

KUIDAS MÄÄRATA EESTI KEELE SÕNAVARA TUNDETOONE?

Ene Vainik

Ülevaade. Artiklis antakse ülevaade sõnavaraüksuste tundetoonide määramise teoreetilistest alustest ja praktilistest võimalustest. Esmalt kirjeldatakse ülesande olulisust, siis teoreetilisi lähte-eeldusi ja vaetakse neist tulenevaid meetodilisi võimalusi. Juhtumianalüüsina kirjeldatakse eesti keele põhisõnavara sõnastiku märksõnade tundetoonide hindamise meetodikat, katse läbiviimist ja tulemusi ning võrreldakse neid varasemate eesti keele emotsioonisõnavara uurimustega. Artikkel näitab, kuidas eesti lingvistika kontekstis uudset üleannet lahendada meetodiliselt võimalikult korrektselt, kuid arutleb ka küsitavuste ja perspektiivide üle.*

Võtmesõnad: emotsionaalne valents, meetodika, sõnavara, afektiivsus, põhisõnavara sõnastik, eesti keel

1. Sissejuhatus

Asjaolu, et on tekkinud vajadus pöörata tähelepanu eestikeelsete sõnadega tüüpiliselt kaasnevatele emotsionaalsetele lisatähendustele – nn tundetoonidele – tuleneb sellest, et on ülesandeid, mille lahendamiseks ei piisa keele tähistusfunktsiooni uurimisest, tarvis läheb teadmisi ka keele ekspressiivsest funktsioonist ja emotsioonide väljendumisest (Foolen 1997, Hübler 1998).

Üks selliseid ülesandeid on eestikeelse kõnesünteesi loomulikustamine (Altrov, Pajupuu 2008). Selleks, et ettelõetavatele lausungitele oleks võimalik lisada inimlikku emotsionaalsust, on ühest küljest tarvis uurida ja modelleerida emotsioonide prosoodilist väljendumist (Iida jt 2003, Ververidis, Kotropoulos 2006), teisalt aga õpetada süsteem kirjaliku teksti emotsionaalsust või selle puudumist määrama (Vainik 2010, Pajupuu jt, ilmumas). Katsetes määrata teksti, lõigu, lause ja isegi osalause emotsiooni on üksmeelselt leitud, et taolise süsteemi osaks peab olema sõnade tüüpilisi emotsionaalseid lisatähendusi esitav sõnastik (ülevaadet vt Vainik 2010). Eestikeelsete sõnade tundetoonide ehk emotsionaalsete valentside määra-

* Artikkel on valminud sihtfinantseeritava teadusteema SF0050023s09 ja Eesti Keele Instituudi baasrahastusprojekti "Eesti keele sõnavara afektiivsuse määramise raamtööd aastal 2010" toetusel.

misest on vaja kõnelda just nüüd, kuna Eesti Keele Instituudis on koostamisel uut tüüpi eesti keele põhisõnavara sõnastik (nn põhisõnastik), mis koondab ligikaudu 4000 kõige sagedasemat ning aktiivsele keeleõppijale kõige tarvilikumat eesti keele sõna koos mitmesuguse informatsiooniga selle tüüpilise kasutamise kohta (vt Kallas, Tuulik 2011). Uurimused on näidanud, et sellisest sõnavara hulgast piisab haritud kõrgtasemel keeleoskaja vajadusteks (Pajupuu jt 2009). Kuna süntesaatori poolt etteloetav tekst ei ületa keerukuse poolest haritud eesti keele oskaja taset, siis on loomulik mõelda, et põhisõnastik võiks olla põhjaks ka kõnesünteesi loomulikustamise eeltingimuseks olevale emotsiooni määramisele tekstis (Pajupuu jt, ilmumas).

Teisest küljest on nähtud vajadust teavitada sõnavara tüüpilistest tundetoonidest ka teise keele õppijat. Kakskeelsete isikutega tehtud uurimused on näidanud, et kuigi üldine teise keele oskus on neil kõrgtasemel, võib just emotsionaalse väljenduse olukordades tekkida äkki sobivatest sõnadest puudus ja võidakse kergesti emakeelele ümber lülituda (Pavlenko 2008b). Eesti keele põhisõnavara sõnastik on mõeldud pakkuma keeleõppijast kasutajale aktiivset tuge ka sellistes olukordades.

Nii lähtubki ülesanne määrata põhisõnastiku sõnavara eriomased tundetoonid kahest praktilisest vajadusest ning kavatsusest need vajadused ühitada. Ühitamine tundub realistlikuna, kuna mõlemad ülesandepüstitused – nii kõnesünteesiks vajalik sõnavara märgendus kui aktiivse keeleõppija teavitamine tundetoonidest – lähtuvad ühisest eeldusest, et vähemalt teatud osa keele sõnavarast ei kanna mitte üksnes tähendusi ja/või funktsioone, vaid edastab ka emotsionaalset informatsiooni. Lisaks eeldatakse, et too emotiivne varjund on teataval määral konventsionaalne ning seega peaks seda põhimõtteliselt saama kindlaks teha ja ka sõnastikus fikseerida. Ülesandepüstituse praktilise poole pealt näeb asi välja nii, et põhisõnastiku koostajad on kokku pannud märksõnastiku (tulevane maht üle 4500, artikli kirjutamise hetkel 3015) ja määrata tuleb neid iseloomustavad tundetoonid. Küsimus on selles, kuidas teha seda teoreetiliselt läbimõeldud alusel ning meetodiliselt võimalikult korrektselt.

