suhtes, vahest ka rootsi keelele iseloomuliku mitmuse tunnuse (-a) tõttu, mida kiputakse ka ladina vastetele lisama. Tõenäoliselt on osa neist vigadest sarnased eesti meditsiinikeeles, osa aga tänu eesti keele kuulumisele flektiivsete keelte hulka ja soome-ugri keelerühma hoopis teistsugused. Eesti keele käändelõppude lisamine ladinakeelsetele terminitele tundub autentsete toimetamata tekstide puhul olevat üks suure variatsiooni allikas. Seda aspekti oleks võimalik autentsete meditsiinitekstide korpuse olemasolul objektiivselt uurida.

### 2.1.2 Inglise meditsiinikeele arengulugu

Inglise teadus- ja meditsiinikeel on arenenud üle aastatuhande ning saavutanud praeguseks omalaadse *lingua franca* staatuse (Berghammer 2008; Wulff 2004; The PLoS Medicine Editors 2006): **inglise keel ei ole ainult keel, milles avaldatakse valdav osa meditsiinialaseid teadustekste, vaid on muutunud ka valdkonna rahvusvaheliseks suhtluskeeleks.** Areng on olnud siiski palju pikaajsem, kui arvata võiks. Nii käsitletakse Helsingi Ülikooli projektis *Scientific Thought Styles: The Evolution of Early English Medical Writing* (Taavitsainen ja Pahta, käimasolev projekt) inglise meditsiinikeele arengut keskajal korpuspõhiselt. Asjakohane keelekorpus on jagatud kolme suurema perioodi vahel: 1375–1500 (suurus praegu: 86 teksti, 495 322 sõna, avaldatud), 1500–1700 (suurus: 1 800 000 sõna, valmimas) ja 1700–1800. Uurimus käsitleb teadusmõtte ja selle ühe väljendajana meditsiinikeele lähenemist tavakeelele kohalikus keeles kirjutamise kaudu; kuidas sõnavara ja koodivahetuse muutused kajastasid ajastu mitmekeelsust ja intertekstuaalsust inglisekeelsetes tekstides. Esimesed teadaolevad inglisekeelsed teadustekstid on pärit anglosaksi perioodist, meditsiinitekstide inglise keele pidevat arengut on võimalik jälgida alates XIV sajandist (Taavitsainen ja Pahta 2009: 1). Voigtsi ja Kurtzi elektroonses andmekogus on aga ligi 300 anglosaksi ajast (IX–XII saj) pärit kohalikus keeles teadusteksti, sh meditsiinitekste, millest enamik on pärineb XI sajandist (Taavitsainen ja Pahta 2009: 9). Osa neist kirjutistest on tõlked ladina keelest, osa aga ilmselt originaalkirjutised. Kõige varasemad säilinud meditsiinitekstid Inglismaal on õigupoolest kirjutatud anglonormanni prantsuse, mitte inglise keeles (Taavitsainen ja Pahta 2009: 11). Tuleb märkida, et ka mõnevõrra hilisemal perioodil esines meditsiinitekstides veel tolaegse anglonormanni prantsuse ja inglise keele läbisegi kasutamist. Tõlketeaduse seisukohalt on huvitav, et **keskaegsete inglisekeelsete meditsiinitekstide tõlkimisel tekkisid samad probleemid, mis esinevad kaasajalgi.** Väga

varasest ajajärgust on võimalik eristada sõna-sõnalt tõlkimist (*word for word*) sisulisest tõlkest (*sense for sense*) ning suur variatsioon tõlketekstide vahel on juba sel ajal tingitud lähteteksti sisu mõistmisest tõlkija poolt ning tema valitud tõlkestrateegiast. Meditsiinitõlked skaala sõna-sõnalt tõlkimise osas on äärmiselt lähtetekstitruud, järgides piinliku täpsusega ka selle süntaktilist struktuuri ning sõnavara. Mõned nendest tekstidest on nii ladinakeelse lähteteksti struktuuri lähedased, et neid on ilma originaalita võimatu mõista. Seevastu skaala teises otsas olevad tõlkijad, kes edastasid eelkõige lähteteksti sisu – seda vabalt sõnastades, liigendades ja toimetades – jõudsid originaaltekstile tegelikult märksa lähemale (Taavitsainen ja Pahta 2009: 13–14).

Ladina keel oli Läänemaailma teadusekeeleks kuni keskaja lõpuni, mil teadustekste hakati järjest enam tõlkima kohalikesse keeltesse. Teadusmõtte rahvakeelsemaks (ja seega ka arusaadavamaks) muutmise eestvedajaks oli Inglismaal meditsiin. See on iseenesest mõistetav, sest peale selle, et meditsiin oli teadus, oli see ka ametioskus, millega mindi tavaliste inimeste juurde. Diagnoosimine ja ravi põhinesid emakeelsel suhtlusel. Meditsiinitekstide tõlkimine kohalikku keelde algas juba XIV saj teises pooles ning XV saj lõpuks olid inglisekeelsete tekstidega esindatud kõik tolleaegsed meditsiinitekstide žanrid: kirurgiatekstitid, spetsialiseeritud tekstid, ravimite kirjeldused ja *materia medica* tekstid (Mäkinen 2006: 2).

Muidugi sõltus rahvakeelsete tekstide lugemine lugejaskonnast. Kirjaoskuse kohta keskaegsel Inglismaal on vasturääkivaid andmeid. Ülikooliharidus vastas põhimõtteliselt kõrgtasemel kirjaoskusele ning tähendas ka ladina keele oskust. Samas on alust arvata, et enamik aristokraatidest (nii mehed kui naised) olid väga hea kirjaoskusega, mis tõenäoliselt hõlmas ka ladina keele aluseid. Linnastumise kasvu, keskklassi, koolide ja uute ametite tekkega levis kirjaoskus kiiresti. Kõige optimistlikumate hinnangute alusel oskas XV saj Inglismaal enamik

kõigi ühiskonna klasside esindajatest lugeda ja kirjutada, tagasihoidlikumate hinnangute järgi suutis seda umbes 30% elanikkonnast. Meditsiinitekstide selge sihtrühma moodustasid meditsiinikutsete esindajad. Tuleb arvestada, et sugugi kõigil neist ei olnud ülikooliharidust ega ladina keele oskust, nt aadrilaskmise või hambaraviga võisid tegelda ka habemeajajad, diagnoosida ja ravimeid määrata apteekrid, igasugu günekoloogilisi probleeme ravida ämmaemandad jne. Meditsiinitekstitid olid seega suunatud üsna erinevate diskursuste esindajatele ja sihtrühma tuli mingil moel arvesse võtta. Nii nt on tekst „The Knowing of Woman’s Kind in Childing“ suunatud eksklusiivselt naislugejaskonnale, algselt tõenäoliselt ämmaemandatele (Taavitsainen ja Pahta 2009: 15–18).

Inglisekeelsete meditsiinitekstide hulk hakkas kiiresti suurenema koos arstiteaduse arengu ja institutsioonide tekkega umbes 1700. aasta paiku, kui tasakaal nihkus ladinakeelsetelt tekstidelt inglisekeelsete tekstide kasuks. Selleks ajaks, täpsemalt 1660. aastal, oli Inglismaal juba loodud teaduste akadeemia (The Royal Society), mis hakkas 1665. aastal välja andma ajakirja „The Philosophical Transactions”, mis on kõige kauem ilmunud teadusajakiri maailmas (The Royal Society 2010).

**Praegu on valdav osa maailmast jõutud inglise meditsiinikeele ajastusse, mis meenutab ladina meditsiinikeele aega eelkõige selles mõttes, et ka spetsialistide vaheline rahvusvaheline suhtlus on inglisekeelne.** Raske on öelda, millal see muutus täpselt algas, kuid tõenäoliselt on protsessi oluliselt mõjutanud ja kiirendanud maailmasõjad Euroopas. Nii juhib Berghammer (2008) tähelepanu asjaolule, et XIX saj lõpul ja XX saj alguses oli saksa keel väga arvestatav teaduskeel, seda tänu saksa teadlaste (nt Koch, Billroth, Röntgen jt) väljapaistvatele saavutustele. Paljud meditsiiniajakirjad (nt „Acta Medica Scandinavica“) avaldasid saksakeelseid artikleid ning isegi Jaapanis oli saksa ajakirjal „Deutsche Medizinische



Wochenschrift“ väga suur lugejaskond. Teine maailmasõda muutis olukorra täiesti: paljud saksa teadlased emigreerusid, lisaks olid maailmas valdavad saksavastased meeleolud ning ka Saksamaa majanduslikud võimalused olid kesised. Tulemusena kerkis USA maailma juhtivaks majandus- ja teadusjõuks. Tänapäeval on enamik arvestatavaid refereeritavaid meditsiiniajakirju inglisekeelsed ning see tähendab, et teistes keeltes teadustööde avaldamine vähendab oluliselt artikli lugejaskonda. Ühe keele kasutamine on iseenesest eelduseks ühtsetel alustel rahvusvahelisele teadustöö taseme hindamise toimimisele. Autorite seisukohalt on inglise keele kaudu võimalik jõuda palju suurema hulga lugejateni ning algatada huvipakkuva teema rahvusvahelist arutelu või uurimist. Näiteks loobus Pariisis asuv Pasteuri Instituut 1990ndatel oma väljaande „Annales de l’Institut Pasteur“ publitseerimisest prantsuse keeles, sest kui 1973. a esitati vaid 15% käsikirjadest inglise keeles, siis 1987. a oli praktiliselt 100% käsikirjadest inglisekeelsed (Berghammer 2008).

### **Inglise keele levik teiste riikide keeltesse meditsiinikeele kaudu tekitab aga ka muret.**

Näiteks arutleb Hans Nyman oma artiklis selle üle, et kaks keelevaldkonda on olnud rootsi keeles eriti tugeva inglise keele mõju all: infotehnoloogia ja meditsiin. Nii tõdeb ta, et sageli ei tea gümnaasiumilõpetajad enam isegi hapniku rootsikeelset nimetust (*syre*), vaid kasutavad pigem inglisekeelset (*oxygen*). Teaduskeele mõju on hakanud tungima tavakeelde ja pole ka ime, kui suur osa arstiteaduskonna kursuste lektüürist on inglisekeelne, loenguid ja vahel koguni täispikki kursusi viiakse läbi inglise keeles, väideldakse inglise keeles ning ka tähtsamad valdkonna teadusavastused avaldatakse inglise keeles. See tekitab ka ridamisi keeleprobleeme alates sellest, kuidas inglise termineid keelde sulandada – kas kohandada, võtta üle tsitaatsõnadena või tõlkida, kuidas hakkavad suures osas inglise keeles õppinud tudengid tööle rootsikeelsete arstidena ning kas ja kuidas see mõjutab nende suhtlemist

patsientidega (Nyman 1999). Samuti üritab Nyman vastata sellele, miks peaks rootsi keele suhtes rohkem muretsema praeguste inglise keele põhiste terminite ülevõtmiste kui (endiste) klassikalistest keeltest võetud/võetavate terminite pärast. Viimaste suhtes on sajandite vältel väljakujunenud kindlad reeglid ning – mis peasi (ja kehtib tõenäoliselt eesti keelegi kohta) – ladina keele hääldus on palju lihtsamalt kohandatav võrreldes inglise keelega, mille puhul erineb hääldus märkimisväärselt rootsi (ja ka eesti) keele omast. Laias laastus on rootsi keeles (nagu teisteski keeltes) võimalik pakkuda inglise terminile tõlkevaste (nt inglise *orphan drug* (harvikravim) on rootsi keeles *särlekemedel*), kohandada sõna rootsipärase kirjapildiga (tüüpiliselt nt c → k (mida kasutatakse ka meil): *scanner* → skanner) või võtta üle kogu sõna. Alati ei ole võimalik tõlkevastet pakkuda, sest inglisekeelse termini sisu täpselt kajastavat sõna lihtsalt keeles (veel) ei ole. Näiteks rootsi keeles võiks inglise sõna *stroke* (insult) ligilähedane vaste olla *slaganfall* (ajurabandus), kuid tänapäeva kontekstis ei ole see enam täiesti täpne vaste. Kõigil juhtudel, kui tegu ei ole tõlkevastega, tuleb rootsi keeles inglisekeelse sõna puhul mõelda hääldusele. Näiteks inglise *stroke* võiks häälduda nagu rootsikeelne *bok* (raamat) ja võiks muutuda seega sõnaks *strok*. Lisaks tuleks lahendada grammatilise soo probleem (seda probleemi eesti keeles ei ole): kas *en strok* või *ett strok* ning milline võiks tulenevalt olla sõna mitmus: *flera strokar* või *flera strok*? (Nyman 2000). **Seega nõuab isegi pelgalt ühesõnaliste terminite ülevõtmine inglise keelest mõnda teise keelde üsna suurt keelekorralduslikku tööd. Sageli jõuavad inglise meditsiinikeelest oma keelde nii mõnesõnalised püsiühendid kui ka lausa fraasid.**

Tuleb arvestada, et vähemalt seni, kuni juhtivaid teadustöid avaldatakse inglise keeles, on olukord selline, et inglise keelest ülevõetavaid termineid tuleb paratamatult juurde. Näiteks sellised enamikesse keeltesse mingil moel jõudnud sõnad nagu *clearance*, *bypass operation*,

*screening, scanning* jpm. Nii nagu ladina keelegi korral, on nende ülevõtmisel eri Euroopa keeltesse kasutatud erinevaid strateegiaid. Näiteks on *bypass* tsitaatsõnana aktsepteeritud saksa, hollandi ja skandinaavia keeltes, samas prantsuse keeles on oma tõlkevaste *pontage*, poola keeles öeldakse *pomostowanie* ja vene keeles on eelistatud hoopis teist anglitsismi *шунтирование*. Eesti keeles on kasutatakse tegelikkuses nii sõnu *šunt* kui ka *bypass*, mis kajastab ilmselt varsemat seost venekeelse meditsiiniga ja praegust inglise keele mõjuväljas olemist. **Omaette valdkonna moodustavad ka inglisekeelsete lühendite ülevõtmine**, eriti keeltes, kus akronüümide kasutamine ei ole kuigi levinud (nagu eesti keeles). Ka seda probleemi on eri keeltes lahendatud erinevalt, nii kasutatakse nii skandinaavia keeltes ja ka eesti keeles lühendeid AIDS või DNA, mis prantsuse ja vene keeles on vastavalt SIDA ja СПИД ning ADN ja ДНК.

Eesti ja inglise keele vahekorras eesti meditsiinkeeles on samuti kuigivõrd diskuteeritud. Näiteks on Väino Sinisalu 2005. a „Eesti Arsti“ artiklis murelikult küsinud, kas eesti keel säilib teaduskeelena. Olles nõus toleaege arstiteaduskonna dekaani professor Toomas Asseri väitega, et inglise keele pealetungi ei pea kartma, vaid tuleb lihtsalt mõista maailmas toimuvat ning sellega kaasa minna, rõhutab ta, et samal ajal tuleb säilitada ja arendada eesti meditsiinkeelt, anda välja teadusajakirja ja kirjastada meditsiiniraamatuid eesti keeles. Kuna tõhusa arstiabi teeb eelkõige võimalikuks arsti ja patsiendi vaheline suhtlus, siis peab olema võimalik ka patsiendiga tema emakeeles suhelda (Sinisalu 2005).

Kokkuvõtteks võib öelda, et inglise keele muutumisega meditsiini *lingua franca*ks kaasneb järgmisi põhimõttelisi (sh majanduslikke) probleeme:

- teoreetiliselt on inglise keelt emakeelena kõnelejalatel meditsiinis teatav eelis nende ees, kellele see on võõrkeel (Berghammer 2008);

- mitte-inglisekeelsed riigid peavad lahendama kõik kulud, mis on seotud inglise kui meditsiinikeele õpetamise ja tõlkimisega (Berghammer 2008) ning tegelema keelekorralduslike küsimustega (Nyman 1999, 2000).

### 2.1.3 Eesti meditsiinikeele teke ja areng

Eesti meditsiinikeel on üsna noor, selle alguseks võib tõenäoliselt pidada Friedrich Reinhold Kreutzwaldi tööd, eelkõige tema 1879. aastal ilmunud „Kodutohtrit“ (Kreutzwald 1900). Suur meditsiiniteadmiste populariseerija ja Eesti arstkonna ühendaja oli Peeter Hellat, kelle „Terwise õpetuse“ esmatrükid ilmusid Peterburis ajavahemikus 1891–1894. Aktiivne eesti keeles kirjutaja oli veel Juhan Luiga (nt „Langetõbi“ 1904, „Närwinõrkus“ 1908, „Liikwa joomine ja müümine maal“ 1912).

Soomes olid „Kodutohtri“ ilmumise ajaks juba ilmunud mõned soomekeelsed meditsiinitekstid, sh esimesena 1788. a Christfrid Gananderi „Maamiehen Houne- ja Koti-Aptheeki eli Tieto ja neuwwo tarpeellisimmista Lääkityxistä”. Viiskümmend aastat hiljem ilmus Elias Lönnroti „Suomalaisen talonpojan Koti-Lääkäri“, mille väljaandmine tõenäoliselt on mõjutatud Kreutzwaldi varem ilmunud teosest. Vanim ainult soome meditsiinikeele sõnavara käsitlev teos on Johan Fredrik Ticklénin väitekirj (1832. a), kus oli esitatud 500 soomekeelset vastet ladina terminitele. Meditsiinis kaitses soome keeles esimesena väitekirja Edvard Winter 1860. aastal („Vaimon synnytimestä ja raskauden tilasta“). Kõige esimesed soomekeelsetest arstiteaduslikud sõnad on kirjas 1692. a kaitstud Henrik Hasselqvisti väitekirjas „*Dissertatio Medica de Ossibus humanis*“ ning osa neist on siiani ka kasutusel (nt Polwi Lumbio ja Curcun Cansi) (Haarala 1999).

Suhteliselt tüslik on otsida andmeid eesti keeles kaitstud meditsiinidoktori väitekirjade kohta, kuid kataloogiotsingute tulemusel tundub, et esimesed eestikeelsed väitekirjad paigutuvad aastasse 1923. Sellel aastal esitasid oma väitekirjad meditsiinidoktori kraadi saamiseks Aadu Lüüs („Sugemed vigaste laste kohta Eestis: Tartu Ülikooli arstiteaduskonnale esitatud väitekirja arstiteaduse doktori astme omandamiseks. *Les enfants infirmes en Esthonie.*), kelle väitekirja kokkuvõte ja osa teksti olid prantsusekeelsed; Peeter Hansen („Katselised uurimised ohatise viruse üle: väitekirja arstiteaduse doktori auastme omandamiseks“); Juhan Vilms („Kõhreglükogeeni püsivusest mõnesuguste glükogeeni vähendavate tegurite puhul: eksperimentaalne uurimus 6 joonisega kahel tabelil ja 1 tabeliga katsete ülevaateks: väitekirja arstiteaduse-doktori auastme omandamiseks“) ja Feliks Lepp („Pleuravedeliku bakterioloogiline ja eksperimentaalne uurimine: väitekirja arstiteaduse doktori aukraadi omandamiseks“). Väitekirjade pealkirjad on esitatud selleks, et paremini edasi anda juba tollaegse meditsiinikeele märkimisväärsed ja kiiret arengut: „Kodutohtri“ ilmumisest ei olnud ju kuigi palju aega möödas, kuid keelekorralduslik mõju eesti keelele on juba näha.

Eesti meditsiinikeelt on kahtlemata väga tugevalt mõjutanud saksa ja vene keel ning need mõjud on käinud käsikäes riigikordade ning õpetuskeelte vahetumisega Tartu Ülikoolis. Eesti arstiteaduses on olnud ajajärke, kus õpetatud on ainult saksa keeles, ainult vene keeles, ainult eesti keeles ja eesti ning vene keeles korraga. Neist kõige lühem on siiani tegelikult olnud aeg, mil arste on õpetatud eesti keeles. Õigupoolest Eesti meditsiinikeele arenguloos tuleks lisaks muile Euroopa meditsiinikeele teket mõjutanud perioodidele (kreeka, ladina, riigikeelte, inglise keele perioodid) lisada eraldi veel saksa ja vene keele perioodid, mille mõju eesti meditsiinikeelele on olnud vahetu. Eripäraseks teeb asja seegi, et õpetus ja teadustöö nendes kahes keeles on üleminekuperioodidel käinud Eestis sageli ka paralleelselt.

Omakeelse suhtluse tekkele meditsiiniringkondades aitas kaasa erialaühenduste teke.

Esimesena neist registreeriti 1912. a Tallinnas Põhja-Balti Arstide Selts, mille registreerimise üks tingimusi oli, et nimes ei tohi esineda sõna „Eesti“. Selts arendas viljakat tegevust ning andis välja ajakirja „Tervis“. Seltsi juhiks sai Henrik Koppel, kes on juhendanud ka üht varasematest eestikeelsetest meditsiinidoktori väitekirjadest, 1924. a Tartu Ülikoolis Ernst Saareste kaitstud tööd „Kõrvahaigused ja kuulmisrikked Tartu koolilastel: väitekiri arstiteaduse doktori aukraadi omandamiseks“. 1918. a mais asutati Tartu Eesti Arstide Selts ja 1919. a Tallinna Eesti Arstide Selts. Samal ajal tegutsesid Tartus ja Tallinnas ka saksa, vene ja juudi arstide seltsid. Eesti Arstide Liidu eelkäija Eesti Arstiseltside Liit asutati 28. veebruaril 1921. aastal.

Väga suure töö eestikeelse meditsiiniterminoloogia korrastamisel tegid ära Albert Valdes ja Johannes Voldemar Veski. Nii ilmus nende koostööna juba 1924. a „Kogu eestikeelseid arstiteaduslisi oskussõnu“, 1929. a Albert Valdese sulest aga „Eestikeelseid anatoomilisi oskussõnu“ ajakirja „Eesti Arst“ osana.

Ajaloolist arengut ja ka meie ressursse arvestades on mõistetav, et eestikeelsed arstiteaduslikud õppematerjalid katavad (õigupoolest on alati katnud) õpitavat vaid osaliselt ning väga palju tuleb õppida muukeelsete õpikute vahendusel, viimastel aastakümnetel eelkõige inglisekeelsete materjalide järgi. Inglise meditsiinkeel erineb siiski oluliselt saksa ja vene omast, seega on sellel tuntavaid erisusi ka eesti meditsiinkeelega.

Näiteks ei ole inglise meditsiinkeeles ladina keele kasutamine kaugeltki nii ulatuslik kui Eestis. Seetõttu on uuritud, kui vajalik on üldse ladina keele õppimine meditsiini õppimisel. Uuring viidi 2003./2004. aasta sügissemestril läbi arsti- ja kehakultuuriteaduskonna ning

veterinaariatudengite seas. Selles osales 114 tudengit (68% arstiteaduskonnast) ning uuring näitas, et tudengite suhtumine ladina keele õppesse oli pigem positiivne ning kolmveerand usutletuist leidis, et ladina keele kursus peab arstiteaduskonnas olema kohustuslik aine ning üle poole nägid ladina keelt koguni rahvusvahelise meditsiinikeelena (Rein 2004).

