

# Sõnaraamat keeleteaduses: generatiivse leksikoni teooria

**Margit Langemets**

*Eesti Keele Instituut*

Selles kirjutises vaadeldakse sõnaraamatu (leksikograafia, leksikoloogia) suhet teoreetilise keeleteadusega, mis omakorda on tänapäeval seotud arvutilingvistikaga või vähemalt arvutis olevate tekstikogudega. Samuti tutvustatakse üht leksikaalsemantilist teooriat – generatiivset leksikoni –, mis on saanud tunnustust mitmelt tuntud leksikograafiakorüfeelt ja leidnud kasutamist just arvutilingvistikas.

## Leksikograafia ja lingvistika

Üsna üldiselt on peetud lingvistika peaeesmärgiks eeskätt keele grammatika kirjeldamist, mitte “aina rafineeritumate sõnastike koostamist” (Rajandi 1999 (1968)). Leksikograafia (leksikoloogia) ja lingvistika vahel on keeleteaduse ajaloos olnud pigem isatapjalik lõhe kui sild (viimane on tekkinud alles uuemal ajal) ning ikka kangastub inimeste silme ette pilt sõnaraamatust kui igavesest ootajast riivil või arvutis. On pakutud, et üldkeele sõnaraamat pole lingviste huvitanud vähemalt kahel põhjusel: esiteks on kiputud seda pidama liiga kommertslikuks ettevõtmiseks, mida toodetakse ilma mingisugusele lingvistilisele teooriale toetumata, ja teiseks, kuna see on raamat sõnadest, siis teatava ajani jäi see (nagu semantikagi) üldise uurimishuvi alt täiesti kõrvale – leksikoloogiat lihtsalt ei loetud lingvistika haruks (Béjoint 1994). Sõnastikud peegeldavad küll hästi inimeste kõige vahetumat kogemust keelest, nimelt seda, et sõnadel on tähendus, ning ka kasutajad soovivad lihtsalt “vaadata sõna järele” (isegi eri tüüpi sõnaraamatutest otsitakse kõige enam just sõna tähendust<sup>1</sup>), ent näiteks omaaegsete generativistide seisukoht oli

---

<sup>1</sup> Seda kinnitavad prantsuse ja inglise sõnaraamatute kasutajate uurimistulemused: nii ükskeelse emakeelse, ükskeelse võõrkeelse kui ka kakskeelse sõnastiku puhul on esimesel kohal tähenduse otsimine (Svensén 1993). Võimalik, et eesti keele vormirohkus tõstab esimeseks morfoloogilise info järelevaatamise, seda vähemalt õigekeelsussõnaraamatus.

karm: “grammatikale ei anna need semantilised väited midagi juurde” (Rajandi 1999 (1968)).

Leksikograafiat saadab müüt, mis jutustab visalt, kuidas prakti- list sõnaraamatutööd tehakse ilma teooriat kasutamata ja kuidas see töö, mis sõnaraamatuartikliks vormub, ei saa mingil juhul olla teaduslik. Sellele on kaasa “aidanud” ka tegijad ise, sest täpselt samamoodi nagu Eestis on ka mujal täheldatud leksikograafide endi tagasihoidlikkust artiklite ilmutamisel, sest nad on “täiesti hõivatud sõnaraamatute koostamisega, selle asemel et selle üle teoretiseerida” (Kilgarriff 1992). Kaheldakse ka nn leksikograafi intuitsioonis, mille najal keeleandmeid analüüsitakse, mis siis, et tihti jõutakse nüüdsete arvutivõimaluste abil samasuguste tulemusteni.

Lingvistilise teooria mõttes iseloomustatakse kõigile teada suuri üldkeele sõnaraamatuid kui Saussure'i eelseid teoseid, mis kesken- duvad pigem üksiksõnale kui kogu leksikaalsele ja lingvistilisele süsteemile. Niisugust tüüpi esindab näiteks ka Saussure'i kaasaegne, brittide “Oxford English Dictionary” (ilmus vihikutena 1884–1928), kõnelemata kõigist varasematest ja paljudest hilisematest leksiko- graafilistest suurtöödest. Saussure'i üldideed, et sõna tuleb vaadelda selle paradigmaatilistes ja süntagmaatilistes seostes, alles hakatakse teadlikult sõnaraamatutesse juurutama. F Hausmann kritiseerib meieaegseid üldkeele sõnaraamatuid just selle poolest, et seal on liiga vähe lausemalle, kollokatsioone, sünonüüme ja antonüüme, morfosemantilistest paradigmatel rääkimata (Béjoint 1994).

Ka lingvistid ise on sõnaraamatust eemale hoidnud. H. Béjointi hinnangul muutus olukord Ameerikas varem kui Euroopas – oli lingviste, kes kuulusid kirjastuste kolleegiumidesse (L. Bloomfield, Ch. C. Fries jt). Ent samas luhtus täielikult katse saada 1961. aastal ilmunud “Webster's Third New International Dictionary, Un- abridged” koostamise juurde lingvisti: kõik, kelle poole pööruti, keeldusid, võibolla ebakindlusest testida praktikas oma teooriaid mitte ainult mõnede sobivate näidete, vaid kogu sõnavara ulatuses. Niisamuti on Eestis, suurte sõnaraamatutega kaasnev “töö vähene uudsus, argine süsteemipärasus tõrjuvad midagi omalaadset teha soovivaid noori kõrvale” (Rätsep 1987) – seletussõnaraamatu mant on pagenud grammatikaurijad Mati Erelt, Ülle Viks, Joel Sang, Ene Vainik.

Üldiselt on teoreetiline lingvistika sõnaraamatut puudutanud parimal juhul vaid riivamisi. Uuemad suunad sobituvad üldkeele

sõnaraamatuga halvasti või ei sobitu üldsegi mitte: generativistid on ise öelnud, et nende teoorial pole leksikograafia suurt midagi pakku (Béjoint 1994).