Artikli esimene pool on pühendatud teoreetilistes lähteoletustes orienteerumisele ning neist tulenevate meetodite sobivuse vaagimisele. Juhtumianalüüsi osas kirjeldatakse eesti keele põhisõnavara sõnastikule rakendatud konkreetset meetodit, esitatakse väljavõtte hindamiste tulemustest ning võrreldakse neid varasemate uurimustega. Lõpuosas arutletakse rakendatud meetodi tugevate ja nõrkade külgede üle ning pakutakse välja võimalusi, kuidas tundetoonide määramist veelgi eesmärgipärasemaks muuta.

2. Teoreetilised ja meetodilised küsimused

2.1. Millisest teoreetilisest taustast lähtuda?

Nähtuse kohta, et loomuliku keele sõnadel on kalduvus edasi anda mitte üksnes ideesid, vaid ka emotsionaalset või hinnangulist suhtumist, on kasutatud erinevaid nimetusi – konnotatsioon, tundetoon, emotsionaalne valents. Käesolevas artiklis kasutatakse peamiselt eestikeelset nimetust tundetoon, mingil määral aga ka

terminit emotsionaalne valents. Valents kui metafoorne (keemiast pärit) termin võeti psühholoogias kasutusele 1930-ndatel aastatel sündmuste ja nähtuste kas külgetõmbavuse (positiivne valents) või eemaletõukavuse (negatiivne valents) tähistamiseks. Peamiselt kasutatakse seda terminit tunnete, afektide, käitumiste (nt liginemine vs. vältimine), eesmärgi (saavutamine vs. mitte-saavutamine jmt) hedoonilisest toonist rääkimiseks, ehkki välistatud pole ka hinnangud muudelt alustelt ja nende mitmekihilisus (Colombetti 2005). Enne tundetoonide praktilist määramist ja sobiva meetodi valimist on hea end teoreetiliselt positsioneerida küsimuse suhtes, kas ja kuidas stabiilselt saavad tundetoonid suvalise sõnavaraüksuse juurde üldse kuuluda.

Traditsiooniline lingvistiline lähenemine on sõnaga kaasneva tundetooni suhtes kõige skeptilisem: nende olemasolu küll mööndakse, kuid seda peetakse sõna mõistelise tuuma “sentimentaalseks juurdekasvuks”, mida on suure variatiivsuse tõttu pea võimatu määratleda ja millest teadusele kasu pole (Sapir 2009: 50–51). Leksikograafias lähtutakse objektiivsetest tõenditest keele tegeliku kasutamise kohta kas kirjalikus või suulises suhtluses ja mitte selle “potentsiaalid”. Öeldakse, et sõnaraamatus tuleb kirjeldada nähtusi, mis on sagedased ja laialt levinud (esinevad korpuses eri tekstitüüpides ja kuuluvad keele “tuuma”) (Atkins jt 2008). Traditsioonilisest puhtlingvistilisest ja leksikograafilisest tähenduse kirjeldusest on tundetoonide kirjeldamine ja määramine jäänudki välja arvatavasti just neile omistatud subjektivsuse ning muutlikkuse tõttu. Üksnes sõnastikes kasutatavad stiilmärgendid (nt *halvustav*, *hellitlev*) viitavad mõningal määral sõnade emotsionaalsele lisaväärtusele ja tüüpilisele kontekstile (Voll 2002).

Psühholoogias seevastu ei kahelda, et ka üksiksõna nagu iga muugi stiimul võib esile kutsuda emotsionaalseid reaktsioone ja et neil reaktsioonidel võib olla mingil määral stabiilne ja kinnistunud iseloom; arutletakse pigem selle üle, kuidas üksiksõnade iseloomulike valentside põhjal määrata lause ja kogu teksti valentsi (Bestgen 1994). Selle, lähtekohalt bihevioristliku mõtteviisi klassikaks on C. E. Osgoodi, G. J. Suci ja P. H. Tannenbaumi ([1957] 1975) uurimus, milles esmakordselt kirjeldati nn semantilise diferentsiaali meetodit ja selgitati välja semantilise ruumi kolm universaalset dimensiooni: *hinnang*, *tugevus* ja *aktiivsus*, mille suhtes paigutuvad väidetavalt niihästi keele kõik sõnad kui ka mittedõnalised mõisted – nimed, pildid, narratiivid – ja üldse kõik, mida küsimustiku ja polaarsest adjektiividest moodustatud skaalade abil mõõdetakse.

Psühholingvistilise lähenemise järgi seevastu ei eeldata, et emotsionaalsus peaks olema potentsiaalselt keele iga sõna iseloomustav parameeter. Küll aga leidub iga inimese jaoks siiski hulk sõnu, mille kasutamisega emotsionaalne konnotatsioon võib kaasneda. Väidetavalt kujutavad emotsionaalsed keelendid endast isiku mentaalses leksikonis eraldi klassi, erinedes nii abstraktsetest kui konkreetsetest sõnadest abstraktsuse, ettekujutatavuse ja konteksti kättesaadavuse osas (Pavlenko 2008a). Kuulumine samasse emotsionaalsesse kategooriasse võib olla viisiks, kuidas on mentaalse leksikoni üksused omavahel seotud. See oletus sobib kokku ka kognitsiooni ja emotsioonide seoseid selgitava nn assotsiatiivse võrgustiku teooriaga (Bower, Forgas 2000: 94). Selle kohaselt praimivad sõnad, kui nad on antud isiku jaoks emotsionaalselt tähendusrikkad, selle emotsiooniga seostuvate assotsiatsioonide ja mälestuste meenumise.

