

# EESTI VERBNET'I LOOMISE VÕIMALIKKUSEST

Indrek Jentson

**Ülevaade.** Artiklis käsitletakse eesti keele jaoks verbide leksikaalse andmebaasi loomise ideed, lähtudes inglise ja juba ka teiste keelte jaoks loodud (Liu, Chiang 2008) samalaadse andmekogu VerbNet koostamise põhimõtetest ja senistest kogemustest. Vaatluse alla võetakse verbide klassifitseerimine nii nende tähenduse kui ka predikaat-argument-struktuuri järgi ning analüüsitakse võimalusi täiendada VerbNet'i andmestikku eesti keelele omaste tunnustega. Kirjeldatakse leksikaalsete kirjade loomist ühe mitmetähendusliku tegusõna näitel ja esitatakse eksperimendi tulemused sõnatähenduse ühestamise kohta sellise leksikaalsete kirjade abil teostatud semantilise analüüsi käigus.\*

**Võtmesõnad:** keeleressursid, arvutilingvistika, arvutileksikoloogia, semantiline märgendus, eesti keel

## 1. Sissejuhatus

Seoses arvutustehnika jõudsa arenguga viimasel paaril aastakümnel on hakatud pöörama suuremat tähelepanu selliste keeleressursside loomisele, mis oleksid arvutite poolt kasutatavad mitmesuguste loomuliku keele töötlemise probleemide lahendamisel. Olulisteks tugisammasteks on mitmesugused korpused ja mitmed spetsiifilise eesmärgiga loodud teadmusbasisid nagu WordNet (Miller jt 1990, Miller 1995), FrameNet (Baker jt 1998), VerbNet (Kipper Schuler jt 2000) ja PropBank (Kingsbury, Palmer 2002). Nende andmekogude üks eesmärk on pakkuda lähteandmeid loomulikus keeles kirjapandud tekstide semantilise analüüsi võimaldamiseks. Kahjuks peab tõdema, et nimetatud ressursid on inglise keele kesksed ning kui eesti keele kontekstis on samas suunas liigutud ainult WordNet'i loomisel (Kerner jt 2010) ja mingis osas ka FrameNet'i andmekirjete defineerimisel (Õim jt 2010), siis VerbNet'i sarnast andmekogu eesti keele jaoks veel ei ole.

VerbNet'i<sup>1</sup> loomise üheks põhjuseks oli äratundmine, et WordNet'i andmestikust on jäänud välja oluline semantilise info allikas – predikaat-argument-struktuur (Kipper Schuler jt 2000). VerbNet'i koostajad valisid verbide klassifitseerimiseks

\* Artikli valmimist on toetanud Haridus- ja Teadusministeerium (projekt EKT12). Autor tänab anonüümseid retsensente asjalike soovitude ja konstruktiivse kriitika eest.

<sup>1</sup> VerbNet <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> (15.02.2013).

WordNet'i klassifikatsioonist hoopis erineval põhimõttel koostatud Beth Levini inglise keele verbide käsitluse (1993), kus grupeerimise aluseks on just argumentide struktuuri sarnasus ja verbide käitumine erinevate teiseenduste korral. Erinevalt WordNet'ist ja FrameNet'ist on VerbNet'i andmestruktuuris koht ka viidetele, mis võimaldavad luua seoseid eelnimetatud andmekogudega. Shi ja Mihalcea (2005) tõstsid oma töös erinevate andmekogude kombineerimisest semantilise analüüsi parendamiseks esile, et selline andmekogude ühendamine tõstab märgatavalt semantilise analüüsi kvaliteeti. Viimasel ajal ongi näha, et erinevate lähenemiste ühendamisele pööratakse üha enam rõhku (Palmer 2009).

Eesti keeles on verbe analüüsinud ja grupeerinud Huno Rätsep juba 1970-ndate lõpus oma töös eesti keele lihtlausetes tüüpidest (1978). Tema tööst nähtub, et eesti keeles kannavad lausete tähenduse ja ülesehituse koha pealt olulist infot ka verbi juurde kuuluvate argumentide käänded. Nende märgendamiseks ei ole inglise keele jaoks koostatud VerbNet'is praegu küll vahendeid, kuid VerbNet'i andmestruktuuris on olemas koht nende lisamiseks.

