

**TARTU ÜLIKOOL
FILOSOOFIATEADUSKOND
ÜLDKEELETEADUSE ÕPPETOOL
ARVUTUSLINGVISTIKA ERIALA**

Kristel Uiboaed

SÕNASELETUSTE GENEREERIMINE TESAURUSE INFO PÕHJAL

Bakalaureusetöö

Juhendaja Kadri Vider, M.A.

TARTU

2005

SISUKORD

SISUKORD	2
SISSEJUHATUS	4
1. SÕNAARTIKLI ÜLESEHITUS	5
2. SÕNAARTIKLIS SISALDUV INFO	6
2.1 MÄRKSSÕNA.....	6
.....	7
2.2 GRAMMATILINE INFO SÕNAARTIKLIS.....	7
2.2.1 <i>Morfoloogiline info</i>	7
2.2.2 <i>Süntaktiline info</i>	8
2.3 SÕNADE STILISFÄÄR JA KASUTUSINFO.....	9
2.4 TÄHENDUSE SELETAMINE.....	9
2.4.1 <i>Seletuste tüübid</i>	10
2.5 ETÜMOLOOGIA.....	12
2.6 FRASEOLOGISMID.....	12
2.7 NÄITED.....	12
2.8 VIITED.....	13
3. TESAURUSEST	14
3.1 WORDNET.....	14
3.2 EUROWORDNET.....	15
3.3 EESTI KEELE TESAURUS (TEKSAURUS).....	15
4. SEMANTILISED SUHTED	17
4.1 SÜNONÜUMIA.....	17
4.2 HÜPERONÜUMIA JA HÜPONÜUMIA	19
4.3 ANTONÜUMIA.....	21
4.4 MERONÜUMIA.....	22
4.5 OSALUS-ROLLISUHTED.....	24
4.6 PÕHJUSLIKKUSE SUHE.....	26
4.7 OSASÜNDMUSE SUHE.....	27

4.8 SEISUNDISUHTED.....	27
4.9 KLASSI-ESINDAJA SUHTED NING MÄÄRATLEMATUD SUHTED.....	28
4.10 XPOS-SUHETEST.....	29
5. SÕNASELETUSTE GENEREERIMINE TESAURUSE INFO PÕHJAL.....	30
5.1 SELETUSTES TEKKINUD PROBLEEMID.....	32
5.1.1 Morfoloogia.....	35
5.2 MILLINE VÕIKS OLLA PROGRAMM?.....	36
KOKKUVÕTE.....	37
KIRJANDUS.....	38
SUMMARY.....	40
LISA 1.....	41
LISA 2.....	43
LISA 3.....	46
LISA 4. DISKETT	

SISSEJUHATUS

Selles bakalaureusetöös käsitletakse sõnaseletusi sõnaraamatus, sõnaartikli ülesehitust ning selles sisalduvat infot (1.–2. ptk). Samuti antakse ülevaade teaurusest (3. ptk) ja semantilistest suhetest (4. ptk) – eelkõige teauruse kontekstis. Töö eesmärgiks on vaadelda, kas ja kuidas oleks võimalik teaurust ja selles sisalduvat informatsiooni kasutades genereerida sõnaseletusi, kui loomulikud sellised seletused oleks, millised probleemid tekivad ning kuidas oleks neid võimalik parandada – sellest koosnebki töö praktiline pool (5. ptk).

Töö teoreetiline pool tugineb suuremas osas Margit Langemetsa magistritööle (vt Langemets 2000) ning “Eesti kirjakeele seletussõnaraamatu” seni ilmunud vihikutele (vt EKSS 1988 – 2004), samuti nii eesti keele teauruse kui WordNeti ja EuroWordNeti kohta ilmunud artiklitele.

Töö praktilise poole aluseks on tekstifail (vt lisa 4. diskett) eesti keele teauruse infoga, kus sõnad on jaotatud sünohulkadega omakorda hulkadesse nii, et nad on seal juba koos kõigi oma seostega teauruses. Algsed, ingliskeelsed suhete nimed on asendatud eestikeelsetega (vt lisa 4. diskett). Seejärel on vaadeldud kõiki selliselt tekkinud ridu eraldi ning analüüsitud, kui loomulikud oleksid sel viisil loodud sõnaseletused. Antud on ülevaade probleemidest, mis tekkisid ning on esitatud mõned ettepanekud, mida teha selliste probleemide lahendamiseks. See võiks olla eeltöoks programmile, mis genereeriks automaatselt sõnaseletusi teauruse info põhjal.

1. SÕNAARTIKLI ÜLESEHITUS

Sõnaartikkel on sõnaraamatu põhiüksus. “Sõnaartikkel on sõnaraamatu organiseerimise printsiip, mille mõte on näidata, mis info (mingitelt alustelt lähtudes) kuulub kokku” (Langemets 2000: 30). Tavaliselt algab sõnaartikkel märksõnaga, mis selle juurde kuuluva ainesega moodustavadki sõnaartikli. Näiteks “Eesti kirjakeele seletussõnaraamatus” (edaspidi EKSS) eristatakse märksõna juures ühteist erinevat osa. Loomulikult ei ole kõik need osad esindatud kõigi märksõnade juures, leksikaalse info hulk sõltub konkreetsest sõnast. Märksõnal võib olla ka mitu tähendust. Tavaliselt esitatakse sellised sõnad erinevate märksõnadena, EKSS-s näitks araabia numbritega:

sidend, -i² s. 1. el. vooluringi teat. osa

1. lgv. lause element (sidesõna, siduv ase- v. määrõna, partikkel), mis vormistab grammatilise seose lauseosade v. osalausete vahel.

Märksõnale järgneb leksikaalne informatsioon – nii grammatiline kui ka süntaktiline. Lisatakse näitelauseid, et anda ülevaade vastava sõna kasutusvaldkonnast ja -kontekstist. Kui märksõna esineb kollokatsioonides või on osa mingist fraseologismist, siis lisatakse märksõnale ka vastavad püsiühendid.

2. SÕNAARTIKLIS SISALDUV INFO

Tavaliselt algab iga sõnaartikkel sõnaraamatus märksõnaga, millele järgneb erinev informatsioon selle artikli kohta. Sõnaartiklis sisalduva info võib jagada kuueks: ortograafiliseks, foneetiliseks, grammatiliseks, pragmaatiliseks ja semantiliseks infoks, lisandub veel illustreeriv materjal (Langemets 2000: 28). Sõnaartiklite esitamise viis sõltub sellest, millise sõnaraamatuga on tegemist. Infot võib esitada märksõnadena tähestikulises järjekorras, aga kui tegemist on mõistelise sõnaraamatuga, esitatakse informatsioon mõisteliselt. Järgnevalt antakse ülevaade sõnaartikli ülesehitusest ja selles sisalduvast informatsioonist. Ülevaade põhineb EKSS-I (vt EKSS) ja Margit Langemetsa magistritööl (vt Langemets 2000).

2.1 MÄRKÕNA

Sõnaartikkel algab märksõnaga. Märksõna vorm on eesti keeles tavaliselt verbidel *ma*-infinitiiv ja substantiividel ainsuse nominatiiv (nii ka EKSS-s). Eesti-soome sõnaraamatutes on tavaliselt märksõna vormiks verbidel *da*-infinitiiv. Kui sõna ainsuses ei kasutata, on märksõnaks vastava sõna mitmuslik vorm (*bikiinid*, *prillid* jt). Kui märksõna on antud ainsuslikult, kuid seda sõna tavaliselt kasutatakse mitmuses, siis lisatakse märksõnale märgend *pl*. Märksõna esitab sõna ortograafilise info. EKSS-s antakse sõnaartiklis ka sõnamoodustuslikku infot, märkides näiteks liitsõnapiiri:

tähendus|funktsioon lgv. ...

tööjõu|defitsiit maj. ...

Märksõnas võidakse osutada ka sõna fonoloogilistele omadustele (välde, rõhk jmt). Kuna eesti keeles on sõna kirjalikult sarnane hääldusele, siis fonoloogiline info ei ole primaarne. Häälduse märkimine on oluline näiteks tsitaatsõnade puhul. EKSS-s on hääldus märgitud nurksulgudesse:

charleston [tšarlston], ...

curriculum vitae [kurriku- viitä]⁶, ...

Eraldi märksõnadena antakse tavaliselt ka homonüümid. Näiteks EKSS-s tähistatakse homonüüme eraldi sõnaartiklitena ning neile lisatakse ilma punktita rooma number:

siire, siirde¹⁹ **I** s. **1**. siirdumine; üleminek. ...

siire, siirde¹⁹ **II** s. siirmine, ülekanne. ...

2.2 GRAMMATILINE INFO SÕNAARTIKLIS

2.2.1 Morfoloogiline info

Tavaliselt antakse eestikeelsetes sõnaraamatutes muutuvate märksõnade järel nende sõnade **tüübinumbrid** ning olulisemad vormid. Sõnade tüübistik on esitatud kas sõnaraamatu alguses või lõpus, EKSS-s on sõnatüüpide tabel sõnaraamatu algusosas. Tüüpsõna näitab, millise sõna eeskujul sõna käänata või pöörata. Tavaliselt liitsõnadele tüübinumbreid ei lisata, nii nagu EKSS-s, kuna liitsõna muutmise seisukohalt on oluline põhisõna, mis määrab ka liitsõna muutmisparadigma. Märksõnale on lisatud tema **sõnaliigimärgend**:

ekraan, -i²² s. ...

korrekselt adv. ...

vahajas, -ja, -jat² adj. ...

EKSS-s ei ole sõnaliigimärgendit verbidel. Kõik sõnad tüübinumbriga 37 kuni 49 on verbid (EKSS).

“**Seisundivormid** on teatavate nimisõnade üksikvormid, mis häägustavad subsantatiivide ja adverbide vahelist piiri” (Langemets 2000: 43). Ka seisundivorme märgitakse sõnaraamatutes, aga EKSS-s selle kohta eraldi märgend puudub, küll aga selgitatakse selliste sõnade tähendust:

- kulul** postp. [gen.] arvel. **1.** kellegi rahadega, kellegi ülalpidamisel. ...
- 2.** kellegi huve kahjustades, kedagi v. midagi mingis suhtes piirates. ...
- 3.** kedagi v. midagi pilgates, kellegi v. millegi üle v. kohta. ...