Üldjoontes sarnane on ka kognitiivses lingvistikas omaks võetud nn entsüklopeedilise semantika kontseptsioon (vt nt Evans, Green 2006, Langacker 1987). Viide entsüklopeediale ei tähenda siin mitte autoriteeti ja üldkehtivust, vaid semantiliste tegurite hõlmavust. Üksiksõna vaadeldakse sisendsõlmene (ingl *access node*) tervesse teadmiste ja kogemuste võrgustikku, milles tippudena esinevad entiteedid ja kaartena ühendavad neid omavahelised suhted (vt Langacker 1987). Entsüklopeediline semantikakäsitlus ühendab teadmisi ja kogemusi, mille subjekt on omandanud, sõltumata omandamise viisist, ajast ja "õigsusest" ning hõlmab kindlasti mõningat varieerumist isikuti. Veelgi lähemal mõistete subjektiivsetele representatsioonidele seisab uusempiritsistlik simulatsiooniteooria. Väidetakse, et sõnavaraüksus kutsub kognitiivsete representatsioonide tasandil esile tähistatava situatsiooni simulatsiooni, selle osalise "läbimängu" aju multimodaalse kogemuse representatsioonide "teatris". Mingi sõna "töötlemisel" aktiveeritakse samade tajumodaalsustega seotud keskused, mis on kasutusel sama sõnaga tähistatavas situatsioonis osalemisega või selle vaatlemisega. (Barsalou 2003, Evans, Green 2006, Feldman Barrett 2006, Gallese, Lakoff 2005.) Paljude, kui mitte kõigi loomuliku keele mõistete olemuslikuks osaks on situatsioonile "läbimängus" antav hinnang selle kohta, kas seda nähtust tuleks pigem vältida või on see huvipakkuvalt kasutootav. Need hinnangud võivad ilmneda nt teadvustamata liikumistendentsina kas objekti suunas või sellest eemale.

Kokkuvõttes võib öelda, et lähenemised erinevad selle poolest, kas sõnade emotsionaalseid omadusi peetakse pigem osaks keele (juba realiseerunud) struktuurist või osaks inimese mentaalsest struktuurist, kus see esineb alles potentsiaalina. Vastusest sellele küsimusele olenevad meetodid, mida sõna tähenduse ja kaastähenduste määramisel soovitatakse kasutada. Esimesel juhul (traditsiooniline ja leksikograafiline lähenemine) arvatakse sõnavaraüksuse tähendus ja sellega kaasnev tundetoon sugenevat kasutuskontekstist ja seega tuleb tähenduste määramisel võtta arvesse eelkõige andmeid keele(kasutuse) seest. Teisel juhul (psühholoogiline, psühholingvistiline, kognitiivne ja simulatsiooniteooria) usutakse pigem, et omadused kuuluvad sõnavaraüksuse juurde ka ilma lingvistilise kontekstita, seosed esinevad tähenduste ja kaastähenduste representatsioonide vahel pigem assotsiatsioonidena ja seega on nende tuvastamiseks õigustatud kaasata ekstralingvistilist informatsiooni (Valitutti jt 2004). Käesolevas artiklis kirjeldatud uurimuse aluseeldused sobivad kokku kõigi lähenemistega peale traditsioonilise ja leksikograafilise.

2.2. Millist meetodit valida?

Nagu öeldud, lähtutakse käesolevas uuringus teoreetilisest seisukohast, et sõnad on abstraktsed sümbolid, mille omadused on tuvastatavad ka ilma konteksti ja kasutussituatsioonita. Sõnavara tundetoonide määramise ülesanne kujutab endast ülesannet eksplitseerida seos, mis on sõnal kui stiimulil ja selle emotsionaalsel tähendusel kas siis semantilisel, kogemuslikul või tajude simulatsiooni tasandil. Järgnevas tutvustatavad meetodid tuginevad valdavalt semantilistele seostele ja inimeste kas eksplitsiitsetele või implitsiitsetele teadmistele, ainult kaks viimasena tutvustatavat meetodit kasutavad emotsioonidega kaasnevaid kogemuslikke ja aju tasandil esinevaid multimodaalsuse efekte.

Introspektsioon on meetod, mille puhul uurija tugineb enda teadmistele, kogemustele ja keelepädevusele. Introspektsioon on aktsepteeritav uurimismeetod näiteks semantikas (vt Huumo 2008). Samas, kuivõrd tundetoonid on osaliselt ekstrasemantiline nähtus, millest on haaratud mitte üksnes mõisteliste seoste võrgustik, vaid ka individuaalsed emotsionaalsed kogemused ning tajusimulatsioonid, ei pruugi ühe uurija sisevaatlus anda piisavalt usaldatavat tulemust. Lisaks eri uurijate hinnangute võimalikule mittekokkulangemisele ei pruugi anda korratavat tulemust ka sama uurija introspektsioonilised hinnangud eri aegadel. Nende karide ületamiseks kasutatakse nt psühholoogias kahe sõltumatu eksperdi meetodikat ja usaldatakse seda osa tulemustest, kus hinnangud kattuvad (vt nt Aavik, Allik 2002). Võidakse kasutada ka nn domineeriva arvamuse meetodit, mispuhul eksperte on kolm ja arvesse läheb see arvamus, mida on kaks (vt Pennebaker jt 2001). Introspektsiooni usaldatavust saab tõsta ka korduvate hindamiste seeriat kasutades, kus vahepeale jäävad piisavalt pikad ajavahemikud ning arvesse lähevad sarnaselt langetatud otsused.

