

LIHTLAUSE SEMANTIKA: TEOREETILINE KONTSEPTSIOON JA ARVUTIANALÜÜSI VÕIMALUSED*

HALDUR ÕIM, HEILI ORAV, PIIA TAREMAA

1. Sissejuhatus: lausesemantika ja lause semantiline analüüs

Lausesemantika on keelekäsitluses põhimõtteliselt vana ala (näiteks, kui arvestame ka keskaega). Tänapäevases teoreetilises keeleteaduses on see aga uus ala, millest ollakse teadlikud küll vähemalt generatiivse grammatika algusaegadest (Katz, Fodor 1963), kuid erilist progressi ei ole siin olnud võimalik täheldada. Sõnasemantika arenguga ei anna võrreldagi: kui palju on ilmunud viimase kohta raamatuid ja lisaks on kogu maailmas tuntud leksikaal-semantilised andmebaasid, nagu WordNet või FrameNet (ja nende analoogid erinevate keelte jaoks), mis on leksikaalse semantika arendused. Põhjus on tegelikult arusaadav. Sõna (või laiemalt leksikoniüksus) on midagi, mis eeldatavasti eksisteerib inimese püsimälus. Aga lause on suhtlusüksus, mis pannakse kokku sõnadest mingi teate/teadmise edasiandmiseks. Kui lause on öeldud ja vastu võetud, siis vastuvõtja (ja küllap ka ütleja) „kahe kõrva vahel” säilib teate sisu (= lause tähendus), aga mitte lause ise oma morfoloogilises-süntaktilises vormis (kui seda mingil põhjusel pähe ei õpita). Laused ei moodusta leksikonile mingit analoogi, isegi kui keegi kasutab täpselt sama lauset mitu korda. Laused on keelevormid, mis käibivad suhtluses, neil on vorm ja sisu, aga (pikaajalises mälus) säilib vaid sisu.

Viimastel aastatel on lausesemantika teoreetilise käsitluse problemaatikale lisandunud arvutilingvistide-keeletehnoloogide huvi, sest noid ei huvita niivõrd üksiksõnad kui võrd tekstid, ja need koosnevad lausetest. Lause on teadmiste (ja muidugi ka arvamuste, suhtumiste jne) edastamise üksus nii teoreetilise keeleteaduse kui ka arvutilingvistika jaoks. Teoreetilise käsitluse osas peaksid niisiis mõlemad problemaatikale ühtmoodi lähenema. Ja rakenduslik arvutilingvistika/keeletehnoloogia võib pakkuda mõistlikke kitsendusi, mis aitavad ka teoreetilistes käsitlustes probleeme mõttekalt allprobleemideks jagada ja neid järjestada.

Semantilise analüüsi tähtsus on keeletehnoloogias väga kiiresti kasvamas. Selle oodatumateks rakendusteks on näiteks masintõlge ja semantiline veeb, aga samuti on see vajalik paljude teiste keeletööstust eeldavate programmide osana (mitmesugused infootsiprogrammid, nt nn faktiotsiprogrammid). Semantilise analüüsi vajalikkust neis valdkondades on alati tunnetatud, nüüd on see vajalikkus kiiresti muutumas võimalikkuseks (vt nt Calzolari 2008: IV).

* Töö on valminud sihtfinantseeritavate teemade SF0180078s08, SF0180056s08, ETF-i grand nr 7149 ja riikliku programmi „Eesti keele keeletehnoloogiline tugi” raames projekti „Lihtlause semantiline analüüs” toel.

Arvutisemantika (semantiliste andmebaaside, ontoloogiamudelite) arengu kõrval on vaieldamatult kaasvõtmeteguriks olnud teoreetilise semantika areng. Huvitaval viisil on siin olulisimat rolli mänginud kaks pealtnäha sõltumatut, aga tegelikult sisuliselt vägagi seotud arengut: esiteks predikaadikesksete lausestruktuuride arendamine ja teiseks nimisõnade ja nimisõnafraaside (kui predikaadi tüüpiliste argumentide) semantika arendamine. Predikaadid on põhimõtteliselt relatsioonilised üksused, mis eeldavad argumente, argumentideks taas sobivad prototüüpsed nimisõnad mitterelatsiooniliste keeleüksustena. Kuid ainult predikaatide ja argumentide kombinatsioonidena sünnivad semantilises mõttes laused, kus kombineerijaks on süntaks.

Lisaks on teoreetilises keeleteaduses – ilmselt funktsionaalse lähenemise üha kasvava mõju ühe tagajärjena (vt Õim, Tragel 2007) – keelepädevuse kui keeleteaduse uurimisobjekti „loomulikud” piirid nihkumas keeletarvituse, meie teema puhul keeleväljendite töötlemise, moodustamise ja mõistmise käsitlemise suunas.

Ka eesti keele puhul on semantiline analüüs saanud reaalselt võimalikuks seetõttu, et ühelt poolt on lausete süntaksianalüüs ja teiselt poolt leksikaal-semantiline andmebaas (thesaurus, WordNet) saavutanud taseme, kus nende ühitamisel on võimalik semantilist analüüsi teha (eesti keele lausete süntaktilise ja semantilise esituse sidumisest vt Müürisep, Orav, Õim, Vider, Kahusk, Taremaa 2007; eesti keele tesauruse kohta vt nt www.cl.ut.ee/ressursi/teksaurus). On käivitatud riiklik programm „Eesti keele keeletehnoloogiline tugi”, mille raames on rahastatud projekti „Lihtlause semantiline analüüs”. Projekti eesmärgiks on luua semantilise analüüsi arvutiprogramm, mis analüüsib üksiklauseid, kuid esimese sammuna ainult lihtlauseid, s.o lauseid, kus on üks predikaat, pole lauselühendeid ega osalauseid (kuid see on muidugi vaid metodoloogilistest kaalutlustest lähtuv kitsendus, lõppeesmärk on teksti semantiline analüüs eeldusel, et tekst koosneb lausetest).

2. Artikli eesmärgid

Eesmärgiks ongi esitada meie lähenemise ja analüüsiprogrammi aluseks oleva lausesemantika kontseptuaalne mudel, millele saab edaspidi rajada tekstide semantilise analüüsi.

Arvutianalüüsis on eeltoodud põhimõtetest lähtudes esimese ülesandena mõttekas kitsendada võimalikku temaatikat, sest lausesemantikas ja seda enam tekstisemantikas ei saa ignoreerida valdkonna spetsiifikat, mille kui taustteadmise arvestamist lause/teksti mõistmine (aga seda ju semantiline analüüs modelleerib) peab kajastama. Me ei pea siin silmas valdkonda tegevusala mõttes (sport, põlevkivitööstus jmt), vaid valdkonda keelelises (semantilises) mõttes, milles kehtivad teataval määral ühtsed süsteemsed ja selles mõttes huvipakkuvad semantilised seaduspärasused (ja mille käsitlemisest on ka praktilist tulu).

Üks valdkondi, mille puhul lausesemantikast on nii teoreetilist kui praktilist tulu, on ruumilise liikumisega seotud keeleväljendite semantika. Ruumis paiknemine ja liikumine, objektide liigutamine on olulised nii teoreetilises kui ka praktilises plaanis. Teema on piisavalt lai, hõlmates põhimõtteliselt nii verbisemantikat ja lause grammatilist vormistust kui ka kõiki lause-

liikmeid, s.t liikumissündmuse võimalikke osalisi (seega nimisõnasemantikat eelkõige). Praktilises plaanis võimaldab ruumis paiknemise ja liikumise-liigutamise domeeni läbitöötamine luua keeletehnoloogilisi rakendusi virtuaalsele või reaalsele masinatele, mis on võimelised ruumis liikuma.

Artikkel keskendub eelkõige projektis esilekerkinud üldlingvistilistele probleemidele. Esitame oma lähenemise metodoloogilise tausta. Nagu eespool osutatud, on selles meie jaoks kaks võtmeteemat: esiteks lause kui terviku semantiline struktuur ja selle esitus (erinevalt predikaadi, näiteks verbi kui leksikaalse üksuse semantilisest esitusest) ja teiseks nimisõnade ja nimisõnafraaside kui lause predikaadiga seotud struktuursete põhikomponentide semantiline struktuur ja esitus.