Tuleb tõdeda, et kõik eelnev käsitleb meditsiinikeelt õpetus-, teadus- ja valdkonna üldkeelena, samas autentse tööolukorras kasutatavat meditsiinikeelt ei ole Eestis süstemaatiliselt uuritudki. Ka mujal maailmas alles hakatakse teemaga tegelema. See on iseenesest paradoksaalne olukord, sest viimasel paaril aastasajal on vähemalt Euroopas haiguslugude näol kirjutatud palju kordi rohkem autentset meditsiiniteksti, kui seda on ilmunud teadusväljaannetes või raamatutena. Juba see asjaolu õigustaks autentse meditsiiniteksti uurimist, kuid praegu on autentsete tekstide vastu tekkinud huvi ajendatud eelkõige praktilisest vajadusest: terviseandmete elektroonses loomisest ja käsitlemisest ning üha sagedamini tekkivast vajadusest tõlkida olemasolevaid terviseandmeid teistesse keeltesse.

## **2.2 Kommunikatsioon meditsiinis, meditsiinikeel ja e-tervise ajastu**

Meditsiinikeel erineb tavakeelest nii spetsiifilise sõnavara kui ka kommunikatsioonisituatsiooni eripära poolest. **Meditsiinikeelt võib vaadelda allkeelena (*sublanguage*)**. Tasub meeles pidada, et see termin ei kajasta algselt Zellig Harrise poolt kasutusele võetuna hoopiski üld- ja allkeele vahelist hüperonüümia suhet, vaid väljendab matemaatilise allsüsteemidele (*subsystems*) iseloomulikku terviku ja osa suhet (Somers 2003: 285). Kittredge ja Lehrbereri (1982) definitsiooni järgi saab allkeele mõistet kasutada nende lausekogumite kohta, mille leksikaalsed ja grammatilised piirangud peegeldavad piiratud hulka objekte ja nendevahelisi seoseid ning mis on (iseloomulikul viisil) kasutusel diskursuse teatud domeenis. Määratluse rõhuasetus on sellel, et allkeelt kasutatakse diskursuse mingis

ühes kindlas domeenis. Seega tekib allkeel domeeni spetsialistide omavahelise suhtluse käigus. **Allkeele sees tekib mitte ainult võhikule arusaamatu terminoloogia, vaid ka tavakeele sõnad omandavad hoopis uusi konnotatiivseid ja kontekstuaalseid tähendusi, mis peegeldavad domeenis olevaid kontseptuaalseid käsitlusi** (Somers 2003: 284).

Meditsiinkeelt mõjutavad tugevalt kommunikatsiooniakti eesmärk ja selle osapooled: arsti ja patsiendi vahelises suhtluses kasutatav keel erineb märkimisväärselt arstide omavahelises suhtluses kasutatavast, arsti vastuvõtul kasutatav väljenduslaad haigusloo teksti omast, kuigi edastatav teave võib sageli olemuslikult olla enam-vähem sama.

Oluline vahe on ka kirjaliku teksti ning suulise kõne vahel. Üks meditsiinkeele suulise kõne variantidest esindabki pigem kirjalikku keelt – nimelt dikteeritud meditsiinitekstid (nt operatsiooniprotokollid, diagnostikaprotseduuride kirjeldused, patoloogi lahanguprotokollid jne), mis on sageli algselt suulised, ent järgivad ikkagi kirjaliku teksti eesmärki ehk neis esineb tavalisele suulisele kõnele ebaloomulik struktureeritus ja süstematiseeritud sõnakasutus.

Käesolevas töös käsitletakse ainult tööolukorras toimivas arstide vahelises teabevahetuses kasutatavat kirjalikku meditsiinkeelt, sest need tekstid erinevalt suulisest kõnест on kättesaadavad ning enamik terviseandmeid luuakse Eestis kohe kirjalikena.



### 2.2.1 Terviseandmed, meditsiinkeel ja e-tervis

Omaette teemaks võiks pidada seda, et avaldatud, st teaduse ja hariduse, meditsiinkeel tervikuna on palju vähem allkeel kui tööolukorras kasutatav autentne keel, sest teaduse ja haridusega seotud meditsiinkeel on õigupoolest suunatud suhtlusele teiste diskursuse domeenidega ehk osaliselt meditsiinisüsteemist väljapoole. **Autentsed meditsiinitekstdid on tänini peaaesjalikult rangelt spetsialistide vahelise suhtluse vahendid**, kuigi ka siin on märgata teatavaid nihkeid. Juba kas või see, et elektroonsed terviseandmed Eestis on tehtud patsiendile nähtavaks, on samm selles suunas, et keel peab muutuma arusaadavamaks ka teistes valdkondades tegutsevatele inimestele. Eesti e-tervise kontseptsioon on tugevalt patsiendikeskne (<http://www.e-tervis.ee/>) ning eeldab ka patsiendi aktiivset osalust oma tervise eest vastutamisel ja terviseandmete kasutamisel.

Meditsiinkeele areng on viimasel paaril aastakümnel olnud tihedalt seotud vajadusega talletada, edastada ja analüüsida (patsiendi) terviseteavet elektroonselt. Patsiendi iga haigusjuhu või arstile pöördumisega seotud andmed salvestatakse nagu ennegi tema haigusloos, kuid lugu ise on elektroonne. Nii nagu paber kandjal sõnastike muutumisel elektroonseteks on elektroonne terviseteave samuti avanud täiesti uusi võimalusi: saab käsitleda suuri andmehulki, võimalik on tuletada andmete vahelisi seoseid, võrrelda neid teiste andmetega, vaadelda andmeid ajas jne.

Elektroonse ajastu algus meditsiinis on saanud suure – kui mitte peamise – tõuke majandusest ning ettevõtlusest. Haiglate infosüsteemid tekkisid pigem vajadusest pidada arvet tulude ja kulude üle kui otseselt patsiendi haiguskulu jälgimise vajadusest. Teiseks haiglate infosüsteemide otsustavaks arenguteguriks oli kahtlemata info- ja meditsiinitehnika kiire areng, mis oli käsitletavate andmehulkade suhtes murranguline. Seda eriti

diagnostikaerialadel, nt radioloogias ja laborimeditiinis. Radioloogias tekkis võimalus talletada ja analüüsida digitaalset kujutist ning laborimeditiinis koondus üha enam uuringuid automaatanalüsaatoritele, mille tõttu kasvas plahvatuslikult süsteemi läbilaskevõime tervikuna. Selleks, et tohutut andmehulka tõhusalt käsitleda (nt leida kõigi andmete seast (kiiresti) vajalik teave või vajadusel tõlkida olemasolevaid andmeid), **on vaja nii andmete kui ka nendevaheliste seoste standardset esitust, teatud reegleid andmebaasi haldamiseks ning asjakohaseid leksikone.** Iseenesest see pool terviseandmete käsitlest ei erinegi nii väga leksikaalsete andmebaasidega seotud probleemidest (Orav 2006). Standarditult piiratud hulga andmete esitus tekitab aga kohe ka küsimuse nende automaatse käsitlese võimalikkusest. Üheks piiratud ja piiritletud ning standarditud andmehulgast tulenevaks võimaluseks on masintõlge (Somers 2003: 288–294).

### **2.2.2 Radioloogi vastus: meditsiiniline narratiiv ja struktureeritud tekst**

Meditsiiniandmed võivad olla esitatud arvudena, kujutistena, struktureeritud või vabatekstina. Arvude korral on salvestamise ja andmebaasi ülesehitus kõige lihtsam: kindlatele andmeväljadele sisestatakse teatud arvvärtusi. Kui tegu poleks nii tohutu andmehulgaga, mis vajab ka teatud kontrollimehhanisme, siis võiks ju põhimõtteliselt läbi ajada ka lihtsalt Excel-tüüpi tabelarvutusprogrammiga. Näiteks on relevantsses kontekstis kliinilist tähendust omavad arvvärtused mitmed laboriuuringute tulemused, patsiendi hingamissagedus, pulss, vererõhu näit, antropomeetrilised andmed, vanus jpm. Arvandmete puhul esineb suhteliselt harva olukorda, kus neid peaks leiuna eraldi tekstina kirjeldama või interpreteerima, tavalahendustes näidatakse ära asjakohased referentsvärtused ning nende suhtes otsustab raviarst ise, kas konkreetse patsiendi näit jääb normaalvariatsiooni piiridesse.

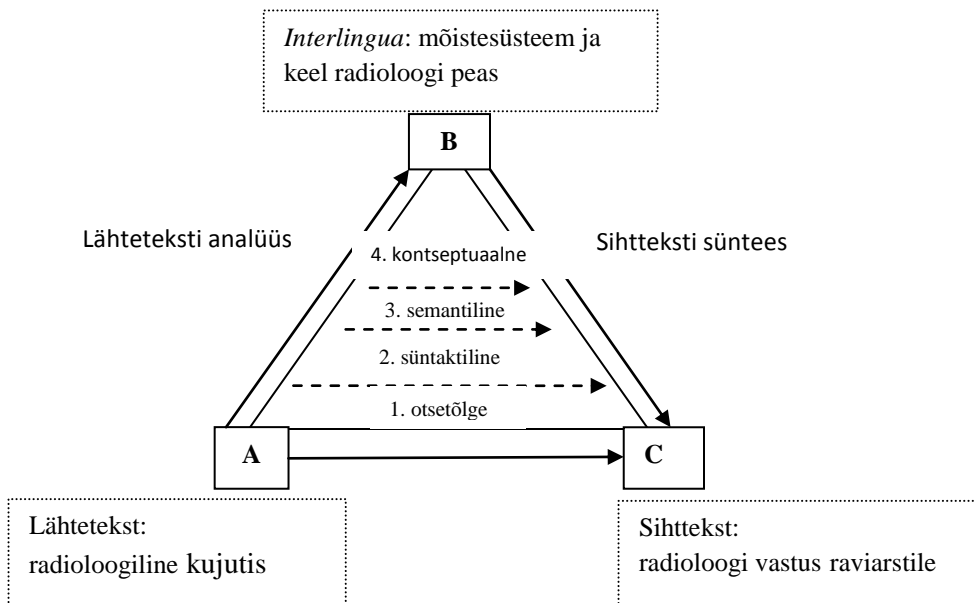
Kujutised võivad olla radioloogilised pildid (nt röntgenogrammid), mitmesugused spektrogrammid, elektrokardiograafia väljund, raku- või koepreparaatide fotod jpm. Tavaliselt kaasneb kujutistega see probleem, et nende mõistmine nõuab erikvalifikatsiooni. Sisuliselt võibki radioloogi pidada tõlgiks, kes vahendab raviarstile kujutise sisu ja tähenduse talle arusaadavas (sõna)keeles. Kui mõelda sellest situatsioonist kui omalaadsest tõlkemudelist, siis võiks seda olukorda kujutada Vaquois' kolmnurgast inspireeritud joonisega (vt joonis 1A), lihtsalt lähteteksti esimene tasand ei ole sõna tasand, vaid pilt.

Joonisel 1A on näha, et raviarst võib ka ise kujutist tõlgendada, mis keelelises tõlkesituatsioonis vastab olukorrale, kus lugeja valdab lähteteksti keelt. Üldjuhul soovitakse siiski kujutise tõlgendamiseks spetsialisti abi ehk radioloog annab oma kirjalikus vastuses kujutise kirjelduse ja selle tõlgenduse.

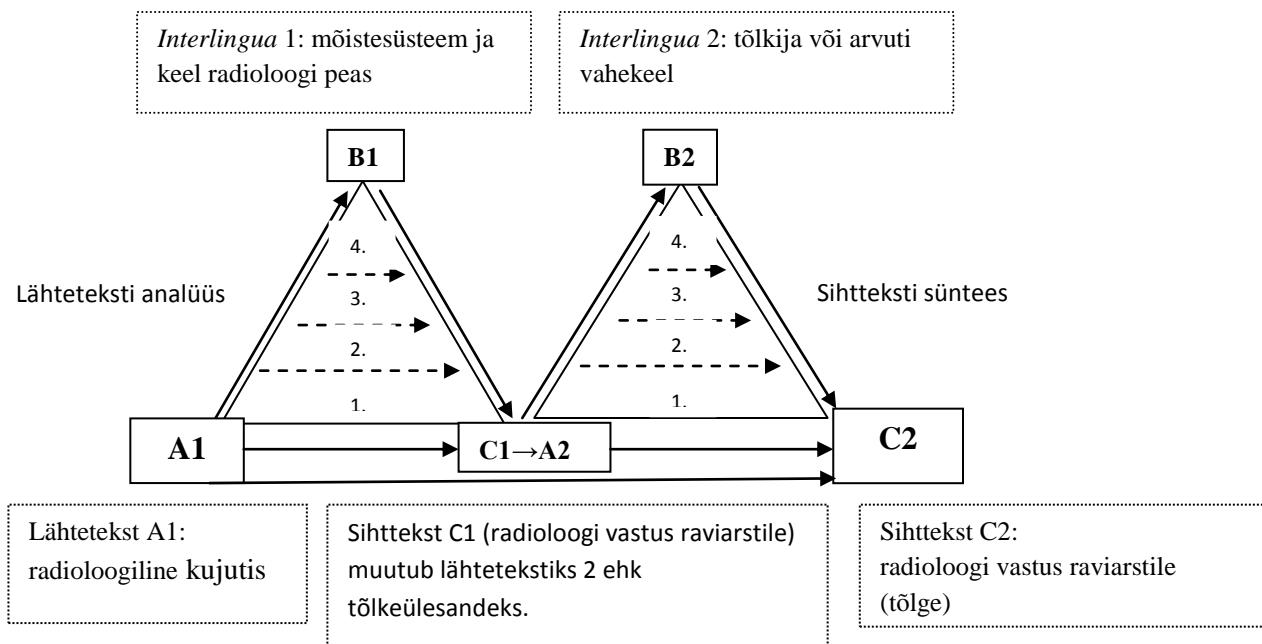
Joonis 1A ja 1B selgitavad, miks on väga harv see olukord, kus samast kujutisest on olemas mitmekeelseid vastuseid – lihtsalt alguses puudub enamasti vajadus sama kujutist teises keeles kirjeldada ja juhul, kui seda tehakse, siis ei pruugi seda enam teha sama radioloog, mis omakorda lisab teatavat variatsiooni. Praeguseks ei ole olemasolevaid kirjelduste tõlketekste olnud nii palju, et saaks mõelda representatiivse paralleelcorpuse koostamisest (vt jaotis 2.3.1 „Keeletehnoloogiaressursid“), kus sama pilt ja/või kirjelduse tekst oleks tõlgitud samaaegselt mitmesse keelde, mida saaks kasutada tõlkealgoritmide väljatöötamiseks.

**Joonis 1. Radioloogiline kujutis ja selle tõlge → radioloogi vastuse tõlge**

**A. Radioloogi vastus kui kujutise tõlge**



**B. Radioloogi vastuse tõlge**



Pildiagnostikas on viimastel aastakümnetel toimunud tohutu arenguhüpe: radioloogiatehnika võimaldab peale konventsionaalse röntgenkujutise nt ülitäpseid metabolismi uuringuid nuklearmeditsiinis, kujutised on digitaalsed ning radioloog võib neid kirjeldada mis tahes arvutist. Kuid **radioloogi vastus ja selle struktuur on sajandi vältel iseenesest üllatavalt vähe muutunud** (Reiner 2009). Esimene radioloogiline kirjeldus pärineb 1896. aastast Wilhelm Röntgenilt. Põhimõtteliselt kujutab radioloogi vastus endast tänini teatud žanrinõuetele vastavat proosateksti ning selliseks võib see lõppväljundis (vastuses raviarstile) ka jääda, iseasi kas tekst andmebaasis (andmete sisestamisel) peab tingimata olema proosavormis vabatekst. **Selleks, et raviarst vastust üheselt mõistaks peab selle sõnum (tekst) olema selge, tõetruu, usaldusväärne, lühike, täielik, järjepidev, teavet vahendav ja nõustav, õigeaegne ja standarditud** (Reiner 2009). Ei ole kahtlust, et esimese kaheksa nõude täitmine on iga kvalifitseeritud spetsialisti tunnus, kuid ajagraafikus püsimise ning standardse kirjeldusviisiga tekib vabateksti kasutamisel paratamatult probleeme. **Vabatekstis kirjeldamine takistab andmete standardset esitust, mis omakorda on tõkkeks standardsele andmeesitusviisile ning hiljem seega nii andmekaeve kui ka tõlkerakendustele.**

Elektroonsed andmebaasid kätkevad endas palju enam võimalusi kui paberkandja. Nii nagu paberkandjal sõnaraamat on palju väiksema funktsionaalsusega kui elektroonne, kus saame sõnu otsida mitte ainult tähestikulise järjekorra või valdkonna järgi, vaid teha päringuid ka muudest aspektidest lähtudes (märksõna ise, grammatiline info, seletus, kollokatsioonid, näitelaused jne). Selleks peab info olema arvutile formaalselt identifitseeritav ehk asjakohaselt liigendatud ja märgendatud. Tänu liigendusele on sõnastiku esitatud materjal ka arvuti poolt „loetav“ ning mitte lihtsalt paberkandjal raamatu asendaja (Orav 2006). Täpselt sama peab

paika ka teabe kohta, mis on radioloogilises kirjelduses: võime esitada andmed paber kandja koopiana või liigendada nad selliselt, et see võimaldaks edaspidi uusi rakendusi, sh põhjalikumat ja mitmesuguseid analüüse võimaldavat andmekäivet (nt mitte lihtsalt teatud kuupäeva uuringu kirjelduse täistekst, vaid nt hoopis teatud ajavahemikust kõigi samalaadsete leidudega patsientide diagnoosid vms). Ka andmevahetuse seisukohast oleks ühetaolise liigenduse ning standarditud (teatud moel kodeeritud) leksikonide kasutamine asjakohane. Viimaseid on meditsiinis ka arendatud, nt SNOMED-CT (*Systematized Nomenclature of Medicine-Clinical Terms*, vt <http://www.ihtsdo.org/snomed-ct/>) või UMLS (*Unified Medical Language System*, vt <http://www.nlm.nih.gov/research/umls/>).

Teatud mõistepõhised süstematiseeritud leksikonid on seega meditsiinis juba kasutusel. Miks siis vabateksti asendumine liigendatud andmeesitusega läheb nii vaevaliselt? Tuleb arvestada, et arsti teadmised ja oskused põhinevad lisaks õpitud meditsiiniteadmistele ka oma kogemusel. Viimane tekib paljude seoste analüüsimisel alles aastate jooksul. Just selle osa struktureerituna esitamine on äärmiselt keeruline. Optimaalne andmete struktuur on lihtsalt suur probleem: mida keerulisem on kirjeldatav olukord, seda rohkem on varieeruvaid andmeid, mida on üha keerulisem liigendada. Ka võimaldab loomulik keel sageli paremini rõhutada ja tõlgendada fakte kategoriseerides neid raskusastme järgi, kirjeldades kausatiivseid seoseid, esitades hüpoteese või pakkudes alternatiivseid tõlgendusi. **Samuti tuleb arvestada, et vastuse loomise ajal võib etteantud struktuur erineda oluliselt viisist kuidas tõlgendav arst haigusjuhtu tervikuna käsitleb ja seega mõjutada tema mõtlemist.** (Seda kutsutakse inglise keeles kaunis tabavalt „*reporting into structure*“ ehk siis vastamine liigenduse järgi. **Täpselt samamoodi võib liigendus ja vastuse kuju mõjutada selle lugeja mõtlemist haigusjuhust tervikuna.** (Seda kutsutakse inglise keeles „*reading structure*“ ehk siis

lugemine liigenduse järgi.) See rõhutab sisulise liigenduse erakordset olulisust tervikpildi säilitamise seisukohast. Põhimõtteliselt on väljundi suhtes kindlasti parem, kui see on kujul, millisel seda ollakse harjutud lugema (Sistrom 2005). See aga tähendab, et **uute rakenduste loomisel tuleks pragmaatika seisukohast kindlasti pöörata oluliselt suuremat tähelepanu keelele (morfoloogiale, süntaksile, semantikale), milles vastused on kirjutatud.** Seni pole seda aspekti eesti keeles kuigi palju uuritudki. Esimene taoline katse on olnud seoses tõlkemudeli loomisega projekti Baltic eHealth käigus. Projekti sisuks oli piiriülene radioloogia tervishoiuteenuse allhange, kus radioloogiuuringud tehti Taanis, kuid kujutist kirjeldati Eestis või Leedus. Ühe osana projektist loodi algoritm, milles põlve röntgeniuuringud kirjeldati liigendatud tekstina ning pakuti inglise- ja taanikeelsetele lausetele ka eestikeelsed tõlkevasted. Eestikeelsete tõlkevastete otsimine põhines mitme radioloogi ühisel arvamusel. Eestis koordineeris projekti dr Peeter Ross (Schønnemann 2008; Ross 2008).

2007. a pakkusid Johnson jt võimaliku lahendusena nn **struktureeritud narratiivi mudeli** (Johnson 2008), kus meditsiiniandmed olid liigendatud nii, et põhimõtteline n-ö suur tekstiliigendus (jaotised, paragrahvid, lõigud, lisad) märgendati XML-keeles (*Extensible Markup Language*) ja selles märgendati ka teksti väiksemad osad (tähendused, seosed, kategoriseerivad intensifikaatorid). Tegu oli küll olemuslikult juba varem kasutatud osalise liigendusega süsteemiga, mis sundis arste kirjutama teatud struktuuris (suur liigendus), kuid nendele andmeväljadele võis kirjutada vabateksti (ingl *semi-structured*). Erinevus seisnes selles, et ka vabateksti osa märgendati kasutades NLP-vahendeid (loomuliku keele töötlemise vahendeid; *Natural Language Processing*) ning saadi tulemusena vabatekst, milles olid identifitseeritavad nii terminitega kirjeldatud faktid (haigusseisundid, ravimid, protseduurid), intensifikaatorid (raskusaste, tõenäosus jne) kui ka faktide vahelised seosed (kausatiivsed,

ajalised jne). Kõik elemendid olid seotud standarditud koodidega, et hõlbustada eri arvutirakenduste kasutamist. Samas lisati ka teatud sõnavara, mis võimaldas kirjeldada seoseid faktide vahel, mis oli oluline konteksti mõistmiseks, tõlgendamiseks ja info sünteesiks (Johnson 2008). **Radioloogilise kirjelduse esitamisel struktureeritud andmetena peab lähtuma sellest, et loodud struktuur võimaldaks kirjelduse teksti osadest kokku panna lugejale harjumuspärasena.** Sisuliselt tähendab see aga lisaks optimaalsele liigendusele ka arvutile mõistetava leksikoni ja mõistepuu loomist, teatud reeglite (algoritmi) kasutamist ning harjumuspärase keele säilitamist vastuses.