Introspektsiooni implitsiitsete aspektide esile toomiseks on lingvistikas käibel tehnikad, mida on nimetatud mõnevõrra erinevalt: propositsiooniline meetod (Pavlenko 2008a), lausete kontrasteerimise meetod (Cruse 2000), süntaktilis-grammatilised testid (Orav 2006). Nende olemus seisneb selles, et uuritav sõna pistetakse mingisse etteantud lüngaga lausesse, mille vastuvõetavust uurija siis oma introspektsioonile tuginedes hindab. Nii võiks üritada testida näiteks positiivset tundetooni lausega “X on hea ja ilus” ja negatiivset lausega “X on kole ja paha”. Sõna *sõda* näiteks ei sobi esimesse lausesse, aga sobib teise. Küsimus on selles, et mida võimaldaks taoline test teada saada sellist, mida me sõja kohta niigi ei teaks? Sõna *vaidlus* näiteks taolise testiga ära määrata ei õnnestu – kumbki lause poleks eriti mõttekas, kuigi põhimõtteliselt mitte võimatu. Taolised tehnikad põhinevad asjaolul, et sõna eri tähendused on keelesüsteemis sõna teiste tähendustega erinevates suhetes. Kui tundetoon (positiivsus vs. negatiivsus) on osa tähendusest, võib loota seda taoliste testidega määrata. Kui tegu on aga pragmaatilise konnotatsiooniga, mis tähendusele n-ö hinnangulise suhtumisena “külge kleepub”, pole taolistest testidest abi.

Küsitlusmeetodi puhul välistatakse hinnangute subjektiivsus, kaasates hindamisprotsessi rohkem inimesi; enamasti kasutatakse mitte-eksperte, et saada teada n-ö argitasandi teadmist. Küsitlused võivad olla avatud või suletud küsimustega, kirjalikud või suulised. Näiteks nn välimeetod näeb ette vähemalt 25 kuni 30 inimese kaasamist ja suulist intervjuerimist (Sutrop 2001). Korralikult koostatud ja läbiviidud küsitluste puhul on tulemuste kehtivus ja usaldatavus väljaspool kahtlust, ainsaks puuduseks on selle meetodi ressursimahukus. Olemuselt küsitlus on ka semantilise diferentsiaali meetod (Osgood jt 1975 [1957]). Katseisikutel (vähemalt 30, parim 100) lastakse sõnu/mõisteid/nähtusi hinnata polaarsetest adjektiividest koostatud 7-palliste skaalade suhtes. Ühe üksuse hindamiseks läheb vaja minimaalselt 12 skaalat, kuivõrd iga väidetavalt universaalse dimensiooni (hinnang, aktiivsus, tugevus) mõõtmise jaoks läheb tarvis 4 skaalat, mis peavad olema kohandatud vastavalt uuritavale objektile. Semantilise diferentsiaali kriitikana on välja toodud, et selle usaldatavust mõjutab sõnade polüseemia ja sotsiaalne soovitus – isikud kalduvad andma hinnanguid, mida nad peavad soovitatavateks (Heise 1970). Teise puudusena võib mainida selle ressursinõudlikkust, mistõttu selle rakendamine

nii suuremahulisele ainestikule nagu seda on põhisõnastik võtaks tohutult aega ja tööjõudu. Ainuüksi minimaalse skaalade ja katseisikute arvuga tuleks langetada ligi 1,5 miljonit hinnangut. Veelgi suurejoonelisem on patenteeritud süsteem¹, mille on välja töötanud W. Chase ja Connotative Intelligence Corporation². Selles süsteemis kasutatakse 100–200 mõningase ettevalmistusega isikust koosnevat hindajate “tiiki”, millest siis mõne konkreetsema sõnakogumi jaoks valitakse juhuslikult, aga esindatuse printsiipi järgides (mehed-naised, vanuseastmed) 24 hindajat. Hindamised toimuvad mitmes astmes ja väga diferentseeritud kategooriate süsteemi järgi. Tulemused esitatakse visuaalsete profiilidena. Hindajatel on hindamiseks spetsiaalne tarkvara, mille vahendusel nad hindamist sooritavad. Positiivsed ja negatiivsed emotsioonid ei välista üksteist, vaid profiil kajastab kogu arvamuste spektrit.