Esimese teema käsitlemisel keskendume argumentstruktuuridele ja nende raames semantilistele rollikategoriatele kui väidetavalt universaalsetele süntaksi ja semantika vahetavatele reguleerivatele (funktsionaalsetele) lause tasandi kategooriatele. Kuid lause kui suhtlusüksuse käsitlus toob meid vältimatult ka lausega edastatava (lause vastuvõtjani jõudva) täieliku info kajastamise probleemi juurde ehk teisiti öeldes: lause vahetust tähendusest tehtavate võimalike järelduste kajastamise juurde.

Teine teema viib meid küsimuse juurde, kuidas käsitleda ja kirjeldada niisuguste mittepredikatiivsete leksikaalsete üksuste nagu füüsilisi objekte tähistavate sõnade (*laud, puu, kivi, järv*) tähendusi, seejuures esitatuna nii, et need kirjeldused oleksid kasutatavad arvutianalüüsis ehk siis nii, et nendega oleks võimalik teha formaalseid operatsioone (näiteks tuletada järeldusi), ja muuhulgas nii, et nimisõna võimalike laiendite tähendused oleksid sellesse struktuuri integreeritavad (näiteks fraasi *ümmargune kivi* semantilises esituses peab sõna *ümmargune* esitus olema integreeritud sõna *kivi* esitusse sarnaselt kivi teiste omadustega).

Lause semantilise analüüsi juures on ühelt poolt vaja arvestada süntaktilise esituse kui süntaktilise analüüsi moodulist saadud sisendi kuju ja teiselt poolt meie semantilise esituse põhimõtteid ja formaalseid nõudeid. Viimases on käibel mitu teoreetilist enam-vähem võrdset esitusviisi, meie lähtume freimisemantikast, laiendades seda sõnasemantikalt lausesemantikale (analüüsi lõpptulemuseks ja teoreetilise esituse mõttes kirjeldusüksuseks on *lause freim*).

3. Teoreetiline mudel

3.1. Üldteoreetiline taust

Meie lähenemise ja huvide kontseptuaalseks raamiks ja taustaks sobib – vähemalt põhipunktide osas – ilmselt kõige paremini Ray Jackendoffi praegune kontseptuaalne semantika.¹ Tema kontseptsioonis on olemas ja teoreetiliselt piisavalt hästi läbi töötatud eristused, mis meid huvitavad: lausesemantika alused ja nimisõnapõhiste struktuuride (nimisõnafraaside, kaas-

¹ Konkreetne käsitlus ja näited on siin võetud raamatust „Foundations of Language. Brain, Meaning, Grammar, Evolution” (Jackendoff 2002), kuid R. Jackendoff on seda suunda teoreetilises keeleteaduses arendanud pikemat aega, kusjuures algselt püüdis ta vähemalt põhimõtteliselt jääda Chomsky generatiivse grammatika raamesse, ehkki sellega diskuteerides (nt Jackendoff 1997). Hiljem on ta selle seose käsitlemisest konkreetse keelekirjeldusmudeli osas – ehkki mitte inimese keelevõime kui niisuguse osas – lihtsalt loobunud.

sõnafraaside) semantilise käsitluse alused. Selliseid tervikkäsitlusi on teoreetilises keeleteaduses väga vähe. Konkreetsete detailide osas on meie käsitlus – meie eesmärgist (arvutianalüüsist) tulenevalt – R. Jackendoffi omast küll mitmeti erinev.

Ray Jackendoff eristab keele tähendusstruktuurides kõigepealt kaht tüüpi struktuure, mida ta nimetab vastavalt kontseptuaalseks struktuuriks ja ruumiliseks struktuuriks (Jackendoff 2002: 345–346). Seejärel toob ta sisse kolmanda semantilise struktuuri tüübi kvaalistruktuurid (Jackendoff 2002: 369 jj), mis lõikuvad kontseptuaalsete ja ruumiliste struktuuridega, kuid millel on keele semantilise tasandi suhtes oma vaatenurk.

Kontseptuaalsed struktuurid moodustuvad diskreetsetest, s.o mõisteliselt piiritletud ja defineeritavatest kategooriatest, mis tüüpiliselt on hierarhiliselt korrastatud. Olulisimad on hüponüümia-hüperonüümia hierarhiad (taksonoomia) leksikaalse semantika kontseptuaalse alusena ja predikaatide-argumentide struktuurid (koos rollistruktuuridega) lausesemantika kontseptuaalse alusena. Need struktuurid on semantikas suhteliselt hästi tuntud. Struktuurid, mida R. Jackendoff nimetab ruumilisteks struktuurideks, on märksa vähem tuntud ja selles mõttes huvitavam struktuuritüüp, seda eriti meie ruumilise liikumise käsitluse jaoks. R. Jackendoff iseloomustab seda struktuuritüüpi järgmiselt: „Ruumiline struktuur tegeleb füüsilise maailma mõistmise kodeerimisega: mitte lihtsalt selle momentlähilõigete esitamisega, vaid objektide ajas kestvate kuju, liikumise, ruumilise paiknemise (ja võimalik, et ka nendevaheliste mõjutuste) integreerimisega” (Jackendoff 2002: 346). Ruumiline struktuur on inimesekeskne, ta integreerib nägemis-, kuulmis- ja kompimissüsteemi, lisaks somatosensorse (oma keha asendi tajumise) süsteemi infot. Näiteks kui me näeme mingit asja, siis võimaldavad ruumilised struktuurid meil seda enda suhtes lokaliseerida, otsustada näiteks, kas ja kuidas seda on võimalik käega haarata või sellega midagi muud teha (Jackendoff 2002: 346). See struktuur võimaldab otsustada, kas mingit eset saab nimetada tooliks, diivaniks või taburetik, kiviks või palliks. Niisiis on ruumiline struktuur mõeldud eelkõige füüsilise maailma objektide – asjade, ka ainete – käsitlemiseks semantikas (kusjuures nende asjade hulka kuulub ka inimese enda keha).

Kuid R. Jackendoff on hoolikas rõhutama ruumilise struktuuri abstraktsust olemust. „On oluline mõista, kui abstraktnen on hüpoteetiline ruumilise struktuuri tase. Ehkki ruumiline struktuur on olemuselt geomeetiline, ei ole see kujutluslik....: seda ei saa käsitada struktuurina, mis kodeerib „pilte” või „kujutusi” ajus” (Jackendoff 2002: 347). Ruumiline struktuur on abstraktnen struktuur, mille põhjal võib genereerida konkreetseid, aga teatavaid fikseeritud omadustega kujutluspilte.²

R. Jackendoff ei väida, et kontseptualiseeritud maailma mõistmine jaguneb üheselt kontseptuaalsete struktuuride ja ruumiliste struktuuride vahel (Jackendoff 2002: 347). On kattumispirkondi ja ilmselt on veel teisigi tasandeid. Neil kahel tasandil (esitussüsteemil) on ühisosa, mis võimaldab neil oma-

² Ruumiliste struktuuride geomeetrilise olemuse rõhutamine on iseenesest huvitav. Geomeetria teadusena tegeleb tervikkujunditega, aga seejuures mitte nende konkreetsete eksemplaridega: geomeetriat huvitavad näiteks kolmnurk üldise kategooriana ja selle allkategooriad (täisnurkne, võrdhaarne jne kolmnurk), aga mitte konkreetne ülesjoonistatud kolmnurk. Just see tervikkujunditega tegelemine eristab geomeetriat näiteks algebrast (mis sobiks siis kontseptuaalsete struktuuride formaalseks analoogiks). Samal ajal on selge, et see „geomeetrilisus” ei takista esitamast ruumilisi struktuure freimidena.

vahel suhelda, et kujundada mingist situatsioonist või sündmusest tervikpilt (ja tuletada mõlemat tüüpi järeldusi). Nii on füüsilisi objekte tähistavad sõnad tüüpilised kvantifitseeritavad (kvantoritega haaratavad) üksused, kvantorid kuuluvad aga kontseptuaalsete struktuuride juurde, tehisaaju nagu näiteks tooli, taburetti, diivanit või voodit tähistavatel sõnadel on semantikas nii hüperonüümia–hüponüümia dimensioon (kontseptuaalne struktuur) kui ka osa–terviku dimensioon (ruumiline struktuur), ja need on omavahel ilmselgelt seotud.