1970.–80. aastatest alates on leksikograafia ja lingvistika suhted tublisti soojenenud, seda kahtlemata suuresti tänu arvutilingvistika püüdlustele analüüsida ja sünteesida loomulikku keelt. Kui juba vägisi kippus levima arvamus, et kaasaegne lingvistika ei paku arvutianalüüsi jaoks midagi kasulikku, siis kontrastiks varasema aja süntaksikesksusele ilmus keelekirjeldusmudelitesse semantika. Sellega seoses tuli välja töötada semantilise analüüsi ja semantilise esituse vahendid (Koit 1996) ning peagi leiti, et semantilist infot on võimalik ammutada ka olemasolevatest sõnastikest.

Tänapäeval ollakse üldiselt ühel meelel, et teoreetiline lingvistika (eriti semantika) peaks pakkuma leksikograafia erinevaid vaateid keelele ja sõna tähendusele ning olema lingvistilise info esitamise aluseks teatavat tüüpi sõnaraamatus. Paljutöötavaimaks lingvistiliseks toeks on peetud prototüüpide teooriat ja semantiliste tunnusoonte teooriat, ent mõlemad on siiski seni praktikas läbi proovimata (Béjoint 1994). Vahendustööd lingvisti ja sõnaraamatuinimese vahel peaks tegema metaleksikograaf, aga leidub ka inimesi, kes on korraga nii lingvistid, leksikograafid kui ka arvutispetsialistid, näiteks B. Atkins, N. Calzolari, Ch. Fillmore, D. Geeraerts ja J. Pustejovsky ja mõned teised (Swanepoel 1994). Nemad on moodsa leksikograafia ja leksikoloogia käekäiku suunanud ja mõjutanud kõige enam.

Kokkuvõtteks võib nentida, et ühelt poolt läheb leksikograafia igal pool maailmas vältimatult seda teed, mis on äriliselt edukas, ent selle kõrval loodetavasti säilib tahe (teadmised) ja võimalus (inimesed, aeg, raha) edendada leksikograafiat ka lingvistikateooriate suunal.

## **Sõnaraamat ja tähendus**

Tähendust, eriti seda, mitu tähendust sõnal on, kuidas neid tõestada, eristada ja seletada, on uuritud keelefilosoofide ja semantikute poolt vähemalt Aristotelesest alates. Tavainimesele on asi selge: tuleb vaid avada sõnaraamat ja järele vaadata – enam levinud ehk põhitähendus on harilikult kohe esimesena kirjas. Lingvistide, eriti aga filosoofide jaoks on küsimus pikka ja põhjalikku arutlemist väärt. Enam-vähem ühel meelel ollakse, et tähendust väljendatakse enamasti keele abil, ent mis on sõna tähendus? Jätan siinkohal kõrvale filosoofide

tähendusõpetuslikud arutlused ja püüan, kui see üldse võimalik on, piirduda keeleteadusega, tegelikult veelgi kitsamalt, sõnaraamatuga.

On teada, et deskriptiivne lingvistika “vihkas” tähenduse kriteeriumi (Rajandi 1999 (1968)), sest pidas keele tähendusliku külje formaalset kirjeldamist võimatuks, kuigi loogikas olid formaalsed semantikateooriad ammu olemas. Semantilise komponendi ilmumist keelekirjeldustesse on põhjalikult seletanud H. Õim (Õim 1974, 1976), sealhulgas täheldades, et generatiivse semantika mudel – enne mõte (semantilised esitused), siis väljendus (pindstruktuurid, foneetiline esitus) – on “intuitiivselt loomulikum ja kergemini mõistetav” kui Chomsky vastupidine rida.

1970. aastatel ei huvitanud keeleteadust otseselt see, mis sünnib inimeste peades, kui nad kõnelevad, vaid keel kui “füüsiline reaalsus”: sõnad, laused, tekstid (Õim 1974). Tänapäeval peab eriti kognitiivne lingvistika keelt inimtunnetuse lahutamatuks osaks, mitte iseisevaks nähtuseks ega süsteemiks. Keelenähtusi üritatakse seletada universaalsete kognitiivsete sündmuste ja võimete põhjal, grammatika on üksnes inimese teadmiste struktureeritud varamu (Vainik 1992; Hanks 2000). Sõnaraamat on lingvistilise kirjelduse esitamise koht, teooria ülesanne oleks seletada suhteid lingvistiliste üksuste ja tähenduste vahel. Kui keel on pigem protseduurne kui deklaratiivne, nagu väidab kognitiivne lingvistika, siis on tähendused pigem sündmused kui objektid ning sõnaraamatu seletused pigem “potentsiaalsete tähenduste” korrastatud loendid kui sätestatud tähendused (Hanks 2000).

Sõna tähenduse uurimine on hiljuti uuesti elustunud, eriti palju tegeldakse mõiste visandamise ja kirjeldamisega, polüseemia ja konteksti interpreteerimisega. Raskeimini käsitletav on polüseemia, kus enim mõjutab erinevaid lahendusi just kontekst. Traditsiooniline polüseemia käsitus on enamasti deskriptiivne tähenduste loetlemine, uuemad suunad otsivad teatavat tüüpi sisemisi (loogilisi) struktuure ja reegleid, mis juhivad sõna tähenduse loomist ja kasutamist.

Arvutilingvistika huvitub sõnaraamatu semantilisest infost kahel küllalt erineval viisil (Wilks jt 1996). Esiteks, kui lähtutakse *teoreetilisest* huvist loomuliku keele semantiliste struktuuride vastu, siis kasutatakse sõnaraamatut kui keelelist abivahendit või andmekogu. Ja teiseks, kui lähtutakse *praktilisest* huvist ületada keeletöötuse “pudelikaela”, siis püütakse sõnaraamatu seletustest automaatselt

eristada formaalsed tähenduse ja teadmiste esitused. Need kaks suunda võivad eksisteerida koos, aga ka teineteisest sõltumatult.