Eraldi meetodite grupi moodustavad füsioloogilised inimkatsed, näiteks tuginemine naha galvaanilise reaktsiooni muutuste tuvastamisele. Kui sõna on emotsionaalse laenguga, siis nahareaktsioon muutub (vt Pavlenko 2008a ja tema viiteid). Jällegi peab usaldatava tulemuse jaoks olema katseisikuid hulgi. Moodne meetod on ka ajulainete mõõtmine (EEG). Selle rakendamine sõnade tundetoonide määramiseks tugineks mõistete multimodaalsuse ja simulatsiooniteooriatele. Sõnade emotsionaalsete tähendustega kaasnevaid ajulaineid on mingil määral uuritud Hiinas (Skrandies, Chiu 2003). Enne eestikeelsete sõnade mõõtma asumist oleks vaja aga ulatuslikke katseid ja võrdlusmaterjali selle kohta, kas ja mil määral ajulainete muster sõnaliste stiimulite peale muutub. Ka selle meetodi jaoks on vaja palju katseisikuid ja lisaks on katseseadmed kallid.

Eesti keele põhisõnavara sõnastiku märksõnade (üle 4500 üksuse) tundetoonide määramisel jäävad esialgu paratamatult kõrvale füsioloogilised inimkatsed, nimekirja teisest otsast jääb vähese usaldatavuse tõttu vaatluse alt välja lihtne introspektsioon, ressursimahukuse tõttu jääb esialgu kõrvale ka semantiline diferentsiaal ja sellel põhinevad tehnikad. Juhtumiuuringus lähtusime meetodi ressursinõudlikkuse ja usaldatavuse optimaalsest suhtest ja rakendasime põhisõnastikule domineeriva arvamuse meetodit pisut modifitseeritud kujul, mida kirjeldatakse täpsemalt allpool.

2.3. Mida ja kui täpselt määrata?

Kui ülesandeks on tundetoonide määramine konkreetse sõnastikuprojekti raames, siis tuleb paratamatult ka küsida, millisel kujul peab olema tulemus. Kas iga märksõna kohta tuleb näidata täpselt, millise emotsiooniga ta tüüpiliselt seostub? Või on piisav, kui näidata ära, kas sõna kaldub olema valdavalt positiivse või negatiivse tundetooniga? Kas tundetoonide esitamisel põhisõnastikus lähtuda psühholoogilisest realistlikkusest, keeleõppija vajadusest või keeletehnoloogilise rakenduse (tekstist emotsiooni tuvastamise) huvidest?

Psühholoogilistes uuringutes on näidatud, et diskreetsed emotsioonikogemused (ja -kategooriad nagu hirm, rõõm, uhkus) pole psühholoogiliselt primitiivsed. Baasilisemad on emotsiooni meeldivust ja tugevust iseloomustavad skaalad valents (hinnangulisus) ja aktivatsioon (Feldman Barrett 2004). Inglise keele kohta on sellistel eristustel põhinevaid nn afektisõnastikke ka tehtud. Üsna varsti pärast semantilise diferentsiaali meetodi leiutamist (Osgood jt 1957) olid selle meetodiga

¹ Patenteeritud USA-s “System for identifying connotative meaning, kood: US 6,389,415 B1”.

² <http://www.connotative.com/EmoDict.htm> (25.04.2011).

mõõdetud 1000 inglise keele kõige sagedasema sõna profiilid valentsi, tugevuse ja aktivatsiooni suhtes (Heise 1965). Suuremat hulka märksõnu (4000) sisaldab C. Whisselli (1989) sõnastik, milles on sõnade mõju mõõdetud seitsmepallisel skaalal meeldivushinnangute ja aktivatsiooni tugevuse suhtes.

Keeleõppija jaoks, eriti ühiskondades, milles leidub erineva etnilise ja rassilise või religioosse grupikuuluvuse taustaga isikuid, on tähtis teada võõras kultuuris sõnadega kaasneva võivaid konnotatsioone, mis on seotud pigem kultuuriliste väärtushinnangutega ja stereotüüpidega kui individuaalpsühholoogiliste eristustega³. Kakskeelsetes sõnastikes pole märksõnade tundetoonide läbivat märkimist siiski seni teadaolevalt kasutatud. Tavalisem on stilistiliselt markeeritud keelendite äärmärkimine stiilimärgenditega nagu *halvustav*, *hellitlev*, *pidulik* vmt. Kuna valdav osa keele sõnavarast on siiski tähendusvarjundi poolest neutraalne, oleks tundetooni läbiv märgendamine sõnastiku kasutajale pigem koormav lisainformatsioon, mida tal suhteliselt harva vaja läheb. Küll on sellest kasu keeleõppijale, kes oma väljendusoskust soovib lihvida.

Keeletehnoloogilistest vajadustest lähtuvates lähenemistes kirjeldatakse sõnavara afektiivseid lisatähendusi kompleksse kategooriate süsteemi abil. Näiteks nn Harvardi nimekirja⁴ iseloomustavast 182-st karakteristikust on tervelt 20 seotud emotsionaalselt relevantsete eristustega (positiivne ja negatiivne, toetav ja vägi-valdne hoiak, nõrk, tugev, võim, alistumine, valu, mõnu, tunne, emotsionaalne, erutus, voorus, pahe, rõhutamine, vähendamine). Teise näitena nn Clairvoyance'i afektisõnastikus koosneb märksõna (arv 3772) kirjeldus viiest väljast: a) lemma; b) lihtsustatud sõnaliik; c) afektiklassi määrang (peensusteni välja arendatud, 86 erinevat); d) sõna kesksuse kaal antud afektiklassi liikmena (vahemikus 0–1); e) sõna intensiivsuse kaal antud afektiklassi liikmena (0–1). Sõnastik on valdavalt koostatud ja märgendatud käsitsi, toetudes koostajate keelepädevusele (Grefenstette jt 2006: 96).