Sõnaartiklis antakse infot ka adjektiivide ja adverbide **võrdluskategooriate** kohta. EKSS-s viidatakse sõna algvormile, kui märksõnaks on komparatiivi või superlatiivi vorm.

rohkem adv. **1.** komp. < palju; ...

Samuti märgitakse nimisõnade **loendatavust ja loendamatus**. Loendamatud on tavaliselt abstraktsed substantiivid. Eesti keele sõnaraamatutes pole sõnade loendatavust ja loendamatus eriti kombeks märkida. Kui märksõnal on märgend *sg/ains*, siis mõnikord võib see tähendada ka loendamatus (Langemets 2000: 46).

2.2.2 Süntaktiline info

Sõnaraamatu ülesandeks on vahendada ka süntaktilist infot. Sõna kasutamise seisukohalt on oluline teada selle **rektsiooni**. EKSS-s märgitakse rektsiooni mitmel moel. Adpositsioonide rektsiooni on märgitud nii, et vastava märksõna juurde on lisatud nurksulgudesse kääne, mida selle adpositsiooniga peaks kasutama:

all I. postp. ... [gen.] ...

ilma I. prep. [abess.] ...

kuni ... II. prep. [term.] ...

Muudel juhtudel edastatakse info rektsiooni kohta näitelausetes. Üritatakse anda võimalikult palju ja erinevaid näitelauseid, nii et näha oleks sõna erinevad kasutusviisid. Nii on näitkeks verbi *arvestama* juures teiste seas selline illustratiivne materjal *Peab arvestama kohalikke tingimusi, rahvuslikke traditsioone, harjumusi. Temaga võib igas olukorras arvestada*. Kui aga sõnal või üksiktähendusel on üksainus kindel rektsioon,

märgitakse see ümarsulgudesse tähenduse seletuse lisana (EKSS). Samuti on oluline märkida verbide laiendeid ning nende vorme, mis tavaliselt ilmnevad näitelausestes.

EKSS-s on **verbikategooriad** esindatud pöörde-, tegumoe-, aja- ja kõneliigikategooria.

2.3 SÕNADE STIILISFÄÄR JA KASUTUSINFO

Lisaks semantilisele informatsioonile on sõnaraamatutes esindatud ka pragmaatiline info. Pragmaatilist infot edastatakse kõikvõimalike märgendite kaudu. Sellised märgendid tähistavad stiili, eriala, märksõna murdelisust jmt. Selline informatsioon on keelekasutuse seisukohalt väga oluline.

Järgnevalt on toodud näited mõndadest stiilmärgenditest, mis on esindatud EKSS-s.

eestkostetav, -a² s. jur. ...

ennatama³⁷ 1. murd. ...

erustama³⁷ hrv. ...

kõrva|puhuja¹ s. van. ...

kõrva|põletik med. ...

mokerdama³⁷ murd. ...

muidu|mees nlj. ...

mõist, -u²² s. hrv. ...

mää¹⁴ s. 1. lstk. ...

2.4 TÄHENDUSE SELETAMINE

Märksõnadele järgnevad sõnaraamatutes sõnaseletused. Seletused peaksid avama märksõna sisu, seletus on märksõna semantiline info. Tavaline on, et tundmatut üritatakse seletada tuntu abil ehk keerulisema sõna sisu avatakse vähem keeruliste sõnade abil. Eesti keele kohta ei ole välja töötatud spetsiaalset defineerimissõnavara, nii nagu näiteks *Longman Dictionary of Contemporary English* kasutab vaid kindlat defineerimissõnavara.

Need definitsioonides kasutatavad sõnad pole mitte keele kõige sagedasemad sõnad vaid sõnaseletuste kõige sagedasemad sõnad. (Langemets 2000). EKSS-s avataksegi sõna tähendus tavaliselt lühikese kirjeldava seletusena:

jenka⁶ s. Soomest pärinev seltskonnatants. ...

johtuma³⁷ tulenema, põhjustatud olema. ...

Oskussõnade puhul võib olla vajadus kasutada ka entsüklopeedlisi elemente. EKSS-s on lisatud taime- ja loomanimedele ka nende ladinakeelsed nimed ning keemilistele elementidele nende sümbolid ja mõõtühikute juures on vastavad lühendid. Partiklite seletused on aga tunduvalt pikemad.

2.4.1 Seletuste tüübid

Parafras ehk ümberütlemine. See on märksõna teiste sõnadega ümberütlemine, võib pidada ka mingil määral sünonüümiks, kuid ühe sõna asemel väljendatakse mitme sõnaga:

suur|uluk suur loom. ...

Sünonüümid. Ka sünonüümide abil on võimalik sõnu seletada. Märksõnale lisatakse tähenduse poolest lähedane või sarnane lihtsam sõna, mis peaks aitama avada märksõna sisu. EKSS-s on lisatud sünonüüme ka kirjeldavale seletusele. Näiteks:

sumadan, -i, -i¹⁰ **1.** van. kohver. ...

Antonüümid. Ka antonüümid võivad olla abiks märksõna tähenduse mõistmisel. Tavaliselt lisatakse antonüümid sõnaseletuses sünonüümidele või kirjeldustele. EKSS-s tuleb erandkorras ette ka seda, et kirjeldavat seletust on raske leida ning seetõttu on seletuseks vaid antonüüm ning sõnatähendus peaks avanema järelkult vaid tema antonüümi kaudu.

Definitsioonid. Keelemärgi tähenduse väljendamiseks tuuakse välja mõiste distinktiiv- ehk eritunnused. Sõnu seletatakse semantiliste suhete abil, näiteks hüponüümia abil. (Langemets 2000: 56).

Ekstensionaalsed definitsioonid püüvad määratlema mõiste ulatust ning seega on rohkem esindatud terminoloogiasõnaraamatutes kui üldkeele sõnaraamatutes (Langemets

2000: 56). Üldkeele sõnaraamatutes on selline teaduslik seletamine pigem taunitav. Teine äärmus on jällegi seletuste liigne lihtsustamine, mille eeliseks võib pidada seda, et sel viisil esitatud seletus peaks kõigile täiesti arusaadav olema.

sümbioos, -i²² s. biol. eri liiki organismide vastastikku kasulik v. tarvilik kooselu.

...

Tulenemistähendused. Et selgitada märksõna sisu, lisatakse talle vastav tuletusalus (Langemets 2000: 57). See on väga levinud seletusviis ka EKSS-s. Seletust ei avata otse, vaid osutatakse, et tähenduse mõistmine on võimalik teiste sõnade kaudu. Eriti kasutatakse sellist seletusviisi *us*-liiteliste abstraktsubstantiivide puhul.

suviselt adv. <suvine. ...

Reegel. Märksõna järele lisatakse reegel, kuidas seda sõna keeles kasutatakse (Langemets 2000: 57).

2.5 ETÜMOLOOGIA

Etümolooogia näitab sõnade päritolu. Kui märksõna mõistmise seisukohalt on etümolooogia oluline, siis vastav informatsioon ka esitatakse. Siiski tavalised sõnaraamatud etümolooegiat eriti ei märgi. Etümolooogia kohta saab täpsemat infot eraldi etümolooogiasõnaraamatutest.

2.6 FRASEOLOGISMID

Fraseologismide tähendus ei tulene enamjaolt neid moodustavate sõnade tähendusest, seega on vaja neid eraldi seletada. Fraseologismide kõrval on ka teisi püsiühendeid, neid seletatakse EKSS-s tavaliselt muu näitematerjali hulgas ja abil. Eraldi fraseoloogiasõnaraamatutes on fraseologismid tervikuna märksõnadeks. EKSS-s on need tavaliselt sõnaartikli lõpus ning fraseologismi algust tähistab eriline märk (Δ). Fraseologismide varieeruvate osiste vahel on tilde ning püsiühendi fakultatiivne osa on ümarsulgudes (EKSS).

Δ juuksekarva otsas ~ küljes rippuma ~ olema ...

Δ vaim tuleb [kellelegi] peale; vaim on [kellelgi] peal ...

2.7 NÄITED

Näitelaused lisatakse märksõnale selleks, et näidata konteksti, milles seda sõna kasutatakse. Tavaliselt on need lühilaused, mis peaks andma pildi vastava sõna kasutuspiirkonnast. Eriti oluline on näidetes välja tuua kollokatsioonid – sõnad, mis tavaliselt koos esinevad. Näidetena võib kasutada ka tsitaate (ka EKSS-s on seda tehtud). Sel juhul on kindlasti lisatud tsitaadi autor (vahel lisatakse ka muud bibliograafilised

andmed, kuid EKSS-s on märgitud vaid autor) tsitaati eraldab muust tekstist vastav märgend. (EKSS).

2.8 VIITED

Kuna pole otstarbekas sarnast infot samas sõnaraamatus mitu korda trükkida, siis on sageli vaja suunata sõnastiku kasutaja sõnaartikli juurde, kus vajalik informatsioon peaks olema. EKSS-s on viitamist kasutatud palju fraseologismide puhul. Suunatakse fraseologismi ühe või teise osa juurde, kus selle tähendust täpsemalt seletatakse. Näiteks:

vaene, vaese⁴ **1.** adj. ... Δ **vaimust vaene** vt. vaim.

tühi, tühja³³ **1.** adj. ... Δ **tühi koht** vt. koht. ...

3. TESAURUSEST

Tesaurus on mahukas leksikaalsemantiline andmebaas, kus sõnad on omavahel seotud mõisteliselt, semantiliste suhete kaudu. Tesauruses on ühendatud seletussõnaraamat, mõisteline sõnaraamat, sünonüümi- ja antonüümisõnaraamat (Vider, Orav 1998). Tuntuim arvutitesaurus on WordNet, mille eeskujul on tehtud EuroWordNet – mitmekeelne leksikaal-semantiline andmebaas, millesse on kaasatud ka eesti keel. Eesti üldkeele tesaurust hakati koostama 1998. aastal (täpsemalt vt ptk 3.3).