Selleks, et kasutada predikaat-argument-struktuure eesti keele semantilise analüüsi ülesannetes, nagu näiteks sõnatähenduste ühestamise probleemide lahendamisel, on autor seisukohal, et eesti keele jaoks tuleb luua oma VerbNet, milles on ära kasutatud senise VerbNet'i loomise ja laiendamise kogemusi (Kipper Schuler jt 2006) ning mida on täiendatud eesti keelele omaste nüanssidega. Järgnevalt vaatleme, kuidas seda teha saab.

## **2. Inglise VerbNet'i leksikaalse kirje koostamise põhimõtted**

VerbNet'i andmestik koosneb kirjetest, millest igaüks kirjeldab üht verbiklassi. Iga verbiklass hõlmab endas ühesuguse argument-struktuuriga ja sarnase süntaktilise käitumisega verbe – klassi liikmeid – ning võib sisaldada ka alamklasse, moodustades nii klasside hierarhia. Klassi nimetus ja number annavad infot nii klassi liikmete üldisema tüübi, tähenduse kui ka klassi järjenumbriga ja hierarhia taseme kohta – näiteks PUT-9.1 ('panema') või PUT\_SPATIAL-9.2-1 ('panema kusagile spetsiifilisel viisil', näiteks *riputama*). Klassi kuuluvate verbide argumentide fikseerimiseks on loetletud vastavad semantilised rollid, mille juures võivad olla ka valikukitsendused määratlemaks rollide peenemaid nüansse. Süntaktilise käitumise iseloomustamiseks on igas verbiklassis esitatud üks või mitu süntaktilist freimi. Igas sellises freimis on ära toodud üks näitelause, lausemall ja freimi semantika kirjeldus. Lausemall näitab üht lubatud lauseehitust, esitades verbi, selle semantiliste rollide nimetustega tähistatud argumentide ja muude kohustuslike lauseliikmete järjekorda. Kui on vajadus spetsifitseerida süntaksist tulenevaid kitsendusi, siis lisatakse need lausemalli juurde. Freimi semantika kirjeldamiseks kasutatakse valitud hulka semantilisi predikaate ning ajaline dimensioon tuuakse sisse sündmust tähistava muutujaga ja sellele muutujale rakendatud operaatoritega.

### 3. Eesti ja inglise keele verbide võrdlus

Järgnevalt vaatame, kas ja kuidas on võimalik olemasolevaid inglise VerbNet'i andmeid eesti keele jaoks ära kasutada. Lähtudes seisukohast, et verbi tähendus ja argumentide struktuur ei sõltu kasutatavast keelest (Dorr 1997), võib valida ühe eestikeelse verbi, leida sellele vastav ingliskeelne verb ja püüda teisendada vastavad inglise keele jaoks koostatud VerbNet'i kirjed eesti keelele sobivaks. Erinevate keelte verbide omavahelise vastavuse kindlakstegemiseks on otstarbekas kasutada WordNet'i andmeid, sest paljudel juhtudel on valitaval eesti keele verbil mitmeid eri tähendusi ja iga erinev tähendus viib meid üldjuhul erineva inglise keele verbi juurde.

Valime meid huvitavaks eestikeelseks verbiks *istuma*. Sellel sõnal on Eesti Wordnet'i<sup>2</sup> andmetel 4 tähendust (näide 1).

- (1) *istuma\_1*, *istet võtma* → ingl *sit* (sit%2:38:00::)<sup>3</sup>  
*istuma\_2*, *kinni istuma* → ingl *serve* (serve%2:42:00::), *do* (do%2:42:02::)  
*istuma\_3*, *sobima*, *passima*, *klappima* → ingl *fit* (fit%2:42:02::)  
*istuma\_4* 'istuvas asendis olema' → ingl *sit* (sit%2:35:00::)

Leitud ingliskeelsetele vastetele vastavad VerbNet'is<sup>4</sup> tõepoolest erinevad klassid (näide 2).