Kuivõrd põhisõnastiku tundetoonide määramise ülesanne lähtub kahest eri vajadusest, siis peab see olema: a) piisavalt informatsiooniküllane, et rahuldada keeletehnoloogilise rakenduse vajadusi; b) keeleõppijale soovi korral käepärast, ent mitte koormama teda ülearuse informatsiooniga. Nendele tingimustele näib enim vastavat psühholoogilisele realismile iseloomulik kirjelduste granulaarsuse tase, kus ei eristata konkreetseid emotsioonikategooriaid (nagu hirm, viha, rõõm, armastus jne), kuid eristatakse positiivset ja negatiivset tundetooni ning aktivatsiooni taset, s.t neutraalse ja tundetoonilise keelendi eristamist.

Põhisõnastikku koostatakse Eesti Keele Instituudi veebipõhises sõnastikusüsteemis **ΞLex**. Põhisõnastiku kasutajaliideses tuleb keeleõppija jaoks leppemärgiga (nt emotikoniga, tähtlühendiga) näidata märksõna potentsiaalne tundetooni väärtus (positiivne, negatiivne või vastuoluline), neutraalsust kui vaikeväärtust pole vaja näidata. Keeletehnoloogilise rakenduse (tekstist emotsiooni tuvastamise) tarbeks sisaldugu andmebaasis aga rohkem välju. Lisas 1 on esitatud vaade põhisõnastiku toimetamiskeskonnast, kus "semantikagrupile" järgneb "emotsiooni-grupp" ehk tundetooni määrangute tööväli. Nagu näha, on selles lahtreid rohkem kui üksnes neutraalsus vs. tundetoon (ning võimalus määrata viimase iseloomu kas positiivseks, negatiivseks või vastuoluliseks). Vajadust ülejäänud väljade järele selgitatakse artikli viimases osas.

³ <http://www.connotative.com> (25.04.2011).

⁴ <http://www.wjh.harvard.edu/~inquirer/Home.html> (16.04.2010).

3. Eesti keele põhisõnavara sõnastiku tundetoonide määramine

Määramise hetkel koosnes põhisõnastik 3015 märksõnast, puudu olid ase- ja sidesõnad. Peamiselt määrati nimisõnu, verbe, adjektiive, määrsõnu, numeraale – sõnu, mis enam-vähem mahuvad kategooria “täistähenduslikud sõnad” alla. Sõnastiku märksõnade tähendusjaotus polnud määramise faasis veel tehtud, kuid see pole määratavate tundetoonide rakendusi arvestades ka oluline, kuivõrd ei arvuti, mis teksti automaatselt analüüsima hakkab, ega eesti keelt võõrkeelena õppija, kes tekstis tundmatut sõna kohtab, ei tea sissejuhatajateks, et tegu võiks olla polüsemse sõnaga ja milline alltähendustest just antud tekstis on relevantne. Lähtuti põhimõttest, et kui sõnal on tundetoon kas või ühes alltähenduses, siis arvuti/keeleõppija jaoks on otstarbekas teda sellest hoiatada. Tundetoon on miski, mis võib, aga ei pruugi sõna kasutamisega kaasneda.

3.1. Meetod

Kasutati nn domineeriva arvamuse meetodit (Pennebaker jt 2001), selle erinevusega, et hindajaid oli neli ja domineeriva arvamuse väljaselgitamine selle tõttu pisut komplitseeritum.

Hindajad. Hindajad olid eksperdid: kõrgharidusega lingvistid, kes omavad mõningaid taustateadmisi ka emotsioonidest ja nende avaldumisest mitmesugustel keeletasanditel. Kõik olid eesti keelt emakeelena kõnelejad ja eeldati, et nad jagavad meie kultuurile vaikimisi omast väärtuste ja emotsionaalsete hinnangute süsteemi. Hindamisel soovitati juhendada mitte isiklikest emotsioonidest, vaid kultuuriomasest normatiivsest hinnangust. Ekspertide töö oli tasustatud.

Hinnangud. Kõiki hinnatavaid sõnu tuli hinnata ühel järgnevast neljast viisist: p – positiivne, n – negatiivne, o – neutraalne; p/n – kui ei olnud neutraalne, kuid ka lihtsalt positiivne ja negatiivne ei tundunud sobivat (sisaldas mingit vastuolu, võimalik, et sõna oli näiteks polüsemne vmt).

Läbiviimine. Märksõnastik esitati elektroonse andmetabeli veergudes. Igaüks hindas omaette. Koormus jaotatud ühe nädala peale, keskmiselt tehti 600 hindamist päeva kohta. Kaks hindamiste seeriat, vahemik 2 kuud, rutiini vältimiseks sõnad teisel katsel pöördjärjestuses.

Andmetöötlus. Tabelis teisedati kogutud emotsionaalsushinnangud numbrilisteks väärtusteks järgnevalt: p = 1; n = -1; n/p = 0; neutraalsete hinnangute lahtrid jäeti tühjaks. Tulemusest filtreeriti välja kokkulangevad hinnangud ja märgiti need suhtelise üksmeelsuseindeksiga (SÜ) järgnevalt: 1 – absoluutselt domineeriv arvamus, s.t kõigi nelja hindaja arvamus langes kokku; 0,75 – domineeriv arvamus (kolm ühe vastu); 0,6 – suhteliselt domineeriv arvamus (kaks kokkulangevat arvamust kahe omavahel lahkneva arvamuse vastu); 0,5 – vastuoluline arvamus (arvamused jagunevad pooleks kaks kahe vastu); 0,25 – ülivastuoluline arvamus (kõik neli on erineval arvamusel).