Otseselt sõna tähendusega tegelevad leksikograafid, kes päev-päeva kõrval peavad a) otsustama, mida mingi sõna tähendab, mitu tähendust, tähendusvarjundit või erisugust tarvitust sel on, ja b) kõike seda sõnaraamatutähendusena üldmõistetavalt kirjeldama. On paratamatu, kuigi mõneti absurdne, et leksikograafid peavad seletama ka sõnu, mille tähendusest pole neil tegelikult õrna aimugi. Sageli küsitaksegi, kuidas leksikograaf ikkagi teab, mida mingi sõna tähendab? Ja kui ta teab, siis kuidas ta seda ammendavalt väljendab? Leksikograafi kolm elutähtsat abivahendit on:

- a) hästi disainitud, mahukas, autentset keeleainestikku sisaldav tekstiarhiiv (tekstikorpus)<sup>2</sup>;
- b) tema enda kui emakeele kõneleja teadmised ja intuitsioon; ning
- c) varem ilmunud sõnaraamatud ja grammatikad.

Eesti keele käsiraamatu semiootilise lühimääratluse järgi on leksikaalne tähendus “märgiga vastavusse seatud sisu” (Erelt jt 1997: 58). Sõnaraamatu tähendus on üldjuhul denotatiivne, aga kirja pannakse ka enam levinud konnotatsioonid (eesti seletussõnaraamatus nt *kana* naise kohta või *poisike* mehe kohta<sup>3</sup>), ja seletatakse pigem püsivat tavatähendust kui kontekstis loodavat-tekkiivat juhutähendust, kuigi ka siin on piiri tihti raske tõmmata.

## Tähenduse seletus

Sõnaraamatutähendus ehk, täpsemini, tähenduse seletus, kirjeldab (ühe) sõna tähendust (teiste) sõnade abil. Seletuseks on tavaliselt vabas vormis tekst – harilikult definitsioonid, ümberütlevad lühifraasid (ehk perifraasid), sünonüümid –, mille teooriat, inimlik ebatäielikkus ja ebasüsteemaatilisus on põhjustanud uskumatult palju kriitikat (vt näiteid allpool). Mõned sõnaraamatud pakuvad sõna tähenduse ja kasutuse täpsustamiseks rohkelt lisainfot, nii grammatilist kui ka stilistilist, ka autentseid katkeid tekstikorpusest, pildimaterjali, etimoloogilist infot jpm. Tihedalt tähendusekomponendiga seotud on tuletised, liitsõnad ja mitmesõnalised ühendid (ühend- ja väljend-

<sup>2</sup> Eesti seni parimad allikad on Eesti Keele Instituudi eesti kirjakeele arhiiv (4,2 mln sõnasedelit) ja Tartu Ülikooli eesti kirjakeele korpus TÜKK (1990. aastate korpuses on ca 1 mln sõnet).

<sup>3</sup> Piret Vollilt saadud näited.

verbid, fraseologismid, kollokatsioonid), mis otseselt justkui sõna tähendusse ei kuulu.

Enamikul juhtudel pole seletuses kasutatava keele/sõnastuse kohta kehtestatud eksplitsiitseid piiranguid, ent on siiski ka näiteid, kus sõnaseletuste süstemaatilist käsitlust kergendab spetsiaalselt välja töötatud defineerimissõnavara<sup>4</sup> või definitsiooni avarust vältida üritavad piirangud.<sup>5</sup>

Üldine seletamise põhimõte on, et tundmatut selgitatakse tuntu abil ehk tundmatu/keeruline objekt (märksõna) tehakse mõistetavaks tuntud või vähem keeruliste sõnadega. Märksõnal ja tähenduse seletusel on iseenesest üks ja sama tähendus, nad on vastastikku asendatavad. Tähenduse seletus (lad *definiens*) on oma "laiali laotatuse" tõttu tihti arusaadavam kui seletatav sõna (lad *definiendum*) ise. Ent ikkagi jääb küsimus: mida ja mismoodi "laotada"? Juba inglise leksikograafia isa Samuel Johnson tegi 18. sajandil tähelepanekuid, et pea võimatu on lingvistilistes terminites väljendada väga peeni tähendusnüansse, samuti näiteks sõnu, mis tähistavad loomi. Mis on elevandi mõte? Kas lisada pilt? Missugune pilt on õige? Kuidas definitsioon tegelikult "töötab", sellest palju ei teata. Mõned seletused tunduvad lihtsamad kui teised. Kas asi on lühiduses? Mõnikord on pikemad analüütilisemad seletused mõistetavamad, aga need on tihti jälle entsüklopeedilisemad.<sup>6</sup> Milline peaks olema seletuse ideaalne sisu?

Ükskeelsete sõnaraamatute seletused on pärinud iseäranis karmi kriitikat väga paljude lingvistide poolt. Olgu siinkohal näiteks toodud väike valik (viitamata väited pärinevad H. Béjointilt (Béjoint

<sup>4</sup> Teadaolevalt ainus sõnaraamat, kus see ka järjekindlalt läbi on viidud, on inglise ükskeelne *Longman Dictionary of Contemporary English* (LDOCE): välja töötatud (ehk nn kontrollitud) defineerimissõnavara koosneb ca 2000 sõnast, mis sageduselt esinevad igas kolmandas sõnaseletuses. Need sõnad ise on aga ca 6 korda polüseemilisemad kui ülejäänud, nn mittekontrollitud sõnad (ca 25 000) (Wilks jt 1996: 99).

<sup>5</sup> Seletuse täpsuse huvides ja mitmetähenduslikkuse vältimiseks kõrvaldati *Collins Cobuild English Language Dictionary* (COBUILD) sõnaraamatust ähmased lühendid nagu *etc.*, konventsioone murrab ka esitusviis, mis täislausetega kirjeldatud sõnaseletusi lubab lugeda nagu tavalist inglise keelt.

<sup>6</sup> Entsüklopeedilise ja lingvistilise info vahekorda on lahanud Enn Veldi (vt Veldi 2001).