- (2) *sit* (sit%2:38:00::) → ASSUMING\_POSITION-50, PUT\_SPATIAL-9.2-1  
*serve* (serve%2:42:00::) → SPEND\_TIME-104  
*fit* (fit%2:42:02::) → FIT-54.3, FUNNEL-9.3-1  
*sit* (sit%2:35:00::) → SPATIAL\_CONFIGURATION-47.6, PUT\_SPATIAL-9.2-1

Võtame lähema vaatluse alla tähendusele 'istuma\_1, istet võtma' leitud vaste klasside kirjeldused ASSUMING\_POSITION-50 'asendit sisse võtma' (näide 3), PUT\_SPATIAL-9.2-1 'panema spetsiifilisel viisil' (näide 4). Analüüsisid klasside semantilisi rolle ja lausemalle võib tõdeda, et klass ASSUMING\_POSITION-50 (vt näidet 3) on tähenduslikust küljest õige vaste, kuna AGENT võtab sisse istuva asendi, ja klass PUT\_SPATIAL-9.2-1 (näide 4) ei saa sobida selle tõttu, et eesti keele seisukohast elutud asjad üldjuhul istet ei võta.

- (3) ***assuming\_position-50*** 'asendit sisse võtma'<sup>5</sup>

Liikmed (25):

..., *rise*, *sit*, *sit\_down*, *slouch*, ...

Rollid:

AGENT [+ELUS]

KOHT [+KOHT & -REGIOON]

Freimid:

NP V

Näide: "The dog flopped." 'Koer heitis pikali.'

Süntaks: AGENT V

<sup>2</sup> Eesti Wordnet <http://www.cl.ut.ee/ressursid/teksaurus/> (01.10.2012).

<sup>3</sup> See sümbolite jada on WordNet'i tähendusviide (ingl sense key), mille täpsema kirjelduse leiab WordNet'i dokumentatsioonist veebilehel <http://wordnet.princeton.edu/man/senseidx.5WN.html> (10.12.2012).

<sup>4</sup> VerbNet, vt sit <http://verbs.colorado.edu/verb-index/index/S.php> (15.02.2013).

<sup>5</sup> Näide on eestindatud. Vt VerbNet: [http://verbs.colorado.edu/verb-index/vn/assuming\\_position-50.php#assuming\\_position-50](http://verbs.colorado.edu/verb-index/vn/assuming_position-50.php#assuming_position-50) (15.02.2013).

Semantika: EITUS(POSITSIOON(ALGUS(SÜNDMUS), AGENT, POS))  
LIKUMINE(TOIMUMISEL(SÜNDMUS), AGENT)  
POSITSIOON(LÖPP(SÜNDMUS), AGENT, POS)<sup>6</sup>

NP V PP.KOHT

Näide: “*The dog flopped in the corner.*”  
‘Koer heitis nurgas pikali.’

Süntaks: AGENT V {{+LOC}} KOHT

Semantika: EITUS(POSITSIOON(ALGUS(SÜNDMUS), AGENT, POS))  
LIKUMINE(TOIMUMISEL(SÜNDMUS), AGENT)  
POSITSIOON(LÖPP(SÜNDMUS), AGENT, POS)  
PREP(SÜNDMUS, AGENT, KOHT)

(4) **put\_spatial-9.2-1** ‘panema spetsiifilisel viisil’<sup>7</sup>

Liikmed (8):

..., *rest, sit, stand, ...*

Rollid (pärinevad ülemklassist *put\_spatial-9.2*):

AGENT [+ELUTU]

TEEMA [+KONKREETNE]

SIHT [+KOHT & -REGIOON]

Freimid:

NP V PP.SIHT

Näide: “*The books lean against the shelf.*”  
‘Raamatud toetuvad riuli vastu.’

Süntaks: TEEMA V {{+loc}} SIHT

Semantika: PREP(SÜNDMUS, TEEMA, SIHT)  
POSITSIOON(SÜNDMUS, TEEMA, POS)

NP V ADVP

Näide: “*The books lean there.*” ‘Raamatud toetuvad seal.’