Nagu mis tahes muuski majandusharus on meditsiinis kriitilise tähtsusega optimaalne ressursikasutus, mis tähendab radioloogias muu hulgas ka seda, et teenust tuleb suunata sinna, kus on parasjagu vaba tööjõudu (Ross 2008, 2010). See tähendab, et **piiriüleste tervishoiuteenuste allhangete hulk Euroopa Liidus suureneb. Seega muutuvad järjest aktuaalsemaks ka radioloogiliste kirjelduste tõlkimisega seotud küsimused.** Pealegi sagnevad Euroopa Liidus seoses vaba tööjõu ja inimeste liikumisega olukorrad, kus inimesed satuvad haiglatesse või uuringutele ka mujal kui kodumaal või vajavad uuringute kirjeldusi teises riigis esitamiseks mõnes muus keeles. Kui tulevikust globaalselt mõelda, siis ideaalis peaksid liikmesriikide e-Tervise süsteemid mingil hetkel olema ühendatavad suureks terviseandmete võrgustikuks, millest on andmeid võimalik kätte saada ükskõik missuguses liikmesriigi keeles (KOM (2004) 356).

Nagu öeldud meenutab teksti liigendamise ja leksikoni loomise protsess radioloogia andmebaasi tegemisel mõnevõrra ka lingvistilist ülesannet. Kui läheneda sellele keeletehnoloogilisi võimalusi arvestades, siis on piiratud sõnavara ja kaunis täpselt defineeritud mõistete korral võimalik luua asjakohane masintõlget võimaldav rakendus.



## **2.3 Lingvistika, meditsiin ja tõlkimine**

Veel mõnikümmend aastat tagasi ei nähtud Eestis lingvistika ja meditsiini vahel muud seost kui see, et meditsiini teatud valdkondadel, nt anatoomial, füsioloogial, kirurgial ja stomatoloogial oli roll kõneaparaadi kirjeldamises, kõnedefektide määratlemises ning kõneaparaadi ravis. Nii on Eduard Vääri kirjutanud kirjeldades lingvistika seost teiste erialadega järgmist: „Lingvistikal on sidemeid ka meditsiini e arstiteadusega, mis on kompleksteadus inimese tervise kaitsest ja tugevdamisest, haigustest, nende diagnoosimisest, vältimisest ja ravist. Kokkupuuted piirduvad häälde- ja kuulmiselundite ehituse, talitluse ja raviga – seega lingvistikasse kuuluva foneetikaga.“ (Vääri 1983). Teistpidi – seda, kuidas võiks lingvistika võimalusi kasutada meditsiini kui eriala edendamiseks, ei küsitutki.

Lingvistika areng viimastel aastakümnetel on olnud sama kiire ja murranguline kui meditsiinilgi. Huvitaval kombel on mõlema eriala korral suur arengupotentsiaal kätkevad andmete elektroonse töötamise ja analüüsi võimaluste murrangulises paranemises, andmebaaside kui oluliste ressursside ning arvutitehnika ja tarkvara kui vajalike vahendite arengus.

### **2.3.1 Keeletehnoloogiaressursid**

Eesti keele arendamise strateegias on keeletehnoloogia kui tegevusvaldkonna ülesandeks eesti keeletehnoloogilise toe arendamine sellisele tasemele, et eesti keel oleks võimeline funktsioneerima tänapäeva infoühiskonnas, ühena Euroopa Liidu keeltest (Muischnek 2003: 7). Juba praegu on olemas eesti keele morfoloogiline analüsaator ja ühestaja, suur korpuseks organiseeritud tekstide kogum ning tohutu hulk praktilist oskusteavet, kuidas suuri tekstihulki analüüsida. Seoses e-tervise arenguga on haiglate andmebaasides kogunenud suur hulk autentset meditsiiniteksti, mida saaks isikustamatult kasutada väärtusliku keeleressursi –

meditsiintekstide korpuse loomiseks, millest võiks olla kasu peaaesjalikult andmekaeve ja tõlkeprobleemide lahendamisel.

### **2.3.1.1 Tekstikogu ja korpus**

Keelekorpus on kirjaliku või suulise kõne kogum. Suvaliste tekstide kogumist e tekstiarhiivist eristab korpust see, et tekstid on valitud eesmärgipäraselt selliselt, et nad annaksid tõepärase pildi uuritavast keelest. Seega on korpus loomuliku keele tekstide kogu, mis on koostatud iseloomustamiseks keele hetkeseisu või muutumist (Muischnek 2003: 9). Arvutiajastul on korpus tavaliselt polüfunktsionaalne elektroonsel kujul olev tekstikogu. Korpus on kindlate kriteeriumide järgi valitud ja arvutisse ühtses elektroonses formaadis viidud tekstide kogum.

**Lihtsalt tekstikoguna talletatakse tekste seepärast, et igaüks neist omaette on väärtus ning nende hulgast võib vastavalt vajadusele valida selle teksti, mis huvi pakub.** Sisuliselt on haiglate andmebaasid praegu suured tekstikogud, mille moodustavad patsientide haigusjuhte kirjeldavad haiguslood. **Korpuse tekstid on aga väärtuslikud just kõik koos** (Muischnek 2003: 9). Korpuse tekstide valikul on vaja arvesse võtta palju aspekte, nt millise valdkonna ja/või ajajärgu tekstid valida, kas vaja tervik- või osatekste, kui palju peab arvestama tekstide saamisega seotud organisatoorseid ja tehnoloogilisi probleeme jne. Oluline on ka korpuse representatiivsus keelevariandi suhtes, mida ta peab esindama (Hennoste 2000).

Korpust saab kasutada:

- keele uurimisel e lingvistikas – analüüsides keelt, et kindlaks teha tema omadusi;
- sõnaraamatute koostamisel;
- keeletöötlusvahendite väljatöötamisel, treenimisel või testimisel arvutilingvistikas ja keeletehnoloogias (Kaalep 1998; Muischnek 2003).

**Korpusest on kasu ainult siis, kui vajaliku info saab sealt suhteliselt lihtsalt ja kiiresti kätte.** Selleks peab sageli alustama info lisamisest korpusesse. Et korpus ei oleks pelgalt elektroonne tekstiarhiiv, tuleb tekstidele lisada info nende ülesehituse kohta (peatükid, pealkirjad, lõigud, laused jne) ning morfoloogilise ja süntaktilise analüüsi andmed. Info lisamist korpusesse nimetatakse korpuse märgendamiseks. Märgendada saab automaatselt, käsitsi või neid kahte meetodit kombineerides (Muischnek 2003).

Eesti kirjakeele korpus (vt <http://www.cl.ut.ee>) on väga väärtuslik keeleressurss, kuid paraku ei pruugi see olla piisav lahendamaks kõiki allkeelte uurimisega või rakendustega seotud probleeme. Korpus sisaldab suurel hulgal ajakirjanduslikke tekste ja kuigivõrd ilukirjanduslikke, on olemas väike vana kirjakeele ning ka murrete korpus, tegemisel on dialoogikorpus ning alustatud on ka veebi jututubades kasutatava keele kogumist.

Paralleelkorpustena on olemas Euroopa Liidu õigusloome tekstid. Samuti tuleb arvestada, et eesti keele morfoloogiline analüsaator ja ühestaja on treenitud olemasoleva korpuse materjali peal, mis ei sisalda autentseid meditsiintekste. **Küsimus on seega selles, et eesti keele korpuses olevad tekstid ei ole autentsete meditsiintekstide suhtes representatiivsed, sest kirjalike meditsiintekstide eripära *a priori* sisaldab palju sellist, mida tavakeeles üldse ei kohta** (nt eripärane leksikon) või vähemalt sellisel hulgal ei kohta (nt nimisõnafraasid lausete asemel, eituse ja ladinakeelsete tsitaatsõnade sage kasutus, sage ja ebakonventsionaalne lühendamine jms).

Põhimõtteliselt on isegi see, kas ja kui palju erineb autentne, s.o tööolukorras kasutatav, meditsiinikeel tavakeelest või avaldatud teadustekstidest (nt „Eesti Arstis“ või väitekirjades) ebaselge ja vajaks kindlasti eraldi uurimist. Seda enam, et inglise ja saksa autentse meditsiinikeele uurimisel on jõutud üsna vastandlikule tulemusele: kui sakslaste uurimuses sai

saksa ajalehetekstide peal treenitud morfoloogiline ühestaja ka meditsiintekstis morfoloogilise analüüsi ja sõnaliikide määramisega hakkama (Wermter 2004), siis inglisekeelse meditsiinteksti puhul oli olukord sootuks vastupidine, mida uurijad ise seostasid eelkõige nimisõnafaaside rohkusega tekstis verbifraasidega võrreldes (Campbell 2001). Arvestades, et eesti kirjakeele korpus ei sisalda autentseid meditsiintekste (patoloogi või radioloogi kirjeldust, operatsiooniprotokolli, epikriisi vms), on raske oletada, et selles sisalduv materjal võiks meditsiinikeelt igakülgselt kajastada. Iseasi on muidugi see, et eesti keele morfoloogia ja süntaksi reeglid peaksid kirjalikus meditsiintekstis teatud mööndustega ikkagi kehtima.

### **2.3.1.2 Tõlketeadus, tõlkimine ja korpused**

Korpuspõhise analüüsi vaieldamatu eelis on, et ta kajastab tegeliku keelekasutuse reegleid. Korpusete kasutamine tõlketeaduses ja ka tõlkimisega seotud rakenduslikel eesmärkidel on uus suund. Tõlketeaduses eelnes kaks olulist muutust: arusaam alast nihkus preskriptiivselt käsitluselt deskriptiivsele ning mikro- (s.o keeletasandiga) piiratult makrotasandile (sotsiokultuurilisele). Tulemusena on ka tõlketeaduse fookus nihkunud lähte- ja sihtteksti seoste käsitluselt tõlgetele endile. 1991. a käis R. Bell välja idee tõlketekstide kogu loomisest, et süstematiseeritult uurida tõlkekeelt. See idee äratas kohe suurt huvi ning hakati koostama tõlketekstide korpusi eesmärgiga analüüsida tõlkekeele kaudu tõlke kui lõppsaaduse ja tõlkimise kui protsessi vahelist seost. P. Bakeri arvates on tähtsaim ülesanne korpuslingvistika kasutamisel tõlketeaduses selgitada tõlketeksti kui vahendatud kommunikatsioonisündmuse olemus (Xiao 2009). Selle valdkonnaga on seotud ka tõlkimise õpetamine ja selle sisu, kõikvõimalike tõlkeabivahendite (masintõlge, tõlkemälud, leksikonid ja terminikogud) loomine, tõlke kvaliteedi hindamine jpm.

Tõlketeaduses kasutatakse

- ükskeelseid vs. mitmekeelseid korpusi;
- paralleel- vs. võrreldavaid korpusi;
- võrreldavaid vs. võrdlevaid korpusi;
- üldkeele vs. spetsialiseeritud korpusi.

Ükskeelsed korpused (*monolingual corpora*) on algselt mõeldud ühe keele uuringuks ning tõlketeadusliku rakendamise eesmärgil luuakse tavaliselt võrreldava valimi moodustamise alusel (sarnase tekstivaliku- ja hulgaga) kaks allkorpusi, millest üks koosneb selle keele algtekstidest ja teine sellesse keelde (enamasti mingist ühest keelest) tõlgitud tekstidest. Võrreldavate korpusete korral on see tõlkekeele eripära uurimiseks suurepärase lahendus.

Mitmekeelne (enamasti küll kakskeelne) korpus (*bilingual, multilingual corpus*) koosneb mitmes keeles olevatest tekstidest ning võib olla kas paralleelkorpus (*parallel corpus*) või võrreldav korpus (*comparable corpus*). Mitmekeelne võrreldav paralleelkorpus on eriti kasulik keeltevaheliste erinevuste uurimiseks. Paralleelkorpus sisaldab nii lähte- kui ka sihtteksti paralleelselt, võrreldava korpusena on tegu siis, kui selle komponendid või allkorpused on võrreldavad samadel alustel valimi moodustamise tõttu. Paralleelkorpused võivad olla keelte suhtes vaadelduna ühesuunalised (nt sisaldada ainult eestikeelseid lähtetekste ja inglisekeelseid sihttekste või vastupidi), kahesuunalised (sisaldavad nii lähte- ja sihttekste mõlemas keeles) ning mitmesuunalised (st sama teksti tõlked mitmesse keelde). Selles kontekstis on ka Euroopa Liidus samaaegselt mitmesse keelde tõlgitud tekstid sobivad paralleelkorpusena. Paralleelkorpus peab olema teatud tasemel joondatud (nt dokumendi, lõigu, lause või sõna tasemel), selleks et teda saaks kasutada tõlkeuuringutes. Joondamine võib väga erinevate keelte puhul osutuda probleemseks. Ühe keele eri variante sisaldav korpus (nt *International Corpus of English, ICE*) ei saa olla võrreldav tähenduses, et ta valimi koostamise

alused oleksid tõlketekstide kogumiga samad, kuid võrdlevate uuringute tarbeks sobib selline korpus küll (Xiao 2009: 237–242).

Üldkeele korpused (*general corpora*) erinevad spetsialiseeritud korpustest (*specialised corpora*) katvuse poolest ehk selle poolest, mis žanrid või valdkonnad neis on hõlmatud.

Tasakaalustatud üldkeele korpustel on tõlketeaduses kahtlemata suur väärtus. **Kuid spetsialiseeritud korpused on tõlkeuuringutes eriti tähtsad, sest sisaldavad rikkalikult spetsiifilist terminoloogiat ja keelekasutust ning on ka erakordse praktilise kasutusväärtusega tehnika-, sh meditsiinitekstide tõlkijaile.**

Eri tüüpi korpuste kaudu saab otsida vastuseid erinevatele küsimustele: paralleelkorpused on kasulikud mõistmaks, kas ja kuidas vahendati ühes keeles esitatud mõte/tähendus teise keelde. Seda tüüpi korpused on aluseks ka statistilisele või näitepõhisele masintõlkele ning kakskeelsete leksikonide ja tõlkemälude loomiseks. Võrreldavad korpused on kasulikud selleks, et aidata tõlkijal paremini mõista valdkonda ning parendada tõlke soravust, terminivalikut ja keeles kinnistunud või idiomaatiliste väljendite kasutamist. Ka ükskeelsed lähte- või sihtteksti korpused on tõlketeaduse seisukohast väga olulised, sest nad parandavad tõlkija jaoks nii teksti keelelist mõistmist kui ka arusaama selle (kultuurilisest) taustast ning on tõlkijatele heaks võrdlusmaterjaliks. Neid võib kasutada ka kombinatsioonis paralleelkorpusega selleks, et koostada tõlke hindamise korpus.

### **2.3.1.3 Meditsiinkeele korpused**

Meditsiinkeele korpuste kirjeldusi kuigi palju erialakirjanduses ei ole. Helsingi Ülikooli projekti tulemusel on olemas keskaegsete inglisekeelsete tekstide korpus, millele on juba eespool viidatud (Taavitsainen ja Pahta). Samuti on mõnede meditsiiniväljaannete tekstikogud

olemas korpustena, sh nt ka Kaarel Vesikise koostatud Eesti Arsti 2002–2004 aastakäikude korpus (ca. 712 000 sõna, <http://www.cl.ut.ee/korpused/segakorpus/eestiarst/>). Probleem aga on suuremgi, sest meditsiiniajakirja tekstid ei käsitle tööolukorra tegelikku keelekasutust, vaid juba toimetatud ja ka muile kui eriala spetsialistidele mõeldud tekste. Autentse meditsiinikeele korpustest on väga vähe erialast kirjandust.

Õigupoolest on leida ainult üks viide suuremale autentsete meditsiinitekstide annoteeritud (märgendatud ja süntaktilise informatsiooniga) kogule, nimelt on Poolas loodud autentsete meditsiinitekstide korpus (KorMediIS), mida käsitletakse poola keele korpuse (IPI PAN korpuse) ühe allosana (Godlewski 2006). Korpusesse valiti tekstid haiguslugude eri allosadest: esialgne diagnoos, anamnees, objektiivne leid, diagnostiliste uuringute kirjeldused, lõplik diagnoos, ravi ja epikriis, kokku 2 miljonit sõnet. Korpuse loojad on rõhutanud, et autentne meditsiinitekst on keeleliselt väga erinev nii üldkeelest kui ka meditsiinikeelest tervikuna. Peale terminoloogia, mis ehk polegi kuigi suur probleem ja lahendatav asjakohase leksikoniga, esines tekstides väga palju lühendeid, ladinakeelseid fraase, ebaharilikku lauseehitust ning suurel hulgal kirjavigu (Godlewski 2006). Kõik see selgitab, miks selliste tekstide edukaks morfosüntaktiliseks märgendamiseks rakenduslikel eesmärkidel (sh tõlge) peaks asjakohaseid keeletehnoloogiavahendeid treenima just nende tekstide peal.

Praktilistest kaalutlustest lähtudes on vajalik teada autentse keelekasutuse eripära. Ainult nii on võimalik mõelda tulemuslikule tõlkimisele olenemata sellest, kas pidada silmas inim- või masintõlget. **Keelt tundmata ei saa ei tõlkida ega tõlkerakendust luua.**

Praegu ei ole veel olemas sellist valdkonda nagu meditsiiniline lingvistika. Kuid arvestades meditsiinikeele arengut viimastel aastakümnetel, võiks meditsiini-, keele- ja infotehnoloogia

teadmisi ja oskusi ühendav alleriala olemas olla küll. Tööd sellel alal jätkuks Henrik R Wulffi arvates nii ajaloolastele kui ka lingvistidele (Wulff 2004). Lisaksin, et ka arstidele, infotehnoloogidele ja tõlkijatele.



### **3 KASUTATUD MATERJALID JA VAHENDID**

#### **3.1 Uurimisobjekt**

Uurimisobjektiks olid lüüsisamba magnetresonantstomograafia uuringute (MRT-uuringute) kirjeldused. MRT-uuringud olid tehtud AS Ida-Tallinna Keskhaiglas vahemikus 15.11.2007–07.02.2008 ja 17.11.2009–13.01.2010. Isikustamata MRT-uuringute kirjeldused sain juhendajalt dr Peeter Rossilt. Uuringute kirjeldused sisaldasid kõiki MRT-uuringuid, st peale lüüsisamba uuringute ka muude piirkondade omi. Kirjeldused võeti andmebaasist välja kogu ajavahemiku järgi tehtud päringu alusel, neis ei ole mingeid muid andmeid peale uuringu kirjelduse teksti ning kuupäeva. Patsientide isikuandmeid uuringukirjeldused ei sisalda ja teksti põhjal ei ole neid võimalik seostada ühegi konkreetse patsiendiga. Ajaline vahe on seletatav tervishoiuteenuste kodeerimise muutusega vaadeldud perioodil, mis ei võimaldanud andmebaasist vahepealse ajavahemiku kohta koondpäringut teha ja seetõttu ei ole neid kirjeldusi käesolevas töös kasutatud. Lihtsustamise eesmärgil ning selleks, et mõnevõrra parandada olemasoleva tekstimaterjali representatiivsust, on uurimuses keskendutud vaid lüüsisamba MRT-uuringutele. Nimelt on lüüsisamba patoloogia üldiselt suhteliselt hästi piiritletud ning seetõttu võis eeldada põhiliste haigusseisundite radioloogilise pildi teatavat ühetaolisust ning sagedast kirjeldamist. Sellepärast tuli algfailist kõige pealt teha tööfail, mis koosnes ainult lüüsisamba MRT-uuringute kirjeldustest. Uuringute kirjelduste all tuleb käesolevas töös mõista radioloogilist leidu kirjeldavat vabateksti. Uuringu muid parameetreid (st milliste seadmetega tehtud, võimalikke menetlusaasjaolusid jms) saadud failid üldiselt ei sisaldanud. Algfaili ja sellest tehtud uuritava faili võrdlus on tabelis 1.

**Tabel 1. Kõikide MRT-uuringute ja lülisamba MRT-uuringute töofaili iseloomustus**

<b>Kõiki MRT-uuringuid sisaldav algfail</b>		<b>Lülisambauuringuid MRT-uuringuid sisaldav esialgne teisendusteta töofail</b>	
Sõnesid	87 824	sõnesid	34 215
Tähemärke	648 119	tähemärke	257 906
tähemärke tühikutega	735 419	tähemärke tühikutega	291 923
Lõike	1919	lõike	642
Ridu	9004	ridu	3498
Lehekülgi	125	lehekülgi	50

### **3.2 Inglisekeelsetes lülisamba MRT-uuringute kirjeldustes kasutatavad laused**

Valik inglisekeelseid lülisamba MRT-uuringute kirjeldustes kasutatavaid tüüplauseid on saadud radioloogist töö juhendaja dr Peeter Rossi kaudu ning esitatud lisas 1.

### **3.3 Info- ja keeletehnoloogiavahendid**

Tekstitötluseks ja tabelarvutuseks kasutasin MS Office 2007 programme Word ja Excel, teksti keeleliseks analüüsiks programmi AntConc 3.2.1w 2007.

## **4 MEETODID**

### **4.1 Korpuse korrastamine: lülisamba MRT-uuringute tööfaili teisendused enne keelelist analüüsi**

Lülisamba MRT-uuringute kirjeldused olid ühetaoliselt liigendamata ning sisaldasid palju teksti, mis ei olnud seotud lülisamba (võimaliku) patoloogia kirjeldusega ning palju ebaharilikke ja ühtlustamata lühendeid. Seetõttu tuli enne lingvistilist analüüsi:

- korrastada iga faili liigendus ja lausestus;
- eemaldada failidest kõik ebaoluline ehk jätta alles vaid lülisamba patoloogia kirjeldamise seisukohalt relevantne tekst;
- ühtlustada tekstides kasutatud lühendid.

Selleks pidin faili manuaalselt korrastama, sest mis tahes programmi koostamine kõigi kolme eesmärgi lahendamiseks oluks lihtsalt palju töömahukam ning olemasoleva teksti hulk võimaldas seda tööd teha käsitsi.

#### **4.1.1 Lülisamba MRT-uuringute kirjelduste struktuuri ühtlustamine**

Algses tekstide struktuuris oli kasutatud erisugust liigendamist (tekstis esines erinevaid sisulisi osi). Ka osade järjestus oli kirjeldustes erinev. Viimane asjaolu ei pruugi arvutilingvistilisel analüüsil olla iseenesest takistuseks. Arvutil on lihtsalt võimatu aimata, kust üks osa algab või teine lõpeb, kui see on tekstis tähistamata. Teksti kergema kontrollitavuse huvides korrastasin siiski ühetaoliseks ka kirjelduse osade järjestuse. Lisaks varieerusid kirjeldustes sama teavet sisaldavate osade nimetused. Ühtlustamise eesmärgil kasutusele võetud kirjelduse vabateksti liigenduse skeem on esitatud tabelis 2 koos eelnevalt tekstis esinenud vastetega.