1994), kes on omakorda refereerinud Weinreichi, Apresjani, Lakoffi, Mel'čuki jt):

- puuduvad teoreetilised põhjendused; ei arvestata rahvalikke seletusi;
- aetakse segi info maailma asjade ja lingvistilise märgi kohta;
- tähistamata on süntaktilised mallid;
- ebajärjekindlad terminid ühe sõnaraamatu piires;
- ei tehta vahet tähenduse ja kontekstiga seotud tarvituste vahel;
- tähistamata on konnotatiivsed väärtused;
- seletused on liiga ebamäärased, et neid saaks kodeerida; polüseemsete sõnade puhul jääb identifitseerimata, mis tähendust seletuses parasjagu kasutatakse;
- "Ilmsete, aga ikka vaid ebatäpsete kirjeldavate semantiliste vaatluste nimetamine definitsioonideks on muidugi täiesti põhjendamatu." (Rajandi 1999 (1968)); seletus ületab entsüklopeedilisuselt mõlemad eestikeelsed entsüklopeediad (Mäger 1971);
- eriti jäme viga on see, kui ta [= leksikograaf] mingit konkreetset tarvitusjuhtumit käsitab tähendusena (Mäger 1971);
- ühendi kogutähendus omistatakse ühele komponendile (Mäger 1971);
- vigaselt seletatakse isegi võõrsõnu (Mäger 1971).

Kritiseeritud on ka ringseletusi ( $A=B$ ,  $B=C$ ,  $C=A$ ), ent samas on osutatud, et niisuguseid ringe ei saa ega tulegi kunstlikult vältida, sest need väljendavad meie mõtlemise struktuure. Vaid kahest liikmest koosnevaid ringseletusi tuleb pidada leksikograafiliseks läbikukkumiseks – ei anna need ju midagi tähenduse mõistmise ja seletamise heaks.

M. Kay ei usu üleüldse, et "abstraktne" sõnaraamat – see, mis väidab end olevat teoreetiliselt "neutraalne" – võiks kõigi oma puudujääkidega realselt peegeldada mõtlemise struktuure (Wilks jt 1996). Ja B. T. Atkins küsib: "Kas on mingit põhjust – peale traditsiooni või kohusetundliku optimismi – uskuda, et sõnaraamatu artikkel annab reaalse pildi sellest, mida emakeele kõneleja teab sõna semantilistest omadustest?" (Béjoint 1994).

## James Pustejovsky: "Jah!"

Eelmises osas püstitatud murelikule küsimusele annab oma generatiivse leksikoni teooriaga kindla jah-vastuse Chomsky ambitsioonikas õpilane James Pustejovsky, kelle meelest on sõnaraamat üks ütleмата väärtuslik varaait ja kes veenab, et "sõna on väga kasulik koht suure hulga info edastamiseks"<sup>7</sup> ja "suur osa lause struktuuri puudutavast infost on kõige paremini kodeeritud leksikaalsest perspektiivist lähtudes" (Pustejovsky 1995: 5). Eeldades, et sõnaraamatu siht on seletada sõnade tähendusi, usub ta, et just seal on keele semantiline struktuur eksplitsiitselt nähtav ning võrreldav teadmiste esituse semantilise struktuuriga. Eriliselt rõhutab Pustejovsky, et ollakse jõutud punkti, kus lingvistika ja arvutileksikoloogia ilma vastastikuse infota enam hakkama ei saagi. Nende (vastandlike) uurimislade "ühinemine" (ingl *wedding*) on tema arvates nii tähtis, et peatselt on võimatu teha (arvuti)lingvistilisi uuringuid ja keeletöötlust, toetumata elektroonilistele leksikograafiliste kogudele. Kusjuures selle sünteesi keskmesse jääb sõna tähendus, leksikaalne semantika.

Ühtlasi toonitab ta oma kahte metodoloogilist lähtekohta:

- a) semantikat ja süntaksit tuleb vaadelda koos, nende lahutamine kukutaks uuringu läbi;
- b) sõnade tähendused peaksid kuidagimoodi peegeldama kognitiivse süsteemi mõistelisi süvastruktuure ja valdkonda (ingl *domain*) või ümbrust, kus sõna kasutatakse.

Järgnevalt püüan põgusa ekskursina kirjeldada generatiivse leksikoni teooria põhimõtteid, mõisteid ja väga lühidalt ka selle kasutamist maailmas.

## Generatiivse leksikoni teooria

Generatiivse leksikoni teooria (edaspidi GL) vastandub ühelt poolt generatiivsele grammatikale ja teiselt poolt traditsioonilisele ükskeelsele sõnaraamatule. Kui generatiivne grammatika analüüsib keelepäeva lause struktuuri ja on üldiselt alati hoidunud seda seostamast

---

<sup>7</sup> Intervjuu James Pustejovskyga. Ilmunud Hollandi tudengite arvutilingvistika ajakirjas Ta!. Internetis: <http://www.coli.uni-sb.de/~mineur/Itv/Pustejovsky.html>

“semantilist laadi tähelepanekutega” (Rajandi 1999 (1968)), siis generatiivne leksikon keskendub just tähenduste süstematiseerimisele ja modelleerimisele. Kui traditsiooniline sõnaraamat kirjeldab mitmetähenduslikkust kõikide tähenduste ja tähendusvarjundite eksplitsiitse loetlemise teel, siis generatiivne leksikon püüab seda liiasust ja ka lõplikkust vähendada ja neid tähendussuhteid reeglita. GL põhineb loogikal ja selle formaalne läbitöötatus võimaldab ka semantilist arvutitöötlust: luuakse sõnade tuumkirjed (ingl *core entries*) ja reeglid, mille abil põhitähendusest genereeritakse teisi, ka uusi, potentsiaalseid tähendusi.

GL-i teooria “ristiisaks” on peetud D. A. Cruse’i (1986 ilmus tema raamat “Lexical Semantics”), kes kirjeldas nähtust, kuidas leksikaalset vormi võib siduda mõnede tähendustega, mis pigem on potentsiaalsed kui eksplitsiitselt sõnaraamatus esitatud ja kuidas kontekst võib stimuleerida reeglid, mis genereerivad vastava tähenduse. Uurijate tähelepanu on hakanud köitma polüseemse sõna tähenduste vahelised semantiliste suhete *ennustatavad regulaarsused*, mis omakorda võiks pakkuda malle või mustreid mitmetähendusliku sõna kirjeldamiseks sõnaraamatus.<sup>8</sup>

GL vaatab lausungi konteksti kui nähtust, mis koosneb paljudest erinevatest *generatiivsetest faktoritest*, mis kokku seletavad ära viisi, kuidas inimesed loovad kontekste ja manipuleerivad nendega teatavate kitsenduste piires.