Süntaks: TEEMA V SIHT <+adv\_loc>

Semantika: PREP(SÜNDMUS, TEEMA, SIHT)  
POSITSIOON(SÜNDMUS, TEEMA, POS)

Järgnevalt vaatleme üht võimalikku süntaktiliste freimide hulka sõnale *istuma* klassis ASSUMING\_POSITION-50 ‘asendit sisse võtma’ (näide 5). Kuna selle klassi alla peaksid kuuluma verbid, mis vastavad sellisele kirjeldusele nagu ‘tegevus, mille tagajärjel tegija võtab sisse tegevusega määratletud positsiooni’, siis verbi *istuma* puhul vaatleme neid lauseid, kus AGENT istub mingisugusele KOHALE<sup>8</sup> (määratud lokaaladverbiaaliga, siin latiiitse kohamäärusega) – lause peab vastama küsimusele “istub kuhu?”. Selleks, et eristada tähendust ‘istuma\_1, istet võtma’ tähendusest ‘istuma\_4 (‘istuvast asendis olema’), peab freimile lisama kitsenduse, et rollis KOHT olev lauseliige peab olema kas allatiivis või illatiivis või genitiivis koos sobiva adpositsiooniga. Huno Rätsepa järgi on siin kõige õigem kasutada substitutsiooniklassi *intraalokaalne direktsionaal* (tähistatakse sümboliga “Di”, vt näidet 5), mis katab kõik eespool nimetatud variandid (Rätsep 1978: 45).

<sup>6</sup> Esitatud semantilist formalisatsiooni saab lugeda järgmiselt: sündmuse (antud näites pikali heitmise) alguses ei ole agent (koer) positsioonis POS (pikali); sündmuse toimumise ajal agent liigub; sündmuse lõppedes on agent positsioonis POS.

<sup>7</sup> Näide on eestindatud. Vt VerbNet: [http://verbs.colorado.edu/verb-index/vn/put\\_spatial-9.2.php#put\\_spatial-9.2-1](http://verbs.colorado.edu/verb-index/vn/put_spatial-9.2.php#put_spatial-9.2-1) (15.02.2013).

<sup>8</sup> Antud näiteverbi tähenduse ‘istuma\_1, istet võtma’ puhul oleks põhjendatud semantilise rolli SIHT kasutamine, kuid Inglise VerbNet’i andmetes on vaadeldavas verbiklassis kasutusele võetud üldisema tähendusega semantiline roll KOHT just verbiklassi teisi liikmeid silmas pidades.

On huvitav märkida, et õige käände kasutamine võib olla sõltuvuses rollis KOHT kasutatavast asjast, nagu näiteks: *istus pingile / tugitooli / mootorrattale / autosse*. Ehk siis on olemas asju, kuhu saab ainult sisse istuda ja asju, kuhu saab ainult peale istuda. Samas tuleb tõdeda, et nii peent eristamist ei saa VerbNet'i andmestikku kasutades ette kirjutada, sest see on spetsiifiline iga üksiku verbi puhul ja VerbNet'i kirje keskendub siiski verbide klassile. Selline iga verbi jaoks sobivate atribuutide ja nende süntaktilise käitumise määratlemine vajab täiendavat andmekogu.

(5) NP V PP.location

Näide: "Mees istus pingile."

Süntaks: AGENT <+nom> **V** KOHT <+Di>

Semantika: EITUS(POSITSIOON(ALGUS(SÜNDMUS), AGENT, POS))

LIKUMINE(TOIMUMISEL(SÜNDMUS), AGENT)

POSITSIOON(LÕPP(SÜNDMUS), AGENT, POS)

PREP(SÜNDMUS, AGENT, KOHT)

Freimis (5) olevas lauses võib kolme lauseliiget järjestada kuuel erineval viisil, mis tähendab, et kõik permutatsioonid on lubatud. Võib muidugi kõik sõnajärjed eraldi freimidena üles lugeda, kuid sellistel juhtudel on otstarbekam täiendada freimi kirjapanemise viisi. Permutatsioon võiks tähistada funktsiooniga "Perm", nagu näites (6).

(6) Perm(NP V PP.location)

Näide: "Mees istus pingile."

Süntaks: Perm(AGENT <+nom> **V** KOHT <+Di> )

Semantika: EITUS(POSITSIOON(ALGUS(SÜNDMUS), AGENT, POS))

LIKUMINE(TOIMUMISEL(SÜNDMUS), AGENT)

POSITSIOON(LÕPP(SÜNDMUS), AGENT, POS)

PREP(SÜNDMUS, AGENT, KOHT)

Lähtudes Huno Rätsepa poolt pakutud lihtlausetüübist 2.2.126 (N+nom V<sup>1</sup> (Di) (V<sup>2</sup>+ma)) (Rätsep 1978: 100), saab lisada veel ühe freimi (7):

(7) Perm(NP V PP.location S\_INF)

Näide: "Peeter istus laua taha sööma."