**Tabel 2. Lülisamba MRT-uuringu kirjelduse teksti liigendus**

<b>Pärast korrastamist:</b>	<b>Enne tekstis kujul (kirjapilt muutmata):</b>
UURINGU AEG (kuupäev kujul xx/yy/zzzz)	Palju variante.
UURING	Eraldi tähistamata, võis eelneeda või järgneeda kliinilistele andmetele.
KLIINILISED ANDMED	Kl. andmed, Saateandmed, Kliinik, Kliinilised andmed, Kliin. andmed, l. andmed, Kliendi andmed Kliinilised andmed võisid alustada kirjeldust või järgneeda uuringu nimetusele. Ei pruukinud igas kirjelduses esineda.
LEID (Leid1, Leid2 jne)	LEID, Leid, LIED Enamasti järgnes uuringu leiu kirjeldus kliinilistele andmetele, kui need olid esitatud. Tekstis võis olla enam kui üks eraldi väljatoodud leid, mida kajastasin ka ühtlustatud teksti struktuuris.
KOKKUVÕTE	KOKKUVÕTE, Kokkuvõte, Arvamus, ARVAMUS Kui oli eraldi välja toodud, siis lõpetas kirjelduse. Mõlema nimetuse all (kokkuvõte ja arvamus) esitati radioloogilise kujutise põhjal lühidalt radioloogi vastus kliinilisele küsimusele, mis oli uuringule saatmise põhjuseks, arvamus praeguse seisundi kohta või ettepanek edasiste uuringute ja/või ravimenetluste suhtes.

Kui kirjelduses puudus mõni liigenduse osadest, siis lisasin osa nimetuse ikkagi ja tähistasin puuduva teksti järgmiselt: puuduv uuringu kirjelduse osa – ZZZ, puuduv kliiniliste andmete osa – WWW, puuduv kokkuvõtte osa – YYY. Viimast esines sageli, mis tähendas, et kokkuvõtte oli kirjutatud leiu kirjelduse sisse või selle lõppu. Tähisega YYY on tähistatud ka lülisamba patoloogia seose puudumise tõttu eemaldatud kokkuvõtte kirjeldus. Kuna mõnes kirjelduses esines mitu leidu, siis esitasin selle osa vastavalt nummerdatuna: Leid1, Leid2, jne. Leidu kui üldjaotust tähistasin suurtähelise sõnaga „LEID“.

Kirjelduses lugesin lülisamba patoloogia seisukohalt ebaoluliseks kogu teksti, mis kirjeldas mingit muud anatoomilist piirkonda – näiteks eesnäärme pahaloomulise protsessi puhul on

suhteliselt tavaline, et veendutakse selles, kas patsiendil esineb ka luumetastaase, sh lülisambas. Sellise uuringu kirjelduse puhul eemaldas in uuritavast tekstist kogu eesnäärme anatoomia ja patoloogia kirjelduse ning jätsin alles ainult asjakohase, s.o lülisamba patoloogiaga otseselt seotud teksti.

#### **4.1.2 Lülisamba MRT-uuringute kirjelduste vabateksti keeleline korrastamine**

Arvestades võimalikku vajadust analüüsida tekste tulevikus ka arvutilingvistiliselt, korrastasin iga kirjelduse nii, et:

- kõik laused või fraasid algavad suure algustähega;
- kõik laused või nendega võrdsustatud fraasid lõpevad lause lõpumärgiga (käsitletud tekstides üldjuhul punkt, vahel ka küsimärk);
- kirjavahemärkide, sh side- ja mõttekriipsude kasutus vastaks „Eesti keele käsiraamatus“ (Erelt 2007) esitatule (kogu teksti ulatuses oli kasutatud ainult sidekriipsu, st ka seal, kus pidanuks olema mõttekriips (nt tähenduses „kuni“);
- lühendite keskel ega lõpus lause keskel ei ole punkti (et vältida selle tõlgendamist arvuti poolt lause lõpuks); lühendi keskel on punkt asendatud sidekriipsuga; v.a sõnas „kontrastaine“, mis oli kirjeldustes algselt lühendatud kaldkriipsu kasutades k/a, et vältida segiminekut sõna „käesolev“ uue lühendiga (k-a))
- osalaused on eraldatud komade või semikoolonitega;
- kõikide lauset või fraasi lõpetavate punktide järel on üks tühik;
- kõik sõnad on eraldatud tühikutega;
- kõik ilmsed näpuvead on korrigeeritud (selguse huvides failis sinises kirjas).

Peale selle on ühtlustamise eesmärgil parandatud kirjavead ja ühtlustatud kasutatud lühendid, nii et kõigis kirjeldustes oleksid need kasutatud ühetaoliselt. Paranduste tegemisel on lähtutud „Ladina-eesti-vene sõnaraamatust“ (I: 1982; II: 1983), „Meditšiinisõnastikust“ (2004) ning käsiraamatust „Radioloogia“(1995). Kui ladinakeelsetel vastetel olid lubatud paralleelvormid, siis valisin eelpool mainitud kirjanduse alusel ikkagi välja ainult ühe vaste ning ühtlustasin teksti vastavalt. Tehtud parandused on esitatud koondina tabelis 3.

**Tabel 3. Teksti parandused ja ühtlustamine**

<b>Tekstis</b>	<b>Parandus või ühtlustatud variant</b>
<i>annulus fibrosus</i>	<i>anulus fibrosus</i>
antero-lateraalne	anterolateraalne
CA	Ca ( <i>cancer</i> )
<i>cauda eq, cauda equina, c-equina</i>	<i>cauda equina</i>
<i>conus medullaaris, conus medulaaris, Conus medullaris, conus medularis</i>	<i>conus medullaris</i>
CT	KT ( <i>kompuutertomograafia</i> )
<i>diagn</i>	<i>dgn</i>
<i>diam, dm</i>	<i>diameeter</i>
<i>Dx, D</i>	<i>dex</i>
<i>dura kott, durakott</i>	duurakott
fassett-liiges	fassetlliiges
filum terminale	filum terminale
foraamen, firaamen	foramen
kasvur	kasvaja
<i>lig flavumi`te, lig flavumi`de, lig flavumide</i>	<i>lig flavumite</i>
Modic I, II, III	Modic 1, 2, 3
MRT müelograafia	MRT-müelograafia
MR signaal	MR-signaal
MRT natiivuuring...	MRT-uuring: natiivuuring...
MR uuring, MR-uuring, MRT uuring, MRT	MRT-uuring
närvikompress	närvikompressioon
<i>os ischi</i>	<i>os ischii</i>
postero-lateraalne	posterolateraalne
<i>proc</i>	<i>processus</i>
ruutcm	cm <sup>2</sup>
S	<i>sin</i>
<i>spondylolistees</i>	<i>spondülolistees</i>
Tarlov'i, Tarlowi, Tarlow'i tsüst	Tarlovi tsüst
torako-lumbaalosa	torakolumbaalosa

Lisaks on lülisamba MRT-uuringute failis ühtlustatud lülisamba lülide lühendamine.

Algfailides oli kasutatud mitmeid variante, nt L5 vs. L 5; L3–5 vs. L3–L5; L6,7 vs. L6/7.

Kõige rohkem lühendite variatsioone kasutati lülisamba torakaalosa kirjeldamiseks, kus olid kasutusel ka Rooma numbrid (nt Th IV) ning kasutati lühendeid Th, TH ja ka lihtsalt T. Ka oli tüüpiline, et suuremate lülinumbrite puhul pandi lühendi Th järelle enne numbrit tühik (nt Th 11), väiksemate numbrite korral aga tühikut enamasti ei pandud (nt Th5) ning erinev märgistus võis olla kasutusel sama kirjelduse piires.

Lülisamba lülide ühtlustatud märgistus on järgmine. Lülisamba tservikaalosa lülide tähis on **C**, torakaalosa tähis on **Th** (sest lühendid T1 ja T2 on kasutusel MRT-uuringu parameetrite kirjeldamiseks), lumbaalosa tähis on **L**, sakraalosa tähis on **S**. Lüli number kirjutatakse tühikuta teda tähistava tähe taha, nt **C1**, **Th11**, **Th1**, **L2**. Kui näidatakse teatud hulka lülisamba sama piirkonna lülisid haaravat protsessi, siis on kirjaviis **C1–C5**; kui mingit lülivahemikku haaravat protsessi, siis **C6/C7**. Kui haaratud olid eri piirkondade lülid, siis on esitanud kõigi nende tähised eelpool öeldust lähtuvalt, nt **L5/S1** või **C5–C7/Th1–Th4**. Kald- ja mõttekriipsu järel pidasin oluliseks lüli tähistav täht alles jätta, sest nii on hiljem lihtsam päringuid teha.

Kõik kuupäevad on viidud kujule kuupäev/kuu/aasta (xx/yy/yyyy, enne erinevaid tähistusi xx/yy/yyyy või yy/yyyy või xx/yy). Kaldkriipsu kasutamist kuupäevas olen eelistanud selleks, et vältida punktide kasutamisest tulenevat segiminekut lauselõpu märgiga.

### **4.1.3 Korpuse liigendus ja märgendus**

Käesolevas töös on tekstid organiseeritud ükskeelseks spetsialiseeritud korpuseks (Hunston 2006: 14). Tulevikus on sellele vajadusel võimalik lisada muid tekste, sh ka teiste anatoomiliste piirkondade ja teiste radioloogiliste uuringute tekste. Seega peab korpuse liigendus tulevikus arvestama ühelt poolt lülisamba MRT-uuringute kirjelduste liigendust ja

teisalt lülisamba MRT-uuringute paigutust teiste radioloogiauuringute seas. Kuna korrastatud failis oli tegu selge struktuuriga, siis praegu ei olnud esmase lingvistiliseks analüüsi eesmärgil vajadust täiendavat märgendust lisada (nt lõikude või lause lõpu märgendust).

## **4.2 Lülisamba MRT-uuringu kirjelduse kliiniline struktuur**

Analüüsisin tekstipõhiselt olemaolevaid kirjeldusi selleks, et mõista uuringute võimalikku kliinilist liigendamist. Lülisamba anatoomia ja patoloogia kirjelduste analüüsil lähtusin asjakohastest mulle kättesaadavatest materjalidest (käsiraamatud „Radioloogia“ 1995; „Radiographic Anatomy” 2000; „Spinal Imaging” 2008; „MRI Parameters and Positioning” 2003; „Getting Started in Clinical Radiology” 2006). Analüüsi tulemused on esitatud käesoleva töö tulemuste osas.



## 5 TULEMUSED

### 5.1 Lülisamba MRT-uuringute keeleline analüüs

#### 5.1.1 Tekstides esinevate sõnade sagedus võrrelduna eesti kirjakeele korpuse sõnasagedustega

Tekstis esinevate sõnade sageduse leidmiseks kasutasin programmi AntConc 3.2.1w.

Programm annab küll kõigi tekstis esinevate sõnavormide (sõnede) loendi, kuid ei seosta neid algvormiga, mis tuli teha käsitsi eraldi. Põhimõtteliselt oluks ehk mugavam kasutada UNIXis tehtud skripte ja eesti keele morfoloogilist analüsaatorit ning ühestajat, millega võib-olla saanuks tulemuse mõnevõrra väiksema vaevaga. See ent oleks eeldanud veel suuremat algse materjali korrastamist ning tegelikult ka väikse leksikoni lisamist, et vähemalt kuidagi ületada lühendite ja ladinakeelsete sõnade ühestamisega seotud probleeme.

Sõnade sageduse määramiseks kasutasin korrastatud faili, kust olin eemaldanud liigenduse jaotiste pealkirjad (st uuring, kliinilised andmed, leid, kokkuvõte), sest need iseenesest ei olnud tekstispetsiifilised ja kordusid igas kirjelduses. Vaid üks kord tekstis esinevaid sõnu oli 1685, mis moodustas ligi poole (52%) kõigist teksti sõnadest.

Tabelis 5 on enim esinenud sõnade sagedus võrrelduna varem avaldatud eesti kirjakeele ajakirjandus- ja ilukirjandustekstide korpuste sõnade sagedusega. Lülisamba kirjelduste teksti sõnaarv oli 3211 ja sõnede arv 24 892, ühetähelised lühendid (põhiliselt lülisamba piirkondade tähised) eemaldasid programmis AntConc 3.2.1w käsitlemise järel failist käsitsi. Paksus kirjas on lülisamba uuringute kirjelduste sõnad, mis esinevad ka teistes korpustes. Tabelis on verbid paremaks eristamiseks kaldkirjas. Ladinakeelsed sõnad on kaldkirjas ja ülakomade vahel.

**Tabel 4. Enim esinenud sõnade sagedus lülisamba MRT-uuringute kirjeldustes võrrelduna Eesti 1990. aasta ilukirjanduslike ja ajakirjandustekstide korpusetes olevate sõnade sagedusega (Hennoste 2000)**

	Lülisamba kirjelduste tekstid			Ajakirjandustekstid			Ilukirjandustekstid		
	Sõna	Sagedus	%	Sõna	Sagedus	%	Sõna	Sagedus	%
1.	<i>olema</i>	1046	4,2	<i>olema</i>	16228	4,22	<i>olema</i>	28237	4,62
2.	<i>ei</i>	846	3,4	<i>ja</i>	10881	2,83	<i>ja</i>	18469	3,02
3.	<i>ja</i>	778	3,13	see	7618	1,98	tema	18235	2,98
4.	disk	664	2,67	et	4366	1,14	see	15240	2,49
5.	kõrgus	371	1,49	<i>ei</i>	3864	1,00	mina	13261	2,17
6.	uuring	369	1,48	mina	3439	0,89	<i>ei</i>	10360	1,70
7.	närvijuur	347	1,39	tema	3173	0,83	et	7538	1,23
8.	MRT	344	1,38	<i>ka</i>	3096	0,81	kui	5833	0,96
9.	muutus	333	1,34	Eesti	2968	0,77	<i>mis</i>	5621	0,92
10.	lülisammas	327	1,31	<i>mis</i>	2916	0,76	sina	5505	0,90
11.	signaal	317	1,27	kui	2842	0,74	aga	4350	0,71
12.	lülilikeha	304	1,22	oma	2049	0,53	oma	4120	0,67
13.	ahenemine	266	1,07	<i>saama</i>	1889	0,49	siis	3655	0,60
14.	osa	249	1,00	<i>ning</i>	1739	0,45	ise	3403	0,57
15.	spinaalkanal	235	0,94	aasta	1730	0,45	<i>ka</i>	2992	0,49
16.	tase	229	0,92	aga	1728	0,45	kes	2819	0,46
17.	parem*	219	0,88	kes	1667	0,43	nagu	2814	0,46
18.	kerge	201	0,81	<i>tulema</i>	1396	0,36	nii	2589	0,42
19.	oluline	196	0,79	<i>pidama</i>	1387	0,36	<i>tulema</i>	2401	0,41
20.	seljaaju	183	0,74	või	1260	0,33	<i>saama</i>	2474	0,41
21.	veesisaldus	181	0,73	nii	1235	0,32	<i>ütleva</i>	2419	0,40
22.	ega	177	0,71	<i>võima</i>	1210	0,31	<i>ning</i>	2416	0,40
23.	degeneratiivne	174	0,70	kõik	1172	0,31	<i>minema</i>	2378	0,39
24.	<i>ka</i>	171	0,69	üks	1153	0,30	kõik	2360	0,39
25.	<i>mis</i>	170	0,68	siis	1143	0,30	või	2321	0,38
26.	langus	169	0,68	ise	1126	0,29	mees	2169	0,36
27.	lõpp*	167	0,67	<i>tegema</i>	992	0,26	<i>pidama</i>	2122	0,35
28.	väljakummumine	164	0,66	teine	960	0,25	miski	2101	0,34
29.	protrusioon	158	0,63	aeg	940	0,24	veel	2096	0,34
30.	ax	141	0,57	veel	856	0,22	üks	2072	0,34
31.	signaalimuutus*	139	0,56	inimene	815	0,21	<i>tegema</i>	1933	0,32
32.	<i>sedastama</i>	131	0,53	palju	780	0,20	kas	1839	0,30
33.	difuusne	118	0,47	uus	756	0,20	<i>teadma</i>	1677	0,27
34.	fassetliiges	117	0,47	võtma	727	0,19	nüüd	1641	0,27
35.	ilmne	117	0,47	juba	721	0,19	teine	1637	0,27
36.	sag	107	0,43	suur	712	0,19	aeg	1610	0,26
37.	<i>ning</i>	105	0,42	<i>andma</i>	711	0,19	juba	1605	0,26
38.	<i>cauda</i>	105	0,42	<i>jääma</i>	692	0,18	<i>hakkama</i>	1499	0,25
39.	mähis	103	0,41	kuid	687	0,18	<i>võima</i>	1480	0,24
40.	patoloogiline	103	0,41	kas	686	0,18	<i>vaatama</i>	1477	0,24
41.	ekstrusioon	103	0,41	välja	683	0,18	<i>võtma</i>	1440	0,24
42.	' <i>equina</i> '	101	0,41	kus	658	0,17	naine	1420	0,23
43.	' <i>conus</i> '	93	0,37	vaid	657	0,17	käsi	1419	0,23
44.	<i>madalduma</i>	93	0,37	<i>ütleva</i>	656	0,17	välja	1400	0,23
45.	ca	89	0,36	pool	638	0,17	<i>jääma</i>	1380	0,23
46.	<i>langema</i>	89	0,36	nagu	628	0,16	küll	1334	0,22
47.	piir	84	0,34	<i>minema</i>	625	0,16	<i>nägema</i>	1329	0,22
48.	mm	83	0,33	üle	623	0,16	<i>inimene</i>	1298	0,21
49.	natiiv	74	0,30	Tallinn	620	0,16	mitte	1283	0,21
50.	cm	73	0,29	töö	615	0,16	<i>tahtma</i>	1238	0,21

\**parem* tähenduses parempoolne; *lõpp* – õigupoolest lõpp-plaadid (sidekriipsu tõttu arvestas programm eraldi sõnana); *signaalimuutus*: vääriks eraldi uurimist, peaksin enamikus lausetes õigemaks pigem lahku kirjutatud varianti – signaali muutus (mitte eraldi mõistena; sama kehtib ka signaali intensiivsuse ja signaali tõusu kohta, seega parandatud võiks sõna „signaal“ kasutuskordade arv märgatavalt suurenedada).

### 5.1.2 Tekstides esinevad lühendid

Tekstid sisaldasid väga palju lühendeid, millest osa kordus sageli. Enamik lühenditest ei olnud tavalise kirjakeele lühendid, vaid meditsiinilise terminoloogia lühendid või harvem ka kiirustamisest tingitud lühendid. Lühendid olid kas sõna lühendamise tulemus või (enamasti) suurtähelised akronüümid. Algtekstides võis lühendite lõpus olla punkt, kuid see võis ka puududa. Korpuse tegemisel kaasneb sellega kaks probleemi: punkt lühendi lõpus ei eristu arvuti jaoks lause lõpumärgist ning ilma eritähistuseta (märgendamata) lühendit käsitleb arvuti lihtsalt sõnana. Selline „sõna“ ei käitu aga enam eesti keele morfoloogia ja süntaksi reeglite järgi. Üheks võimalikuks lahendiks on teha lühenditest arvutile eraldi leksikon. Selleks tuleb aga tekstides esinevad lühendid üle vaadata ning otsustada, millised neist on olulised (kas neid on lühenditena vaja) ning millisel kujul neid peaks lühendama (kas on teisi samasuguseid lühendeid jms). Samuti tuleb loobuda kõigist korduvatest või ambivalentsetest lühenditest, sest arvutil ei ole võimalik tähenduse üle konteksti põhjal otsustada. Õigupoolest, ei pruugi ka teksti tõlkija ja/või lugeja alati aru saada, mida on mõeldud. Kokku esines alguses tekstis **104** erinevat lühendit (vt tabel 5).

Lühendid on lihtsustuse mõttes esitatud üldjuhul grammatilist sugu (ladinakeelsed), arvu ja käänat ning eespool kirjeldatud lüüsisamba lülikehade ja -vahemike lühendamist arvestamata. Samuti ei ole eraldi lisatud haiglate nimede lühendeid (nt ITK, PERH, LTKH), sest õigupoolest peaks eelneva ravi koht vajaduse korral olema kirjeldatud mujal ning teiseks ei ole faktil, mis haiglas konkreetselt patsient varem viibis, patoloogia visuaalse kirjeldamise seisukohalt otsest tähtsust.

**Tabel 5. Lülisamba MRT-uuringute kirjelduste algtekstides kasutatud lühendid**

Nr.	Lühend	Mõiste	Kommentaar
1.	ad	<i>adenoma</i>	
2.	anam	anamnees(is)	
3.	bil; bilat	bilateraalne	kaks lühendit, üks mõiste
4.	c Douglas	<i>cavum Douglasi</i>	
5.	ca	1. <i>cancer</i> 2. <i>circa</i>	kaks mõistet, üks lühend
6.	can	<i>canalis</i>	
7.	c-equina	<i>cauda equina</i>	
8.	cerv	<i>cervicalis</i>	
9.	chr	<i>chronicus</i>	
10.	compr	<i>compressio</i>	
11.	craniovertr	<i>craniovertebralis</i>	
12.	dex	<i>dexter</i>	esineb ka suurtähtlühendina
13.	dgn, diagn	diagnoos	kaks lühendit, üks mõiste
14.	diam, dm	diameeter	kaks lühendit, üks mõiste
15.	dif. diagn.-lt	diferentsiaaldiagnostiliselt	
16.	discogen	diskogeenne	
17.	dün	dünaamika, dünaamikas; dünaamiline,	
18.	etiol	etioloogia(s)	
19.	fr	<i>fractura</i>	
20.	i/v	1. intravertebraalne 2. intravenoosne (-selt)	kaks mõistet, üks lühend
21.	interver, intervert.	intertvertebraalne	kaks lühendit, üks mõiste
22.	k/a	kontrastaine	
23.	kl, kliin	kliiniline	kaks lühendit, üks mõiste
24.	l, lig	<i>ligamentum</i>	kaks lühendit, üks mõiste
25.	l-sõlm, l/s	lümfisõlm	kaks lühendit, üks mõiste (vt ka alljärgnev)
26.	l/s	1. lümfisõlm 2. lumbosakraal- 3. lülisamba	kolm mõistet, üks lühend
27.	l/sacr, l/s	lumbosakraal-	kaks lühendit, üks mõiste
28.	m	<i>musculus</i>	
29.	met, mts	metastaas, metastaasid	
30.	neopl	neoplasma	
31.	obj	objektiivselt	
32.	org	orgaaniline	
33.	op	operatsioon, opereeritud	
34.	ov	ovarium	
35.	pat	patoloogia	
36.	pl	<i>plexus</i>	
37.	proc	<i>processus</i>	
38.	postop	postoperatiivne	
39.	prol	<i>prolapsus</i>	
40.	prost	<i>prostata</i>	
41.	pt	patsient	

42.	pulm	<i>pulmonis (Ca pulmonis)</i>	
43.	reg	<i>regio</i>	
44.	rö	röntgen(-uuring)	
45.	rö-gr	röntgenogramm	
46.	sag	sagitaalne	
47.	seg	segment	
48.	sin	<i>sinister</i>	
49.	sol	<i>solutio</i>	
50.	SpA	spondüloartriit	
51.	spin	spinaal-, <i>spinalis</i>	võib olla nii eesti- kui ka ladinakeelse väljendi osa
52.	staph	<i>staphylococcus</i>	
53.	<i>status post</i>	millegi järgne seisund	
54.	tu	<i>tumor</i>	
55.	ut	<i>uterus</i>	
56.	v, vert	<i>vertebra</i>	kaks lühendit, üks mõiste
<b>Suurtähelised lühendid</b>			
1.	AP		
2.	AVN	avaskulaarne luunekroos	
3.	CRV	C-reaktiivne proteiin	
4.	CT, KT	kompuutertomograafia	kaks lühendit, üks mõiste
5.	D	parem	esineb ka sõnast moodustatud lühenditena
6.	DDD	<i>Degenerative Disc Disease</i>	
7.	ILE	interlaminektoomia	
8.	JNB	jämenõelabiopsia	
9.	KNS	kesknärvisüsteem	
10.	KT	kompuutertomograa	
11.	LP	lumbaalpunksioon	
12.	M	<i>morbus</i>	
13.	MR	magnetresonants-	
14.	MRT	magnetresonantstomograafia	
15.	MTS	metastaas(id)	esineb ka väiketähtlühendina
16.	NB	<i>nota bene</i>	
17.	NSAID	<i>Non-Steroidal Antiinflammatory Drugs</i>	
18.	PSA	prostata-spetsiifiline antigeen	
19.	PK	pahaloomuline kasvaja	
20.	S	vasak	esineb ka sõnast moodustatud lühenditena
21.	SIL	sakroiliakaalliiges(ed)	
22.	UH	ultraheli	
23.	USG	ultrasonograafia	
<b>Töenäoliselt kiirustamisest sündinud lühendid</b>			
1.	isel	iseloomulik	
2.	komp	kompressioon	
3.	koord katsud	koordinatsioonikatsud	
4.	kr	krooniline	
5.	kraniospin	kraniospinaalne	

6.	mamm	<i>mammae</i>	
7.	nimmepiirk	nimmepiirkond	
8.	par	parem	
9.	patol	patoloogiline	
10.	rad val	radikulaarsed valud	
11.	radik	radikulaarne	
12.	temp	temperatuur	
13.	vas	vasak	
14.	ül	ülemine	
15.	ülem	ülemine	
<b>Tavalühendid</b>			
1.	a	1. aasta 2. alates	üks lühend, kaks mõistet
2.	k.a	käesoleva aasta	
3.	max	maksimaalne	
4.	min	minimaalne	
5.	nn	nii nimetatud	
6.	sh	sealhulgas	
7.	vt	vaata	
<b>Lühenditena kasutatud märgid</b>			
1.	+	ja	
2.	<	väiksem	vahel intensiivsuse näitamiseks kahekordselt
3.	>	suurem	vahel intensiivsuse näitamiseks kahekordselt

### 5.1.3 Süntaksi eripära tekstides

Tekstide üheks eripäraks oli nimisõnafaaside rohkus. Sõnade sageduse tabelist on näha, et tekstides esines vähe isegi üldkeeles enim kasutatavaid tavaverbe (pidama, saama, tegema, võima jne). See viitab kaudselt asjaolule, et tekstides ei kasutatud ainult täislauseid, vaid ka kirjeldavaid nimisõnafaase. Selle asjaolu täpsustamiseks arvestasin verbi vormidest tekstides esinenud kesksõnade sageduse. Partitsiipe võib vahel süntaktiliselt funktsioonilt analüüsida pigem atribuutideks ning seega adjektiiviks, sest nad vastavad sageli küsimusele *missugune?* (nt „diski prolapsile sobiv leid“ või „uuritud piirkonnas“ jne). Kesksõnu esineski tekstides palju, kokku 1025 korda, mis moodustas 4,12% kõigist sõnedest.