3.1 WORDNET

"Semantilist andmebaasi, mis keskendub mõistele ja semantiliste suhete kaudu tema semantilisele väljale, võib nimetada tesauruseks" (Vider jt 2000). Tesauruses on esmatähtsad semantilised suhted, morfoloogiline ja süntaktiline info pole primaarsed. Eesti keele tesauruse eeskujuks ongi WordNet (WN). WordNeti sooviti algselt luua selleks, et testida mõnede psühholingvistiliste teooriate paikapidavust. 1985. aastal otsustati Princetoni ülikoolis G. Milleri juhtimisel luua inglise keele leksikaalsemantiline andmebaas. (Vider jt 2000).

WN-i põhiüksuseks on sünohulk (ingl. *synonym set, synset*), mille moodustavad ühte mõistet (*concept, meaning*) väljendavad sünonüümsed (sama tähendusega) sõnad ja sõnaühendid. Sünohulga suurim erinevus sünonüümisõnastiku sünonüümireast on see, et sünohulk võib olla ka üheliikmeline, sest tesauruse ülesandeks on **mõistete** esitamine (sünonüümisõnastik aga esitab keeles leiduvaid sünonüüme). Seega tesauruses seovad semantilised suhted sünohulkadena esitatud mõisteid. (Vider jt 2000).

Suurema osa sõnadest moodustavad verbid ja substantiivid. WordNet seab sõnade esitamisele teatavad piirangud, kuna kõik sõnad peavad olema täistähenduslikud. Päringut tehakse siiski sõnastikuvormi kaudu. (Vider, Orav 1998).

3.2 EUROWORDNET

EuroWordNet on mitmekeelne leksikaal-semantiline andmebaas. WordNetist erineb see põhiliselt oma mitmekeelsuse poolest. Eri keelte tesaurused on ühendatud keeltevahelise indeksi (ILI) kaudu. EuroWordNeti peamiseks eesmärgiks pole testida erinevaid psühholingvistilisi hüpoteese, põhiline rakendus on mitmekeelne tekstiotsing (Vider, Orav 1998). Oma mitmekeelsusest tingitult on EWN-s kahte erinevat tüüpi mooduleid: keelest sõltuvad ja keelest sõltumatud moodulid (Vider jt 2000).

3.3 EESTI KEELE TESAURUS (TEKSAURUS)

Eesti keele tesaurus on samuti *wordnet*'i tüüpi. Eesti üldkeele tesaurust on koostatud Tartu Ülikooli arvutilingvistika uurimisrühmas 1998. aastast alates. Eesti keele tesaurus on enamusel sünohulkadest olemas viited ingliskeelsele WordNetile. Eesti üldkeele tesauruse koostamisel on järgitud Princetoni WordNeti ja EuroWordNeti põhimõtteid. Samuti on eesti üldkeele tesaurus üks kaheksast EuroWordNet-2 projekti tulemusena saadud ja ELRA kaudu levitatavast *wordnet*-tüüpi tesaurusest. (TEKsaurus).

Eesti üldkeele tesaurus on sarnane oma eeskujudele. Sõnad on esitatud sünohulkadena, igal sünohulgal on oma number. Igal sõnal sünohulgas ehk igal sünohulga liikmel on oma tähendusnumber. Üks sõna võib esineda mitmes sünohulgas ning tal on number ka siis kui, ta on sünohulgas ainuke sõna. Suurema osa eesti keele tesaurusest moodustavad verbid ja substantiivid, vähem on adjektiive ja väike osa sõnadest on proopriumid. Osadele sõnadele on lisatud ka seletused, mis on võetud EKSS-st. Leksikaalsed allikad eesti keele tesauruse koostamisel on olnud järgmised: "Eesti kirjakeele seletussõnaraamatu" ilmunud vihikud,

KeeleWebi ühispäring, mis haarab ka: eesti keele tesauruse (Filosoft), Asta Õimu “Sünonüümisõnastiku” ja “Antonüümisõnastiku”, Tartu Ülikooli eesti kirjakeele korpuse (TÜKK), WordNeti, EuroWordNeti, Paul Saagpaku “Sünonüümisõnastiku” (pole elektroonilisel kujul), mitmesugused kakskeelsed sõnastikud nagu inglise-eesti ja eesti-inglise sõnastikud. (TEKsaurus).

4. SEMANTILISED SUHTED

Tesauruses on sõnad esitatud sünohulkadena ning omavahel seotud semantiliste suhete kaudu. Täendusuhed esitavad assotsiatsioone, mis moodustavad keerulise semantilise võrgustiku: teades, kus sõna selles võrgus asub, suudame eristada paremini ta tähendust (Vider, Orav 1998). EWN-i loomise käigus on välja töötatud ka testid, mille abil on võimalik kontrollida tähendustevahelisi suhteid. Esialgu on olemas sellised testid vaid noomenite ja verbide jaoks. Järgnev ülevaade katab suurema osa semantilistest suhetest, mis on tesauruses esindatud. Samuti ei ole eraldi kirjeldatud suhteid verbide ja substantiivide kohta nii nagu seda on tehtud EuroWordNeti dokumentides. Näidetes on kasutatud sõnu, mis on selle bakalaureusetöö praktilise poole aluseks (vt lisa 4. diskett). Eestikeelsed testlauseid on osaliselt võetud kadri Videri ja Heili Orava artiklist (vt Vider, Orav 1998). Teise osa testlausete eeskujuks on EWN dokumentides olevad testlauseid (vt Alonge 1996, Climent jt 2005).

4.1 SÜNONÜÜMIA

„Sünonüümia ehk samatähenduslikkus tähendab kahe või rohkema vormilt erineva lekseemi denotatsiooni identsust“ (Karlsson 2002: 250). Sünonüümia on üks kesksemaid suhteid tesauruses. Kõik sõnad jagunevad sünohulkadesse. Tesauruse põhimõte on, et kõik tähenduselt sarnased sõnad oleksid koos, seega kuuluksid samasse sünohulka, hoolimata nende kasutusvaldkonnast, stiilist, murdelisusest jmt. Tavaliselt peetakse sünonüümideks sõnu, mida saab üksteisega asendada. Selliseid täielikke sünonüüme on keeles siiski vähe, tavaliselt on sõnadel erinevad tähendusvarjundid ning nad pole üksteisega täielikult

asendatavad, ilma et tekstis midagi ei muutuks. Tesauruses on sünonüümidele seatud nõue, mille kohaselt ühte sõna peab olema võimalik asendada teisega nii, et ei muutu öeldu tõeväärtus. Eraldi märgendit tesauruses sünonüümide jaoks pole, kuna sarnased sõnad ongi juba jaotatud sünohulkadesse. Sünonüümia omaduseks on veel see, et sõnad on asendatavad mõlemat pidi. (Vider, Orav 1998).

Näiteks:

Kui miski/keegi on X, siis on ta ka Y.

Kui miski/keegi on Y, siis on ta ka X. (Vider, Orav 1998).

Kui miski on tugevus, siis see on ka sitkus.

Kui miski on sitkus, siis on see ka tugevus.

Kui miski on teaduskond, siis on see ka fakulteet.

Kui miski on fakulteet, siis on see ka teaduskond.

Kui keegi on üliõpilane, studioosus, siis on ta ka tudeng.

Kui keegi on tudeng, siis on ta ka üliõpilane, studioosus.

Verbidevahelise sünonüümia test võiks eesti keeles olla selline:

Kui miski/keegi X-b, siis ta ka Y-b.

Kui miski/keegi Y-b, siis ta ka X-b.

Kui keegi lamab, lebaskleb, lamaskleb, siis ta ka pikutab.

Kui keegi pikutab, siis ta ka lamab, lebaskleb, lamaskleb.

Kui keegi uurib, siis ta selgitab (midagi) välja, uurib järele.

Kui keegi selgitab (midagi) välja, uurib järele, siis ta uurib.

Kui miski asetseb, siis ta ka paikneb, on, asub.

Kui miski paikneb, on, asub, siis ta ka asetseb.

Lisaks sellele eristatakse ka lähisünonüümiaid ning sõnaliikidevahelist sünonüümiaid – sünonüümia erinevate sõnaliikide vahel (tavaliselt on see siiski lähisünonüümia), tesauruses põhiliselt verbide ja noomenite vahel. Tavaliselt on selliste sõnade vahel mingi morfoloogiline seos, näiteks eesti keeles esineb selline suhe enamjaolt verbidest tuletatud *mine*-liiteliste substantiivide ja verbide vahel. Tesauruses (nii eesti keele tesauruses kui EWN-s) on lähisünonüümia märgendiks *near_synonym* ja sõnaliikide vahelise sünonüümia

märgendiks *xpos_near_synonym*, eestikeelne vaste neile suhetele on aga samasugune – *peaaegu samatähenduslik on*. Tegelikult on ka sünohulga liikmed enamasti lähisünonüümid, kuna täielikke sünonüüme on keeles väga vähe. Eelnevalt mainitult aga on tesauruses sünonüümidele seatud nõudeks tõeväärtuse muutumatus sõnade üksteisega asendamisel.

4.2 HÜPERONÜÜMIA JA HÜPONÜÜMIA

Hüponüümia ja hüperonüümia on hierarhilised alistussuhted. Kõrgema kategooria sõna on hierarhiliselt alama mõiste hüperonüüm ning madalama kategooria sõna on oma hüperonüümi suhtes hüponüüm. Mõlemad suhted on asümmeetrilised ja transitiivsed, s.t et nendes suhetes olevad objektid ei ole vastastikku asendatavad ning hierarhiliselt kõrgema objekti esindaja on näitkes hüperonüüm kõigi hierarhiliselt madalamate objektide suhtes. (Climetn jt 2005). Seda suhet testides saab eristada:

- 1) noomenite/verbide ja
- 2) noomenite/verbide ja liiki/klassi/tõugu/sorti väljendavate noomenite/verbide vahelist hüperonüümia ja hüponüümia suhet.

Esineb ka lähihüperonüümiat ning lähihüponüümiat. EWN-s on hüperonüümia ja hüponüümia märgendid vastavalt *has_hyperonym* ja *has_hyponym* eesti keelde tõlgitakse need jällegi järgnevalt: *on teatud liiki* (noomenite puhul) ja *üks liik on*.