Väärrib osutamist, et ruumilised struktuurid on ilmselt valdkond, kus kõige selgemini tuleb esile probleem, mis on tuntud vajadusena teha vahet ühelt poolt mingit objekti tähistava keeleväljendi keelelisel tähendusel (see kuulub uurimisele keeleteaduslikus semantikas) ja teiselt poolt objektiga seotud „maailmateadmusel” (see kuulub üldisesse ontoloogiasse). See on üks alasid, kus puhtlingvistilise semantika, nn ontolingvistilise semantika ja keele (tekstide) mõistmise modelleerimist taotleva arvutilingvistilise ning keeletehnoloogilise semantika huvid ja arusaamade kokkuviimine on ehk kõige keerulisem (üldteoreetilisest seisukohast vt nt Ontolinguistics 2007, keeletehnoloogia osas nt Kaljupärv 2007; ka A&A 2008).

3.2. Lausesemantika: argumentstruktuurid ja semantilised rollid

Kui meid huvitavad süntaksi ja semantika suhted lause tasandil, siis on (praeguste üldiste arusaamade järgi) vältimatu tegelda argumentstruktuuridega (predikaadid ja nende argumentid ning viimaste „peale” kihistuvad rollstruktuurid). R. Jackendoffi järgi on argumentstruktuuride teooria iga grammatikateooria üks kesksemaid teemasid (Jackendoff 2002: 132). Ja seda kinnitab viimastel kümnenditel selle temaatika kohta ilmunud publikatsioonide hulk ning tegelikult võib rääkida ka tervete teoreetiliste kontseptsioonide esilekerkimisest.³ Tinglikult võib eristada kaht probleemi: esiteks argumentid ja argumentide tüübid „omaette” võetult ning teiseks argumentide semantilised rollid lause süntaktilise ja semantilise struktuuri ühenduslülidena. Reaalses keeleanalüüsis ei ole neid võimalik lahus hoida, aga probleemidest ülevaate saamiseks tasub neid eraldi käsitleda.

3.2.1. Argumentstruktuurid. Argumentstruktuuridest saab ehk kõige selgema ülevaate James Pustejovsky käsitluse kaudu tema raamatus „The Generative Lexicon” (Pustejovsky 1995).⁴ Selle järgi on argumentstruktuur üks keele neljast semantiliselt relevantsest struktuuritüübist, teised on sündmusstruktuur, kvaalstruktuur ja leksikaalne pärilusstruktuur (Pustejovsky 1995: 61). Neist käsitleme siinses artiklis lähemalt veel vaid kvaalstruktuuri. Sündmusstruktuuride all mõtleb J. Pustejovsky selliseid lausetega väljendatavaid semantilisi struktuure nagu SEISUND, PROTSESS, TEGEVUS jt; leksikaalse pärilusstruktuuri tuumiku moodustavad tuttavad hierarhilised seosed nagu hüponüümia ja hüperonüümia, meronüümia jt. R. Jackendoffi eespool toodud liigituse (kontseptuaalse struktuuri ja ruumilise struktuuri) mõttes paigutuvad argumentstruktuur ja sündmusstruktuur ning suures

³ Hea ülevaate erinevatest lähenemistest ning edasisi viiteid saab kogumikust „Semantic Role Universals and Argument Linking” (2006); vt ka Õim, Trägel 2007.

⁴ Tasub märkida, et J. Pustejovsky on teoreetiliselt pigem arvutilingvistika kui puhta teoreetilise lingvistika esindaja, kuid mitme probleemi käsitlemisel toetub temale ka R. Jackendoff.

osas ka leksikaalne pärilusstruktuur ilmselt kontseptuaalse struktuuri alla, kvaalistruktuur, nagu öeldud, lõikub nii kontseptuaalse struktuuri kui ka ruumilise struktuuriga.

Argumentstruktuuris võib J. Pustejovsky järgi eristada vähemalt nelja tüüpi argumente: 1) ehtsad argumentid, 2) vaikumisiargumentid, 3) varjatud argumentid ja 4) ehtsad adjunktid.

Ehtsad argumentid on sellised, mis realiseeruvad antud predikaadi juures kontekstivaba lause puhul obligatoorselt: *John ehitas maja*. Vaikumisiargumentid on sellised predikaadi laiendid, mis loomuldasa (semantilisel) kuuluvad antud predikaadiga tähistatava tegevuse või situatsiooni juurde viimase teatava komponendina, kuid mille süntaktiline realisatsioon pole obligatoorne: *John ehitas maja tellistest*. On selge, et midagi ehitada saab ainult mingist materjalist (ja semantikas võime seega järeldada, et maja ehitamisel kasutati mingit materjali), kuid predikaadi *ehitama* puhul lause süntaktilises vormis ega laiemalt isegi mitte tekstis seda täpsustama ei pea. Varjatud argumentid on predikaadiga tähistatud tegevuse komponendid, mis on inkorporeeritud predikaadi tähendusse ja mida „puhtal kujul” süntaksis ei saagi seetõttu väljendada. Nii kuuluvad *jooksma* tähendusse vahendina 'jalad' ja *vaatama* tähendusse 'silmad', nii et laused *Ta jooksis jalgadega* või *Ta vaatas mind silmadega* on hälbivad. Küll saab selliseid argumente väljendada, kui neid on kuidagi spetsifitseeritud (näiteks täienditega): lause *Ta vaatas mind oma kurbade silmadega* on normaalne. Lõpuks (ehtsad) adjunktid on komponendid, mis fikseerivad sündmuse või situatsiooni teatavaid parameetreid, kuid erinevalt vaikumisiargumentidest ei seostu individuaalsete leksikaalsete üksustega (kui predikaatidega), vaid üksuste semantiliste klassidega. Tüüpilisimad näited on koha- ja ajalaiendid tegevus- ning protsessilauses: *Ma sõin teisipäeval nurgapealses baaris kaks biifsteeki*. Süntaksi seisukohalt oleks ka lause *Ma sõin kaks biifsteeki* täielik korrektne lause.

See liigitus lähtub süntaksist, kuid selle juured on semantikas: kuidas liigendatakse mingi sündmuse või situatsiooni struktuuri. Süntaks asetab sellele struktuurile oma relevanttsuse filtri, kusjuures relevanttsus lähtub predikaadist: see filter määratleb predikaadiks oleva üksuse „süntaktilise käitumise”, tehnilises mõttes süntaktilise argumentiraami lause piires. Predikaadi semantilises kirjes (meie puhul freimis) tuleb aga esitada kõik semantilisel võimalikud argumentid (ja nende rollid). Süntaksi ja semantika seose kajastamiseks tuleb erinevad argumentitüübid selles freimis varustada erinevate märgenditega. Eriti käib see varjatud argumentide ja adjunktide kohta. Nii peab *vaatama* kirjes esitama vahendina vastava märgendiga varustatult *silmad* (leksikaalsemantilise üksusena, millele vastab mujal leksikonis ka omaette kirje), mida saab *vaatama* freimieksemplari (s.o konkreetse lause predikaadi freimi) koosseisus samas lauses või ka hiljem tekstis spetsifitseerida, isegi kui algse lause süntaktilises struktuuris seda üksust pole. Lausete järjend *Ta vaatas mulle otsa. Ta silmad olid sügavsinised* on ju normaalne eesti keele tekst. Niisamuti peab adjunktidele vastavate argumentide nagu näiteks koha- ja ajamääratluste lisamise võimalus olema eksplitsiitselt predikaadi semantilises kirjes märgitud, osutades seejuures, missuguselt hüperonüümilt (predikaatide semantilisel klassil) mingid adjunktid koos rollidega ja semantiliste piirangutega päritakse.