Pustejovsky vaatleb alustuseks just substantiive ja adjektiive (ingl *nominals*), sõnu, mis senimaani on tema väitel verbide kõrval põhjendamatult vähese tähelepanu osaks saanud, tohtides olla vaid “augutäiteks” verbifreimides. Uuritakse ka põhjussuhete semantikat (eriti kausatiivsust), semantiliste tüüpide kitsendavat mõju süntaktilistele konstruktsioonidele, sündmuste (ja osasündmuste) formaalset käsitlemist, polüseemiat üldiselt.

## Loogiline polüseemia

Viisid, mil moel sõnad tähendusi kannavad, on erinevad. Lisaks Weinreichi pakutud kontrastiivsele ja laiendavale polüseemiale defi-

<sup>8</sup> Kõiges selles on ka kaheldud. Näiteks Lakoff peab polüseemiat motiveerituks, mitte suvaliseks ega ette ennustatavaks: iga uus tähendus on seletatav vähemalt ühe eelnenud tähenduse seosega metafoori või metonüümia kaudu (Béjoint 1994).

neerib Pustejovsky loogilise polüseemia<sup>9</sup> kui niisuguse laiendava mitmetähenduslikkuse, kus leksikaalne kategooria ei muutu ning sõna eri tähendustel on osaliselt kattuv, sõltuv või jagatud tähendus. Näiteks nimisõnade süstemaatilised tähenduse vaheldused<sup>10</sup>:

loendatav–loendamatu, nt *lammas*

- a *Valge, must lammas.* [1. imetaja]  
 b *Sõin lõunaks lammast.* [1. ... || lambaliha]

• anum–sisu, nt *pudel*

- a1 *Täitis pudeli veega.* [.. (vedeliku) anum]  
 a2 *Annad sa titele rinda* [.. (vedeliku) anum]  
*või pudelit 'pudelist piima'*  
 b1 *Poiss on hakanud* [|| ka piltl. kõnek. alkoholipudel, selle  
*pudeli poole piiluma* sisuks olev alkoholne jook]  
 b2 *Jõin paar pudelit õlut.* [|| pudelitäis]

otsene–kujundlik, nt *uks, aken*

- a1 *Vaatan aknast välja* [1. ... avaus; selle raamid, ka aknalaud]  
 a2 *Aknal on lilled;* [1. ... avaus; selle raamid, ka aknalaud]  
*Rinnutas aknal 'aknalaua'*  
 b *Aken Euroopasse* \_\_\_\_\_ [1. ... || piltl.]

• toode–tootja, nt *Postimees*

- a *Postimees vallandas mitu toimetajat.*

<sup>9</sup> Sedasama nähtust on kirjanduses nimetatud mitmel eri moel: süstemaatiline polüseemia (*systematic polysemy*, Nunberg & Zaenen), regulaarne polüseemia (*regular polysemy*, Apresjan), edasiviide/edasisuhe (*deferred reference*, Nunberg), semantilise ülekande reeglid (*semantic transfer rules*, Leech), tähenduse ülekanne (*sense transfer*, Sag), konnektorid (*connectors*, Fauconnier), tähenduse laiendused (*sense extensions*, Pustejovsky, Briscoe, Copestake), leksikaalne võrgustik (*lexical networks*, Lakoff), allregulaarsused (*subregularities*, Wilensky), leksikaalsed implikatsioonireeglid (*lexical implication rules*, Ostler & Atkins) (vt Nunberg, Zaenen 1992). Samade näidete toel illustreeritakse tihti ka metonüümiat, kus asja (entiteedi) enese nimetuse asemel kasutatakse üht selle omadust (atribuuti).

<sup>10</sup> Nurksulgudes on vastavad tähenduse seletused "Eesti kirjakeele seletussõnaraamatus" Kõik näitefraasid on samast pärit. Nagu näha, on sõnaraamat pidanud kasutama mitut eri moodust tähenduse täpseks väljendamiseks. On võimalik, et mõned näitemärksõnad ise (nt *lammas*) on esialgu inglise keele "kallakuga", sest eesti keele polüseemia uurimine on alles algusjärgus.

b *Laual vedeleb Postimees.*• taim–toit, nt *pirn*

a <i>Mahlane <u>pirn</u></i>	[1. pirnipuu vilji]
b1 <i><u>Pirnid</u> on õites</i>	[2. pirnipuu]
b2 <i><u>Pirn</u> on üks kõvemaid puiduliike</i>	[2. ..    selle puit]

protsess–tulemus, nt *katse, eksam*

a1 <i>Haige tegi <u>katseid</u> istukile tõusta.</i>	[1. püüe, üritus midagi teha, sooritada v. saavutada]
a2 <i>Perenaine tõi <u>katseks</u> suure tüki õunakooki.</i>	[1. ..    katsetus, proov ..]
b1 <i>Sportikooli astujatel tuleb sooritada kehalised <u>katsed</u>.</i>	[2. hrl. pl. võimete v. teadmiste ning oskuste kontroll]
b2 <i>See teooria põhineb arvukatel <u>katsetel</u>.</i>	[3. uurimise eesmärgil mingi nähtuse esilekutsumine v. mõjustamine, eksperiment]

• koht–inimesed, nt *linn, Tartu*

a1 <i>Jalutasime <u>linnast</u> välja.</i>	[1. .. asula]
?a2 <i><u>Linn</u> 'linlikkus, linna elulaad' meelitab noori.</i>	[1. .. asula]
b1 <i><u>Linn</u> 'linnavalitsus' otsustas ..</i>	[1. .. asula]
b2 <i>Seda teab terve <u>linn</u> 'kogu linna rahvas'</i>	[1. .. asula]
a3 <i>Jõudis <u>linnast</u>, <u>linna pealt</u> tagasi.</i>	[1. ..    kõnek. kesklinn, vastandatuna äärelinnale, kodule, töökohale]

Loogilist polüseemiat kandvate sõnade õige tähendus on tuvastatav üksnes konteksti abil. Pustejovsky väidab, et kuigi tavapärase leksikoni ülesehitusest piisab kontrastiivse mitmetähenduslikkuse väljendamiseks, on selle abil võimatu haarata polüseemia tegelikku olemust. Et kompositsioonilisust säilitada, tuleb rikastada kompositsiooni väljendamise semantikat.