Verbidevahelist hüponüümiat nimetatakse **troponüümiaks**. Princetoni WordNetis ongi tehtud vahet troponüümia ja hüponüümia vahel. Verbide puhul ei kehti päris sama test, mis noomenite puhul, semantiline erinevus kahe verbi vahel pole samal viisil kirjeldatav kui kahe hüponüümia suhtes oleva noomeni vahel. Troponüümiat kahe verbi vahel võiks väljendada järgnevalt: *tegema mingit tegu X, on sama, mis tegema teatud viisl mingit tegu Y*. Troponüümiat võib pidada ka teatavaks sisaldumisvahekorraks või tuleneemisvahekorraks. Näiteks iga üldisema mõiste X suhtes olev troponüüm Y sisaldab ka X-i omadusi, on samal ajal samuti X või igast Y-st tuleneb ka X. Näiteks: *naima, naist*

võtma, kosima on troponüümid verbi *abielluma* suhtes, kuid kui *naida, naist võtta, kosida*, siis ka *abiellutakse; kosimisest tuleneb ka abielu*. Seetõttu esindab troponüümia ka teatud sisaldumissuhet/-vahekorda või tulenemissuhet – s.t sellises suhtes olevad verbid on alati ajaliselt sama ulatuvusega ja on omavahel teatavas sisaldumis- või tulenemisvahekorras. Kuna tesauruses esitatakse mõistete mitte üksikute sõnade vahelisi suhteid, siis sellest tulenevalt võib üks sõna olla teise suhtes troponüüm ühes sünohulgas, kuid teises sünohulgas võib sama sõna olla ka teistes suhetes näiteks osalus- või kaasnevussuhetes. (Fellbaum 1993).

Eesti keele tesauruses tuleb selline vahe välja suhete nimedes. Kui noomenite puhul on hierarhiasuhete nimed *on teatud liiki* ja *üks liik on*, siis verbide puhul vastavalt *on teatud viisil* ja *üks viis on*. EWN-s on aga suhte nimi mõlema jaoks sama.

Testid nende suhete jaoks (nii noomenite hüponüümia kui verbide troponüümia) on esitatud järgnevalt.

1) Noomenite/verbide vaheliste suhete jaoks:

X on teatud Y ~ Kui see on X, siis peab ta olema ka Y (Vider, Orav 1998).

Vaidlus, vaidlemine on arutelu, diskussioon, mõttevahetus.

Noorus on iga, vanus.

Perioodiline väljaanne on trükis, trükitoode, üllitis.

Liikumapanek on tegu.

kaaluühik on ühik, mõõtühik.

Nappima on teatud viisil puuduma, puudu olema.

Abielluma on teatud viisil liituma, ühinema.

2) noomenite/verbide ja liiki/klassi/tõugu/sorti väljendavate noomenite/verbide vaheliste suhete jaoks:

X on teatud liiki/klassi/tõugu/sorti Y (Vider, Orav 1998).

Laupäev on teatud liiki nädalapäev.

Õppeasutus on teatud liiki haridusasutus.

Kollasus, kollane on teatud liiki spektrivärv.

Sonaat on teatud liiki kompositsioon, heliteos, muusikateos, helitöö, helind.

Mähkima, mässima on teatud viisil varjama, katma.

Läbi käima on teatud viisil suhtlema, lävima.

EWN-s eristatakse hüperonüümia ja hüponüümia puhulgi suhteid sõnaliikidevaheliselt nende märgendid on vastavalt *has_xpos_hyperonym* ja *has_xpos_hyponym*. Eesti keelde on need suhted tõlgitud samamoodi kui tavalisedki hüperonüümia ja hüponüümia (nii oli ka lähisünonüümia puhul).

4.3 ANTONÜÜMIA

Antonüümia on semantiline suhe, mis seisab leksikaalsemantilise paradigma liikmete – antonüümide vahel. Ühe liikme tähendus vastandub teise liikme tähendusele. (Õim 1998). Antonüümia on sümmeetriline suhe. Tegelikult võib antonüümiat pidada ka sarnasussuhteks, kuna vastandatavatel sõnadel peab olema midagi ühist, neid vastandatakse vaid ühe kindla tunnuse alusel, näiteks on mees ja naine mõlemad inimesed, kuid on samas ka antonüümid. Antonüümid on võimalik jagada kolme rühma: 1) relatiivsetel vastandustel põhinevad, 2) komplementaarsel vastandusel põhinevad ning 3) vektorvastandusel põhinevad antonüümid (Õim 1998). EWN-s on antonüümia suhte märgendiks *antonym* (eestikeelne vaste *vastand on*) ning lähiantonüümia märgendiks *near_antonym* ja *xpos_near_antonym* (eesti keeles *peaaegu vastand on*). Viimane neist on jällegi EWN-s eristatud sõnaliikidevaheliselt, eesti keeles muidugi on kahel viimasel vahet tehtud, kuid tõlgitud ühtemoodi, nii nagu oli ka kõigi eelnevate suhetegagi. Testlause antonüümia jaoks võiks olla noomenite puhul selline:

X ja Y on mõlemad teatud liiki Z-d, aga X on Y antonüüm.

Välismaa ja kodumaa on mõlemad teatud liiki maad, aga välismaa on kodumaa on antonüüm.

Lootusetus ja lootus on mõlemad teatud liiki tunded/tundmused, aga lootusetus on lootuse antonüüm.

Ratsionaalarv ja irratsionaalarv on mõlemad teatud liiki reaalarvud, aga ratsionaalarv on irratsionaalarvu antonüüm.

Verbide puhul võiks testlause olla järgnev:

Kui keegi/miski teeb tegu X, siis ta ei tee tegu Y.

Kui keegi/miski teeb tegu Y, siis ta ei tee tegu X.

Kui keegi võtab, siis ta ei anna.

Kui keegi annab, siis ta ei võta.

Kui keegi eitab, ütleb ei, vastab eitavalt, siis ta ei jaata.

Kui miski sulgub, läheb kinni, siis ta ei avane ega lähe lahti.

4.4 MERONÜÜMIA

Meronyümia on osa ja terviku suhe. Osa on terviku suhtes meronyüm ning tervik osa suhtes holonyüm. Meronyümia tähised on EWN-s järgmised: *has_meronym* (osa on) ja *has_holonym* (on osa). Tegelikult võiks meronyümiat nimetada terveks suhtegrupiks, kuna tal on palju alaliike (alljärgnevalt on esitatud kõik meronyümi liigid, nende märgendid EWN-s ja eestikeelne tõlge, samuti testlauseid ja näited). Tesaures on meronyümia suhe jagatud viieks:

- 1) liikme-grupi meronyümia (*has_holo_member; has_mero_member, on liige; üks liige on*),
- 2) aine-osa meronyümia (*has_holo_portion; has_mero_portion, on annus; üks annus on*),
- 3) komponendi-terviku meronyümia (*has_holo_part; has_mero_part, on osa; osa on*),
- 4) aine/materjali meronyümia (*has_holo_madeof; has_mero_madeof, on osa materjalist; üks_osa_(materjalist)_on*),

- 5) paiknemist täpsustav meronüümia (*has_holo_location; has_mero_location, on osa kohast; üks osa (kohast) on*). (Vider, Orav 1998).

Testid ning näitelaused iga suhteliigi kohta võiks olla sellised:

- 1) X on liige Y-st.
Y üks liige on X.
Aadlimees on liige aadlist.
Aadli üks liige on aadlimees.
Üliõpilane, tudeng, studioosus on liige korporatsioonist.
Korporatsiooni üks liige on üliõpilane, tudeng, studioosus.
Komsomol, kommunistlik noor on liige komsomolist, Üleliidulisest Leninlikust Kommunistlikust Noorsooühingust.
Komsomoli, Üleliidulise Leninliku Kommunistliku Noorsooühingu üks liige on komsomol, kommunistlik noor.
- 2) X on annus Y-st.
Üks annus X-st on Y.
- 3) X on osa Y-st.
Y osa on X.
Teaduskond on osa ülikoolist.
Ülikooli osa on teaduskond.
Seeria on osa seriaalist, teleseriaalist.
Seriaali, teleseriaali osa on seeria.
Varvas on osa jalast.
Jala osa on varvas.
- 4) X on osa (materjalist) Y.
Y üks osa (materjalist) on X.
Nikotiin on osa sigaretist, suitsust.
Sigareti, suitsu üks osa on nikotiin.
Süsinik on osa söest.

Söe osa on süsinik.

Sulam, metallisulam, metall on osa obadusest, aasast, metall-aasast.

Obaduse, aasa, metall-aasa üks osa on sulam, metallisulam, metall.

5) X on osa kohast Y.

Y üks osa (kohast) on X.

Lava, estraad on osa kohast teater.

Teatri üks osa (kohast) on lava.

Ülikool on osa kohast ülikoolilinn

Ülikoollinna üks osa on ülikool.

Lava, saunalava on osa kohast saun.

Sauna üks osa on lava, saunalava.

Mõneti on teistest meronüümia alaliikidest erinev aine-osa meronüümia, kuna 1) tervik on alati olemas enne osa või ainet, 2) tavaliselt pole ainetele-osadele võimalik anda leksikaalset tähistajat, millele oleks võimalik otseselt viidata, 3) ainetel ja osadele seatud piirid pole tavaliselt motiveeritud, 4) osadel ja ainetel tavaliselt puudub kindel suhe oma tervikutega (Climent jt 2005). Sellest tingitult tekivad ka semantiliste suhete põhjal loodud sõnaseletustes probleemid, mis töö praktilises pooles ka ilmnevad (vt ptk 5).

4.5 OSALUS–ROLLISUHTED

“Keeled on rikkad tuletussuhete poolest, mis genereerivad nimisõnadest tegusõnu ja vastupididi See on tugev sõnadevaheline semantiline suhe. Kõige üldisemalt võib seda nimetada osalussuhteks” (Vider, Orav 1998). Osalus-rollisuhetes olevate sõnade tähendused on tugevalt seotud vastava verbi tähendusega. EWN-s on osalussuhte märgendiks vastavalt *involved* ja *role*, eesti keele tesaurususes on vastavad suhted tõlgitud *kaasneb, mängib rolli*. Osalussuhe jaguneb omakorda viieks alasuhteks: tegija, vahendi, koha, suuna ja tegevusobjekti osaluseks. Osalus-rollisuhete jaoks üldiselt võiks test olla selline:

X-ga kaasneb Y.