3.2.2. Semantilised rollid. Semantika seisukohalt esitavad argumendid niisuguse sündmuse või situatsiooni osalisi, kus sündmuse-situatsiooni tüübi ja üldstruktuuri (liigenduse) määrab predikaat. Ent igal osalisel (argumendil) on selles struktuuris kindel roll. Rolli mõiste tõstis teoreetilise keeleteaduse tähelepanu fookusesse teatavasti Charles Fillmore (1968). Idee ise küpses varem, näiteks viidatakse tihti Jeffrey Gruberi väitekirjale (Gruber 1965; 1976). Kuid Fillmore'i käändegrammatika ilmumisest peale on see teema püsivalt kuulunud teoreetilise keeleteaduse aktuaalsete probleemide loendis. On ilmunud suur hulk erinevaid käsitusi, nagu on ka nähtusele endale antud erinevaid nimetusi: temaatilised rollid, teeta-rollid, semantilised rollid; mõningates käsitlustes tehakse neil terminitel vahet, üldiselt aga on tegu sama nähtuse erinevate nimetustega (vt nt Semantic Role Universals and Argument Linking 2006). Charles Fillmore ise on ideed edasi arendanud freimiseantika raames ja realiseerinud FrameNeti projektis (www.icsi.berkeley.edu/framenet/). FrameNeti üldideed on ka meie lauseanalüüsiprojekti ja eelkõige selle leksikoni – freimileksikoni – aluseks, kuid semantiliste rollide määratlemises lähtume teistest põhimõtetest. Meie jaoks ei ole semantilised rollid vajalikud mitte ainult predikaadi (verbi), vaid ka predikaadist lähtuva lause kontseptuaalset struktuuri organiseerivate elementidena ja seetõttu peavad nad olema märksa üldisemad kui FrameNeti individuaalsetest verbidest lähtuvad. Rollide termineis peab olema võimalik määratleda nende „all” esineda saavate argumentide semantilisi kategooriaid ja sõnastada järeldusreegleid, ning seda võimalikult üldiselt. Seetõttu oleme jäänud ka termini *semantilised rollid* kui ehk kõige neutraalsema juurde, mis seejuures esindab meie arvates keeleliselt kõige paremini kontrollitavat ja semantiliselt (üldistuste mõttes) kõige mõttekamalt tasandit, mille sisu ja määramiskriteeriume me oleme praeguse kompetentsi juures suutelised piiritlema.⁵

Rollimõistetega seoses on – nii üldiselt kui ka (ja eriti) empiiriliselt orienteeritud lähenemiste puhul, ja seda meie oma on – kahtlemata olulisimad kaks probleemi. Esiteks, kuidas rolle sisuliselt, mõisteliselt määratleda, nii et iga rolli sisu oleks võimalikult üheselt selge „iseeneses”, aga ka vastandatuna teistele samas süsteemis kasutatavatele rollidele, s.t et tekstis oleks iga teatava rolli esinemus semantiliste-mõistelisete tunnuste põhjal tuvastatav; teiseks, missuguste (võimalikult üldiste) reeglite abil seostada semantilised rollid nende väljendamiseks keeles kasutatavate morfosüntaktiliste vahenditega. Esimene on semantika siseprobleem ja sel peaks perspektiivis olema suhteliselt universaalne lahendus, teine puudutab keele semantilise ja vormilise struktuuri vahekorda ning on suurel määral keelespetsiifiline (analüüsiprogrammis on nad muidugi vältimatult seotud).

Põhimõttelisest aspektist on olulisim meeles pidada, et rollimõisted on metakeele üksused. Kui kasutame nende tähistamiseks loomuliku keele sõnu (nagu *Agent*, *Kogaja*, *Vahend* jne), siis peame jälgima, et *automaatselt* ei tuleks metamõistesse loomuliku keele sõna semantilisi komponente, mis võivad kontekstist sõltuvalt varieeruda. Metakeele mõiste tuleb defineerida ja ainult see definitsioon määrab mõiste sisu ja piirid. Loomuliku keele sõnad on metakeeles vaid inimestele mõistete kasutamise mugavuseks mõeldud vorm. Näiteks füüsikas on kasutusel terminid *jõud*, *kaal*, *mass*, *kiirus* või *kii-*

⁵ Rollide hierarhiate kohta vt nt Van Valin 1999; teemat on korduvalt käsitletud ka eestikeelses kirjanduses, nt Lindström, Tragel 2007.

rendus, mille sisu on määratletud nende mõõtmise ja arvutamise protseduuridega ja mille puhul seetõttu ei tule kellelgi pähe hakata vaidlema selle üle, kas mingit parameetrit mingil konkreetset juhul sobib nimetada kiiruseks või kiirenduseks. Rollimõistete puhul see kahjuks nii ei ole. Ikka ja jälle vaieldakse selle üle, kas mingi väljendi konkreetse esinemusega tähistatud entiteeti sobib nimetada näiteks *Agendiks*, *Põhjuseks*/*Põhjustajaks* või *Instrumentiks* (*Tuul liikkas vaasi ümber. Auto pööras kõrvaltänavasse: mis rolli täidavad tuul või auto?*). Probleem on hästi teada ja lahendusena on pakutud isegi võimalust, et rollide tähistamiseks ei kasutataks loomuliku keele sõnu, vaid rollid nummerdatakse: R1, R2, R3 jne (vt näiteks Primus 2006). See aga tähendaks muidugi, et iga roll tuleks defineerida mingite semantiliste tunnuste komplektina (nagu foneem fonoloogias), seejuures nii, et kõikide keelte kõigile antud rollitähisega märgendatud struktuuridele oleksid rakendatavad samad semantilise töötluse, näiteks järeltule tuletamise reeglid, mis moodustavad osa selle rollimõiste sisust; ja teiseks peaksid mainitud relevantssed tunnused olema iga vastava (tekstis esineva) keeleüksuse, näiteks noomeni-fraas või selle peasõna – nimisõna – tähenduses tuvastatavad. Kummakski ei ole loomuliku keele semantikakäsitlused tänapäeval täies mahus suutelised ja nii lepitakse poolel teel lahendusega. Olulisemate rollide nagu *Agendi* või *Patsiendi* määratlemisel opereeritakse teatavate suhteliselt abstraktsete tunnustega, ehkki erinevatel autoritel võivad need erineda. Üldiselt lähtutakse aga nii rollide valikul kui ka nende määratlemisel keeleteadlase-keelekasutaja intuitsioonist, mis vormistatakse kriteeriumideks suhteliselt vabas sõnastuses (näiteks liikumise puhul *Sihtkoht* või *Locto* nagu meie mudelis: koht, kus liikumine lõpeb). Seetõttu ei ole üllatav, et isegi niisuguses aktiivselt käsitletud valdkonnas nagu ruumiline liikumine ei ole ühtset arusaama sellest, missuguseid rolle eristada ja mis on ühe või teise rolli sisu/maht. Meie semantilise analüüsi mudeli praeguses variandis oleme jäänud ühetasemelise (mitte hierarhilise) rollide loendi juurde. Tähistustena ei kasuta me siiski eesti keele sõnu, vaid inglise keelest lähtuvaid lühendeid (*Ag* = *Agent*, *Loc-from* = *Lähtekoht*, *Locto* = *Sihtkoht/Lõpp-punkt* jne), mille määramise kriteeriume sõnastades oleme püüdnud saavutada, et need oleksid võimalikult üheselt mõistetavad. Kuid tuleb kohe osutada, et selle ühesuse osas oleme jälginud ennekõike meid huvitavat sündmuse (tegevuse, protsessi) makrotüüpi, milleks on ruumiline liikumine-liigutamine.