Süntaks: Perm(AGENT <+nom> **V** KOHT <+Di> EESMÄRK <+v\_ma> )

Semantika: EITUS(POSITSIOON(ALGUS(SÜNDMUS), AGENT, POS))

LIKUMINE(TOIMUMISEL(SÜNDMUS), AGENT)

POSITSIOON(LÕPP(SÜNDMUS), AGENT, POS)

PREP(SÜNDMUS, AGENT, KOHT)

TEGEVUS(PÄRAST(SÜNDMUS), AGENT, EESMÄRK)

Selleks, et fikseerida iga verbiklassi atribuudid, eriti just süntaktilised freimid, tuleb loomulikult leida kõigi klassi liikmete atribuutide ühisosa. Kui leidub selliseid erinevusi, mille alusel saab osasid verbe täiendavalt grupeerida, siis võib kaaluda ka inglise keele VerbNet'ist tuleneva klassihierarhia täiendamist.

## 4. Milliseid ülesandeid saab VerbNet'i abil lahendada?

Alustades lihtlause uurimist morfoloogilise analüüsiga, on üheks järgmiseks võimalikuks sammuks leitud verbi määratlemine VerbNet'is. Selleks tuleb leida kõik uuritava verbi esinemised klassihierarhias ja hinnata iga vastava klassi freimi sobivust uuritava lausega. Kõige paremini sobiva freimi leidmise läbi oleme kindlaks teinud ka verbi viite WordNet'ile ning sisuliselt lahendanud sõnatähenduste ühestamise ülesande (Brown jt 2011).

Kui on kindlaks tehtud lausele vastav VerbNet'i freim, siis sellele infole toetudes on võimalik läbi viia nende lauseliikmete semantiline märgendamine vastavalt verbiklassis loetletud semantiliste rollide kasutusele freimi lausemallis.

Selleks, et saada aimu VerbNet'i andmestikul põhineva sõnatähenduse ühestaja loomisega kaasnevatest probleemidest, viidi läbi katse Keeleveebi ilukirjanduse korpuse andmetega.<sup>9</sup> Katse eesmärgiks oli verbi "istuma" jaoks koostatud VerbNet'i kirje abil selle verbi ühe tähenduse esinemise määratlemine testandmetes. Kuna määratluse korrektsust tuli kontrollida käsitsi, siis sai esialgu piiratud ühe juhuliselt valitud andmefailiga korpusest. Selleks osutus teos "Seitsmes rahukevad", mis sisaldas 110 lauset meid huvitava verbiga *istuma*. Laused olid eelnevalt automaatselt märgendatud morfoloogilise analüsaatoriga ja osalause piiride määrajaga.

Tähenduse 'istuma\_1, istet võtma' määratlemiseks kasutati algoritmi, mis kontrollis lausemallile (vt näidet 6) vastavate käänetega sõnade olemasolu kontrollitavas lauses ning arvestas seejuures osalause piiride ja käskiva kõneviisiga. Täheendusega 'istuma\_1' verbe oli 110 lause hulgas 39, neist õigesti määrati 30 ja 9 jäi leidmata. Samas leiti 6 teise tähendusega verbi, mis väideti kuuluvat ka tähenduse 'istuma\_1' alla, aga mis seda tegelikult ei olnud. See annab tehtud katse saagiseks 76,9% ja täpsuseks 83,3%.

Analüüsides nii vale määratlusega lauseid kui ka lauseid, milles õige tähendus jäi leidmata, selgus, et mitmel juhul oli tegemist andmetöötluse eelmise etapi vajakajäämistega. Esines nii morfoloogilisest analüüsist kui ka osalause piiride määramisest pärit ebatäpsusi, mis tõid kaasa algoritmi töö tulemuste ebakorrektsuse. Sisulisest küljest näitas katse seda, et ettevalmistatud VerbNet'i kirjes olid puudulikult kirja pandud subjekti käänete kitsendused, kuna arvesse jäi võtmata võimalus, et subjekti on hulgafras. Seega, täiendav töö eesti keele süntaksireeglite kirjapanemisel VerbNet'i lausemallidesse võimaldab luua parema täpsusega töötava rakenduse.