X mängib rolli Y-s.

Võitjaga kaasneb võit.

Vaatamises, vahtimises mängib rolli nägema.

1) Tegija osaluse test:

X-ga kaasneb tegija Y.

X-s mängib tegijana rolli Y-s.

Abielluma kaasneb tegija abikaasa, kaasa.

Tsensor mängib tegijana rolli tsenseerima.

Lugemises mängib tegijana rolli lugeja.

2) Vahendi osaluse test:

X-ga kaasneb vahend Y.

X-s mängib vahendina rolli Y-s.

Süstima, pritsima kaasneb vahend süstal.

3) Koha osaluse test:

X-ga kaasneb koht Y.

X-s mängib kohana rolli Y.

Olema, asetsema, asuma, paiknema kaasneb koht asukoht, asupaik, paik, koht.

Riigipiir mängib kohana rolli piiritsoonis.

4) Suuna osaluse kohta on eesti keele tesauruses olemas vaid tegevuse sihtkoha osalus (*involved_target_direction, role_target_direction*). EWN-s on olemas lisaks sellele veel lihtsalt suunaosalus (*involved_direction*) ja lähtekoha suuna osalus (*involved_source_direction, role_source_direction*).

5) Tegevusobjekti osaluse kohta võiks test olla selline:

X-ga kaasneb tegevusobjekt Y.

X-s mängib tegevusobjektina rolli Y-a.

Lugemisega kaasneb tegevusobjekt lugeja.

Lugema mängib tegevusobjektina rolli lugemises, lugemismaterjalis, loetavas.

4.6 PÕHJUSLIKKUSE SUHE

Põhjuslikkuse suhet kasutatakse tavaliselt verbide puhul – ühe verbi tähistatavale seisundile osutatakse verbiga, mis selle seisundi põhjustas, kuid põhjuslikkuse suhe võib olla seotud ka teiste sõnaliikidega. Sageli on aga sellistes suhetes olevad sõnad seotud tugevalt verbidega (paljud *mine*-liitelised substantiivid jmt). Eristatakse kolme liiki põhjuslikkust, mis väljendavad kolme ajalist suhet:

- 1) kahe ajaliselt eristatud olukorra põhjuslikkuse suhe;
- 1) kahe ajaliselt osaliselt kattuva olukorra põhjuslikkuse suhe;
- 1) kahe ajaliselt ühtiva olukorra põhjuslikkuse suhe. (Vider, Orav 1998).

Põhjuslikkuse suhe on mingil määral alati seotud verbidega. Sellistes suhetes olevatest verbidest peab üks olema kindlasti kausatiiv ning see suhe ei pruugi kehtida vastastikku. Näiteks *mittenõustumine* põhjustab *vaidluse* või *vaidlemise*, kuid *vaidlus*, *vaidlemine* ei pruugi tuleneda *mittenõustumisest* (Fellbaum 1993). EWN-s on suhete märgendid *causes* ja *is_caused_by* eesti keeles vastavalt *põhjustab* ja *on põhjustatud*. Testid ja näited:

X põhjustab Y-i.

X on põhjustatud Y-st.

Mittenõustumine põhjustab vaidluse, vaidlemise.

Vaidlus, vaidlemine on põhjustatud mittenõustumisest.

Valmistama, tegema põhjustab looma.

Looma on põhjustatud valmista, tegema.

Taotlema, taotlust esitama põhjustab keelduma.

Keelduma on põhjustatud taotlema, taotlust esitama.

Teatamine, informeerimine, teada andmine põhjustab teadma, teadlik olema.

Teadma, teadlik olema on põhjustatud teatama, informeerima, teada andma.

4.7 OSASÜNDMUSE SUHE

“Mitmed verbiga kirjeldatavad situatsioonid jagunevad omakorda väiksemateks situatsioonideks ... , ajaliselt kaasnevateks situatsioonideks ...” (Vider jt 2000). Selle suhte eripäraks on veel see, et suhe ei pruugi kehtida mõlemas suunas. EWN-s tähistatakse seda suhet järgnevalt *has_subevent* ja *is_subevent_of*, eesti keeles vastavalt *osasündmus on* ja *on osasündmus*.

Testid ja näited:

X-i osasündmus on Y.

X on osasündmus Y-st.

Hüvasti jätma, jumalaga jätma osasündmus on kättelema, kätt andma, kätt suruma.

Õppima, koolis käima, koolitust saama osasündmus on õppima, tudeerima.

Valima, selekteerima on osasündmus leidma, välja valima.

Liikumapanek on osasündmus liigutusest, liigutamisest.

Kastmine on osasündmus leivategemisest, leivateost.

Õppima, tudeerima on osasündmus õppima, koolis käima, koolitust saama.

4.8 SEISUNDISUHTED

See suhe on EWN-s spetsiaalselt loodud selleks, et tähistada seoseid noomenite vahel, mis viitavad ükskõik millisele erilisele seisundile, mida väljendab adjektiiv. Tavaliselt on sellistel noomenitel avatud denotatsioonid s.t sellised noomenid võivad viidata ükskõik millisele objektile, mille kohta seisund kehtib (Climent jt 2005). EWN-s tähistatakse selliseid suhteid järgnevalt: *be_in_state* ja *state_of*, eesti keeles vastavalt *on seisundis* ja *on seisundiks*. Testid ja näited:

X on seisundis Y.

X on seisundiks Y-le.

Sitkus, tugevus on seisundis nõrk, jõuetu.

Põhiarv, kardinaalarv on seisundis naturaalarv.

Laupäev on seisundis puhkepäev.

Soolane on seisundiks maitseomadusele, maitsele, mekile, maigule.

Soe on seisundiks temperatuurile.

Värvitu on seisundiks värvile, värvusele.

Vedel on seisundiks konsistentsile.

4.9 KLASSI-ESINDAJA SUHTED NING MÄÄRATLEMATUD SUHTED

Klassi ja esindaja suhted on sarnased hüperonüümia ja hüponüümia suhetele. Viimased aga kehtivad üldisemalt, kuid ka üksikud mõisted kuuluvad mingisse rühma ja selleks on loodud klassi ja esindaja suhted. Nende tähistuseks EWN-s on *belongs_to_class* ja *has_instance*, eesti keeles vastavalt *kuulub klassi* ja *esindaja on*. Tuleb tähele panna, et esitatud testis X peab olema kindlasti pärisnimi. Testid ja näited:

X on üks Y-test.

X kuulub klassi Y.

Yale on üks ülikoolidest (esindab ülemmõistet ülikool).

Times kuulub klassi perioodiline väljaanne.

Yale on üks ülikoolidest.

Kadriorg on üks linnaosa, linnajagu.

Dublin on üks ülikoolilinnadest.

Viimaks on loodud veel seos olukordadeks, kui kõik eelnevad testid ei tööta, aga siiski tundub sellises seoses sõnade vahel olevat mingi tugev semantiline seos. Selle jaoks aga pole veel eraldi suhet loodud. Ka see suhe – samuti kui paljud eelnevad – võib esineda nii samasse sõnaliiki kuuluvate sõnade vahel kui ka erinevate sõnaliigi esindajate vahel. EWN-s on suhte märgenditeks *fuzzynym* ja *xpos_fuzzynym*, eesti keeles on mõlemad ühte moodi –

on hägusalt seotud (näiteks: *vili fuzzynym (on hägusalt seotud)* söödav vili). On näha, et nende sõnade vahel on olemas tugev semantiline seos, kuid pole määratletav ühegagi eelnevatest.

4.10 XPOS–SUHETEST

Seletuste seisukohalt on problemaatilised *xpos*-suhted. EWN-s on väga hästi näha, kus ja miks neid kasutatakse, nimelt siis kui tegemist on suhtega erinevate sõnaliikide vahel. Eesti keeles aga on need suhted tõlgitud kõik samamoodi kui teisedki suhted, tõlkes ei ole näiteks vahet tehtud suhetel *xpos_near_synonym* või *near_synonym*, mõlema jaoks on sama vaste – *peaaegu samatähenduslik on*. Eriti tulevad sellised probleemid esile värvinimetuste puhul, kui sõnade vahel on suhe *kollasus/kollane xpos_near_synonym (peaaegu samatähenduslik on) kollane*. Eesti keeles tuleb selline erinevus välja alles vaadeldavast materjalist. Antud juhul on mõeldud suhet noomeni ja adjektiivivi vahel, kuid suhte nimi ise seda ei väljenda. Tesaaruses on siiski vastavad sõnaliigi märgendid olemas. Kui aga seletada sõna viisil *kollane on peaaegu samatähenduslik sõnale kollane*, siis ilmselt ei avaks selline seletus mõiste tähendust. Siin on mõeldud erinevaid sõnaliike, mis loomulikult ei saaks olla teineteise täistähenduslikud sünonüümid. Tuleks leida viis, kuidas *xpos*-suhted oleksid arusaadavad ka ilma taustmaterjalita. On raske leida sellistele suhetele head eestikeelset vastet, ilma et suhte nimi ei läheks liiga pikaks. Kui lisada igale poole juurde, et suhe kehtib erinevate sõnaliikide vahel, siis teeks see info kasutamise automaatselt palju keerulisemaks ning selline lisamärkus teeks ka seletused kohmakamaks.

5. SÕNASELETUSTE GENEREERIMINE TESAURUSE INFO PÕHJAL

Järgnevalt on antud ülevaade töö praktilisest poolest. Vaadeldakse, kuidas tesauruse info põhjal on võimalik genereerida sõnaseletusi. Kas semantilisi suhteid kasutades loodud sõnaseletused on loomulikud ning kas need on loomulikud kõigile sünohulga liikmetele rakendades? Töö käik on järgmine.