3.3. Nimisõna(fraasi)semantika: kvaalid ja ruumilised struktuurid

Nimisõnafraaside alla paigutame siin fraasid, mille semantiliselt peaks (põhjaks) on nimisõna, seega kuuluvad siia ka kaassõnafraasid (nii *suur mets* kui ka *suure metsa kohal*, nii *lai kraav* kui ka *üle laia kraavi*). Kaassõnad on niisiis funktsionaalselt, semantika seisukohalt võrdsustatud käänetega: fraase *suures metsas*, *suure metsa sees*, *suure metsa kohal*, *suure metsa taga* tuleb semantikas käsitleda ühtemoodi fraasidena, mille peasõna tähistab entiteeti nimega *mets* ja mis (antud juhul) esiteks määratlevad selle entiteedi ühe füüsilise parameetri väärtuse (*suur*) ja teiseks osutavad mingi teise entiteedi paiknemise metsa suhtes. Seetõttu on siin meie esmaseks huviobjektiks nimisõnade semantiline esitus. Ja nimisõnadest huvitavad meid uurimisvaldkonnast tingitult esmajoones need, mis tähistavad (prototüüpeid) füüsilisi objekte, nagu *kivi*, *maja*, *tool*, *pilv* (prototüüpsuse aluseks on

objekti ruumiline piiritletus), aga ka selliseid raskesti piiritletavaid, kuid siiski ruumilist mõõdet omavaid entiteete tähistavad sõnad, nagu *mets*, *kallas*, *õhk*, *taevas*, mis liikumissündmustes tüüpiliselt esitavad liikumise ruumilisi parameetreid (kus, kust, kuhu, kustkaudu jne), mitte liikuvaid objekte endid nagu esimest tüüpi objektid. Keskseks probleemiks on, kuidas – missuguste parameetrite kaudu – vastavate nimisõnafraaside muude elementide (traditsioonilises terminoloogias: täiendite, eespoolses näites omadussõna *suur*) semantilised esitused integreerida peanimisõna semantilisse esitusse. Samuti on oluline probleem see, kuidas lause semantilises esituses siduda kaks nimisõnadega (nimisõnafraasidega) tähistatud entiteeti, mida morfosüntaktilisel tasandil esitavad *king* ja *kapp* niisugustes lausetes nagu *Tüdruk viskas kinga kappi/ kapi peale/ kapi taha* jne. (Etteruttava näitena: eelnenud lausetest tulenev kinga paiknemine kapi suhtes on kriitilise tähtsusega vähemalt võimalike edasiste järelduste tõttu. Näiteks kui järgnevas tekstis öeldakse, et kappi liigutatakse – kapi asukoht muutub –, siis järeldusena tuleneb, et kapi sees või peal oleva kinga asukoht muutub samuti, kapi taga, kõrval või mujal oleva kinga asukoht aga mitte; niisiis peavad need seosed semantilises esituses olema erinevalt tähistatud.)

Võrreldes verbidega (üldisemalt: predikaatidega) on nimisõnade semantikat palju vähem uuritud ja kirjeldatud. Ehkki tänapäevase lingvistilise semantikateooria teedrajavatest töödest on kuulsaks saanud just näiteks nimisõnade *poissmees* (Katz, Fodor 1963) ja *tool* (Katz 1972) semantilised kirjeldused, on nimisõnade semantika käsitlemine suuresti jäänudki selliste illustratiivsete näidete tasemele. Alles viimase kümmekonna aasta sees on seda teemat üritatud käsitleda süstemaatiliselt, s.o omaette sisulisi ja formaalseid aspekte kätkeva problemaatikana, mis põhimõtteliselt erineb predikaatide semantikaga seotud problemaatikast. Käsitleme oma teoreetilises tausta selgitamiseks siin lühidalt – meie semantikamudeli spetsiifikast lähtuvalt – eelkõige füüsilisi objekte ja nähtusi tähistavate nimisõnade semantikat (kogu nimisõnade klassiga seotud probleemistiku ülevaadete kohta vt nt Langemets 2004 ja seal viidatud kirjandust). Üldisi põhimõtteid tutvustasime eespool. Vaatleme siin meid huvitavast aspektist natuke lähemalt R. Jackendoffi ruumiliste struktuuride kontseptsiooni koos teise nimisõnade semantika eripära arvestava teooriaga, nn kvaalide teooriaga. Kvaali mõiste tõi semantikas aktiivsesse kasutusse James Pustejovsky (1995), ehkki ka sellel on oma ajalugu (vt Langemets 2004; terminit *kvaal* kasutaski eesti keeles esimesena ilmselt Margit Langemets). Ka R. Jackendoff tugineb oma kvaalikäsitluses põhiliselt J. Pustejovskyle (Jackendoff 2002: 369 jj).

J. Pustejovsky käsitluses on kvaalid loogilis-formaalses mõttes samasugused relatsioonilised mõisted (kõrgemat järku predikaadid) nagu rollimõisted. Põhimõtteliselt isegi üldisemad, sest ka argumentide rolle oleks võimalik predikaatide kvaalidena vaadelda. Ometi pole see viimane käsitlus semantiliselt mõttekas ja nii J. Pustejovsky ise kui ka R. Jackendoff kasutavad kvaali mõistet just nimisõnade semantikat käsitledes. Kvaalid on eelkõige nimisõnadega tähistatavate entiteetide nagu näiteks asjade, ainete jms iseloomustamiseks kasutatavad dimensioonid.⁶

⁶ Siin on ehk asjakohane meenutada, et kvaali mõiste semantikkasse toomisel oligi üks peamisi tõukejõude praktiline vajadus detailselt analüüsida ja vormistada nimisõnafraaside semantilist struktuuri. Selle taga oli omakorda asjaolu, et 1990-ndate aastate algul kujunes arvuti- ja keeletehnoloogilises semantikas üheks aktuaalsemaks teemaks arvutileksi-

J. Pustejovsky eristab (ja tema eeskujul ka R. Jackendoff) nelja tüüpi kvaale: formaalne kvaal, konstitutiivne ehk moodustajakvaal, agentiivne ehk põhjustamiskvaal ja teeliline ehk otstarbekvaal.

Formaalne kvaal esitab tüüpiliselt entiteedi kuuluvust teatavasse semantilisse tüüpide hierarhiasse (*koer* > *on loom* > *on elusolend* > *on füüsiline objekt*; *labidas* > *on tööriist* > *on füüsiline objekt*; *suremine* > *on seisundimuutus* > *on muutus*) ehk siis leksikaalse semantika terminoloogias hüponüümia-hüperonüümia hierarhiat ja selles mõttes semantikas suhteliselt hästi tuntud suhet (R. Jackendoffi käsitluses on see kontseptuaalse struktuuri tüüpnaide). Tasub märkida, et see hierarhia võib erinevat tüüpi entiteetide puhul saada erineva sisu semantiliste järelduste mõttes, mis mingist hierarhia lülist on tuletatavad. Näiteks koera puhul järelduvad sellest, et ta on füüsiline objekt, muuhulgas sellised omadused, nagu et võib ruumis liikuda (kukuda, libiseda) kusagilt kuhugi, ning et tal on teatav kuju ja kaal jne; sellest, et ta on elusolend, järeldub muuhulgas, et ta sünnib ja sureb ja elusoleku ajal sööb midagi; sellest, et ta on loom, järeldub näiteks, et tal ilmselt on kehaosad, nagu jalad, pea, kere jne. Nende tunnuste võimalikud väärtused (alus-tades näiteks võimalikust kujust ja kaalust kuni selleni, mida ta sööb) määratletakse aga juba sõna *koer* enda kirjes.

Moodustajakvaal koondab endasse entiteedi sisestruktuuri esitavad dimensioonid. Esmajärjekorras seostub see muidugi leksikaalses semantikas hästi tuntud osa-terviku struktuuriga (meronüümia). Kuid lisaks kuulub keele – semantika – seisukohalt nimisõnaga tähistatavate entiteetide sisestruktuuri terve rida muid mõistelisi „osi”, parameetreid, millel võib olla nii-või teistsuguseid väärtusi: meeltega tajutavad omadused, nagu suurus, kuju, värvus, lõhn, kaal, kõvadus jne, materjal, millest entiteet koosneb/on tehtud (*savist lauanõud*, *mullakamakas*), konsistents (vedel, tahke) jne. Asjade füüsilised osad kuuluvad samasse ritta ja just siin lõikuvad R. Jackendoffi ruumilise struktuuri teooria ja kvaaliteooria. Struktuursed ja funktsionaalsed osad ei iseloomusta sugugi ainult niisuguseid artefakte nagu tool, ka mäel või jõel on osad, rääkimata taimedest ja loomadest.