Piirid sõna fikseeritud tähenduste vahel on tihti liiga jäigad. Alati pole selge, mis on õige tähendus antud kontekstis, tihti sisaldab üks tähendus teist. Niisugused verbid on näiteks *küpsetama* (ingl *bake*), *keetma* (ingl *cook*) ja *praadima* (ingl *fry*), mis kasutuses võivad tähendada kaht tüüpi tegevust: sama asja seisundi muutust (a1, a2) või millegi uue tekkimist (b):

a1 <i><u>Küpsetatud</u> kaalikad, kartulid, õunad</i>	(seisundimuutus)
[1. toorest toiduainet (kuivas) kuumuses toiduks valmistama]	
a2 <i>Päike <u>küpsetas</u> selja pruuniks</i>	[1. ..    piltl.
(päikesekuumuse kohta:) väga tugevasti mõjuma, praadima]	

b *Ena küpsetas sünnipäevaks kringli* (loov tegevus)  
 [1. toorest toiduainet (kuivas) kuumuses toiduks valmistama]

Probleem on, et kattub just tähenduse tuumkomponent (ingl *core*) ja ainuüksi valikuliste kitsendustega pole võimalik tagada õiget tähendust. Sõna eri tähenduste tuum- ja perifeersete komponentide osalise kattumise tõttu on traditsiooniliste sõnaraamatute tähenduste lame lineaarne loetlev esitus ebaadekvaatne (Atkins 1991).<sup>11</sup>

## Semantiliste tüüpide süsteem

GL püüab keelt kirjeldada nii, et saaks hõlmata leksikaalset loovust ja tähenduse laienemise nähtusi. New Mexico ülikooli 1990. aastate projekti käigus loodi leksikaalsete struktuuride baas<sup>12</sup>, kus semantilised teadmised on sõnadega seotud. Leksikaalse üksuse  $\alpha$  struktuuri kirjeldab neljast semantilise esituse tasemest koosnev valem:

$$\alpha = \langle A, E, Q, I \rangle$$

1. ARGUMENDISTRUKTUUR (A) kirjeldab leksikaalse üksuse loogilisi argumente (arv, tüüp ja süntaktiline esitus). Sellest tasemest üksi ei piisa leksikaalse üksuse semantiliseks iseloomustamiseks, ent see on vajalik komponent, just argumendi struktuur on loogiline lähe semantikasse. (Lastekeele uurijad on oletanud, et see on määravaim tegur verbi tähenduse omandamisel (Pustejovsky 1995).)

2. SÜNDMUSESTRUKTUUR (E) kirjeldab sündmuse tüüpi (olukord, protsess, üleminek) ja sisemisi loogilisi osasündmusi. Nimi-sõnade puhul: mis sündmusele nimisõna osutab.

3. *QUALIA*-STRUKTUUR (Q) kirjeldab leksikaalse üksuse tähendust. Just selle taseme info on kõige otsesemalt sõnaraamatuga seotud. Pustejovskyt on inspireerinud Moravciku käsitlus (vt Moravcik 1975), mis näeb generatiivsust Aristoteelse teoorias olemise ja sündmuste põhjuste (kr *aitia*, lad *causa*) kohta. Nii rõhutab generatiivse leksikoni teooria Pustejovsky väitel just sõna tähenduse neid külgi, mis seni on leksikaalse semantika formaalsetes mudelites

<sup>11</sup> See kriitika ei kehti eesti sõnastike (nt EKSS) ja selle eeskujuks olnud vene sõnastike puhul, millel on oma peen süsteem alltähenduste ja tähendusvarjundite märkimiseks.

<sup>12</sup> Kirjed tekitati automaatselt ükskeelse *Longman Dictionary of Contemporary English*'i (LDOCE) põhjal, mis sisaldab 55 000 sõna-arklit.

tähelepanuta jäänud. Tähenduse olemuslikke külgi (lad *qualia*<sup>13</sup>) kirjeldavad seletusviisid, kus on tähistatud erinevad suhted:

- KOOSTIS (OLEMUSLIKUD KOMPONENDID) (ingl *constitutive*, lad *causa materialis*<sup>14</sup>) = objekti (asja) moodustajasuhe oma olemusega (materjal, kaal, osad, komponendid);
- KUJU (ingl *form*, lad *causa formalis*) = objekti formaalsed tunnused (orientatsioon, suurus, kuju, mõõtmed, värvus, asend);
- EESMÄRK (ingl *telic*, lad *causa finalis*) = objekti eesmärk või funktsioon (tegija eesmärk tegevust sooritades, sisemine funktsioon või tegevust täpsustav eesmärk);
- TEGIJA (SÜND) (ingl *agentive*, lad *causa efficiens*) = objekti suhe looja, loomise v tekkimisega (looja, kunstlikult tehtu, loomulik, põhjuslik ahel).

Mõnes mõttes võib *qualia*'t kujutleda kui leksikaalse üksusega seotud omaduste või sündmuste hulka, mis kõige paremini annab edasi sõna tähendust. Näiteks erinevused sõnade *romaan* ja *sõnaraamat* vahel on seletatavad selle abil, mida nende objektidega tehakse. Mõlemad on tüüpiliselt raamatud [KUJU = raamat], aga nende kasutus erineb: *romaan* on *lugemiseks* [EESMÄRK = lugemine], *sõnaraamat* *kasutamiseks* [EESMÄRK = kasutamine].

See eristus pole muidugi ainus viis neil mõistritel vahet teha. Erinevad on nad ka sisemise struktuuri poolest: *romaan* teksti struktuur on harilikult jutustav [KOOSTIS = jutustav], *sõnaraamat* aga on sõnade korrastatud või süsteemipärane loend [KOOSTIS = korrastatud/süsteemipärane]. Ka on nad eri viisil tekkinud: *romaan* üldiselt kirjutatakse [TEGIJA = kirjutamine], *sõnaraamat* koostatakse [TEGIJA = koostamine].