Tulemuste arvessevõtt. Märgendamise suhtes olulise domineeriva arvamusena (positiivne, negatiivne, neutraalne) läksid arvesse need, mille SÜ oli suurem või võrdne 0,6. Vastuoluliseks märgiti juhtumid, mille SÜ = 0,5 või vähem ning need

juhtumid, kus hindajad olid üksmeelsed tundetooni vastuolulisuse suhtes. Domineerivad arvamused selgitati kahes katses eraldi ja siis võrreldi neid omavahel.

Lõplik märgendusetepanek. Lähtuti seisukohast, et neljase valikvastuse puhul on piisav tuvastusprotsent 50% (Vogt jt 2008). Pärast kahe katse võrdlust tehti järgnev ettepanek: stabiilselt positiivne, negatiivne ja neutraalne – märkida need, mille keskmine üksmeelsuseindeks kahes katses jäi vahemikku 0,6–1; vastuoluliseks märkida need, mis a) said stabiilselt üksmeelseid hinnanguid vastuolulisuse kohta (keskmine indeks vahemikus 0,6–1) ja b) said ebastabiilseid hinnanguid (üksmeelsuseindeks väiksem kui 0,6) kahe katse ja isikute lõikes.

3.2. Tulemused ja arutelu

Tabel 1 iseloomustab domineerivalt positiivseid, negatiivseid, neutraalseid ja vastuolulisi hinnanguid saanud arvamuste jaotumist eri katsetes, arvamuste muutumist kahe katse vahel ning üldkokkuvõttes (märgendusetepanek). Numbrid iseloomustavad üldisi hinnangute proportsioone ega ütle midagi üksikute sõnade ega üksikute hindajate hinnangute stabiilsuse/muutlikkuse kohta. Tabel 1 näitab põhisõnastiku märksõnade tundetoonide distributsioonide üldist stabiilsust: laias plaanis pisut enam kui 60% sõnadest on täiesti neutraalsed ega vaja seega põhisõnastikus tundetooni märgendit. Ülejäänud 40% sõnadest jagunevad positiivse, negatiivse ja vastuolulise märgendi vahel enam-vähem võrdsetes proportsioonides. Enim muutlikkust näitab üles vastuoluliste hinnangute kategooria. See on põhjustatud asjaolust, et üks hindajatest-ekspertidest loobus teises katses võimalusest märkida tundetooni vastuolulisust. Kuivõrd see kategooria on n-ö “prügikast”, kuhu sattusid lisaks üksmeelsele vastuolu hinnangule ka need sõnad, millede osas kas katseisikute või hindamisseeariate lõikes oli enim varieeruvust, siis lõplikes märgendusetepanekutes selle ühe hindaja strateegiamuutus arvatavasti olulist mõju ei avaldanud. Küll aga osutus õigeaks otsus kasutada kolme eksperdi asemel nelja eksperti, kuna praeguse protseduuri korral kõnealune “eksperiment eksperimendis” lõpptulemusi (võrreldes esimese katse proportsioonidega) oluliselt siiski ei muutnud.

Tabel 1. Tundetoonide distributsioon hindamise eri seeriates

| Tundetoon | I hindamine | | II hindamine | | Muutus (n I – n II) | | Märgendusetepanek | |
|--------------|-------------|-------|--------------|-------|---------------------|--------|-------------------|-------|
| | n | % | n | % | n | % | n | % |
| neutraalne | 1978 | 65,61 | 2091 | 69,35 | 113 | 5,71 | 1898 | 62,95 |
| positiivne | 411 | 13,63 | 488 | 16,19 | 77 | 18,73 | 373 | 12,37 |
| negatiivne | 366 | 12,14 | 420 | 13,93 | 54 | 14,75 | 356 | 11,81 |
| vastuoluline | 260 | 8,62 | 16 | 0,53 | -244 | -93,85 | 388 | 12,87 |

Tabel 2 esitab näiteid sõnade kohta, mis said eelpool kirjeldatud domineeriva arvamuse meetodi alusel määratud kas neutraalseks, positiivseks, negatiivseks või vastuoluliseks. Iga veeru ülemine ots sisaldab sõnu, mille määrangute suhtes valitses kõige suurem üksmeel (SÜ = 1,00 tähendab seda, et kõik neli eksperti olid samal arvamusel). Arvatavasti valitseb nende sõnade suhtes tõesti kultuuriline

hoiakute konsensus, s.t positiivne on nende sõnade tundetoon, mis tähistavad meil kultuuriliselt üldaktsepteeritud positiivseid väärtusi, omadusi, tegevusi, emotsioone ning negatiivne nende sõnade tundetoon, mille denotaatide suhtes on kultuuriliselt tõrjuvad hoiakud. Positiivseks hinnatud sõnade skaala teises otsas (SÜ = 0,60, s.t suhteliselt domineeriv arvamus kaks sarnast kahe erineva vastu) leiduvad sõnad, mis ei paikne ilmselt kultuuriliselt positiivsete nähtuste kategooria tuumosas, vaid pigem perifeerias. Siin on konsensus vaid osaline. Sama kehtib negatiivsete tundetoonide skaala alaotsa kohta. Põhjused ilmselt erinevad – võib-olla sõnade poliüseemsus, või hindajate maailmavaade (nt *jumal*), suhtumine kehakaaluküsimustesse (nt *paks*). Samamoodi leiab siit sõnu, mille hinnangulisus on alles välja kujunemas (nt *neeger*).