Teoreetiliselt seisukohalt, aga ka puht tehnilise esitusviisi seisukohalt on oluline osutada nende dimensioonide ja predikaate iseloomustavate argumentide paralleelsusele. Erinevus muidugi on, aga see on pigem süntaktiline: kui tüüpiline predikaat nõuab korrektse kontekstivaba lause moodustamiseks – s.t süntaktiliselt – kindlaid argumente, siis tüüpilised (näiteks asju tähistavad) nimisõnad ei nõua samas mõttes (süntaktiliselt) laiendeid. Kuid semantilisest aspektist on nimisõnad oma võimalike laiendite suhtes niisama selektiivsed nagu predikaadid võimalike argumentide suhtes. Eriti ilmne on see, kui võrrelda neid dimensioone varjatud ja vaikimisi argumentidega. Nii kuuluvad asjade (kivi, palli, puu, pilve) semantilisse struktuuri vaikimisi sellised dimensioonid nagu kuju või suurus, nii et vajaduse korral võime neid asju

konide, konkreetset nn leksikaalsete teadmusbaaside loomine, kasutades maksimaalselt ära olemasolevaid (seletus)sõnaraamatuid, s.t analüüsides automaatselt neis esitatud tähenduste definitsioone ja esitades tulemuse kindlas arvutiga töödeldavas formalismis. Suurem osa neist definitsioonidest – vähemalt nimisõnade omad – on aga nimisõnafraasid. Nii jõutigi tõdemuseni, et ka nimisõnade puhul võib nende võimalike laiendite alusena näha teatavat semantilis-funktsionaalset süsteemi (nagu verbide puhul argumentid ja semantilised rollid), mille kirjeldamiseks võetigi kasutusele kvaali mõiste (vt nt Inheritance, Defaults and the Lexicon 1993).

nende dimensioonide seisukohalt iseloomustada, kuid ainete (vee, liiva, suitsu) struktuuris neid dimensioone ei ole.

Agentiivne ehk põhjustamiskvaal on mõeldud esitama informatsiooni selle kohta, kuidas vaatlusalune entiteet tekib või tekitatakse. Kui tegemist on artefaktiga, siis sinne kvaal fikseerib selle valmistamise protsessi, kui selleks on keeles oma sõna (*maja: ehitama, raamat: kirjutama*); sega-seks jääb, mis on põhjustamiskvaali all kirjas, kui keeles eri sõna ei ole (*laud: ?*, *labidas: ?*, kas üldine *valmistama/tegema* või abstraktne predikaat, mis tagab vajalike järelduste tuletamise?). Teiste, s.o loomulike entiteetide puhul sõltub sinse kvaali määratletus (väärtuse võimalikkus) ilmselt samuti sellest, kas keel kuidagi fikseerib nende „tekkimise”. Taimed üldiselt *kasvavad* ja alg-selt *idanevad*. Teatavate elusolendite puhul on kasutatav *sündima* (ja hiljem samuti *kasvama*), kuid mitte kõigi puhul. Teoreetiliselt on taas tegemist keeleliste tähenduste ja maailmateadmuse vahelise piiri tõmbamisega. Kuid eristus jääb hoopis suuremal määral kui eelmiste kvaalide puhul leksikaalse ega mitte teoreetilise semantika valdkonda. Me võime ju teada, kuidas mäed ja mered tekivad (on tekkinud), kuid sõnade *mägi* või *meri* kasutamisest tekstis ei tulene sellega seotult mingeid erilisi semantilisi järeldusi (erinevalt näiteks sellest, et me teame, et meri koosneb vedelikust, et tal on teatavad mõõdmed jne jne, vt moodustajakvaali iseloomustust).

Teeliline ehk otstarbekvaal on funktsionaalne kvaal ja selles mõttes eelmisega analoogiline. See on obligatoorselt määratletud esmajoones artefaktide puhul: mis on maja, laua, lambi või voodi otstarve? Samal ajal lõikub see moodustajakvaaliga selles mõttes, et kui eristame mingis entiteedis kui tervikus mingeid osi, siis tihti seostub osaga teatav funktsioon (otstarve) terviku jaoks: nii on see jala puhul elusolendi või laua jaoks, katuse korral maja jaoks, isegi kalda puhul jõe jaoks jne. Aga muidugi mitte alati, näiteks mäe tipuga ei seostu mäe kui terviku jaoks semantiliselt mingit otstarvet. Ja taas semantilise ja maailmateadmuse ähmane piir: kas näiteks on tavatekstis võimalik või õigustatud semantiliselt järeldada puu ja selle juurte, okste ning lehtede otstarbesuhet, ehkki teame seda bioloogiast?

Lõppkokkuvõttes on ilmselge, et asjade, ainete, nähtuste jne omadused, mida inimesed tajuvad ja mille järgi neid eristavad, on palju keerulisem süsteem kui siin pakutu. Ometi tuleb millestki alustada, seda eriti siis, kui tegu on arvutianalüüsiga ning teoreetiline keeleteadus midagi paremat ei paku.

3.4. Lausesemantika: järelduste tuletamine

Lause on suhtlusüksus, sellega annab kõneleja kuulajale edasi kindla koguse infot mingi objekti, situatsiooni, sündmuse kohta. Laused konstrueeritakse suhtlussituatsiooni vajadustest lähtuvalt ja nende vajaduste hulka kuulub kindlasti edastatava info jagamine selle vahel, mis on lauses otseselt kodeeritud, ja selle vahel, mis on lausest järeldatav. Siin on lausesemantika ja leksikaalse semantika põhimõttelise erinevuse üks allikaid: leksikaalsed üksused ja nende tähendused on leksikonis (püsimalus) loetletavad-kirjeldatavad, laused ja nende tähendused mitte. Iga lause on uus konstruktsioon, mille semantiline analüüs eeldab selle leksikaal-süntaktilise struktuuri jooksvat „tõlkimist” semantiliseks esituseks ja teiseks sellest vahetust esitusest niisuguste järelduste tuletamist, mis peaksid jõudma kuulaja-vastuvõtja teadmistesse ja seal säilima, nii et selle teksti/suhtluse arengu käigus see info kuu-

luks ühisteadmiste hulka. Niisiis on lausesemantika osana vaja vastavat järelduste tuletamise süsteemi.

Keeleteaduslikus semantikas pole järeldused praegu eriti populaarne teema. Leksikaalses semantikas ei olegi see väga oluline. Näiteks ka FrameNetis, mis on kahtlemata üks viimaste aastate suuremaid leksikaalse ja arvutilingvistilise semantika loikumisalal paiknevaid projekte, pole freimi struktuuris järeldustele vahetult mingit rubriiki eraldatud. Leksikon on staatiline struktuur. Lause on aga dünaamiline struktuur, mis vahendab infot, säilimata ise, sellel on n-ö sisend- ja väljundinfo ja viimase üks kohustuslikke osi on järelduv info. Kui on tegu sidusa tekstiga, kehtib järelduv info järgnevas tekstis samal määral kui lauses vahetult esitatud info. Arvutimudelil on selle arvestamine niisiis vältimatu. Ja problemaatika on tegelikult väga lai, nii selles mõttes, et järeldusi võib olla väga erinevaid tüüpe, kui ka selles mõttes, et (teksti analüüsis) mitte kõik võimalikud järeldused ei ole igas kontekstis olulised. Oma artiklis iseloomustamegi vaid üldisi järelduste tüüpe, mis lause semantilise analüüsiga tegelevas süsteemis olulised on. Need ei haara kõiki teoreetilis-lingvistilises semantikas käsitletavaid järeldusi, kuid teiselt poolt lisanduvad järeldused, mille aluseks on üldine maailmateadmus (ontoloogia).

Meile olulised järeldused võib jagada kolme tüüpi sõltuvalt sellest, missugused lause semantilise struktuuri komponendid (predikaat, argumentid/rollid, argumentide kvaalistruktuur, ontoloogia) on relevantssed.