Kui nüüd võrdluseks vaadata, kuidas nimetatud suhteid kajastab eesti seletussõnaraamat, siis nähtub, et tuvastatavad on pea kõik nimetatud omadused. Ainuke "häda", et üksi tähenduse seletusest ei piisa – olulist lisa tuleb otsida ka näitematerjalist:

*romaan* – kirj. ulatuslik jutustav keerukama süžeeaga kirjandusteos.  
.. *Romaani* kirjutama, lugema, tõlkima. ..

<sup>13</sup> Et paremini mõista: lad *qualis*, *quale* – milline, mis tüüpi, (nii nagu); lad *qualia* – omadused, laadid, väärtused. Sama tüvi esineb ka võõrsõnas *kvaliteet*.

<sup>14</sup> Ladinakeelsed vasted on võetud H. Seideli raamatust (Seidel 1984).

**sõnaraamat** – eri raamatuna väljaantud süsteemipärane sõnastik.

.. *Sõnaraamatut koostama, kasutama.* ..

**sõnastik** – korrastatud sõnavarakoogu, sealhulgas (hrl. väiksem) sõnaraamat.

Näide. Sõnade *romaan*, *sõnaraamat* ja *sõnastik* kirjed "Eesti kirjakeele sõnaraamatus" Alla joonitud on eelnevalt suhetes mainitud sõnad.

Pustejovsky lisab siiski, et niisugused piirid pole absoluutselt selged, ent üldiselt on need siiski olemas. *Qualia* ei ole lihtsalt huvitavate faktide kuhjamine leksikaalse üksuse või objekti kohta, vaid pakub Pustejovsky meelest lähtepunkti semantilise analüüsi jaoks.

4. LEKSIKAALNE PÄRILUSESTRUKTUUR (ingl *lexical inheritance structure*) tuvastab tulenemissuhte ühe leksikaalse struktuuri ja teiste struktuuride vahel (mis millest tuleneb), ükskõik, kuidas see sõnaraamatus on organiseeritud.

### Generatiivsed mehhanismid semantikas<sup>15</sup>

Sõna tähendusele võib läheneda kaht eri teed pidi: primitiivipõhiste (ingl *primitive-based*) või suhtepõhiste (ingl *relation-based*) teooriate abil. Primitiivide pooldajate (Wilks, Katz, Lakoff, Schank) meelest saab sõna tähendust defineerida määratud hulga primitiivide abil. Vastupidiselt sellele väidab suhtepõhine teooria (Quillian, Fodor, Carnap, Brachman), et sõnu pole tarvis primitiivideks lahutada, kui on olemas eksplitsiitselt määratletud ja loogilistele järeldusreeglitele toetuv võrgustik, mille abil seletatakse seoseid leksikaalsete tähenduste ja propositsioonide vahel.

Pustejovsky pakub uudse lähenemise, vaadeldes pigem leksikaalse semantika generatiivseid ehk kompositsioonilisi aspekte kui tähenduse lahutamist (dekompositsiooni) teatavaks hulgaks primitiivideks. Enda sõnutsi esindab ta konservatiivset tähenduse dekompositsiooni, kus leksikaalseid üksusi on vaid minimaalselt lahutatud teatavatesse struktuuridesse (või mallidesse).

Eelmises alaosas kirjeldatud neli semantilise esituse taset on omakorda seotud generatiivsete mehhanismide abil, mis varustavad kontekstis sõna kompositsioonilise interpreteeringuga. Nende generatiivsete operaatorite hulgas on kolm semantilist transformatsiooni – tüübikohustus (ingl *type coercion*), valikuline sidustus/sidumine (ingl

<sup>15</sup> Kuna käesolevasse kirjutisse ei mahu nende keeruliste mehhanismide lähem iseloomustamine, siis toon nad siin vaid nimetamisi ära.

*selective binding*) ja kaaskompositsioonilisus (ingl *co-composition*) – mis hõlmavad semantilist seotust süntaktiliselt erivormiliste väljendite vahel. Pustejovsky põhiline järeldus on, et süntaktilist fraasi ei saa interpreteerida väljaspool süntaktilist ja semantilist konteksti, milles ta esineb.

### Generatiivse leksikoni teooria kasutamisest

GL on viimastel aastatel äratanud küllaltki suurt huvi ja leksikaal-semantikaga tegelejad on mõnelgi korral selle poolt või vastu sõna võtnud. Adam Kilgarriff testis GL-i mehhanismi nende korpuses esinevate sõnade puhul, millele sõnastikud seletust ei andnud: tulemus oli, et selliseid mittestandardseid juhtumeid oli üldse 5% ja neist omakorda vaid 5% oli seletatav GL teooria abil (Kilgarriff 2001). Kilgarriff peab seda küll elegantseks, aga mitte täielikuks leksikaalse produktiivsuse teooriaks.

Produktiivsem on GL kindlasti sõnastikemaailmas. GL-i lihtsustatud versiooni on kasutanud kaks suurt sõnaraamatuprojekti: Acquilex ja SIMPLE. Lihtsustus seisneb selles, et sõnakirjetes kasutatakse üksnes semantilise info esmast esitust (sh *qualia*-struktuuri) ja ei kaasata semantilisi transformatsioone (Badia, Sauri 2000).

Projekti SIMPLE (lõppes 2000) käigus koostati 12 Euroopa keele<sup>16</sup> jaoks semantilised tuumleksikonid, mille aluseks on ühtne semantilise info mudel. Leksikograaf kasutab mingi leksikaalse üksuse tähenduse semantiliseks kodeerimiseks kindlaid skemaatilisi struktuure (mustreid).

SIMPLE'i mudelis on formaliseeritud järgmise info esitus:

- semantiline tüüp (\*)<sup>17</sup> nt kunstlikult tehtud ese;
- valdkond (\*), nt üldkeel, matemaatika;
- leksikograafiline lühiseletus (\*);
- argumentidstruktuur (\*);
- valikulised kitsendused argumenti juures (\*);
- verbi sündmuse tüüp (\*);
- argumentide seos (ingl *link*) süntaktiliste freimidega (välja töötatud eelnenud projekti PAROLE raames);
- tüübihierarhia info;

<sup>16</sup> Olgu need siis ka üles loetud: katalaani, taani, hollandi, inglise, soome, prantsuse, saksa, kreeka, itaalia, portugali, hispaania, rootsi.