## 5. Tulemused

Eestikeelse verbi jaoks vastavat inglise keele verbi VerbNet'i kirjet muutmata kujul kasutada ei saa, kuid on põhimõtteliselt võimalik ära kasutada verbiklasside hierarhia koos klasside semantiliste rollide, nende valikukitsenduste ning freimi semantika kirjeldustega. Verbiklasside liikmed tuleb asendada samatähenduslike eesti keele verbidega ja need varustada viidetega nii Eesti Wordnetile kui ka inglise VerbNet'ile. Eraldi tähelepanu nõuavad süntaktilised freimid, sest erineva süntaksi ja vabama sõnajärje tõttu võib eesti keeles olla verbidel rohkem lausemalle ning kindlasti on vaja võtta kasutusele käänetega seotud grammatilised kitsendused.

Tehtud katsetuste põhjal võib hinnata, et ühe verbi analüüsimine ja vastavate VerbNet'i kirjete vormistamine võib vastavalt eri tähenduste hulgale võtta aega üks kuni kolm päeva. Kogemuste kogunemine kiirendab antud tööd, samas iga järgmise verbi lisamine andmekogusse võib põhjustada vajaduse mõne klassi ümberstruktureerimise järele. Kui võtta eesmärgiks inglise keele VerbNet'iga võrreldavas mahus andmekogu loomine, siis võib hinnata töömahuks ligikaudu 35 inimaastat.<sup>10</sup>

## 6. Kokkuvõte

Artiklis on kirjeldatud VerbNet'i põhimõtteline andmestruktuur, mis on sobiv kasutamiseks ka eesti keele jaoks. Lisatud on VerbNet'i kirjete esitamist illustreerivad näited. Autor on seisukohal, et Eesti VerbNet on kindlasti vajalikuks täienduseks meie keeleressursside kogule, andes omapoolse panuse mitmete keeletehnoloogia alamülesannete lahendamisse. Tuleb rõhutada, et selline andmekogu ei ole valdkonnaspetsiifiline ja on seega kasutatav kõikide valdkondade rakendustes, mis võivad vajada loomuliku keele töötlemisega seotud funktsionaalsust. Kuna verbe ja nende predikaat-argument-struktuure ning tähendusi tuleb analüüsida ning klassifitseerida igäüht individuaalselt, siis on inglise VerbNet'i raames välja pakutud andmestruktuur sobiv verbide semantilise info kirjelduste kogumiseks nii selle keeleressursi võimalikke rakendusi kui ka teistes maades tehtava samalaadse tööga kokkusobivust silmas pidades.

### Viidatud kirjandus

- Baker, Collin F.; Fillmore, Charles J.; Lowe, John B. 1998. The Berkeley FrameNet project. – Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Vol. 1, 86–90.
- Brown, Susan Windisch; Dligach, Dmitriy; Palmer, Martha 2011. VerbNet class assignment as a WSD task. – Johan Bos, Stephen Pulman (Eds.). Proceedings of the Ninth International Conference on Computational Semantics. IWCS 2011. Stroudsburg: Association for Computational Linguistics, 85–94.
- Dorr, Bonnie J. 1997. Large-scale dictionary construction for foreign language tutoring and interlingual machine translation. – Machine Translation 12 (4), 271–322.
- Kerner, Kadri; Orav, Heili; Parm, Sirli 2010. Growth and revision of Estonian WordNet. – Principles, Construction and Application of Multilingual Wordnets. Proceedings of the 5th Global Wordnet Conference. Narosa Publishing House, 198–202.
- Kingsbury, Paul; Palmer, Martha 2002. From TreeBank to PropBank. – Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC-2002). <http://www.lrec-conf.org/lrec2002/> (25.02.2012).
- Kipper Schuler, Karin; Dang, Hoa Trang; Palmer, Martha 2000. Class-based construction of a verb lexicon. – Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence. AAAI Press / The MIT Press, 691–696.
- Kipper Schuler, Karin; Korhonen, Anna; Ryant, Neville; Palmer, Martha 2006. Extending VerbNet with novel verb classes. – Proceedings of 5th International Conference on Language Resources and Evaluation (LREC 2006), No. 2.2.