1. Järeldused, mis lähtuvad predikaadist, sõltumata selle argumentide realiseatsioonist konkreetsetes lauses (eeldusel, et lause on semantiliselt mittehälbig). See tüüp ei iseloomusta tegelikult üksikpredikaate, vaid nende semantilisi klasse. Tüüpnäide on semantilised kausatiivid (A põhjustab mingi muutuse B seisundis/olukorras): järeldub B uus seisund/olukord. Klassikaline kirjandusest tuntud näide on *tapma* (*A tappis B* $\rightarrow$ *B on surnud*), aga sama kehtib näiteks *lukkama* kohta näites *A lukkas B laua alla* $\rightarrow$ *B on laua all*. Oluline on, et predikaadid – nagu siin semantilised kausatiivid – oleksid verbide freimileksikonis vajaliku märgistusega varustatud. Tulevikus on kindlasti üks esimesi samme (arvutianalüüsis) komponentalüüsi järkjärguline arendamine, kus näiteks kausatiivide semantilises esituses on olemas põhjustamist esitav elementaarpredikaat, ja sellega seotakse vajalikud järeldused. Praegu praktiline semantika selleks valmis ei ole, ehkki R. Jackendoff ütleb: „Praegu-seks peaks olema selge, et sõnatähenduste üldistatud käsitlus pole võimalik ilma leksikaalse komponentalüüsi teooriata” (Jackendoff 2002: 377).

2. Enamasti sõltuvad lausetest tehtavad järeldused aga predikaadi ja argumentide kombinatsioonist, kusjuures argumentide puhul on oluline esiteks nende semantiline roll ja teiseks konkreetse üksuse (nt nimisõna) kvaalistruktuur. Eelmises tüübis oli näide verbiga *tapma*. Aga kui verbiks on *maha murdma* (*Hunt murdis meie lamba maha*), siis on olulised nii *Agendi* kui ka *Kogēja* konkreetsemad kvaalid: maha murdmise *Agent* saab ilmselt olla vaid loom, seejuures hammastega, mida kasutab *Instrumendina* [isegi lindude puhul ei kasutata seda verbi (??*Kotkas murdis meie lamba maha*), niisiis kuulub see info *Agendi* moodustajakvaali alla], ja teiseks, mahamurtav *Kogēja* peab samuti olema teatavat liiki loom, mitte suvaline elusolend (vrd **Kass murdis vihmaussi maha*): mõlemat tüüpi info on osa järeldustest, kui arvuti teksti analüüsib. Või võrreldagu lauseid *Torm murdis kase maha*; *Torm mur-*

dis kase ladva maha; Torm murdis kase oksa maha, [kus (*maha*) murdma on teises tähenduses]. Siin on oluline arvestada kase kui teatava puu kvaaliesitust: kui torm murrab maha kase, on see kase kui puu lõpp (formaalne kvaal), kui murrab oksa või ladva, tähendab see kase struktuuri muutust (moodustajakvaal). Need näited on juba lingvistilise ja ontoloogilise info piiri peal: ladva murdumine ja ühe oksa murdumine on kase jaoks ilmselt erinevate tagajärgedega.

3. Ent predikaatide ja argumentide lingvistilisest tähendusest, ka kvaalistruktuuridena esitatult, jääb arvutianalüüsis tihti väheks, et lausest relevantseid järeldusi tuletada. Vaja on lisainfot – seejuures just järeldustena –, mida saab arvestada kogu situatsiooni (näiteks tekstis järgnevate lausete) mõistmiseks. See info kuulub pigem inimeste maailmateadmusesse, sõnade tähenduste lingvistilisest kirjeldusest seda ei leia. Olgu näitena kaks lauset: *Poiss loopis sillalt puuks jõe jõe* ja *Poiss loopis sillalt kive jõe jõe*. Küsimus on, mis neist objektides edasi saab. Jõgi on vesi ja me teame, et puuks ei vaju vee alla, aga kivid kindlasti vajuvad. Niisiis võidakse edasises tekstis kirjeldada, kuidas poiss jälgis puukste, aga mitte kivide jõe pinnal liikumist. Ja see vahendav info saab tekkida vaid järelduste kaudu.

Ontoloogiat arvestava semantilise analüüsi probleem on see, et pole piiri, millest sügavamale ei minda. Seetõttu ongi vaja ise määratleda valdkond, mille sees ning kuivõrd üldontoloogilisel ja kuivõrd erialalisontoloogilisel tasemel kehtivaid seaduspärasusi käsitletakse. See erinevus on üldjuhul olemas kindlasti näiteks liikumise ja liigutamise valdkonnas.

4. Lühidalt esitusviisist

Nagu öeldud, on meie käsitluse teoreetiliseks aluseks – üldise mõistestiku ja põhimõtete osas – kontseptuaalne semantika, esitusviisina kasutame aga freime, mis on freimisemantika „tehniline” kirjeldusvahend. Vastuolu siin ei ole, sest freimid on kirjeldusstruktuuridena vanemad kui freimisemantika keeleteadusliku semantika kontseptsioonina ning mõeldud just teadmiste esitusviisina, mida oleks arvutil mugav töödelda (Minsky 1975). R. Jackendoffi esitusviis eeldab komponentanalüüsi ega ole mõeldud arvutianalüüsiks.

Freim on formaalses mõttes struktuur, mille tuumosa moodustavad kindlal viisil korrastatud atribuutide (tunnuste) ja nende väärtuste paarid (vt ka Õim, Saluveer 2002). Meie mudelis on atribuutideks semantilised rollid (argumentstruktuurides, mille peaks on verb kui potentsiaalne lause predikaat) ja kvaalide alla käivad omadused (nt *Kuju*, *Konsistents*, *Kehaosad* moodustajakvaali all või *Otstarve* teelilise kvaali all). Väärtustena on iga tunnuse puhul fikseeritud leksikonis (semantiliste kategooriate termineis, aga ka konkreetsete sõnadena) võimalikud keelelised üksused: näiteks *kõndima* puhul peab *Agent* olema ELUSOLEND ja *Instrument* peab olema *jalad* (NB! konkreetne leksikaalne üksus) ja siit tuleneb, et *Agendi* väärtuseks oleva ELUSOLENDI kvaalistruktuuris peab moodustajakvaali all olema määratletud atribuut *Kehaosa* väärtusega *jalad*. Põhimõtteliselt on seega oluline osutada, et freimis võib mingi atribuudi väärtuseks olla kompleksne semantiline struktuur: teine freim, hierarhiline kvaalistruktuur. Teiseks moodustavad freimid ise hierarhia: predikaadi/verbifreimi esimeseks elemendiks on tema

hüperonüüm (HYP; ülemmõiste), millelt ta pärib osa rolle koos võimalike väärtustega (*kõndima* HYP: LIIKUMA või selle konkreetsem alaliik), nii et neid rolle ei ole tarvis vastava üksuse kirjes korrata, välja arvatud sel juhul, kui mingi rolli täitjale esitab see üksus erinõudeid. Nii esitab *kõndima* erinõudeid *Agendile* ja *Instrumendile*, nagu osutasime, aga ilmselt mitte lähte-ja sihtkoha määratlusele ega kindlasti mitte ajamääratlustele.

Lausefreimis on predikaadi rollide ja argumentide väärtusteks juba lauses olevad konkreetsed struktuurid, näiteks nimisõna fraasid, kuid viimased on omakorda esitatud peanimisõnade kvaalistruktuuridena. Näiteks (*kõndis külmas vees* on kõndimise koharolli *Loc* (kus kõndis) osutav fraas *külmas vees* esitatud struktuurina, kus argumendiks on *vesi* ja selle kvaalistruktuuris on koos HYP: AINE ja selle all kvaalitunnused ja -väärtused *Konsistents*: VEDEL ja *Temperatuur*: KÜLM.