1. Ette on antud sünohulkadega tekstifail (vt lisa 4. diskett) tesauruse infoga, kus sõnad on seotud kõigi oma sihtmärksõnadega tesauruses. Esialgses tekstifailis on sõnade vahel ingliskeelsed suhete nimed (sünohulka ja selles sisalduvaid sümboleid on seletatud esimeses lisas, vt lisa 1.).

6804@ n#algus(1)#kujunemine(1)@has_hyperonym:502@n#protsess(1)

6804@ n#algus(1)#kujunemine(1)@has_hyponym:8096@n#alge(1)#suge(1)#alged(1)

6804@ n#algus(1)#kujunemine(1)@xpos_near_synonym:622@v#tekkima(1)

1061@ v#varjama(1)#katma(5)@has_hyponym:5131@v#mähkima(1)#mässima(1)

1061@ v#varjama(1)#katma(5)@near_synonym:27@v#katma(2)

*1061@ v#varjama(1)#katma(5)@near_synonym:4980@v#summutama(3)#katma(4)#matma(2)
#neelama(6)*

8862@ a#vana(1)@antonym:8833@a#uus(1)

8862@ a#vana(1)@has_hyponym:8885@a#antiikne(1)#vanaaegne(1)

8862@ a#vana(1)@near_synonym:8884@a#igivana(1)#iidne(1)#vana(3)

2. Ingliskeelsed suhete nimed on asendatud eestikeelsete vastetega (vt lisa 4. diskett) ning vaadeldud, kas sellistest järjenditest on võimalik moodustada normaalseid eestikeelseid lauseid, mis sobiksid sõnaseletusteks.

6804@ n#algus(1)#kujunemine(1)@on_teatud_liiki:502@n#protsess(1)
6804@ n#algus(1)#kujunemine(1)@üks_liik_on:8096@n#alge(1)#suge(1)#alged(1)
6804@ n#algus(1)#kujunemine(1)@peaaegu_samatähenduslik_on:622@v#tekkima(1)

1061@ v#varjama(1)#katma(5)@üks_viis_on:5131@v#mähkima(1)#mässima(1)
1061@ v#varjama(1)#katma(5)@peaaegu_samatähenduslik_on:27@v#katma(2)
1061@v#varjama(1)#katma(5)@peaaegu_samatähenduslik_on:4980@v#summutama(3)
#katma(4)#matma(2)#neelama(6)

8862@ a#vana(1)@vastand_on:8833@a#uus(1)
8862@ a#vana(1)@üks_liik_on:8885@a#antiikne(1)#vanaaegne(1)
8862@ a#vana(1)@peaaegu_samatähenduslik_on:8884@a#igivana(1)#iidne(1)#vana(3)

3. Seejärel on eestikeelsete suhete nimedega materjal läbi vaadatud ning sealt välja korjatud read, mis hästi seletusteks ei sobi.
4. Kõige eelneva põhjal oleks vaja välja töötada programm, mis genereeriks seletusi automaatselt. Niimoodi loodud seletused võiks välja näha taolised:

- *Algus, kujunemine on teatud liiki protsess. Alguse, kujunemise liigid on suge, alge. Algus, kujunemine peaaegu samatähenduslik on tekkima.*
- *Varjama, katma üks viis on mässima. Varjama, katma peaaegu samatähenduslik on katma. Varjama, katma peaaegu samatähenduslikud on summutama, katma, matma, neelama.*
- *Vana vastand on uus. Vana liigid on antiikne, vanaaegne. Vana peaaegu samatähenduslikud on iidne, vana.*

5.1 SELETUSTES TEKKINUD PROBLEEMID

Üks suuremaid probleeme oli see, kui suhe ei kehtinud sünohulga kõigi liikmete kohta või kui kehtis, siis sellisel viisil esitatud seletus ei oleks kuigi informatiivne ega ava sõna tähendust. Näiteks:

1061@ v#varjama(1)#katma(5)@üks_viis_on:5131@v#mähkima(1)#mässima(1)
1061@ v#varjama(1)#katma(5)@peaaegu_samatähenduslik_on:27@v#katma(2)
1061@ v#varjama(1)#katma(5)@peaaegu_samatähenduslik_on:
4980@v#summutama(3)#katma(4)#matma(2)#neelama(6)

Sellises reas on lähisünonüümia sõnade *katma(4)* ja *katma(5)* vahel, kuigi tegemist on sama sõna erinevate tähendustega, ei oleks selline järjend seletuse seisukohalt informatiivne.

Tuleks leida viis, kuidas näiteks vastava programmi abil sellised sõnad seletustest eraldada, täpsemalt, välistada see, et sõna seletuseks kasutatakse sama sõna (kuigi erinevas tähenduses), mida tahetakse seletada.

Samasugune olukord tekib lähisünonüümia puhul. Nii nagu eelmises näites oleks *katma(5)* ja *katma(2)* ning *katma(5)* ja *katma(4)* vahel. Nii on ka lähisünonüümia substantiivi ja adjektiivi vahel. Suhe on tesauruse seisukohalt õige, kuna tegemist on erinevate sõnaliikidega, seega ei saakski tegemist olla täieliku sünonüümia, kuigi seletuseks ei sobiks ka selline variant. Seletuse seisukohalt ei annaks see edasi olulist infot ega seletaks sõna tähendust. Näiteks:

2407@ n#punasus(1)#puna(1)#punane(1)@on_teatud_liiki:2410@n#spektrivärv(1)
2407@n#punasus(1)#puna(1)#punane(1)@peaaegu_samatähenduslik_on:9034@a#punane(1)

Näites on tegemist isegi sama tähendusnumbriga sõnadega, kuigi sõnaliigid on erinevad. Taolised probleemid tekivadki enamasti adjektiividega ning substantiividega, mis on täpselt samakujulised.

Samuti tekib probleem homonüümia puhul, kui sõnadel on küll erinev tähendus, kuid kirja pilt on samasugune. Selliselt esitatud sõnad ei edastaks mingit tähenduslikku informatsiooni. Näiteks:

5843@ v#rentima(2)#rendile võtma(1)#üürima(1)@vastand_on:5842@v#rentima(1)#üürile andma(1)#välja rentima(1)#üürima(2)
5843@ v#rentima(2)#rendile võtma(1)#üürima(1)@on_teatud_viisil:738@v#laenama(1)

Tegemist on erinevate sõna tähendustega, *rentima(2)* ja *rentima(1)* on kindlasti antonüümid, kuid kui mitte teada kummagi sõna tähendust, ei aitaks ka selline seletus.

Semantilistest suhetest tekitabid veel probleeme seisundisuhted. Seletuse seisukohalt on need üsna väheütlevad. Et sellistest seletustest aru saada, peaksid eelnevalt olemas olema teadmised tesaurusest ja selles sisalduvatest semantilistest suhetest. Näiteks:

Tugevus, sitkus on seisundis nõrk, jõuetu.
Põhiarv, naturaalarv on seisundis kardinaalarv.
Laupäev on seisundis puhkepäev jne.

Sellised seletused sõnaraamatus oleksid kummalised ning mitte eriti arusaadavad inimesele, kes tesaurust ei tunne.

Kui vastava mõiste jaoks on tesauruses liiga vähe suhteid, siis ei ole võimaik luua pikka ja põhjalikku seletust, mis avaks vastava sõna sisu. Järgnevad näited on mõned sellistest, mille kohta on tesauruses vaid üks suhe. Sellistest suhetest jääb mõistetava seletuse jaoks natuke väheks. Taolisi üherealisi hulka on kõige rohkem adjektiivide seas, kuna adjektiive on tesauruses üldiselt kõige vähem, siis on see ka loomulik. Esitatud adjektiividel võiks olla näiteks viiteid mingi objekti kohta, mida nad iseloomustavad.

8904@ a#poliitiline(1)@peaaegu_samatähenduslik_on:8905@a#riiklik(1)
9145@ a#sulgpehme(1)@peaaegu_samatähenduslik_on:9139@a#pehme(1)
779@ v#üles leidma(1)@on_osasündmus:631@v#leidma(1)#avastama(1)

Selline probleem on lahendatav tesauruse täiendamisega – sinna uute mõistete lisamisega. Nii tekib rohkem semantilisi seoseid ja on võimalik luua põhjalikumaid seletusi. Seletuse selgus ja põhjalikkus oleneb mingil määral ka sellest, millised on suhted. Nii nagu näideteski, on sageli seletuste seisukohalt kõige vähem informatiivne lähisünonüümia (oleneb ka selliste sünonüümide hulgast ja jällegi konkreetsest mõistest).

Nii nagu keeles üldse, on ka seletuste loomise seisukohalt problemaatilisemad abstraktsed ja väga üldisi mõisteid tähistavad sõnad. Nende jaoks seletuste genereerimiseks on vaja rohkem seoseid, neid on vaja seletada põhjalikumalt, samal ajal kui väga konkreetsete sõnade jaoks piisaks vaid paarist semantilisest suhtest ning üsna lühikesest ja konkreetsest seletusest. Järgnevatest näidetest on näha, et selliste suhete põhjal oleks võimalik moodustada täiesti normaalne ja informatiivne seletus. Näiteks:

4522@ n#ratsionaalarv(1)@vastand_on:4524@n#irratsionaalarv(1)
4522@ n#ratsionaalarv(1)@on_teatud_liiki:4525@n#reaalarv(1)
4522@ n#ratsionaalarv(1)@üks_liik_on:4521@n#täisarv(1)

1438@ n#foto(1)#ülesvõte(1)#päevapilt(1)@on_liige:9639@n#fotoalbum(1)
1438@ n#foto(1)#ülesvõte(1)#päevapilt(1)@on_teatud_liiki:231@n#pilt(1)
1438@ n#foto(1)#ülesvõte(1)#päevapilt(1)@üks_liik_on:10326@n#portree(2)#portreefoto(1)
1438@ n#foto(1)#ülesvõte(1)#päevapilt(1)@üks_liik_on:10518@n#aktifoto(1)

Seletused esitatud kahe sõna kohta võiks olla sellised:

- *Ratsionaalarvu vastand on irratsionaalarv. Ratsionaalarv on teatud liiki reaalarv. Ratsionaalarvu üks liik on täisarv.*
- *Foto, ülesvõte, päevapilt on liige fotoalbumist. Foto, ülesvõtte, päevapilt on teatud liiki pilt. Foto, ülesvõtte, päevapildi üks liik on portree, portreefoto. Foto, ülesvõtte, päevapildi üks liik on aktifoto.*

Taolised read on seletuste seisukohalt juba üsna aktsepteeritavad.