**Tabel 6. Kesk sõnade sagedus lüli samba MRT-uuringute kirjeldustes**

Mineviku kesksõnad				Oleviku kesksõnad	
-nud kesksõna		-tud kesksõna			
madaldunud	94	uuritud	43	viitav	48
langenud	88	teostatud	37	sobiv	39
ahenenud	88	manustatud	18	kontrasteeruv	15
väljendunud	60	süstitud	15	põhjustav	15
lamenenud	49	opereeritud	9	jälgitav	12
nihkunud	22	arvestatud	6	nähtav	10
säilinud	19	tehtud	6	vastav	10
vähenenud	19	antud	5	ulatuv	9
kitsenenud	12	haaratud	5	erinev	8
ülejäänud	12	kirjeldatud	5	kiirguv	6
lisandunud	11	piiratud	4	domineeriv	5
süvenenud	10	tingitud	4	komprimeeriv	5
dehüdreerunud	8	diagnoositud	3	deformeeriv	4
muutunud	7	eemaldatud	3	kaasuv	4
asendunud	6	kaalutud	3	prevaleeruv	4
tõusnud	6	kombineeritud	3	promineeruv	4
paksenenud	4	komprimeeritud	3	süvenev	4
tugevnenud	4	omandatud	2	eelnev	3
deformeerunud	3	põhjustatud	2	eristatav	3
hüpertrofeerunud	3	ahendatud	1	hinnatav	3
intensiivistunud	3	asetatud	1	hoidev	3
kummunud	3	avatud	1	kaarduv	3
piirdunud	3	dislotseeritud	1	küsitav	3
tekkinud	3	distantseeritud	1	märgatav	3
ekstruseerunud	2	infiltreeritud	1	ümbritsev	3
jäänud	2	leitud	1	aimatav	2
kadunud	2	mainitud	1	käesolev	2
kestnud	2	mõjutatud	1	külgnev	2
kukkunud	2	raskendatud	1	looklev	2
suurenenud	2	seotud	1	olev	2
võlvunud	2	soovitatud	1	püsiv	2
väljakummunud	1	teatud	1	väljakummuv	2
asteniseerunud	1	täidetud	1	väljavõlvuv	2
dehüdrateerunud	1	ülalkirjeldatud	1	demüeliniseeruv	1
elavnened	1			eristuv	1
elongeerunud	1			esinev	1
haigestunud	1			ilmnev	1
kujunenud	1			kalduv	1
moodustunud	1			korduv	1

muutnud	1			liigestuv	1
niverdunud	1			mainitav	1
nõrgenenud	1			mõeldav	1
olnud	1			allpaiknev	1
pidurdunud	1			progresseeruv	1
pistunud	1			puuduv	1
protruseerunud	1			sedastatav	1
rebenenud	1			soovitav	1
saanud	1			suurenev	1
sagenenud	1			ägenev	1
sirgendunud	1				
taandunud	1				
tulnud	1				
täitnud	1				
vajunud	1				
võtnud	1				
õhenenud	1				
<b>Kokku:</b>	<b>576</b>	<b>Kokku:</b>	<b>191</b>	<b>Kokku:</b>	<b>258</b>
<b>Kokku kesksõnu:</b>			<b>1025</b>		

Lausetes oli väga palju kasutatud eitust, kuid samas erinevaid eituse konstruktsioone oli üllatavalt vähe. Põhiliselt väljendas eitust patoloogia puudumist ning üsna sageli sisaldas ka fakti, et see puudumine on tõdetud radioloogi hinnangu käigus (nt väljendid „ei sedasta, ei tähelda, ei näe, ei leia, ei arva“ jne, vt tabel 7).

Sageli olid laused ja fraasid väga pikad ning lausete sidumiseks oli kasutatud mõttekriipsu. Viimane viitas enamasti teises lauseosas järgnevale järeldusele ega täitnud lihtsalt kiilu eristamise funktsiooni. Kokku oli 85 lauset (vt lisa 2), milles mõttekriipsu kasutati otseselt järeldusele viitamiseks ja mitte tähenduses „kuni“ või kiilu eristajana. Ootuspäraselt oli lausetes väga vähe anafoore, mis eristab teksti selgelt ajakirjanduslikust või ilukirjanduslikust. Isikulisi asesõnu peaaegu ei esinenudki („ta“ 1, „tema“ 2, „nad“ 2 korda); „see“ esines tekstides 3, „selle“ 5 ja „seda“ 2 korda, „teine“ asesõna funktsioonis 1 kord.



**Tabel 7. Tekstides kasutatud eitused**

Nr	Eitus	Sagedus tekstis	Kommentaar
1.	ei ole	349	Sisuliselt tähendab, et otsitavat pildil ei ole. (St üks tähendus, mitu väljendusviisi.)
2.	ei sedasta	130	
3.	ei ilmne	117	
4.	ei esine	72	
5.	ei tähelda	50	
6.	ei paista	19	
7.	ei leia	14	
8.	ei tule (nähtavale, esile)	11	
9.	ei näe	1	
<b>Kokku samatähenduslikke eitusi</b>		<b>763</b>	
10.	ei põhjusta	24	Tähendab sama, mis „ei suru“.
11.	ei komprimeeri	11	
12.	ei ahenda	9	
13.	ei saa	7	
14.	ei kaasne	6	
15.	ei ulatu	3	
16.	ei visualiseeru	3	
17.	ei avalda	3	
18.	ei suru	3	Tähendab sama, mis „ei komprimeeri“.
19.	ei destrueeri	3	
20.	ei kontrasteeru	2	
21.	ei pea	1	
22.	ei painuta	1	
23.	ei vallandu	1	
24.	ei kummu	1	
25.	ei arva	1	
26.	ei kanna	1	
27.	ei mõjuta	1	
28.	ei anna	1	
29.	ei nimme- ega rinnaosas pole	1	
<b>Kokku peamiselt eritähenduslikke</b>		<b>83</b>	
<b>Kõik kokku</b>		<b>846</b>	

#### 5.1.4 Samatähenduslikud laused

Sama mõtet väljendati tekstides väga sageli eri sõnastuses lausetega, kuigi sisuline teave lauses tundus olevat täpselt ühesugune. Näiteks võib tuua signaali muutuste kirjeldamisest lülisamba luudes (vt tabel 8).

**Tabel 8. Samatähenduslike lausete näide**

	<b>Näitelauseid tekstist (kirjaviis muutmata)</b>	<b>Kommentaar</b>	
<b>I</b>	Luudes signaaliintensiivsus normaalne. Luud ühtlase signaaliga. Luudes signaalimuutusi ei paista. Luudes signaalimuutusi ei sedasta. Luudes signaalimuutusi ei ole.	Kas mõttes on mingit vahet tulenevalt: – jaatuse vs. eituse kasutamisest? – eri tegusõnadest?  Kas ühtlane signaal on sama mis normaalne signaali intensiivsus?	Kas I ja II jaotise väljenduses on põhimõtteline vahe?
<b>II</b>	Luudes patoloogilist signaali muutust ei ilmne. Luudes patoloogilist signaalimuutust ei tähelda. Luudes aktuaalseid signaalimuutusi ei paista. Luudes patoloogilist signaali muutust ei sedasta. Luudes patoloogiliseks sobivat signaalimuutust ei ole.	Lisandub signaali muutuse kirjeldus.  Patoloogilisus ei välista aktuaalsust. Selle lausetes eristamine tähendab, et juhul kui patoloogilisi muutusi ei ole, võib sellegipoolest kliiniliselt tähenduslikke (sest miks neist muidu kirjutatakse) aktuaalseid muutusi olla. Tegelikult ongi kliiniliselt tähenduslikud muutused seotud aktuaalse patoloogia või normi kirjeldamisega.  Ka siin on kasutatud erinevaid tegusõnu.	
<b>III</b>	Luudes on vaadeldavas osas normaalne signaaliintensiivsus. Luud uuritud osas patoloogilise signaalimuutusega. Luudes uuritud piirkonnas aktuaalseid signaalimuutusi ei paista. Luudes uuritud ulatuses patoloogilist signaalimuutust ei ole.	Lisandub teatav üldlokalisatsioon (vaadeldavas osas, uuritud piirkonnas, uuritud ulatuses jne). Põhimõtteliselt uurimata osa ei saagi kirjeldada. Lisaks on nagunii igas kirjelduses täpselt ära näidatud, millises anatoomilisest piirkonnast see on tehtud. Seetõttu ei tundu selle asjaolu rõhutamisel olevat mingit tähendust.	Kas on tegelikkus vahet I ja II jaotises toodud lausetega?
<b>IV</b>	Luulistes struktuurides patoloogilist signaali muutust ei ilmne. Luulistes struktuurides patoloogilist signaali muutust ei esine. Luulistes struktuurides signaali muutust ei ilmne. Luustruktuurides aktuaalseid signaalimuutusi ei leia. Luulistes struktuurides patoloogilist signaali muutust ei esine	Mis vahe on luudel, luustruktuuridel ja luulistel struktuuridel? (Lisaks tundub, et struktuuride puhul võib tegu olla inglise keele interferentsiga, s.o <i>bony structures</i> või <i>bone structures</i> otsetõlkega.)	Kas erineb I, II ja III jaotisest? St kas on võimalik nt öelda: „Luudes signaali muutusi ei ole, aga luulistes struktuurides on signaali intensiivsus langenud?“ Tõenäoliselt mitte.
<b>V</b>	Luustikus koldeid ei ilmne. Luustikus koldeid ei sedasta. Luudes patoloogiliseks sobivat kollet ei ole. Luudes koldeid esile ei tule. Luude struktuur iseärasusteta, koldeid ei sedasta. Luudes veenvat koldeleid ei ole.	Siin on käsitletud kontekstis küsimuseks luustiku ja luude tähenduse olematu vahe.	Võib erineda küll jaotistest I–IV, sest teatud kontekstis viitab otseselt suunatud kolde (nt metastaasi) otsingule. Samas võib väljendada ka lihtsalt fakti, et luudes muutusi ei sedastatud.

## 5.2 Lülisamba MRT-uuringute tekstide terminoloogia ja kliiniliste andmete struktuur

Radioloogilise uuringu tegemisel tuleb radioloogil enne uuringu tegemist otsustada kliiniliste andmete ja kliinilise pildi põhjal, millisest anotoomisest piirkonnast ja millist uuringut tehakse. Käesolevas töös on käsitletud ainult lülisamba MRT-uuringuid, kuid põhimõtteliselt võib radioloogilised uuringud jagada järgmisteks uuringuteks:

- röntgenuurid
- ultraheliuurid
- kompuutertomograafia uuringud
- magnetresonantstomograafia uuringud (MRT-uuringud)
- nuklearmeditsiini uuringud
- menetlusradioloogilised uuringud.

Seega on radioloogiliste uuringute seas vähemalt 6 uuringuklassi ning neist mõningaid tehakse veel ka kontrastainega või ilma (natiivuuringuna) või eristatakse muid lisajaotisi. Seda peaks kõikide radioloogiliste uuringute korpuse liigendamisel või vähemalt andmebaasi koostamisel arvestama. Haigla infosüsteemis ongi eri uuringud üldjuhul salvestatud eri moodulitesse, millised neist võiks säilitada korpuse liigenduses on täiesti eraldi teema.

Lülisamba uuringud on keha üht anotoomilist piirkonda hõlmavad uuringud. Samas tuleb arvestada seda, et lülisamba kirjeldamise fakt ei pruugi vältimatult kirjelduse alguses tehtud uuringu tüüpi kirjeldavas jaotises („UURING“) üldse kajastuda. Tulenevalt luumetastaaside või muude patoloogiliste protsesside esinemisest on nt vaagnapiirkonna MRT-uuringul sageli kirjeldatud ka lülisamba haaratust. Seetõttu tuleb arvestada, et lisaks selgelt sõnastatule

„uuring lüüisambast“ võib lüüisamba patoloogiat olla kirjeldatud ka muude anatoomiliste piirkondade uuringute käigus.

Kuna radioloog sisuliselt tõlgib piltkujutise meile arusaadavasse keelde, siis sõltub tulemus ka kirjelduse tegija n-ö „tõlkijaoskustest“ ja keeletajust. Kuid keelest sõltumata on lisaks olemas radioloogilise uuringu kliinilise struktuuri põhikomponendid, mis peavad mingil moel käsitletud olema. Põhimõtteliselt sisaldab mingi anatoomilise piirkonna radioloogiline kirjeldus vähemalt järgmisi komponente (Schønnemann 2008):

- morfoloogilise muutuse kirjeldus (kohustuslik);
- selle lokalisatsiooni kirjeldus (kohustuslik);
- selle raskusastme kirjeldus (võib ka puududa).

Seega peaks lüüisamba radioloogilise kirjelduse kliiniline sisu kajastama vähemalt järgmisi aspekte (vt tabel 9).

**Tabel 9. Lüüisamba MRT-uuringute leiu kirjelduste kliinilised põhikomponendid**

<i>Jaotis</i>	<i>Näited eestikeelsete tekstide korpuse materjali põhjal, millele leidus vasteid olemasolevates inglisekeelsetes lausetes</i>	<i>Vasted inglise keeles (olemasolevate lausete põhjal)</i>	<i>Kommentaar</i>
<b>Kehapool*</b>	parem / <i>dex</i> parempoolne vasak / <i>sin</i> vasakpoolne	<i>right</i> <i>right-sided</i> <i>left</i> <i>left-sided</i>	Eestikeelsetele kirjeldustele on väga omane ladina keelest tuletatud lühendite <i>dex</i> ja <i>sin</i> kasutamine vastavalt parema ja vasaku poole tähistamiseks, mida inglise keeles üldiselt ei tehta. *Teoreetiliselt peaks kehapoolse jaotises olema

			ka anterioorne-posterioorne ja kraniaalne-kaudaalne.
<b>Anatoomiline piirkond</b>			
<b>Lüli samba anatoomiline piirkond</b>	kraniovertebraalpiirkond  kaelaosa  torakaalosa  nimmeosa/lumbaalosa  <i>conus medullaris cauda equina</i>	<i>craniocervical junction</i>  <i>cervical spine</i>  <i>thoracic spine</i>  <i>lumbar spine</i>  <i>conus medullaris cauda equina</i>	Eestikeelses tekstis esines kaunis sageli omakeelse vaste kasutamine, nt lüli samba kaelaosa puhul oli see praktiliselt eksklusiivne. Nimmeosa nimetati umbes võrdse sagedusega ka lumbaalosaks. Seevastu lüli samba torakaalosa nimetati ainult torakaalosaks. Sakraalosa eelistati ladinakeelset <i>sacrum</i> . Ristluud kasutati kõigis tekstides üks kord. Üldiselt eelistati piirkonda esitada lüli vahemikena. Eestikeelsetes kirjeldustes esines palju omakeelse ja ladina päritolu võõrsõnade läbisegi kasutamist.
<b>Muud seotud piirkonnad</b>	vaagen	ei esinenud näitelausestes [pelvis]	
<b>Anatoomilised struktuurid</b>			
<b>Luud ja teised anatoomilised tugistruktuurid</b>	disk lülid lüli vahemik, fasettliiges	<i>disc,</i> <i>vertebrae,</i> <i>intravertebral space,</i> <i>facet joint</i>	
<b>Pehmed koed</b>	duura  duurakott  <i>lig flavumid</i>	<i>dura</i>  <i>thecal sac</i>  <i>ligamenta flava</i>	Eestikeelsetes kirjeldustele iseloomulik ladinakeelsetele terminitele eesti

	närvijuured seljaaju rasvkiht, rasvkude, rasvüdi, rasv, epiduraalne rasv	<i>nerve roots</i> <i>spinal cord</i> <i>fat planes</i>	käändelõpu ja/või arvu lisamine, inglise keeles ladina kirjapilt. Tavaliselt esimene sõna ladinakeelsest ühendist on eestikeelses kirjelduses alati esitatud lühendina. <i>Fat plane</i> 'il eesti keeles mitmeid vasteid.
<b>Muud struktuurid või nende vahelised ruumid</b>	spinaalkanal subarahnoidaalne	<i>spinal canal</i> <i>subarachnoid space</i>	
<b>Leid</b>			
<b>Normileid</b>	...on normaalne	<i>normal...</i>	
<b>Muutused puuduvad</b>	nt signaali muutusi ei ole; signaali muutusteta ja veel mitmeid variante (vt tabel 7 – ei ole, ei ilmne, ei tähelda, ei paista, ei sedasta, ei näe jne)	<i>no signal alterations;</i> <i>no signal changes;</i> <i>no evidence of...;</i> <i>there is no...</i>	Eestikeelsetes kirjeldustes esineb sageli teatav radioloogi kui kirjeldaja esitamine – [mina] ei näe, [mina] ei sedasta, [mina] ei tähelda. Inglisekeelses tõdetakse lihtsalt fakti: <i>there is no...</i>
<b>Rühimuutus</b>	küfoos lordoos skolioos	<i>kyphosis</i> <i>lordosis</i> <i>scoliosis</i>	
<b>Patoloogiliste muutuste kirjeldus</b>	<i>anterolistees</i>  <i>conus medullaris</i> ja <i>cauda equina</i> muutusi ei ole  (diski) väljakummumine, väljavõlvumine (palju harvem võrreldes esimesega – 2 korda)  diski väljakummuvus	<i>anterolisthesis</i>  <i>the conus medullaris and cauda equina are normal</i>  <i>bulging</i>  <i>disc bulge</i>	

diskide veesisalduse langus	<i>dehydration</i>
diskide kõrguste langus	<i>loss of disc height, loss of height</i>
diskide protrusioonid	<i>disc protrusions</i>
dislokatsioon	<i>displacement</i>
duurakoti kompressioon	<i>compression of the thecal sac</i>
foraminaalsed ekstrusioonid	<i>foraminal extrusions</i>
foraminaalsed protrusioonid	<i>foraminal protrusions</i>
hemangioom	<i>haemangioma</i>
intensiivse signaaliga lesioon	<i>high signal intensity lesion</i>
[tekstis ei esinenud – malsegmentatsioon]	<i>disturbance of segmentation</i>
metastaas	<i>metastasis</i>
müelopaatia	<i>myelopathy</i>
neuroforamenite kitsenemine	<i>foraminal narrowing</i>
närvijuurte kompressioon	<i>nerve root compression</i>
närvijuured pitsunud	<i>nerve root flattening</i>
osteoporoos osteoartriit	<i>osteoporosis osteoarthritis</i>
rasvkude kohati kadunud	<i>partial loss of fat planes</i>
retrolistees	<i>retrolisthesis</i>

	Scheuermanni tõbi	<i>Scheuermann disease</i>	
	seljaaju kompressioon	<i>spinal cord compression</i>	
	spinaalkanali stenoos	<i>spinal canal stenosis</i>	
	spondülolüüs	<i>spondylolysis</i>	
	süringomüeelia	<i>syringomyelia</i>	
	Tarlovi tsüstid	näitelauses ei esinenud [ <i>Tarlov cysts</i> ]	
	paksenemine	<i>thickening</i>	
	Schmorli hernia	näitelauses ei esinenud [ <i>Schmorl hernia (body, nodule)</i> ]	
	lülakeha kompressioonmurd (lülakeha kiilukujuline kokkuvajumine)	<i>wedging</i>	
<b>Muutuse lokalisatsioon</b>			
<b>Lüli või lülivahemiku tasand</b>	unkovertebraalne	<i>at the level of uncovertebral subarticular disc</i>	
<b>Lateraalsus</b>	bilateraalne mõlemapoolne lateraalne parempoolne vasakpoolne tsentraalne ristisuunaline	<i>bilateral both involved lateral right-sided left-sided central traversing</i>	
<b>Anterio-posterioorsus</b>	anterioorne, eesmine dorsaalne posterioorne, tagumine	<i>anterior dorsal posterior</i>	
<b>Lõiketasapind</b>	aksiaalne sagitaalne	<i>axial sagittal</i>	
<b>Vertikaalsus</b>	ülemine alumine	<i>upper lower</i>	



<b>Raskusaste</b>			
<b>Muutuse ulatus</b>	kerge mõõdukas raske, väljendunud Modic 1 Modic 2	<i>slight, mild, minor moderate severe grade I grade II</i>	
<b>Muutuse olemus</b>	degeneratiivne hüpertroofia	<i>degenerative hypertrophia</i>	
<b>Signaali intensiivsus</b>	intensiivne madal	<i>high low</i>	

Praktikas on leid sageli koondatud vaid ühte lausesse või pikka nimisõnafraasi, hõlmates seejuures sageli mitmeid eelpool loetletud aspekte. Sellest tulenevalt võib andmebaasi seisukohalt vajalik olla täpsem liigendus ja mitmeetapilisus, kuigi kirjelduse lugejale võib diagnoosimiseks ja ravi planeerimiseks piisavaks ning harjumuspäraseks osutada vaid üks lause.