<sup>17</sup> Tärniga info on obligatoorne iga kodeeritud tähenduse juures.



Laiemas perspektiivis nähakse võimalust seostada eri keelte semantilised kirjeldused omavahel seotud mitmekeelseks leksikoniks (Villegas jt 2000). Taanlased on SIMPLE'iga seotult eraldi tegelnud püsiühendite ja metafooridega (Nimb, Pedersen 2000).

## Kokkuvõtteks

Ei ükskeelse sõnaraamatu põhjal tuletatud leksikaalsed struktuurid ega kakskeelse sõnaraamatu vasted suuda täielikult kajastada tegelikku rikkalikku keelepruuki, mille kirjalikku varianti registreerivad tänapäeval ükskeelsed ja paralleelsed tekstikorpused. Uurida, täiustada ja katsetada tuleb nii meetodeid, abivahendeid kui ka allikaid, et tabada tabamatut – keelt, mida ühed loovad ja teised mõistavad.

## Kirjandus

- Atkins, B. T. 1991. Building a lexicon: Reconciling anisomorphic sense differentiations in machine-readable dictionaries. – *International Journal of Lexicography* 1, 84–126.
- Badia, T.; Sauri, R. 2000. Towards a filtering of the relevant semantic information from MRDs. – *EURALEX 2000 Proceedings I–II*. Toim Ulrich Heid jt. Institut für Machinelle Sprachverarbeitung, Universität Stuttgart. 651–662.
- Béjoint, H. 1994. *Modern Lexicography: An Introduction*. Oxford UP.
- Hanks, P. 2000. Contributions of lexicography and corpus linguistics to a theory of language performance. – *EURALEX 2000 Proceedings I–II*. Toim Ulrich Heid jt. Institut für Machinelle Sprachverarbeitung, Universität Stuttgart. 3–13.
- Kilgarriff, A. 1992. *Polysemy*. PhD thesis. University of Sussex. <http://www.itri.bton.ac.uk/~Adam.Kilgarriff/publications.html#1992>
- Kilgarriff, A. 2001. Generative lexicon meets corpus data: the case of non-standard word uses. ITRI technical reports. <ftp://ftp.itri.bton.ac.uk/reports/ITRI-01-10.pdf>
- Koit, M. 1996. Arvutuslingvistika arenguloost ja stuudiumist. – *Keel ja Kirjandus* 3, 171–178.
- Erelt, M.; Erelt, T.; Ross, K. 1997. *Eesti keele käsiraamat*. Tallinn: Eesti Keele Sihtasutus.
- Moravcik, J. M. 1975. Aitia as generative factor in Aristotle's philosophy. – *Dialogue* 14, 622–636.
- Mäger, M. 1971. Eesti kirjakeele seletussõnaraamatu proovivihiku puhul. – *Keel ja Kirjandus* 1, 39–46.

- Nimb, S.; Pedersen, B. 2000. Treating metaphoric senses in a danish computational lexicon – different cases of regular polysemy. – EURALEX 2000 Proceedings I–II. Toim Ulrich Heid jt. Institut für Machinelle Sprachverarbeitung, Universität Stuttgart. 679–691.
- Nunberg, G.; Zaenen, A. 1992. Systematic polysemy in lexicology and lexicography. – EURALEX '92 Proceedings I–II. Toim Hannu Tommola jt. Tampere. 388–396.
- Pustejovsky, J. 1995. *The Generative Lexicon*. The MIT Press.
- Rajandi, H. 1999 (1968). Eesti impersonaali ja passiivi süntaks. (Kandidaadiväitekiri 1968.) Eesti Keele Instituudi toimetised, 3. Tallinn.
- Rätsep, H. 1987. Eesti keele uurimise *status quo*: lünki ja väljavaateid. – Keel ja Kirjandus 3, 129–135.
- Seidel, H. 1984. *Aristoteles und der Ausgang der Antiken Philosophie*. Berlin: Dietz Verlag.
- Svensén, B. 1993. *Practical Lexicography: Principles and Methods of Dictionary-Making*. Oxford University Press.
- Swanepoel, P. 1994. Problems, theories and methodologies in current lexicographic semantic research. – EURALEX '94 Proceedings. Toim. W Martin jt. Netherlands. 11–26.
- Vainik, E. 1992. Genereerib inimene, mitte grammatika. – Keel ja Kirjandus 3–4, 156–160, 216–223.
- Verspoor, C. M. 1997. *Contextually-Dependent Lexical Semantics*. PhD thesis. The University of Edinburgh.  
[http://www.cogsci.ed.ac.uk/~kversp/ftp\\_html/thesis.html](http://www.cogsci.ed.ac.uk/~kversp/ftp_html/thesis.html)
- Veldi, E. 2001. Entsüklopeediad ja sõnaraamatud. – Keele kannul. Pühendusteos Mati Ereli 60. sünnipäevaks. Tartu. 322–334.
- Vider, K. Kahusk, N.; Orav, H.; Õim, H.; Paldre, L. 2000. Eesti keele teaurus. – Arvutuslingvistikalt inimesele. Toim T. Hennoste. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 1. 127–152.
- Villegas, M.; Bel, N.; Lenci, A.; Calzolari, N.; Rummy, N.; Zampolli, A.; Sadurni, T.; Soler, J. 2000. Multilingual linguistic resources: from monolingual lexicons to bilingual interrelated lexicons. – Proceedings of the Second International Conference on Language Resources and Evaluation (LREC 2000). Vol I-III. 1193-1200.
- Õim, H. 1974. *Semantika*. Tallinn: Valgus.
- Õim, H. 1976. Elementaartähendused keele semantilises struktuuris. – Keel ja Kirjandus 10–11, 598–605, 675–682.
- Wilks, Y.; Slator, B.; Guthrie, L. 1996. *Electric Words: Dictionaries, Computers, and Meanings*. (ACL-MIT Press Series in Natural Language Processing). Cambridge, Massachusetts, London: The MIT Press.