Seletuste genereerimisel tekitab probleeme **autohüponüümia**. Autohüponüümia tähendab seda, et ühe ja sama sõna erinevad tähendused on teineteise suhtes hüponüümia või hüperonüümia suhetes. Tavaliselt väljendab sellises suhtes olev hüponüüm mingit erilist tähendusnüanssi, mida tema ülemmõistel pole. (Peters, Peters 1998).

5.1.1 Morfoloogia

Tesauruse info põhjal sõna seletusi genereerides ei piisa ainult semantilisest informatsioonist. Arvestada tuleb ka morfoloogilisi seoseid loodud seletuses. Tesauruses on kõik substantiivid ainsuse nominatiivis ja verbid *ma*-infinitiivis ning mõne semantilise seose puhul polegi vaja sõnu muuta, näiteks (*on_seisundis*, *kuulub_klassi*, *on_osa_kohast*, *on_teatud_viisil/liiki*). Sellistes suhetes olevatest sõnadest on automaatselt üsna lihtne moodustada seletusi. Kahjuks aga on sellised suhted TEKsauruses selges vähemuses. Et genereerida kõigist järjenditest normaalseid eestikeelseid lauseid, oleks vaja kasutada abivahendeid, mis suudaksid automaatselt töödelda ja analüüsida ka morfoloogilist informatsiooni. Taolisi abivahendeid võiks vaja minna näiteks sellistes olukordades:

```
6826@ n#värav(1)@on_teatud_liiki:1662@n#takistus(1)#tõkestus(1)#tõke(1)
6826@ n#värav(1)@osa_on:6399@n#väravapost(1)
```

Et moodustada teisest reast normaalne eestikeelne lause, oleks vaja sõna *värav* panna genitiivi (*Värava osa on väravapost.*). Morfoloogilised seosed on esitatud kolmandas lisa (vt lisa 3.).

Kokkuvõtteks võiks öelda, tesaurus ja seal sisalduv info on sõnaseletuste genereerimiseks suhteliselt hea alus. Tekivad üksikud probleemid, mille lahendamine ei ole eriti keeruline.

5.2 MILLINE VÕIKS OLLA PROGRAMM?

Milline võiks välja näha programm, mis genereeriks sõnaseletusi tesauruse info põhjal automaatselt? Seda võiks teha näiteks järjendis, kus iga sünohulk on üks järjendi liige.

1. Programmi tuleks sisse lugeda sünohulga esimene liige ning selle tähenduse number.
2. Leida semantilised suhted ning nende ingliskeelsed nimed asendada eestikeelsetega.
3. Leida suhte sihtmärksõna(d) ning nendega moodustada lause.
4. Sõnad tuleks lauses esitada õigetes käänetes ja pööretes (kasutada abivahendeid morfoloogiliseks töötluks).
5. Eelnevalt mainitult peaks programm eraldama juhud, kus sõna seletatakse sama sõna kaudu.

Millised seosed, milliseid morfoloogilisi seoseid nõuavad, on näha kolmandas lisas (vt lisa 3.).

KOKKUVÕTE

Töö eesmärgiks oli anda ülevaade sõnaseletusest sõnaraamaamatus ning vaadelda, kuidas on tesauruse info põhjal võimalik genereerida sõnaseletusi. Töö praktiline pool võiks olla eeltöök vastava programmi välja töötamisel.

Töö 1. ja 2. peatüki antaksegi EKSS-le tuginedes ülevaade sõnaseletusest sõnaraamatus, selles leiduvast morfoloogilisest, süntaktilisest ja muust infost. 3. peatükis on lühike ülevaade WordNetist ja EuroWordNetist ning eesti keele teasaurusest ehk TEKsaurusest. 4. peatükk on ülevaade semantilistest suhetest eriti just tesauruse kontekstis.

5. peatükk on töö praktilise käigu kirjeldus. Vaadeldes, kuidas tesauruse info põhjal on võimalik luua sõnaseletusi. Esitatud on probleemid, mis tekivad, ning pakutud ka mõned lahendused.

Suurimaks probleemiks on see, et tesaurus pole piisavalt mahukas, et genereerida kõigist sõnadest normaalseid seletusi, probleem on lahendatav tesaurusesse uute sõnade lisamisega. Nii tekib sõnade vahel uusi seoseid ning neid saab kasutada seletuse täpsustamiseks. Samuti tekivad probleemid seletustes, kus semantilised suhted esinevad kahe samasuguse, kuid tähenduselt erinevate sõnade vahel. Need raskendavad seletusest arusaamist ning tuleks seletusest eemaldada.

Üldiselt võib öelda, et tesaurus ja selles sisalduv info on sõnaseletuste genereerimiseks suhteliselt hea alus. Tekkivad probleemid on hõlpsasti lahendatavad.

KIRJANDUS

- Alonge, Antonietta* 1996. Definition of the links and subsets for verbs. Deliverable D006, WP4.1, EuroWordNet, LE2-4003.
- Climent, Salvador; Rodríguez, Horacio; Gonzalo, Julio* 2005. Definition of the links and subsets for nouns. Deliverable D005, WP3.1, EuroWordNet, LE2-4003.
- EKSS* = Eesti kirjakeele seletussõnaraamat, I – VI köide. 1988 – 2004. Tallinn: Eesti Keele Instituut.
- Fellbaum, Christiane* 1993. English Verbs as a Semantic Net; <ftp://ftp.cogsci.princeton.edu/pub/wordnet/5papers.ps>
- Karlsson, Fred* 2002. Üldkeeleteadus. Tallinn: Eesti Keele Instituut.
- Langemets, Margit* 2000. Sõnaraamatu arvutilingvistiline analüüs: väitekiri magistrikaadi taotlemiseks. Tallinn: Eesti Keele Instituut.
- Muischnek, Kadri; Orav, Heili; Kaalep, Heiki-Jaan; Õim, Haldur* 2003. Eesti keele tehnoloogilised ressursid ja vahendid. Arvutikorpused, arvutisõnastikud, keeletehnoloogiline tarkvara. Tallinn: Eesti Keele Sihtasutus.
- Peters, Wim; Peters, Ivonne; Vossen, Piek* 1998. Automatic sense clustering in EuroWordnet; http://www.dcs.shef.ac.uk/~wim/lrec_III.pdf.
- TEKsaurus*; <http://test.cl.ut.ee/ressursid/teksaurus/>.
- TEKsaurus*; <http://test.cl.ut.ee/ressursid/teksaurus/teksaurus.cgi.et>.
- Vider, Kadri, Orav, Heili* 1998. Sõna tasandilt mõiste ruumi. – Keel ja Kirjandus nr 1, lk 57-64.
- Vider, Kadri; Kahusk, Neeme; Orav, Heili; Õim, Haldur; Paldre, Leho* 2000. Eesti keele tesaarus. Arvutuslingvistikalt inimesele. Tartu Ülikooli üldkeeleteaduse õppetooli toimetised 1. Tartu: Tartu Ülikooli Kirjastus, lk 127–152.

Õim, Asta 1998. Antonüümisõnastik. Eesti Keele Sihtasutus.

SUMMARY

Generating Word Explanations Based on the Information in Thesaurus.

This B.A. thesis gives an overview of word explanations in a dictionary. There is an overview of that, which kind of morfological, syntactic and semantic information these explanations give. The theoretical part of the work is primarily based on Margit Langemets's master's thesis and an Explanatory Dictionary of Estonian Written Language. Also has been considered articles about WordNet, EuroWordNet and Estonian thesaurus.

Third chapter gives a brief overview of the WordNet, EuroWordNet and Estonian thesaurus (called TEKsaurus, Estonian WordNet). Fourth part of the work is a description of semantic relations, especially has been cosidered these relations in the thesaurus.

A practical part of the work could be a preparation for the programme that generates word explantions automatically on the mentioned way. In the practical part of the work has been scrutinised, if and how is possible to generate word explanations based on the information in thesaurus. There is a brief overview, which kind of problems occur and how is possible to improve this kind of work. There are also some suggestions, what to do to make this kind of work more efficient. The practical part of the work is based on the text file that contains information of thesaurus.

As a result it seems that thesaurus and information in there is a very good way to make word explanations.

LISA 1.

6804@ n#algus(1)#kujunemine(1)@has_hyperonym:502@n#protsess(1)
6804@ n#algus(1)#kujunemine(1)@has_hyponym:8096@n#alge(1)#suge(1)#alged(1)
6804@ n#algus(1)#kujunemine(1)@xpos_near_synonym:622@v#tekkima(1)

1061@ v#varjama(1)#katma(5)@has_hyponym:5131@v#mähkima(1)#mässima(1)
1061@ v#varjama(1)#katma(5)@near_synonym:27@v#katma(2)
1061@ v#varjama(1)#katma(5)@near_synonym:4980@ v#summutama(3)#katma(4)#matma(2)
#neelama(6)

8862@ a#vana(1)@antonym:8833@a#uus(1)
8862@ a#vana(1)@has_hyponym:8885@a#antiikne(1)#vanaaegne(1)
8862@ a#vana(1)@near_synonym:8884@a#igivana(1)#iidne(1)#vana(3)

6804, 1061, 8862 jne – sünohulkade id-numbrid.

n, v, a – sõnaliigi märgendid vastavalt noomen, verb, adjektiiv.

Järgnevad sünohulga liikmed, iga sõna järel sulgudes olev number on vastava liikme tähenduse number.

has_hyperonym, *xpos_near_synonym*, *has_hyponym*, *near_synonym*, *antonym*,
has_hyponym – semantiliste suhete märgendid.