## 6 ARUTELU

### 6.1 Miks on vaja tõlkida autentseid terviseandmeid?

Tõlkevajadus on seotud kolme asjaoluga. Esiteks liiguvad inimesed Euroopa Liidus palju vabamalt ja rohkem kui varasematel aegadel. Seega ei ole harvad need juhud, kui patsiendi ravi algab ühes riigis ja jätkub teises. Patsiendi terviseteadabest võib oleneda see, kui kvaliteetset arstiabi on talle võimalik anda. Enam ei ole harvad need juhud, kui konsulteeritakse välisriigi kolleege arvuti vahendusel või palutakse neilt konsultatsiooni. Valdavalt toimub selline piiriülene suhtlus inglise keeles, kuigi ka erandeid võib sõltuvalt lähinaaberriikidest esineda.

Teiseks on olukorras, kus Euroopa Liidu tervishoiuteenuse turul on tööjõuressursi jaotus ebahühtlane, tekkinud olukord, kus palju tööd tehakse ka allhanke korras. Eelkõige on selline lähenemine rakendatav diagnostika erialadel, kus spetsialist võib leidu (radioloogilist kujutist, mikroskoobi slide jms) tõlgendada kui kaugest maa tagant tahes. See olukord on leidnud kiiret ärakasutamist radioloogia turul. 2007. a lõppes radioloogia kui tervishoiuteenuse kättesaadavust Balti regioonis parandav edukas projekt Baltic eHealth (<http://www.baltic-ehealth.org/>), mis jätkus kohe uue samalaadse projektiga R-Bay (<http://www.r-bay.org/pages/1>), mille eesmärgiks oli luua radioloogia teenuse e-turg. Projektis osales 11 koostööpartnerit 8 Euroopa Liidu liikmesriigist ning selle käigus selgus, et radioloogiliste kirjelduste tõlkimiseks puudus tõhusa lahendus.

Kolmandaks võib tulevikus Euroopa Liidus ette näha riiklike e-tervise andmebaaside ühendamist nii ressursside juhtimise kui ka epidemioloogiliste ja rahvatervise uuringute eesmärkidel. Seega on tõlkimise probleem tulnud selleks, et jääda.

## 6.2 Kui pädev peab autentse meditsiiniteksti tõlkija olema meditsiinis?

Mis tahes teksti tõlkimine eeldab tõlkijalt nii keelelist pädevust kui ka sügavat ja põhjalikku valdkonna tundmist. Ilma erialase ettevalmistuseta meditsiinis on meditsiiniteksti tõlkimine äärmiselt raske (õigupoolest praktiliselt võimatu, kui tegu ei ole just populaarteadusliku lihtsustatud käsitlusega). Meditsiinitõlget ei saa kuidagi rajada sõnastikes vastete otsimisele ega ekspertide piiramisele, (raamatu)kogemusest ilma süsteemse alusõppeta ei piisa ning arusaama tõlgitava sisust lihtsalt ei teki. Näiteks ühes laboritöö korraldust puudutavas tõlkes räägiti läbivalt „labori aruannete koostamisest“ ja „labori aruannetest“ (*reporting, laboratory reports*), mida tegelikkuses nimetatakse vastuste väljastamiseks ja labori vastusteks. Huvitav, et ka mõte aruannete sagedusest ei aidanud tõlkijat kuidagi: täiesti arusaamatu on ju, miks neid aruandeid nii sageli ja kiiruga (*cito!*) tegema peab, nagu tekstis kirjeldati. Kahjuks ei ole olukord tõlketurul kaugeltki ideaalne ning ikka satub ette kummalisi tõlkeleide, viimasel ajal siiski vähem: meditsiinitekste tõlgivad üha sagedamini arstid. Kuigi ei ole mingit põhjust kahelda selles, et üldjuhul on arsti keelepädevus kordi väiksem kui filoloogil, korvab sügav eriala tundmine tõlke keeleprobleemid ning tubli toimetaja silub konarused.

Autentsete meditsiinitekstide puhul on aga olukord palju keerulisem. Väidaksin käesoleva uurimuse põhjal isegi, et **autentset meditsiiniteksti saab tõlkida ainult see tõlkija, kes suudab seda teksti ise algselt nii lähte- kui ka sihtkeeles luua**. Sellele tõdemusele jõudsin uurides lüüsisamba MRT-uuringute kirjeldusi. Võin tagantjärele öelda, et laboriarstina jääksin nende tekstide tõlkimisega kindlasti hätta. Radioloogiliste kirjelduste tekst on kirjutatud nii spetsiifilises keeles, et ilma selle allkeelee eelneva tundmiseta ei saa sellist teksti emakeeles kirjutada, saati siis tõlkida. Toon siinkohal ühe juhuslikult võetud tavalise uuringu kirjelduse.

## Näide 1

UURING: MRT-uuring lülisamba L-osast.

KLIINILISED ANDMED: *Radiculopathia L4–L5 dex.*

LEID: Lülisamba L-osas intervertebraaldiskides veesisaldus langenud, kõigi intervertebraaldiskide kõrgus madaldunud, L4/L5 intervertebraaldisk puudub – sellel tasemel omandatud osaline luuline blokk tõenäoline. Th8/Th9 tasemel intervertebraaldiski difuusne väljakummumine, millest mõlema lateraalkanali ahenemine, paremal Th8 närvijuurega konflikt tõenäoline. Th9–Th12 tasemel intervertebraaldiskid kerge difuusse väljakummumisega, kuid veenvat kontakti närvijuurtega ei leia. Th12–L3 intervertebraaldiski difuusne väljakummumine, millest mõlema lateraalkanali ahenemine. L3/L4 intervertebraaldiski difuusne väljakummumine, ning paremal intraforaminaalsel tõenäoline protrusioon. Samal tasemel luulist spinaalkanali deformatsioon fassettide ja *lig flavum*ite hüpertroofiast, artroosist. Esineb vasakpoolsete L4, L5 ja S1 ja parempoolsete L4 ja L5 närvijuurte dislokatsioon ventromediaalsele. L4/L5 intervertebraaldisk puudub, omandatud osaline luuline blokk vastavate lülkehade vahel tõenäoline. Spinaalkanali kerge ahenemine hüpertrofeerunud fassettide poolt. Mõlemal pool konflikt L4 närvijuurega tõenäoline. L5/S1 intervertebraaldisk oma anatoomilistes piirides.

KOKKUVÕTE: Degeneratiivsed muutused lülisamba L-osas. Mitmel tasemel mõõdukas spinaalkanali ahenemine, närvijuurte dislokatsioonid.

Tõlkijana näeksin sellise teksti tõlkimisel vaeva nii terminoloogia, sõnastuse kui ka lühendite ning ladinakeelsete väljenditega. Väga keeruline on teada, milline on inglisekeelse autentse radioloogilise kirjelduse tava. Näiteks lause „*Th8/Th9 tasemel intervertebraaldiski difuusne väljakummumine, millest mõlema lateraalkanali ahenemine, paremal Th8 närvijuurega konflikt tõenäoline.*“ on tegelikult lihtsalt pikk nimisõnafraas, kus kõik verbid on välja jäetud ning n-ö üldise meditsiinikeele taustaga (st meditsiinikirjanduse keele põhjal) kirjutaks selle pigem nii: „*Th8/Th9 tasemel esineb intervertebraaldiski difuusne väljakummumine, mis põhjustab mõlema lateraalkanali ahenemist. Paremal on konflikt Th8 närvijuurega tõenäoline.*“ Samas tuleb kohe lisada, et ma ei oska erialase hariduse foonilt kuidagi kommenteerida väljendit „*konflikt närvijuurega*“, mida esines kõikides tekstides kokku 26 (!) korda. Inglisekeelse autentse kirjelduse teksti kohta ei oska midagi oletada, kuid veebiotsinguga selle väljendi suurt kasutust meditsiinikeeles tuvastada ei õnnestunud. (Tuleb muidugi arvestada, et veebis ei olegi autentseid meditsiinitekste kuigi palju võimalik leida.) „*Omandatud luuline blokk*“ võiks tähendada plokkülisid, kuidas oleks seda õige väljendada

autentses inglisekeelses kirjelduses on raske arvata. Fraas „*kuid veenvat kontakti närviuurtega ei leia*“ jätab võimaluse mõelda selle üle, kas seda kontakti tõesti *ei ole* või seda *ei leita*.

Eestikeelne radioloogiline kirjeldus sisaldab palju eripäraseid väljendeid ning on kirjutatud stiilis, mida tekstiloomes ise ei oskaks kasutada. Õigupoolest on kõige lihtsam osa terminid, need üldjuhul ikka leiab või on õpingute käigus ette tulnud. Meditsiinterminite tähendus on enamasti üsna kitsas ning seetõttu on otseselt terminitega seotud mitmetähenduslikkus väiksem kui tavakeele mitmetähenduslikkus. Kuid inglise keelde tõlkimisel tekib muidki küsimusi. Kas diagnoosi peaks kirjutama ladina või inglise keeles, kas lühendid on lubatud ja kas nad on samasugused nagu meil? Kas nt fassetliigeseid lühendatakse ka inglise keeles lihtsalt fassetideks?

Ülaltoodud juhuslik näide oli sisult suhteliselt kergesti mõistetav, tekstide seas oli ka kordades keerulisemaid. Radioloogilise kirjelduse teksti vaadates on selge, et tõlkijat ootavad ees mitte ainult terminoloogiast tulenevad probleemid, vaid palju tõlkimiskäigusi on seotud väga spetsiifilise keelekasutusega.

### **6.3 Milliseid probleeme esineb autentses meditsiinitekstis?**

Selleks, et objektiviseerida autentses tekstis esinevaid probleeme, panin kokku lüüsisamba MRT-uuringutest koosnevast väikesest tekstikogust n-ö pilootkorpuse. Selle käsitlemise käigus tuvastasin mitmeid probleeme, mis on takistuseks teksti tõlkimisel.

#### **6.3.1 Tekstide suur variatsioon – kuidas standardida?**

Üldine väide on, et vabatekstis autentsed meditsiinitekstid järgivad teatud tava ning esinevad lugejale harjumuspärasel kujul (Reiner 2009; Siström 2005). Narratiiv on kriitiline meditsiinilise tõendus põhise hindamisel, raviotsuste tegemisel, meditsiiniteadmiste

vahendamisel ning on sageli täpsem, katvam ja lisateabe esitamisel asjakohasem kui piiratud sõnavaraga struktureeritud tekst. Kindlasti on hästi kirjutatud narratiiv lihtsamini mõistetav, harivam ja isegi veenvam kui struktureeritud andmed. Loomulik keel on kahtlemata kõige väljendusrikkam teabekandja, seda ka meditsiinis (Johnson 2008). Iseasi on, kas vaadanult näidet 1, saaksime seda pidada õnnestunuks loomulikus keeles kirjutatud narratiiviks?

Kõigepealt jäi mulle tekstide läbivaatamisel silma, et nende tekstisisene liigendus oli ebajärjekindel (kuigi kõigis tekstides oli üritatud mingit liigendust järgida), esines väga palju kirjavigu (eelkõige näpuvigu), ladinakeelseid termineid ning väga palju lühendeid. Isegi põgusa vaatamise tulemusel oli selge, et tekst tuleb enne käsitlemist korrastada.

#### **6.3.1.1 Kirjavead**

Üheks suure variatsiooni allikaks olid lihtsalt kirjavead. Iseenesest ei ole kirjavead, eriti näpuvead, kuigi suur probleem. Enamasti mõistame, et tegu on näpuveaga ning saame tekstist ikkagi õigesti aru. Kuid kirjavead muutuvad probleemiks kohe, kui soovime hiljem tekstist midagi üles leida või seda mingil moel arvutis töödelda. Meditsiinkaugele tõlkijale võib aga ka lihtne näpuviga mitmeks päevaks uue termini või õige kirjaviisi väljaselgitamisel tööd anda. Paraku loeb arvuti kirjaveaga sõna lihtsalt teiseks sõnaks. Poola autentsete meditsiintekstide korpuse koostajad pidasid autentses tekstis esinevaid kirjavigu koguni nii tähtsaks, et võtsid kõik vead tekstist välja, kirjeldasid need, analüüsisid ning koostasid asjakohase tekstis esinevate vigade leksikoni (Godlewski 2006). See võimaldas edaspidi parandada neid järgmistes tekstides automaatselt.

Käesolevas töös uuritud tekstides esines näpuvigadest enim klaviatuuril lähestikku asetsevate tähtede segiajamisest tekkinud vigu. Kirjavigadest esines rohkesti kokku- ja lahkukirjutamise

vigu. Eriti palju probleeme paistis tekitavat ühend „*signaali muutus*“, mis esines nii kokku kui ka lahku kirjutatult. Kohati tundus ka mõte muutuvat, sest kas „*patoloogilise signaali muutuseta*“ on sama kui „*patoloogilise signaalimuutuseta*“ ja kas ei peaks olema hoopis „*signaali patoloogilise muutuseta*“? Arvestades, et eesti keeles on liitsõnade loomine väga produktiivne sõnade tuletusviis, oleks tekstides esinenud liitsõnade ja kokku-lahkukirjutamise vigade analüüs huvipakkuv just tähenduse muutumise aspektist. Niipea kui tähendus on vea tõttu lähtetekstis muutunud, on raske eeldada, et see viga sihtteksti ei kandu.

### **6.3.1.2 Lühendid**

Lühendite kasutamine ei ole samuti iseenesest probleem. Lühendid tekstis muutuvad probleemiks hetkel, kui selgub, et sama sõna lühendamiseks on kasutatud erinevaid lühendeid või sama lühendit on kasutatud eri mõistete tähistamiseks, kasutatakse ebataavalisi ning kuskil mujal kirjeldamata lühendeid. Kõiki neid ebajärjekindla lühendamise viise tekstides esines (vt tabel 5 tulemuste osas) ning peale selle oli lühendeid ka lihtsalt tarbetult ning häirivalt palju. Lühendite käsitelamisega jääksid hätta nii tõlkija kui ka arvuti: tõlkijale ei ole lühendid läbipaistvad ning arvuti ei mõista lühendite taga peituvat mitmetähenduslikkust (sellest ei pruugi aru saada ka tõlkija). Probleemi lahendamiseks tundub olevat mõistlik lühendite loetelu põhjal otsustada, millised neist ja millisel kujul kirjutatuna peavad lühenditeks jääma ning seejärel tuleks nad leksikonina kirjeldada ja arvuti kasutamise eesmärgil veel ka märgendada. Lühendid võivad ju lauses sõnadest erinevalt käituda, kuid tähenduse aspektist vaadatuna ei erine nad sõnadest teatava märgisüsteemina millegi poolest. Seetõttu tasuks kaaluda, kas on ikka vaja teksti täissünonüüme, mis ei allu kuigi hästi tavagrammatika reeglitele, nii suurel hulgal juurde tekitada.

### 6.3.1.3 Nimisõnafaasid täislausete asemel

Nimisõnafaas ei ole kuigi tüüpiline nähtus tavalises kirjalikus tekstis, kus enamasti kasutatakse täislauseid. Väga pikad nimisõnafaasid ei ole kuigi iseloomulikud ühelegi loomuliku keele osale. Nende puhul on sageli raske aru saada (ja pea võimatu arvutile selgeks teha), milline sõna millist täiendab. Eesti keeles, kus sõnade järjestus on üldiselt suhteliselt vaba, esineb sageli olukord, kus püsiühendeid moodustavad sõnad asuvad teineteisest suhteliselt kaugel. Kui nüüd ka lausestruktuur enam nende identifitseerimist ei toeta, siis ei pruugi fraasi mõte enam mõistetav olla. Finiitsete e öeldisfunktsioonis verbide kadumine lausest ning rohkete täiendite, sh partitsiip-atribuutide, kasutamine on tegelikkuses vähendanud lauset asendanud fraasi loetavust. Nimisõnafaasi teke on iseenesest mõistetav: radioloog lihtsalt loetleb järjest neid muutusi, mida ta pildil näeb. Üldjuhul ei käsitle vastuse saaja radioloogi kirjeldust sugugi loeteluna. Teatud mõttes on radioloogi kujutise kirjeldamise hetk ehk esimene koht, kus tegelikult oleks kaunis loogiline teksti struktureeritult andmebaasi sisestada, kuid selleks tuleb kõik radioloogilised muutused ja nende seos kliiniliste muutustega määratleda, kirjeldada (ühetaoliselt) leksikaalse üksusena ning andmepuud järgides asjakohasele andmeväljale kanda. Arvutile võimaldaks see loetelu osi leida ja vastuse tekstis vajalikku kohta paigutada. Väljundiks võiks tulemusena olla tunduvalt narratiivsem ja palju väiksema variatsiooniga vabatekst kui praegu, arvutis oleks aga ikkagi struktureeritud andmete esitus.

Omaette huvitav küsimus on, miks ja millal eestikeelsetes autentsetes radioloogiliste uuringute kirjeldustes on hakatud kasutama nimisõnafaase lausete asemel. Meditsiinidiagnoose on pikka aega kirja pandud nimisõnafaasidena. Sageli asendab inglise keeles verbiühendit *to be* lihtsalt koma. Fraasidest pseudolausetel ka suhteliselt ebaloogiline grammatiline struktuur



(Pratt 1969). Allkeeltele on üldiselt iseloomulikud oma süntaktilised konstruktsioonid ning inglise keele mitme valdkonna kohta on tõdetud selliste fraaside kasutamist, seega võib tegu olla ka teatava interferentsiga. Süntaksi erisusi inglise meditsiinitekstides on uurinud Campbell kaasautoritega ning tuvastanud, et meditsiinitekstide süntaks erineb oluliselt inglise üldkeele omast, kuid on seejuures väiksema variatsiooniga ning tõenäoliselt ka lihtsam (Campbell 2001). Oma töös võrdlesid nad Browni inglise keele korpuse (1 000 000 sõna) ja New Yorgis asuva haigla Presbyterian Hospital haiguslugude väljavõtete (1 073 217 sõna) süntaktiliselt analüüsitud ja ühestatud tekste. Nad pöörasid tähelepanu ka asjaolule, et haiguslugude väljavõtted on tavaliselt kirjutatud minevikus, nende inglise keel võib olla lihtsustatud tavakeelega võrreldes ning meditsiininarraatiiv kasutab sageli nimisõna ja eessõna fraase. Nende järeldus oli, et üldkeele tekstidel treenitud keeletehnoloogilised vahendid ei suuda kuigi hästi analüüsida ja ühestada meditsiinitekste. Täiesti vastupidisele järeldusele jõudsid oma töös Wertmer ja Hahn, kes analüüsisid 100 141 sõnest koosnenud erinevatest saksakeelsetest meditsiinitekstidest (haiguslugude väljavõtted, patoloogia vastused, operatsiooniprotokollid) kokku pandud korpust ning jõudsid järeldusele, et üldkeele peal treenitud keeletehnoloogilised vahendid töötasid ka meditsiinitekstides täiesti võrreldava täpsusega (Wertmer 2004). Tulemust on raske seletada muuga kui inglise ja saksa keele autentsete meditsiinitekstide üldkeele süntaksi järgimise erinevusega: tõenäoliselt on saksakeelsed meditsiinitekstit esitatud pigem täislausetena kui nimisõnafrasidena, kuigi siinkohal on see vaid oletus. Samas peab tõdema, et vene keeles kasutatakse fraasipõhist patsiendi seisundi kirjeldamist küll („состояние больного удовлетворительное“).

### 6.3.1.4 Eituse kasutamine

Eituste sagedasele esinemisele inglise meditsiinitekstides on juhitud tähelepanu eelkõige selles kontekstis, et vabatekstist andmekaeve jaoks võib sellest arusaamine, mida õigupoolest eitatakse, osutada keeruliseks. Nii uurisid Chapman jt mis väljendeid millegi eitamiseks autentsetes meditsiinitekstides, sh MRT-uuringute kirjeldustes, üldse kasutati (Chapman 2002). Nende andmed enim kasutatud inglisekeelsetest eitustest MRT-uuringute kirjeldustes on esitatud tabelis 10.

**Tabel 10. MRT-uuringute kirjeldustes enim kasutatud eituste sagedus võrreldes erakorralise meditsiini osakonna haiguslugude (EMO-kaartide), epikriiside, röntgen- ja KT-uuringute kirjeldustega. % näitab, kui suure osakaalu väljend kõigist eituse väljenditest moodustas.**

	MRT %	EMO %	Epikriisid %	Röntgen %	KT %
no	43,4	51,2	41,7	63,2	55,0
denies		21,0	4,9		
without	2,2	5,6	9,3	0,6	6,1
not	3,7	4,9	8,9	1,3	1,0
no evidence	29,5	3,0	5,2	26,7	23,4
with no		2,2	4,6	3,5	3,4
negative for	3,2	4,1	3,5		0,1
denied		1,4	4,8		
to rule out		0,6	2,4	0,1	0,0
no significant	5,7	0,7	0,9	0,1	1,2
without evidence	3,2	0,2	0,5	0,6	3,8
no new		0,1	0,2	0,3	1,6
no abnormal	2,1	0,0	0,0		1,2
no suspicious	2,5		0,0		0,1

Tabelist on näha, et eri liiki autentsete meditsiinitekstide eituste kasutuses on olulisi erinevusi, mida tuleks ilmselt arvestada andmebaaside leksikonide koostamisel. Eitust väljendavaid eesliiteid inglise meditsiinikeeles on uurinud Dzuganova ning näidanud ka eesliidete omavahelist asendatavust teatud juhtudel, mis taas lisab süsteemi tervikuna variatsiooni ning

tõstatab ka küsimuse, millises kontekstis samatähenduslikke eitavaid eesliiteid kasutada (Dzukanova 2006).

### **6.3.1.5 Samatähenduslikud laused**

Sama mõtet väljendavate eri lausete hulk on suur ning see teeb keeruliseks igal eraldi võetud juhul otsustamise, kas laused ikka tegelikult täpselt sama asja tähendavad. Näitena on eespool tulemuste osas toodud tabelis 8 ära kõik laused, mis kirjeldasid luude seisundit. Mitte ainult tõlkimise, vaid ka andmepäringute tegemise suhtes on erakordselt tüsilik, kui sama asja kirjeldatakse erinevate lausetega. Meditsiinitekstis ei ole tegelikult oluline iga hinna eest vältida sõna kordusi ja otsida „ilusamaid väljendeid“, pigem on soovitatav teksti läbivalt kasutada rangelt sama terminoloogiat ja kirjeldusviisi. Peamine on, et kirjeldus oleks tõene ja õigesti (sh arusaadavalt) tehtud.