5. Kokkuvõtteks

Lausete semantiline analüüs (s.t analüüs, mille tulemuseks on lause tähenduse formaalne esitus) on reaalselt võimalik, ehkki teatavate piirangutega. Kõigepealt on muidugi vaja, et oleks olemas lause süntaktilise analüüsi moodul, mille väljund – lause süntaktiline esitus – on semantilise analüüsi mooduli sisendiks; ja teiseks on vaja, et oleks olemas lausetes esinevate sõnade tähendusstruktuure esitav leksikon, mis haakuks süntaktilises esituses kasutatud kategooriatega. Mõlema komponendi osas on eesti keele arvutilingvistiline kirjeldus sellisel tasemel, et saame lausete semantilist analüüsi üritada, mida olemegi juba teinud. On muidugi ka muid kitsendusi. Need puudutavad näiteks lausete struktuuri: me ei tegele liitlausetega, kus osalaused on sõltuvussuhetes, sest see käsitus viiks kohe semantikas osalausetega väljendatavate sündmuste, situatsioonide ja teiste võimalike ontoloogiliste seoste probleemi juurde. Samuti puudutavad kitsendused käsitletavat semantilist valdkonda: peale sõnade ja lausete puhtalt keelelise tähenduse (mille võime leida sõnaraamatust) on arvutianalüüsis vaja arvestada nn maailmateadmuse, praeguses terminoloogias ontoloogiat, varasemas nn naiivteooriaid. Ja lõpuks: kitsenduseks on kindlasti ka see, et praegune teoreetiline semantika ei võimalda minna eriti sügavale sõnade semantilisse struktuuri (nn semantiliste primitiivide probleem). Teoorias on see kuum teema, aga praktiliseks analüüsiks pole ühtki korralikku lahendust.

Kirjandus

- A&A 2008 nr 6 (semantika erinumber).
 Calzolari, Nicoletta 2008. Introduction of the Conference Chair. – Sixth International Conference on Language Resources and Evaluation. Conference Abstracts. Marrakesh, lk III–V.
 Fillmore, Charles 1968. The Case for Case. – Universals in Linguistic Theory. E. Bach, R. Harms (toim). New York: Holt, Rinehart and Winston, lk 1–90.
 Gruber, Jeffrey 1965. Studies in Lexical Relations. Ph.D. Dissertation. Massachusetts: Massachusetts Institute of Technology.

- Gruber, Jeffrey 1976. *Lexical Structures in Syntax and Semantics*. Amsterdam: North-Holland.
- Inheritance, Defaults and the Lexicon. T. Briscoe, A. Copestake, V. de Paiva (toim). Cambridge: Cambridge University Press, 1993.
- Jackendoff, Ray 1997. *The Architecture of the Language Faculty*. Cambridge, Massachusetts–London: The MIT Press.
- Jackendoff, Ray 2002. *Foundations of Language. Brain, Meaning, Grammar, Evolution*. New York: Oxford University Press.
- Kaljujärv, Kaarel 2007. *Attempto Controlled English as a Semantic WEB Language*. (Dissertationes Mathematicae Universitatis Tartuensis 55.) Tartu: Tartu Ülikooli Kirjastus.
- Katz, Jerrold 1972. *Semantic Theory*. New York: Harper and Row.
- Katz, Jerrold, Fodor, Jerry 1963. *The Structure of a Semantic Theory*. – *Language* 39, lk 170–210.
- Langetš, Margit 2004. Mõnda nimisõnade semantikast. – *Keel ja Kirjandus*, nr 10, lk 736–754.
- Lindström, Liina, Trägel, Ilona 2007. Eesti keele impersonaali ja seisundipassiivi vahekorra adessiivargumendi kasutamise põhjal. – *Keel ja Kirjandus*, nr 7, lk 532–553.
- Minsky, Marvin 1975. *A Framework for Representing Knowledge*. – *The Psychology of Computer Vision*. Toim P. H. Winston. New York: Mcraw Hill, lk 211–277.
- Müürisepp, Kaili, Orav, Heili, Õim, Haldur, Vider, Kadri, Kahusk, Nee-me, Taremaa, Piia 2008. From Syntax Trees in Estonian to Frame Semantics. – *The Third Baltic Conference on Human Language Technologies*. October 4–5, 2007, Kaunas. *Proceedings*. Vilnius, lk 211–218.
- OntoLinguistics 2007 = OntoLinguistics: How Ontological Status Shapes the Linguistic Coding of Concepts. A. C. Schalley, D. Zaefer (toim). Berlin–New York: Mouton de Gruyter.
- Primus, Beatrice 2006. Mismatches in Semantic Role Hierarchies and the Dimensions of Role Semantics. – *Semantic Role Universals and Argument Linking. Theoretical, Typological, and Psycholinguistic Perspectives*. I. Bornkessel, M. Schlesewsky, B. Comrie, A. D. Ffiederici (toim). Berlin–New York: Mouton de Gruyter, lk 53–87.
- Pustejovsky, James 1995. *The Generative Lexicon*. Cambridge, Massachusetts: The MIT Press.
- Semantic Role Universals and Argument Linking. Theoretical, Typological, and Psycholinguistic Perspectives*. I. Bornkessel, M. Schlesewsky, B. Comrie, A. D. Ffiederici (toim). Berlin–New York: Mouton de Gruyter, 2006.
- Van Valin, Robert 1999. Generalized Semantic Roles and the Syntax-Semantics Interface. – *Empirical Issues in Formal Syntax and Semantics 2*. F. Corblin, C. Dobrovie-Sorin, J.-M. Marandin (toim). The Hague: Thesus, lk 373–389. <http://wings.buffalo.edu/linguistics/research/rrg.html>.
- Õim, Haldur, Trägel, Ilona 2007. Teoreetilise keeleteaduse arengust mujal ja meil XXI sajandi algul. – *Keel ja Kirjandus* nr 2, lk 98–115.
- Õim, Haldur, Saluveer, Madis 2002. Freimid keelekirjelduses. – *Akadeemia*, nr 12, lk 2663–2678.

Semantics of Simple Sentences: a Theoretical Conception and Some Possibilities of Computer Analysis

Keywords: computational semantics, syntax-semantics interface, conceptual semantics and frame semantics, argument structure, semantic roles, inferences

The paper describes and analyses the main theoretical aspects of the model underlying a computer program for semantic analysis of Estonian simple sentences. A follow-up presenting the main aspects of the formal-technical realization is under way.

Sentence semantics is a relatively new but very intensively developing subject in theoretical linguistics as well as in computational linguistics. In our model we are trying to combine two theoretical conceptions: conceptual semantics (developed by Ray Jackendoff) and frame semantics (originally proposed by Charles Fillmore). The themes discussed are as follows: argument structure of sentences and argument types; semantic roles of arguments; qualia structure (as a means for semantic representation of arguments); inferences as a means for determining the “whole” meaning of a sentence; the problems arising from the need to take into account, along with the “pure” linguistic meaning of language expressions, also world knowledge (domain ontology).

We have restricted our subject area to the domain of motion (self-motion, e.g. walking or flying, as well as caused motion, e.g. throwing or bringing something somewhere). In the treatment of inferences we have so far restricted ourselves to the question “Which entities participating in the motion event do move, i.e. change places?” so that after the described event the program could answer the question “Where is X?”, where X is an entity participating in the motion event described in the sentence. For instance, in the case of a throwing event only the thrown object obligatorily changes its place, but in the case where an object is carried or brought somewhere by an agent the agent necessarily moves, too.

The input to the semantic analysis module is the dependency tree of a sentence. The tree, in turn, is provided by the syntactic analysis module, where the arguments of the predicate have syntactic labels (Subject, Object, Adverbials). Our program provides those arguments with semantic role labels (Agent, Object, Instrument, Locfrom, Locto etc.), and computes the inferences needed. The result of the analysis is represented in the form of a structure we call sentence frame.

Haldur Õim (b. 1942), Dr. Sc., University of Tartu, Institute of Estonian and General Linguistics, Professor Emeritus, haldur.õim@ut.ee

Heili Orav (b. 1969), PhD, University of Tartu, Institute of Estonian and General Linguistics, Research Fellow, heili.orav@ut.ee

Piia Taremaa (b. 1982), University of Tartu, Institute of Estonian and General Linguistics, post-graduate student, piuksu@ut.ee