6804@ n#algus(1)#kujunemine(1)@has_hyperonym:502@n#protsess(1)
6804@ n#algus(1)#kujunemine(1)@has_hyponym:8096@n#alge(1)#suge(1)#alged(1)
6804@ n#algus(1)#kujunemine(1)@xpos_near_synonym:622@v#tekkima(1)

1061@ v#varjama(1)#katma(5)@has_hyponym:5131@v#mähkima(1)#mässima(1)
1061@ v#varjama(1)#katma(5)@near_synonym:27@v#katma(2)

1061@ v#varjama(1)#katma(5)@near_synonym:4980@ v#summutama(3)#katma(4)#matma(2)
#neelama(6)

8862@ a#vana(1)@antonym:8833@a#uus(1)

8862@ a#vana(1)@has_hyponym:8885@a#antiikne(1)#vanaaegne(1)

8862@ a#vana(1)@near_synonym:8884@a#igivana(1)#iidne(1)#vana(3)

6804, 1061, 8862 jne – sünohulkade id-numbrid.

n, v, a – sõnaliigi märgendid vastavalt noomen, verb, adjektiiv.

Järgnevad sünohulga liikmed, iga sõna järel sulgudes olev number on vastava liikme tähenduse number.

has_hyperonym, xpos_near_synonym, has_hyponym, near_synonym, antonym, has_hyponym – suhete märgendid.

LISA 2.

Suhte märgend EWN-s	Eestikeelne vaste	Näide
antonym	vastand on	lubama <i>vastand on</i> keelama
be_in_state	on seisundis	värv, värvus <i>on seisundis</i> värviline
belongs_to_class	kuulub klassi	Aleksander <i>kuulub klassi</i> mees
causes	põhjustab	lubama <i>põhjustab</i> luba
fuzzynym	on hägusalt seotud	kord, puhk <i>on hägusalt seotud</i> moment
has_holo_location	on osa kohast	ülikool <i>on osa kohast</i> ülikoolilinn
has_holo_madeof	on osa materjalist	puu <i>on osa materjalist</i> puu
has_holo_member	on liige	liige <i>on liige</i> kollektiiv
has_holo_part	on osa	koht <i>on osa</i> ruum
has_holo_portion	on annus	mõte <i>on annus</i> mõttetegevus
has_holonym	on osa	ühik <i>on osa</i> hulk
has_hyperonym	on teatud viisil [v]; on teatud liiki [n]	lubama <i>on teatud viisil</i> soostuma; volitus <i>on teatud liiki</i> luba
has_hyponym	üks viis on [v]; üks liik on [n]	soostuma <i>üks viis on</i> lubama; luba <i>üks liik on</i> volitus
has_instance	esindaja on	mees <i>esindaja on</i> Aleksander
has_mero_location	üks osa (kohast) on	ülikoolilinn <i>üks osa (kohast) on</i> ülikool
has_mero_madeof	üks osa (materjalist) on	puu <i>üks osa (materjalist) on</i> puit
has_mero_member	üks liige on	kollektiiv <i>üks liige on</i> liige

Suhte märgend EWN-s	Eestikeelne vaste	Näide
has_mero_part	osa on	ruum <i>osa on</i> koht
has_mero_portion	üks annus on	mõttetegevus <i>üks annus on</i> mõte
has_meronym	osa on	hulk <i>osa on</i> ühik
has_subevent	osasündmus on	otsustama <i>osasündmus on</i> arvama
has_xpos_hyperonym	on teatud viisil [v]; on teatud liiki [n]	taotlema <i>on teatud viisil</i> suhtlus
has_xpos_hyponym	üks viis on [v]; üks liik on [n]	suhtlus <i>üks liik on</i> mõjutama
involved	kaasneb	püsima <i>kaasneb</i> seisund; teavet andma <i>kaasneb</i> informatsioon
involved_agent	kaasneb tegija	kõnelema <i>kaasneb tegija</i> kõneleja
involved_instrument	kaasneb vahend	käskima <i>kaasneb vahend</i> mõjujõud
involved_location	kaasneb koht	asuma <i>kaasneb koht</i> koht, paik
involved_patient	kaasneb tegevusobjekt	rääkima <i>kaasneb tegevusobjekt</i> kuulaja
involved_target_direction	kaasneb tegevuse sihtkoht	minema <i>kaasneb tegevuse sihtkoht</i> koht
is_caused_by	on põhjustatud	luba <i>on põhjustatud</i> lubama
is_subevent_of	on osasündmus	arvama <i>on osasündmus</i> otsustama
near_antonym	peaaegu vastand on	saabuma <i>peaaegu vastand on</i> minema
near_synonym	peaaegu samatähenduslik on	katma <i>peaaegu samatähenduslik on</i> varjama
role	mängib rolli	teadmine <i>mängib rolli</i> teadma
role_agent	mängib tegijana rolli	kõneleja <i>mängib tegijana rolli</i> kõnelema

Suhte märgend EWN-s	Eestikeelne vaste	Näide
role_instrument	mängib vahendina rolli	meelitus <i>mängib vahendina rolli</i> ahvatlema
role_location	mängib kohana rolli	koht, paik <i>mängib kohana rolli</i> asuma
role_patient	mängib tegevusobjektina rolli	arv <i>mängib tegevusobjektina rolli</i> korrutama
role_target_direction	mängib tegevuse sihtkohana rolli	koht <i>mängib tegevuse sihtkohana rolli</i> minema
state_of	on seisundiks	värviline <i>on seisundiks</i> värv, värvus
xpos_fuzzynym	on hägusalt seotud	õis <i>on hägusalt seotud</i> õitsema
xpos_near_antonym	peaaegu vastand on	küsimus <i>peaaegu vastand on</i> vastama
xpos_near_synonym	peaaegu samatähenduslik on	liikuma <i>peaaegu samatähenduslik on</i> kulgemine

(TEKsaurus)

LISA 3.

Selles lisas on esitatud käänded ja pöörded, mida vastavates suhetes olevatest sõnadest seletusi genereerides kasutama peaks. Tabelist on näha ka, milliste sõnaliikide vahel mingi suhe olla saab (eraldi sõnaliiki pole märgitud, kuid näha on, kas esitatud on ainult kääne, pööre või mõlemad).

Kääne/pöör e	Semantiline suhe	Kääne/pööre
nominatiiv; <i>ma</i> -infinitiiv	vastand on	nominatiiv; <i>ma</i> -infinitiiv
nominatiiv;	on seisundis	nominatiiv;
nominatiiv	kuulub klassi	nominatiiv
<i>ma</i> -infinitiiv	põhjustab	genitiiv
nominatiiv; <i>ma</i> -infinitiiv	on hägusalt seotud	komitatiiv; <i>ma</i> -infinitiiv
nominatiiv	on osa kohast	nominatiiv
nominatiiv	on osa materjalist	nominatiiv
nominatiiv	on liige	elatiiv
nominatiiv	on osa	elatiiv
nominatiiv	on annus	elatiiv, partitiiv
nominatiiv	on osa	elatiiv
nominatiiv; <i>ma</i> -infinitiiv	on teatud liiki [n]; on teatud viisil [v]	nominatiiv; <i>ma</i> -infinitiiv
genitiiv; <i>ma</i> -infinitiiv	üks liik on [n]; üks viis on [v]	nominatiiv; <i>ma</i> -infinitiiv

Kääne/pöör e	Semantiline suhe	Kääne/pööre
genitiiv	esindaja on	nominatiiv
genitiiv	üks osa (kohast) on	nominatiiv
genitiiv	üks osa (materjalist) on	nominatiiv
genitiiv	üks liige on	nominatiiv
genitiiv	osa on	nominatiiv
genitiiv	üks annus on	nominatiiv
genitiiv	osa on	nominatiiv
genitiiv; <i>ma</i> -infinitiiv	osasündmus on	nominatiiv; <i>ma</i> -infinitiiv
nominatiiv; <i>ma</i> -infinitiiv	on teatud liiki [n]; on teatud viisil [v]	nominatiiv; <i>ma</i> -infinitiiv
genitiiv; <i>ma</i> -infinitiiv	üks liik on [n]; üks viis on [v]	nominatiiv; <i>ma</i> -infinitiiv
<i>ma</i> -infinitiiv	kaasneb	nominatiiv
<i>ma</i> -infinitiiv	kaasneb tegija	nominatiiv
<i>ma</i> -infinitiiv	kaasneb vahend	nominatiiv
<i>ma</i> -infinitiiv	kaasneb koht	nominatiiv
<i>ma</i> -infinitiiv	kaasneb tegevusobjekt	nominatiiv
<i>ma</i> -infinitiiv	kaasneb tegevuse sihtkoht	nominatiiv
nominatiiv; <i>ma</i> -infinitiiv	on põhjustatud	elatiiv; <i>ma</i> -infinitiiv
nominatiiv; <i>ma</i> -infinitiiv	on osasündmus	elatiiv; <i>ma</i> -infinitiiv
nominatiiv; <i>ma</i> -infinitiiv	peaaegu vastand on	nominatiiv; <i>ma</i> -infinitiiv
nominatiiv; <i>ma</i> -infinitiiv	peaaegu samatähenduslik on	nominatiiv; <i>ma</i> -infinitiiv
nominatiiv; <i>ma</i> -infinitiiv	mängib rolli	inessiiv; <i>ma</i> -infinitiiv

Kääne/pöör e	Semantiline suhe	Kääne/pööre
nominatiiv	mängib tegijana rolli	inessiiv; <i>ma</i> -infinitiiv
nominatiiv	mängib vahendina rolli	inessiiv; <i>ma</i> -infinitiiv
nominatiiv	mängib kohana rolli	inessiiv; <i>ma</i> -infinitiiv
nominatiiv	mängib tegevusobjektina rolli	inessiiv; <i>ma</i> -infinitiiv
nominatiiv	mängib tegevuse sihtkohana rolli	inessiiv; <i>ma</i> -infinitiiv
nominatiiv	on seisundiks	allatiiv
nominatiiv; <i>ma</i> -infinitiiv	on hägusalt seotud	komitatiiv; <i>ma</i> -infinitiiv
nominatiiv; <i>ma</i> -infinitiiv	peaaegu vastand on	nominatiiv; <i>ma</i> -infinitiiv
nominatiiv; <i>ma</i> -infinitiiv	peaaegu samatähenduslik on	nominatiiv; <i>ma</i> -infinitiiv