Kokkuvõtteks võib öelda, et radioloogiliste uuringute standardne esitusviis on väga oluline selleks, et vähendada võimalikku mitmetähenduslikkust ning luua optimaalsed tingimused andmete struktureerimiseks. Mõistlik on kasutada ühesuguseid lühendeid, termineid ja süntaksit. Selleks et mõista, mis konkreetselt vajaks ühtlustamist, on kasulik teha ükskeelne meditsiinitekstide korpus. Tõenäoliselt on andmete standardses esituses osa probleeme sarnased radioloogiliste uuringute kirjelduste ja muude autentsete meditsiinitekstide vahel, osa probleeme võivad olla tekstispetsiifilisemad. Ühtlustamisvajaduse kindlaks tegemise järel on võimalik sõnavara ja selle kasutamist standardida ning selle põhjal luua uued arvutileksikonid. Täiesti selge on ka see, et allkeelega iseloomuliku piiratud leksikoni ning tohutuid andmehulki arvestades võiks tegelikkuses kaaluda masintõlkelahenduste loomist.

## 6.4 Milleks on hea struktureeritud tekst?

Teksti struktuur võimaldab järjestada fakte ning esitada nendevahelisi seoseid. Leksikonist rääkides võivad need seosed olla grammatilised, süntaktilised, semantilised. Praegu võib radioloogiliste vastuste (ja vastamise) puuduste kirjeldamiseks välja tuua järgmised tasandid:

- sisu (kliiniline) tasand, mis on seotud nende faktide ja tähelepanekutega, mida radioloog kirjeldab;
- teabevahetuse tasand, mis on seotud viisiga, mil moel kontekstuaalne teave esitatakse selliselt, et see võimaldaks õige ja õigeaegse kliinilise meetme võtmist;
- analüüsi tasand, mis seotud leiu tõlgendamisega hõlmates pilditöötlust, tehnilisi ja kliinilisi andmeid;
- struktuuri tasand on seotud viisiga, kuidas vastuse tekst luuakse eeldusega, et selle tulemusena sünnib selge ja reprodutseeritav sõnum, mis on arusaadav selle lugejale (Reiner 2010).

Viimasel ajal liigutakse üha patsiendikesksema meditsiini suunas ning seega on avaldatud arvamust, et sõnum peaks olema mõistetav ka erialase ettevalmistuseta. Bruce Reiner on otsesõnu viidanud ka sellele, et reaalsuses ei ole sõnum sugugi alati mõistetav ning radioloogilise kirjelduse vastus tekitab isegi raviarstides ebakindlust edasiste raviostuste suhtes (Reiner 2010). Kahtlemata oleks selle ebakindluse vähendamiseks vajalik püüelda standardsema andmete esitusviisi poole, see aga eeldab omakorda ka standarditult esitatavaid sisendandmeid. Põhimõtteliselt oleks ju võimalik ka praegu proovida NLP-vahendeid tavalise radioloogilise kirjelduse vabateksti analüüsiks. Kuid sellesse teksti on koos vabatekstiga sisse kodeeritud mitmetähenduslikkus, veidrads ja keerulised süntaksistruktuurid ning järelduste tegemine, mis teeb keeletehnoloogiavahendite kasutamise tulemuslikkuse küsitavaks.

Andmebaasil on nagunii protsessipõhine struktuur, st kuskil omaette väljades liiguvad andmed, mis on seotud uuringu tellimise, kliinilise sisendiga (siinkohal patsient), kujutise loomisega, selle töötuse ja kvaliteedis veendumisega, kujutise tõlgendamisega, vastuse kirjutamisega ja suhtluse-teabevahetusega. Juhul kui vabateksti kirjutamise asemel kasutada standarditud struktuuri ja valikuid, oleks kokkuvõttes võimalik luua palju ühtlasemat lõppteksti. Sisuliselt tähendaks see konkreetse allkeele (autentsed radioloogiliste uuringute kirjeldused) omalaadse teadmusbasi loomist. Selleks et see oleks võimalik, tuleb seoses uuringu vastamisega mõista selle kasutuseesmärki (kes, milleks ja millal seda vajab) ja kuidas see mõjutab vastuse teksti. Struktureeritud esitatud andmeid on kergem käsitleda andmebaasis ka selles mõttes, et neid saab seostada vajadusel teiste väljadega – tõlkimise eesmärgi puhul siis asjakohase leksikoni ja grammatikaga. Muidugi on struktureeritud andmed palju paremad mis tahes arvutilahenduste jaoks.

## **6.5 Miks on masintõlge radioloogiliste kirjelduste tõlkimiseks mõistlikum lahendus kui tõlkija palkamine?**

Kõige lihtsam põhjus on see, et autentseid meditsiintekste on väga palju ja nende tõlkimisvajadus järjest kasvab. Teine põhjus on, et meditsiinis – rääkimata sellega seotud autentsest tööolukorrast – pädevaid tõlkijaid ei ole lihtsalt kusagilt võtta. Kolmas põhjus on praeguste kirjelduste suur variatsioon: ühtlustatud viisil esitatud andmetega standarditud leksikoni korral ning piiratud sõnavaraga allkeeles saab arvuti hakkama küll, nt võib tuua Kanadas aastaid toimiva meteoroloogiliste tekstide masintõlkesüsteemi Météo (Somers 2003). Neljas oluline aspekt on, et standarditud teksti saaks käsitleda NLP-vahenditega, seda ka andmekaeve eesmärgil, mis oleks tõlkerakendusega tegelemise lisaväärtus.

## 7 KOKKUVÕTE

Käesolev töö on esimene Eestis, mis käsitleb autentsete meditsiinitekstide lingvistilise analüüsi vajalikkust tõlkerakenduslikel eesmärkidel. Töö käigus on loodud tagasihoidliku mahuga pilootkorpus selleks, et testida korpuspõhise lähenemise võimalikkust ja eeliseid. Töö põhjal võib öelda, et ükskeelne autentsete meditsiinitekstide korpus on vajalik nii masintõlkerakenduste loomiseks kui ka tõlkijale, sest aitab korpuspõhise statistilise lähenemise kaudu paremini välja tuua ning mõista tekstides ja allkeeles tervikuna esinevaid keelelisi probleeme. Osa neist tunduvad olevat universaalsed ja esinevad kirjanduse andmetel veel vähemalt inglise, saksa, poola ja skandinaavia keeltes. Lähtetekstis peituvate tõlkekaride märkamine, mõistmine ja kõrvaldamine on aluseks mis tahes tulemusliku tõlkemehhanismi loomisele. Eriti oluline on lähtetekstide ühtluse suurendamine. Töö üldistusjõudu piirab korpuse väike maht ning väga kitsa valdkonna käsitlemine. Viimast võib siiski pidada representatiivsust ning selle kaudu ka üldistusjõudu oluliselt parandanud teguriks.

Kindlasti edendaks autentsete meditsiinitekstide tõlkevõimaluste uurimist edaspidi ka paralleelkorpus. Kuid lootust, et lähiajal võiks tekkida autentsete meditsiinitekstide suuri paralleelkorpusi, veel ei ole. Samas valmisolekut ja võimalust nende tekstide tõlkimiseks on juba praegu vaja. Haiglate infosüsteemides on tallel mitte ainult suur hulk terviseandmeid, vaid ka tohutu potentsiaaliga hindamatut keeleteavet, mida saaks kasutada lingvistiliseks uurimiseks tekstides esinevate morfoloogiliste, süntaktiliste ja semantiliste eripärade ja seaduspärasuste kirjeldamise kaudu. Meditsiinitekstide eesti keelest võõrkeelde ja võõrkeelest eesti keelde tõlkimisel on lisaks kliinilisele teabele möödapääsmatu ka lingvistilise teabega arvestamine. Alustamine ükskeelsest spetsialiseeritud korpusest võimaldab seda, et tulevikus

on sama allkeele ükskeelsete võrreldavate korpuste olemasolul mitmetes eri keeltes võimalikud ka nende võrdlusuuringud.

Lisaks tõlkimisele on korpuspõhist lähenemist võimalik rakendada lingvistiliseks andmekaeveks. Huvitav on seejuures, et keeletehnoloogiavahendite abil otsitakse, süstematiseeritakse ja osaliselt ka analüüsitakse kliinilist infot. Saksa keele kohta on näidatud, et lingvistiline andmekaeve, eelkõige lemmatiseerimine ja süntaktiline analüüs, võimaldas edukat kliinilist andmekaevet (Volk 2002). Meditsiinitekstide masintõlke eelduste loomine sillutab seega teed ka kliinilisele andmekaevele: mõlemate puhul on tähtis andmete kvaliteet, nende ühtlustatud esitamine (struktuur) ning käsitlemise algoritm.

Tuleb märkida, et kirjalik tõlkimine erialana on palju laiem ja avaram, kui pelgalt tekst teksti haaval tõlkimine. Eriala raamidesse mahuvad ka tekstiloome, tõlketeadus ja lingvistika.

Teoreetilisele keeleteadusele lisaks eristatakse lingvistikas selgelt rakenduslikku suunda – arvutilingvistikat ja keeletehnoloogiat. Tõlketeaduse ja lingvistika rakenduslikuks haruks või vähemalt üheks väljundiks võiks pidada masintõlget, mis on täpselt nii hea või halb, milliseks ta loome. Selleks et autentsete meditsiinitekstide tõlge oleks tulevikus võimalikult hea või üldse võimalik, on meditsiinitekstide analüüsil vajalik arstide, tõlkijate ja lingvistide koostöö.

Terviseandmed on suur väärtus, aga ka suur vastutus. Neid ei tohi lasta tõlkes kaduma minna.

## 8 KASUTATUD KIRJANDUS

Baltic eHealth. (<http://www.baltic-ehealth.org/>, 05.05.2010)

Berghammer, Gabi. 2008. *English as the lingua franca of science: A translator's view on what's lost – and what's gained – in translation*. The Journal of the European Medical Writers Association 17: 213–219

Campbell David A. and Stephen B. Johnson. 2001. *Comparing Syntactic Complexity in Medical and non-Medical Corpora*. Proc AMIA Symp: 90–94

Chapman Wendy W., Will Bridewell, Paul Hanbury, Gregory F. Cooper and Bruce G. Buchanan. 2002. *Evaluation of Negation Phrases in Narrative Clinical Reports*. Proc AMIA Symp: 105–109

COMMUNICATION FROM THE COMMISSION TO THE COUNCIL, THE EUROPEAN PARLIAMENT, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS. 2004. *e-Health – making healthcare better for European citizens: An action plan for a European e-Health Area*. KOM (2004) 356

Dzukanova Bozena. 2006. *Negative Affixes in Medical English*. Bratisl Lek Listy 107: 332–335

Erelt, M., Erelt, T., Ross, K. 2007. *Eesti keele käsiraamat*. Komas trükk. Eesti Keele Sihtasutus, Tallinn

*Getting Started in Clinical Radiology*. 2006. Thieme, Stuttgart & New York. pp 120–167

Godlewski, Grzegorz, Maciej Piasecki and Jerzy Pejcz. 2006. *Corpus of Medial texts and tools*. Proceedings of Medical Informatics and Technologies 2006, Silesian University of Technology

([http://plwordnet.pwr.wroc.pl/main/content/files/publications/226\\_Piasecki\\_Maciej.pdf](http://plwordnet.pwr.wroc.pl/main/content/files/publications/226_Piasecki_Maciej.pdf), 04.05.2010)

Haarala, Risto ja Timo Kivelä. 1999. *Duodecim-seura lääketieteen suomen kielen vaalijana ja kehittäjänä*. Duodecim 115: 456–464

Hennoste, Tiit ja Kadri Muischnek. 2000. *Eesti kirjakeele korpuse tekstide valiku ja märgendamise põhimõtted ning kahe allkeele võrdluse katse*. Arvutuslingvistikalt inimesele. Tartu ülikooli üldkeeleteaduse õppetooli toimetised: 183–217

Hunston, Susan. 2006. *Corpora in Applied Linguistics*. 4<sup>th</sup> printing. Cambridge University Press, Cambridge UK. pp 14–16



- Johnson Stephen B., Suzanne Bakken, Daniel Dine, Sookyung Hyun, Eneida Mendonça, Frances Morrison, Tiffani Bright, Tielman Van Vleck, Jesse Wrenn, Peter Stetson. 2008. *An Electronic Health Record Based on Structured Narrative*. Journal of the American Medical Informatics Association 15: 54–64
- Kaalep Heiki-Jaan. 1998. *Eesti keele ressurside loomine ja kasutamine keeletehnoloogilises arendustöös*. *Dissertationes Philologiae Estonicae Universitatis Tartuensis*, Tartu Ülikool.
- Kreutzwald, Friedrich Reinhold. 1900. *Kodutohter*. Neljas trükk. Echnafenburg'i trükk ja tulu, Jurjew (Tartu).
- Meditsiinisõnastik*. 2004. Teine trükk. Toimetajad Sirje Ootsing, Laine Trapido. Medicina. Tallinn.
- MRI Parameters and Positioning*. 2003. Thieme, Stuttgart & New York. pp 140–164
- Muischnek Kadri, Heili Orav, Heiki-Jaan Kaalep, Haldur Õim. 2003. *Eesti keele tehnoloogilised ressursid ja vahendid*. *Arvutikorpused, arvutisõnastikud, keeletehnoloogiline tarkvara*. Eesti Keele Sihtasutus.
- Mäkinen, Martti. 2006. *Between Herbals et alia: Intertextuality in Medieval English Herbals*. Master Thesis, University of Helsinki.
- Nyman, Hans. 1999. *Engelska hotet mot svenskan i medicinen – oroande eller ofarligt?* Läkartidningen 96: 4720
- Nyman, Hans. 2000. *Engelska termer kan anpassas till svenskan*. Läkartidningen 97: 4609–4610
- Nyman, Hans. 2003. *Medicinspråkets utveckling från Hippokrates till vår tid*. Läkartidningen 100: 339–340
- Nyman, Hans. 2006. *Inte bara patienter ... även latinet kan behandlas fel*. Läkartidningen 103: 4071–4072
- Orav Heili ja Kadri Vider. 2006. *Millist leksikoni vajab arvuti tähenduse mõistmiseks? Keel ja arvuti*. Tartu Ülikooli kirjastus. Lk: 85–97
- Pratt, Arnold W. and Milos G. Pacak. 1969. *Automated processing of medical English*. Proceedings of the 1969 conference on Computational linguistics: 1–23
- Radiographic Anatomy*. 2000. Thieme, Stuttgart & New York. pp 40–79
- Radioloogia*. 1995. Medicina, Tallinn. Lk 342–359; 533–544
- R-Bay. (<http://www.r-bay.org/pages/1>, 05.05.2010)

- Rein, Kaarina. 2004. Ladina keele vajalikkusest arstiteaduses meditsiiniüliõpilaste hinnangute põhjal. *Eesti Arst* 83: 730–737
- Reiner, Bruce I. 2009. *The Challenges, Opportunities, and Imperative of Structured Reporting in Medical Imaging*. *Journal of Digital Imaging* 22: 562–568
- Reiner, Bruce. 2010. *Uncovering and Improving Upon the Inherent Deficiencies of radiology Reporting through Data Mining*. *Journal of Digital Imaging* 23: 109–118
- Ross, Peeter. 2008. *Teleradioloogia hetkeseis ja võimalused*. *Eesti Arst* 87: 731–732
- Ross, Peeter, Ruth Sepper and Hanna Pohjonen. 2010. *Cross-border teleradiology – Experience from two International projects*. *European Journal of Radiology* 73: 20–25
- Schønnemann, Ulrik. 2008. *IT support of radiology outsourcing*. Master Thesis, University of Southern Denmark
- Sinisalu, Väino. 2005. *Kas eesti keel säilib teaduskeelena?* *Eesti Arst* 84: 749–750
- Sistrom Chris L. and Janice Honeyman-Buck. 2005. *Free Text Versus Structured Format: Information Transfer Efficiency of Radiology Reports*. *American Journal of Roentgenology* 185: 804–812
- SNOMEDi veebileht. (<http://www.ihtsdo.org/snomed-ct/>, 04.05.2010)
- Somers, Harold. 2003. *Sublanguage*. in *Computers and Translation: A translator's guide*. John Benjamins Publishing Company, Amsterdam/Philadelphia: 283–285
- Spinal Imaging*. 2008. Thieme, Stuttgart & New York
- Taavitsainen, Irma and Päivi Pahta. *Scientific Thought Styles: The Evolution of Early English Medical Writing*. University of Helsinki, Corpus of Early English Medical Writing (<http://www.helsinki.fi/varieng/CoRD/corpora/CEEM/STSPProject.html>, 15.05.2010)
- Taavitsainen Irma and Päivi Pahta. 2009. *Medical and Scientific Writing in Late Medieval English*. Cambridge University Press. New York. pp 1–18
- The PloS Medicine Editors. 2006. *Ich Weiss Nicht Was Soll Es Bedeuten: Language Matters in Medicine*. *Plos Medicine* 3: e122 (<http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.0030122>, 15.05.2010)
- The Royal Society. *History of the Royal Society*. (<http://royalsociety.org/History-of-the-Royal-Society/>, 15.05.2010)
- UMLSi veebileht. (<http://www.nlm.nih.gov/research/umls/>, 04.05.2010)

Valdes, Albert ja Johannes Voldemar Veski. 1982; 1983. *Ladina-eesti-vene meditsiinisõnaraamat I–II*. Valgus. Tallinn

Veskis, K. 2009. Ajakirja Eesti Arst (2002–2004) korpus (<http://www.cl.ut.ee/korpused/segakorpus/eestiarst/>, 04.05.2010)

Volk, Martin, Bärbel Ripplinger, Špela Vintar, Paul Buitelaar, Diana Raileanu and Bogdan Sacaleanu. 2002. *Semantic Annotation of Concept-Based Cross-Language Medical Information Retrieval*. International Journal of Medical Informatics 67: 1–16

Vääri Eduard. 1983. *Loenguid keeleteadusest*. Tartu Riiklik Ülikool, soome-ugri keelte kateeder, teine trükk. Lk 10

Wermter, Joachim and Udo Hahn. 2004. *Really, Is Medical Sublanguage That Different? Experimental Counter-evidence from Tagging Medical and Newspaper Corpora*. MEDINFO – Proceedings of the 11th World Congress on Medical Informatics: 560–564

Wulff, Henrik R. 2004. *The language of medicine*. Journal of the Royal Society of Medicine 97: 187–188

Xiao, Richard and Ming Yue. 2009. *Using Corpora in Translation Studies: The State of Art*. in Contemporary Corpus Linguistics. Baker, Paul and contributors. Continuum International Publishing Group, London & New York, GB: 237–261

## 9 LISAD

### **Lisa 1. Inglisekeelsetes lülisamba MRT-uuringute kirjeldustes kasutatavad laused**

*Normal craniocervical junction.*

*The discs and the vertebrae appear normal. There is no spinal canal or foraminal narrowing. No evidence of spinal cord compression. No signal alterations in the cervical (thoracic) spinal cord.*

*The discs and the vertebrae appear normal. There is no spinal canal or foraminal narrowing. No evidence of nerve root compression. The conus medullaris and cauda equina are normal.*

*There is straightening of the cervical lordosis.*

*There is straightening of the lumbar lordosis.*

*There is a reversal of the normal cervical lordosis, with slight/mild/moderate/severe kyphosis. The thoracic kyphosis is exaggerated.*

*There is a left/right (mild/moderate/severe) convexity scoliosis.*

*There is (slight) retrolisthesis of A1 on A2.*

*There is slight/grade I/grade II etc degenerative anterolisthesis of A1 on A2.*

*There is slight/grade I/grade II etc anterolisthesis of A1 on A2 due to bilateral spondylolysis.*

*There is mild/moderate/severe wedging of A1 probably due to osteoporosis/consistent with previous trauma/probably due to previous Scheuermann disease.*

*There is possible disturbance of segmentation. It has been assumed that the lowest axial images are at A5/B1. If surgery is contemplated, this should be verified radiographically.*

*There is a benign haemangioma of A1.*

*No signal changes in the vertebral bone marrow. No evidence of metastatic disease.*

*There are signal changes in the vertebrae in keeping with bone metastasis.*

*There is a high signal intensity lesion in the spinal cord (at this level, at A1/A2, at the level of A1) in keeping with cervical/thoracic myelopathy.*

*There is a high signal intensity centrally in the spinal cord extending from A1 to A5 consistent with syringomyelia.*

*There is a high signal intensity centrally in the spinal cord extending from A1 to A5 consistent with wide central canal which is a normal variant.*

*There are minor/mild/moderate/severe degenerative changes from A1 to A4/throughout the (lower thoracic and) lumbar spine with...*

*...dehydration, (loss of height) and circumferential bulging of the discs/posterior disc protrusions (as well as signs of uncovertebral/facet joint osteoarthritis). Note: numerous combinations of those expressions are possible).*

*At A1/A2, there is loss of disc height.*

*At A1/A2, there is hypertrophy of the ligamenta flava. Thickening of the ligamenta flava is noted at A1/A2. Note: „is noted“ may replace „there is“ to avoid repeating the same expression.*

*There is mild/moderate left sided/right sided/bilateral foraminal narrowing at C2/3 and the left/right/both foraminal C3 nerve root/roots are possibly/probably involved.*

*There is severe left sided/right sided/bilateral foraminal stenosis at C2/3 and the left/right/both foraminal C3 nerve root/roots are almost certainly involved. There is severe left sided/right sided/bilateral foraminal stenosis at C2/3 causing compression of the left/right foraminal C3 nerve root/foraminal C3 nerve roots bilaterally.*

*At C2/3, there are signs of uncovertebral and facet joint osteoarthritis, more so on the left. As a result, there is mild/moderate left sided foraminal etc (see previous paragraph).*

*There is a posterior disc protrusion indenting mildly/moderately the dura but there is no evidence of spinal cord compression.*

*There is a posterior disc protrusion impinging mildly upon the anterior aspect of the spinal cord but there is no evidence of compression.*

*The disc bulge obliterates the anterior subarachnoid space but there is no evidence of spinal cord compression.*

*There is a central disc hernia/extrusion/protrusion/ at C2/3 causing compression of the spinal cord. However, there is no evidence of myelopathy at this level.*

*At C2/3, there is a right-sided foraminal/subarticular disc protrusion/hernia/extrusion impinging upon the right/left traversing/foraminal C3 nerve root.*

*There is mild/moderate left sided/right sided/bilateral foraminal narrowing at L3/4 with partial loss of fat planes around the nerve root but there is no definite evidence of nerve root compression.*

*There is left/right sided/bilateral foraminal narrowing at L3/4 with almost complete loss of fat planes around the nerve root and the left/right/both foraminal L3 nerve root/roots is/are almost certainly involved.*

*The disc bulge has a right sided/left sided predominance and as a result, there is left sided/right sided foraminal narrowing... (see previous paragraph).*

*There is left sided/right sided foraminal narrowing with nerve root flattening indicating probable involvement of the nerve root.*

*There is mild/moderate spinal canal narrowing causing mild/moderate compression of the thecal sac.*

*There is spinal canal stenosis causing compression of the thecal sac and the nerve roots within it (as well as both traversing Lx nerve roots).*

*The disc bulge has a right sided/left sided predominance and touches the right/left traversing Lx nerve root but there is no definite evidence of compression.*

*The disc bulge has a right sided/left sided predominance and there is mild displacement of the right/left traversing Lx nerve root.*

*The disc bulge has a right sided/left sided predominance and impinges mildly upon the right/left traversing Lx nerve root.*

*There is a left-sided/right-sided subarticular disc protrusion/extrusion touching the right/left traversing Lx nerve root but there is no evidence of compression.*

*There is a left-sided/right-sided subarticular disc protrusion/extrusion displacing the right/left traversing Lx nerve root.*

*There is a left-sided/right-sided subarticular disc protrusion/extrusion causing compression of the right/left traversing Lx nerve root.*

*There is a left-sided/right-sided subarticular disc protrusion/extrusion and facet joint osteoarthritis has right/left sided predominance. As a result, there is compression of the right/left traversing Lx nerve root.*

## Lisa 2. Mõttekriipsu kasutamine lülisamba MRT-uuringute kirjeldustes järeldeste esitamiseks

(Lauseste kirjaviis muutmata.)