<sup>10</sup> VerbNet'i versioonis 3.2 on 6280 verbi, 270 verbiklassi ja 206 alamklassi (Unified Verb Index 2013).

- Levin, Beth 1993. English Verb Classes and Alternations: A Preliminary Investigation. Chicago: University of Chicago Press.
- Liu, Mei-chun; Chiang, Ting-yi 2008. The construction of Mandarin VerbNet: A frame-based study of statement verbs. – *Language and Linguistics*, 9 (2), 239–270.
- Miller, George A. 1995. WordNet: a lexical database for English. – *Communications of the ACM*, 38 (11), 39–41. <http://dx.doi.org/10.1145/219717.219748>
- Miller, George A.; Beckwith, Richard; Fellbaum, Christiane; Gross, Derek; Miller, Katherine J. 1990. Introduction to WordNet: An on-line lexical database. – *International Journal of Lexicography* 3(4), 235–244. <http://dx.doi.org/10.1093/ijl/3.4.235>
- Palmer, Martha 2009. Semlink: Linking PropBank, VerbNet and FrameNet. – Fifth International Workshop on Generative Approaches to the Lexicon (GL 2009). Pisa, Italy, 9–15.
- Rätsep, Huno 1978. Eesti keele lihtlausete tüübid. Tallinn: Valgus.
- Shi, Lei; Mihalcea, Rada 2005. Putting pieces together: Combining FrameNet, VerbNet and WordNet for robust semantic parsing. – Alexander Gelbukh (Ed.). *Computational Linguistics and Intelligent Text Processing. 6th International Conference, CICLing 2005, Mexico City, Mexico, February 13-19, 2005, Proceedings. Lecture Notes in Computer Science 3406*. Berlin, Heidelberg: Springer-Verlag, 100–111.
- Õim, Haldur; Orav, Heili; Kahusk, Neeme; Taremaa, Piia 2010. Semantic analysis of sentences: The Estonian experience. – Inguna Skadiņa, Andrejs Vasiļjevs (Eds.). *Human Language Technologies: The Baltic Perspective. Proceedings of the Fourth International Conference, Baltic HLT 2010. Frontiers in Artificial Intelligence and Applications 219*. IOS Press, 208–213. <http://dx.doi.org/10.3233/978-1-60750-641-6-208>

## Võrgumaterjalid

- Eesti Wordnet, <http://www.cl.ut.ee/ressursid/teksaurus/> (01.10.2012).
- Keeleveebi ilukirjanduse korpus, <http://www.keeleveeb.ee/> (01.10.2012).
- Unified Verb Index, <http://verbs.colorado.edu/verb-index/> (21.03.2013).
- VerbNet, <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html> (15.02.2013).
- WordNet'i tähendusviite kirjeldus, <http://wordnet.princeton.edu/man/senseidx.5WN.html> (10.12.2012).

**Indrek Jentson** (Tartu Ülikool) on erialalt informaatik. Huvipakkuvaks uurimisvaldkonnaks on semantiline representatsioon.  
[indrek.jentson@ut.ee](mailto:indrek.jentson@ut.ee)



# THE FEASIBILITY OF ESTONIAN VERBNET

**Indrek Jentson**

University of Tartu

The idea of creating a lexical database of verbs for the Estonian language is introduced, based on principles and experience of creating a similar database for the English language – VerbNet. Classification of verbs by meaning and also by predicate-argument structure is placed under focus and the possibilities of improving the data structures with properties distinctive to the Estonian language are analysed. The creation of lexical entries for the database is described using the example of one polysemic verb. The results of an experiment on semantic analysis and sense disambiguation on sentences from the morphologically disambiguated corpus when using defined lexical entries in the database of verbs are described. It is concluded that the possibility of using English VerbNet data structures and verb classification hierarchy for Estonian VerbNet is feasible. The importance of having such a language resource for the Estonian language is stressed.

**Keywords:** language resources, computational linguistics, computational lexicology, semantic labelling, Estonian