1. Eesmine pikiligament selgelt ei visualiseeru, T1 kujutisel on antud piirkonnas signaalitõusuga moodustis – tõenäoliselt ligament rebenenud.
2. Th7/Th8 lülide eesservas vähene signaalitõus STIR ja T2 kujutistel – tõenäoline stressreaktsioon.
3. Th8 lüli *proc spinosuse* keskosas on ristisuunaline defekt, mis visualiseerub mitmes lõigus kõigis uuringu sekventsides – jääb kahtlus murrule, kuigi varasemas KT-uuringus see ei visualiseeru.
4. Mõlemal pool *m psoastes* enam paremal mitmed piirdunud T2 intensiivse signaaliga ja T1 madala signaaliga alad – abtsessideks sobivad muutused.
5. Anterioorsel epiduraalsel L2–L3 lülkehade ulatusel subligamentoosel T2 kujutistel inhomogeense valdavalt intensiivse signaaliga ja T1 madala signaaliga ja perifeerselt kontrasteeruv lisamass – epiduraalsele abtsessile sobiv leid.
6. Veenvat lülkehade nihkeid ei ilmne. L1 lülkehade kõrgese langus, ülemine lõpp-plaat konkavaalne – võimalik varasem kompressioon fraktuur.
7. Uuritud ulatuses lülkehade lõpp-plaadid ebatasased – Schmorli herniad, lülkehadel servmised osteofüüdid, L2–L5 diskid madaldunud, fassetliigete hüpertroofia.
8. Th11–L3 lülkehade lõpp-plaadid kergelt ebatasased – Schmorli herniad.
9. Kl diagnoos – neuroborrelioos?
10. Obj nimme liikuvus taha piiratud, Laseque – bil Ach bil ei vallandu.
11. L1, L2 ja alumiste torakaallülide – Th10–Th12, lõpp-plaadid on ebasiledad, väikeste Schmorli noodulitega ja kohati ka vähest subkondraalset signaalimuutust.
12. Spinaalkanalis S1 kõrgusel keskel veidi vasakul on 1,3 x 0,6 cm vedelikusignaaliga moodustis – Tarlovi tsüst.
13. L2/L3 tasemel dorsaalsel fassetide piirkonnas vedelikusignaaliga ala – lisavedelik fassetliigesesiseselt?
14. L2–L3 diski keskosas on väike intensiivsema signaaliga ala – fibroosrõnga rebend?
15. 10 min käib – valud jalgades.
16. Ka ülejäänud lumbaallülides on väikseid intensiivsemaid koldeid – nii hemangioome kui ka lõpp-plaatide reaktiivset signaalimuutust.
17. Kaelavalud – > paremasse kätte > vasakusse C6/C7 tasemel ka vasakpoolset neuroforamenit.
18. L4 lülkehade dorsaalsel lülkehade toitva arteri ümber ca 2 cm rasvasignaaliga ala – tõenäoliselt tegemist hemangioomiga.
19. Ka siin lisaks *lig flavumite* hüpertroofia, enam paremal, samas vedelikusignaaliga moodustis 6 mm – sünoviaalne tsüst, mis samuti süvendab lülisambakanali ahenemist.

20. Sakrumi keskosa tasandil posterioorselt on 2 vedelikutihedusega moodustist 4 mm ja 9 mm – nn Tarlovi tsüstid.
21. L4/L5 on mõõdukat degeneratiivsed muutused: diskide dehüdratatsiooni tunnused, diskide tsirkulaarse tüüpi väljakummumised; fassetliigestes mõõdukad artrootilised muutused – need muutused olulist spinaalkanali ega lateraalkanalite ahenemist ei põhjusta, seljaaju ega närvijuurte kompressioonile viitavat ei ole.
22. L3–L5 tasemel lülkehade tagaseina kulgu mööda ekstraduraalsel rasvkoe signaaliga alad (promineeruv epiduraalne rasv – epiduraalne lipomatoos?).
23. L5 nihkunud taga ~ 3–4 mm S1 suhtes. Th12 vasakus pediiklis, L2 ja L5 on T1 ja T2 tõusnud signaalga kolded, samas STIR sekventsil nad on hüpointensiivsed – tõenäoliselt hemangioomid.
24. Kõigis lumbaallülides ning sakraallülides ja vaagnaluudes nähtavas osas on hüperintensiivsed kolded T1 ja T2 sekventsidel – tõenäoliselt hemangioomid.
25. MR-müelogrammil L3/L4 tasemel liikvorisignaali puudub – tõenäoliselt artefakt.
26. Th11 lülkehase ~ 0,9 cm suurune T1 ja T2 kujutistel intensiivse signaaliga ala – hemangioomile sobiv muutus.
27. L3 lülkehase praktiliselt kogu lülkehase haarav T1- ja T2- kujutistel intensiivse signaaliga ala, milles patoloogilist kontrasteerumist ei sedasta – hemangioomile sobiv leid.
28. L4/L5 intervertebraaldiski difuusne väljakummumine, väljendunud assümmeetriline (dex > sin) lig flavumite hüpertroofia, paremal tõenäoliselt ka luulist produktsiooni fassetliigeste piirkonnas – sellest mõõdukas assümmeetriline spinaalkanali ahenemine, *cauda equina* deviatsioon vasemale, kompressioon parema L4, L5 ja tõenäoliselt ka S1 närvijuurtele.
29. Võrdluseks on 2005 a uuring pildiliselt ja 2006 a uuringu kirjeldus – vana kolde osas on leid praktiliselt muutusteta.
30. Parema reieluu ülaosas, petrohhanterisel dorsaalsel on ovaalne, lainja kontuuriga inhomogeense struktuuriga, tõenäoliselt vedeliku + rasva signaaliga kolle 2,6 x 1,6 x 2,7 cm, luukontuurilt ei kummu – sobiks tsüstile või fibroossele düsplaasiale, maliigseks ei arva.
31. Parema S1 närvijuure retsessi piirkonnas 1cm suurune vedeliku signaaliga struktuur – Tarlovi tsüst, mis võib komprimeerida S1 dex närvijuurt.
32. Esineb tsentraalne ekstrusioon, lühikesel alal võimalik kontakt mõlema L5 närvijuurega – küsitav.
33. Lülisamba L-osas intervertebraaldiskides veesisaldus säilinud, v-a L3/L4 ja L4/L5 – madaldunud veesisaldusega. Intervertebraaldiskide kõrguse muutust ei sedasta.
34. L4/L5 intervertebraaldisk kerge difuusse väljakummumisega, paremal ekstraforaminaalsel võimalik kontakt L4 närvijuurega – leid küsitav.
35. Luutursele iseloomulikku signaali ei ilmne – tõenäoliselt suhteliselt vana muutus.



36. L1 lülakeha eesmises osas madaldunud kõrgusega, eesmine ülemine apofüüs distntseeritud – põhjuseks tõenäoliselt eesmine Schmorli hernia kaudaalses suunas Th12/L1 tasemel.
37. L4/L5 diski laiapõhjaline väljakummumine, diski tagaservas signaali intensiivsuse tõus – *anulus fibrosuse* rebendile sobiv leid; samas spinaalkanal ja neuroforamenid olulise ahenemiseta.
38. Kogu uuritud ulatuses lülakehade lõpp-plaadid ebatasased – Schmorli herniad.
39. L4/L5 diski laiapõhjaline väljakummumine, diski dorsaalses servas signaali tõusuga triip – *anulus fibrosuse* rebendile sobiv leid.
40. Lülisamba torakaalosas Th5 lülakehas 8mm hästi piirdunud ümarkolle, mis T1 hüpo- ja T2 hüperintensiivse signaaliga – luutsüstiks sobiv.
41. Th12/L1 lülakehade eesosas Modic 1 tüüpi degeneratsiooni – lülakehade anterioorsetel servadel mõõdukad osteofüüdid ja luuturset.
42. L4/L5 intervertebraaldiski tsentraalne protrusioon, samas lokaalne signaali intensiivistumine – fibroosrõnga rebend.
43. Pikemat aega kestnud kaelavalud, ka öövalud – > diskopaatia?
44. L3/L4 põhjustavad diski difuusne väljakummumine, fasettliigete ja flavumite hüpertroofia spinaalkanali vähese kitsenemise – AP-mõõt kitsaima koha peal ca 10 mm.
45. L3–S1 tasemel ka diskide kõrguse madaldumine, eriti L4/L5 – diski kõrgus praktiliselt puudub.
46. L1–L5 erinevas staadiumis lõpp-plaatide degeneratiivsed muutused, fasettliigete hüpertroofia ning diskide degeneratiivsed muutused – veesisaldus langenud, L2/L3 ja L4/L5 ka tugev kõrguse langus.
47. Lihastoonuse puudus jalgadel ja mingi hüpalgeesia – võimalik spinaalne (seljaaju alumise osa vas pat).
48. Põis korras – *conus m sp* sündroomi ei ole.
49. L4/K5 on samadest muutustest tingitud spinaalkanali tugev kitsenemine – AP-mõõt kitsaima koha peal 5 mm.
50. Vastuoluline leid – Th6 muutused dünaamikas endised: luuüdi turse võiks olla beniigsel kompressioonfraktuuril taandunud, samas maliigsed muutused peaks selle aja jooksul olema süvenenud.
51. Ka nimmeosas midagi uut ei ole – Th12/L1 Schmorli herniad, kolme alumise nimmediski degeneratiivsed muutused ei ole süvenenud ega vähenenud.
52. Pead liigutada ei saa – valu kaelas.
53. Vasakul *os ischii* ca 0,8 cm suurune sklerootiline ala – ilmselt luusaareke.
54. Pt-l jalgades nõrkus, paraparees – > al 2006 a sügis.
55. Presakraalvahemikus ka lõpp-plaatide poolset signaali muutus – rasvdegeneratsioon.
56. MRT-uuring – lumbaalosa + SIL.
57. Diskide kõrgus säilinud, v.a L5/S1 – madaldunud.

58. Lülkehade kõrgus on normis, L4/L5 lõpp-plaatidel on degeneratiivset tüüpi signaalimuutust, Th10 lülkehase on väike hemangioom. L4/L5 diski signaaliintensiivsus on T2 kujutistel langenud – degeneratiivne muutus.
59. Kaasuvana on vasakul *tuber ischiaticuse* ning Hamstringi kõõluste ümber vedelikku ja turset – Hamstringi kõõluste tendinoos.
60. Puusaliigete liigesepilud on ahenenud, liigesepindadel on servmised osteofüüdid; dex > sin subkondraalsed lesioonid – artroos.
61. Vasakul on puusaliigese kõrgusel vedelikukogum *m iliopsoasest* mediaalsemal – *iliopsoas* bursiidile sobiv leid.
62. Teostatud tugeva väljaga MRT-uuring peaaugust ja lülisamba kaelaosast natiivis ja kontrastainega kahe mähisega: peaaugu – T2 ax, T2 Flair ax, T1ax; T2 cor; T1 sag; DWI, ADC ax; post k/a: T1 ax,sag,cor. l/s kael: T1 sag, T2 sag, T2 Spair sag; T1 ax, T2 Ax; Post k/a T1 sag, T1 ax).
63. Vasakul SI liiduse eesosas sakraalsel kaudaalsel on väike T2 SPAIR intensiivne ala, milles k/a-ga uuringul ka kerge kontrasteerumine – tõenäoliselt põletikuline fookus.
64. L5/S1 on diski kõrguse langus ja dehüdratatsioon, lülkehades on lõpp-plaatide lähedal degeneratiivset tüüpi signaali muutust – kerged degeneratiivsed muutused.
65. *Sacrumis*, niudeluudes ja reieluu kaeldes kontrasteeruvad lüütilised kolded – antud uuringul ei saa eristada kas tegemist müeloomi või metastaatiliste kolletega – diagnoosi täpsustamiseks mõeldav biopsia luukoldest.
66. T1-l enam L4/L5 vahemiku poolset hüpointensivsemat signaali – luuüdi tursele isel leid.
67. L5 parempoolne ILE. L4, L5 lülkehades luuüdi turset – ilmselt stressreaktsioon L4/L5 muutustele.
68. Pt-l 2007 a. diagnoositud ekstramedullaarne seljaajukasvaja kaela piir-s (*tu reg craniovertr benignum – meningioma psammomatotum*).
69. Op-tud 11/03/08 – C1/C2 laminektoomia.
70. Tagumises koljukoopas basaalsel (osaliselt oksipitaalluus) kuni C2 kaare ja *proc spinosuseni* on ebakorrapärane ~ 2,3 x 3,7x 2,6 cm suurune defekt – op-i järgne muutus.
71. Samas patoloogilist kontrasteerumis, lisamassi – retsidiivile sobivat nähtavale ei tule.
72. C1 kaar on eemaldatud – laminektoomia.
73. Tagumises koljukoopas basaalsel (osaliselt oksipitaalluus) kuni C2 kaare ja *proc spinosuseni* on op-i järgne muutus (C1 kaar on eemaldatud – laminektoomia).
74. Lisaks võrreldes eelmise uuringuga L3/L4 diski kõrguse süvenenud langus, lülkehade lõpp-plaadid on ebatasased ja esineb mõõdukas signaali tõus ning lülkehade kontrasteerumine – äge lõpp-plaadi degeneratiivne muutus, võimalik, et on tekkinud L3/L4 opi järgse defekti tõttu ebastabiilsus ja minimaalne listees; samas ei saa välistada põletikulist muutust (kliinik?).
75. Vasak L5/S1 neuroforamen on oluliselt ahenenud, vasaku L5 – närvijuure kompressioon tõenäoline.

76. Sellel tasemel väljendunud fassetide hüpertroofiast, flavumite paksenemist oluline spinaalkanali ahenemine – vaba kanali mõõt ca 7 mm.
77. Eelmise uuringuga võrreldes turse S1–S2 tasemel – sobiks eeskätt stressreaktsioonile.
78. L4/L5 diski keskosa tasemel spinaalkanali ahenemine – vaba kanali AP mõõt ca 0,8 cm, SD mõõt ca 0,8 cm.
79. Parema fassetliigese läheduses visualiseerub ca 0,8 cm suurune ümarkujuline struktuur – sünoviaalne tsüst?
80. L5/S1 tasemel parema fassetliigese läheduses – > S1 dex juure pitsumine.
81. L4/L5 diski tsentraalne protrusioon, millest duurakotti deformatsioon; spinaalkanali ahenemine – AP mõõt ca 0,9 cm.
82. Pidevad alaseljavalud, teatud asendis – tekib jalavalu (vasak), kõndimisel jalad muutuvad nõrgaks.
83. C5/C6 diski parempoolne paramediaalne protrusioon, mis ahendab ka parempoolset lateraalkanalit ja on kontaktis C6 dex juurega – kompressioon tõenäoline.
84. L3/L4 diski parempoolne intraforaminaalne protrusioon – kontakt parema L3 juurega olemas, pitsumist siiski pole.
85. Samas patoloogilist kontrasteerumist, lisamassi – retsidiivile sobivat nähtavale ei tule.

## 10 SUMMARY

### MONOLINGUAL SPECIALISED CORPUS IN TRANSLATION STUDIES: EXAMPLE OF RADIOLOGY REPORTS

**Eola Valdre**

**Keywords:** history of medical sublanguage, medical Estonian, authentic medical text, corpus studies, computer-assisted translation, magnetic resonance imaging (MRI) of spine, radiology reports: free-text reporting and structured text

v+91 pages, in Estonian. Includes 1 figure, 10 tables, 2 annexes and 51 references.

Current thesis presents a view and an example of how a monolingual specialised corpus can be implemented in translation studies, particularly how to establish deviations of a specific sublanguage from the conventional grammar and vocabulary.

The study is focused on a subset of the authentic medical texts consisting of radiology reports. The need for the translation of authentic medical data is growing rapidly in Estonia due to the increased mobility of people within the European Union, cross-border outsourcing of some e-health based health care services like radiology, frequent communications and consultations with colleagues abroad. The growth is significantly accelerated also by the establishing of the European e-Health Area.

The corpus-based approach was used to determine and evaluate possible deviations from standard Estonian grammar, also noting the differences in vocabulary. The pilot corpus of MRI reports of the spine imaging is the first systematised collection of authentic medical texts in Estonia used for the linguistic analysis. Corpus-based approach has not been used so far to analyse authentic medical texts for translation purposes, this study is the first one done on this subject in Estonia.

A small, yet representative pilot corpus was used for the initial evaluation of the approach. De-identified MRI reports were obtained by querying all MRI reports of a given period. They comprised of 87,824 tokens. However, to overcome issues of the small text volume and questionable representativeness, only reports on spinal imaging were included (about 30% of the whole file). The final text file consisted of 34,215 tokens. After manually removing redundant subsections names, making some rearrangements (determining of the clause boundaries), adding punctuation and correcting the spelling, 24,892 tokens and 3211 words were left. This file was analysed with AntConc 3.2.1w 2007 program for the frequencies of words and expressions. Where applicable, they were compared to previously published frequencies established for the Estonian written media and literature texts.

The main findings are as follows:

1. The free-text of radiology reports is by its nature highly unconventional.
  - 1.1. The text contained many unconventional and redundant abbreviations, meaning that for a single concept there may be several abbreviations and vv. The methods of abbreviating were non-standard: many words were abbreviated simply because of time pressure, and in completely random and unpredictable manner. A list of all abbreviations used in reports was made.
  - 1.2. There were numerous spelling mistakes, many of them due to mistyping.
  - 1.3. The mistakes in writing of compound words were also abundant, sometimes leading to the shifts in the meaning of the sentence.
  - 1.4. Estonian and Latin terms were used in parallel, meaning that some radiologists preferred Latin terms over Estonian and vv. Latin terms were often misspelled.

1.5. The most striking difference from natural language and any conventional written texts was extremely high occurrence of very long noun phrases, often separated by dashes. The latter usually was used to show that the following part of the noun phrase contained some kind of clinical conclusion.

1.6. Compared to the conventional written text, the complex verb phrases and/or full sentences were rare. Besides, there were many participle verb forms. They are often used in Estonian as adjectives and at least all present participle forms indicate the function of an attribute.

1.7. The use of negations was very frequent, at least three times more frequent than for the media or literature texts written in Estonian.

1.8. Anaphors were extremely rare.

2. The variation of reports from linguistic point of view was considerable.

3. The variation in the structure of the reports was also remarkable.

4. Across the texts, there was considerable redundancy in expressing the same meaning.

Based on the analysis of this subset of authentic medical texts, the main conclusion is that the medical language used for the communication strictly between health professionals is highly specific and very unconventional if compared to other written Estonian texts. To be able to deal with all potential translation issues, the sources of variation in the texts should be established and reduced. For this purpose, a bigger and more representative corpus including also other types of authentic medical texts in Estonian is essential. The question of annotation of this corpus remains to be solved in the future, but it is obvious that it will be needed along with specific lexicon for abbreviations and Latin terms, maybe even for

the spelling mistakes, to successfully implement available natural language processing tools. Without any linguistic analysis the task of translating authentic medical texts is nearly impossible to accomplish, both for a human translator or a computer. Establishing monolingual corpora of authentic medical texts in several countries for several languages will create a possibility to use them as comparable corpora in the future.

For the statistical machine translation approaches a large parallel corpus would be desirable, but it is unrealistic that in near future there will be any. However, by analysing linguistically and clinically free texts already existing in digital format in hospital information systems, by cutting down redundancies and setting standards, the variation between and within the texts should considerably diminish. This will be an essential prerequisite for structured reporting. Adequate ontology, lexicon, restrictive grammar, pre-determined and pre-translated collocations together with a statistical approach will make computer-assisted translation of authentic medical texts possible. Considering the amount of health care data, it is the only way to go.

## TÄNUAVALDUSED

*„The Sailor cannot see the North, but knows the Needle can.”  
Emily Dickinson*

Kuigi meid hoiatati seminaris tänuavaldustele keskendumast, soovin väga tänada neid inimesi, tänu kellele julgesin ette võtta enda jaoks uue valdkonna õpingud, ja kõiki neid, tänu kellele need õpingud nii kaugele on jõudnud.

Südamest tänu:

- kõigile kirjaliku tõlke õppekava õppejõududele nende aja ja suuniste eest;
- armsatele kursusekaaslastele kaasteelisteks olemise eest;
- juhendaja prof Krista Vogelbergile minu töösse entusiastliku suhtumise eest;
- juhendaja dr Peeter Rossile tööks vajaliku algmaterjali (sh motivatsiooni), arutelude ning sõpruse eest;
- õppejõududele Katrin Kernile, Kadri Muischnekile, Kersti Undile ja Enn Veldile keele mitmetähenduslikkuse ja võlu avamise eest;
- õppejõududele Toomas Lapile ja Terje Loogusele elava ja meelde jääva sissejuhatuse eest erialasse;
- prantsuse keele õppejõule Elodie Dominique Bailleux'le, rootsi keele õppejõule Tiina Mullamaale ja poola keele õppejõule Dagmara Eliza Saganowskale armastatud kultuuridesse ukse paotamise eest;
- Võru Lauluisa koolile eesti ja inglise keele, aga ka paljude muude tarkuste eest;
- Moskva II meditsiiniinstituudile meditsiiniteadmiste, julguse ja eneseusu eest;
- Ida-Tallinna Keskhaiglale loovas ja vabas töökeskkonnas töötatud aastate eest;
- töökaaslastele Tartu Ülikooli Kliinikumis paindliku töögraafiku talumise eest;
- sõpradele: Tartu tugijõududele Kärdile ja Arnole ning Annelile ja Janekile, nagu ka Tallinna taustajõududele Margele, Eleonorale ja Triinule nende olemasolu eest; Estale prantsuse keele, ooperipisiku ja eluliste mõttevahetuste eest;
- vanematele Ludmilla ja Lembit Kukele ning vend Jüri perele; Aimele ja kõigile sugulastele mu ettevõtmistesse sooja, kannatliku ja toetava suhtumise eest;
- Carrie'le ja Minnile piiritu osavõtlikkuse ja rõõmu eest;
- parimale sõbrale Andresele, kellela kõik oluks lihtsalt tühi ja olemata, armastuse ja kodu eest.