Igal juhul annab hinnangute keskmise üksmeelsuse määr informatsiooni selle kohta, kui võrd keskne on antud sõna positiivsete ja negatiivsete tundetoonidega sõnade kategooria liikmena. See informatsioon tuleks andmebaasi kaasata keeletehnoloogilise rakenduse tarbeks (eeskujuna Grefenstette jt 2006). Lisas 1 toodud tööväljal võiks lahter “tugevus” olla koht, kus näiteks suhtelise üksmeelsuseindeksi näol iseloomustada antud hinnangute tugevust.

Neutraalsete sõnade veeru ülemine ots (SÜ = 1,00) ei tekita ilmselt kelleski küsimusi, skaala teises otsas on aga sõnad, mille neutraalsus pole sama üksmeelselt hinnatud (SÜ = 0,60). Kahe hindaja jaoks neljast iseloomustas neid sõnu siiski mingisugune tundetoon, puudus aga üksmeel. Selles osas on neutraalse kategooria ääreala sõnad võrreldavad vastuolulisi hinnanguid saanud sõnade kategooriaga. Ratsionaalselt mõeldes ei tundu olevat kuigi suurt erinevust sõnade *ärimees* (neutraalne, SÜ = 0,60) ja *jahimees* (vastuoluline, SÜ = 0,80) võimalike konnotatsioonide vahel. Kummagi puhul on ühiskonna suhtumine mingil määral ambivalentne, nende elukutsete esindajate tegevusse mahub teatud huvigruppide poolt nii moraalseks kui ebamoraalseks peetavaid tegevusi. Vastuoluliste tundetoonidega sõnade veeru alumises otsas on sõnad, mille hinnangud olid kõige ebastabiilsemad, varieerudes suuresti nii hindajate kui hindamisseeriade lõikes. Nii näiteks leiab sealt sõnad, mille tähenduses sisaldub implitsiitselt nii negatiivseid elemente (*appi*, *hoiatama* ja *hoiduma* – kõik seotud mingi hädaohuga), kuid samas ka positiivseid elemente (ohu vältimine). Vastuolulisus, mis on sees juba sõna semantikas, peegeldub ka tundetoonile antavates hinnangutes. Osa afektiivse suhtumise ambivalentsist võib tuleneda sellest, et emotsioonidel on valents vähemalt kahes suhtes: kas kogeva subjekti Ego suhtes või kohalike kultuurinormide suhtes. Esimene väljendab emotsiooni hedoonilist kvaliteeti (meeldiv vs. ebameeldiv) ja teine pigem kultuurist tingitud moraalseid hinnanguid (õilis vs. taunimisväärne). Need kaks dimensiooni ei ole alati vastavuses. (Le Guen 2009, Colombetti 2005)

3.3. Võrdlus varasemate uurimustega

Põhisõnastiku märksõnade hulgas oli 34 kokkulangevat sõna nendega, mida 100 mõlemast soost ja eriealist eestlastest katseisikut produtseerisid vastavalt kas siis positiivsete või negatiivsete emotsioonide ja isikuomaduste vaba loetlemise katsetes (vt täpsemalt Vainik 2002 ja Orav 2006). See võrreldav osa moodustab küll kõigest 1,1% põhisõnastiku mahust, kuid mingisugust tulemuste kokkulangevuse

Tabel 2. Näiteid tundetoonilistest ja neutraalsetest sõnadest

| Neutraalsed | SÜ | Positiivsed | SÜ | Negatiivsed | SÜ | Vastuolulised | SÜ |
|--------------------|-----------|---------------------|-----------|---------------------|-----------|----------------------|-----------|
| <i>aastane</i> | 1,00 | <i>aitama</i> | 1,00 | <i>alkohoolik</i> | 1,00 | <i>jahimees</i> | 0,88 |
| <i>edastama</i> | 1,00 | <i>energiline</i> | 1,00 | <i>avarii</i> | 1,00 | <i>tervis</i> | 0,88 |
| <i>eraisik</i> | 1,00 | <i>hoolima</i> | 1,00 | <i>haigus</i> | 1,00 | <i>töötasu</i> | 0,88 |
| <i>ettevõte</i> | 1,00 | <i>inimlik</i> | 1,00 | <i>hirm</i> | 1,00 | <i>armee</i> | 0,80 |
| <i>jalgratas</i> | 1,00 | <i>issi</i> | 1,00 | <i>hukkuma</i> | 1,00 | <i>efekt</i> | 0,80 |
| <i>keetma</i> | 1,00 | <i>edu</i> | 1,00 | <i>hävitama</i> | 1,00 | <i>kindlustus</i> | 0,80 |
| <i>kohver</i> | 1,00 | <i>julgus</i> | 1,00 | <i>inetu</i> | 1,00 | <i>koolilaps</i> | 0,80 |
| <i>kujunema</i> | 1,00 | <i>medal</i> | 1,00 | <i>karistama</i> | 1,00 | <i>kunst</i> | 0,80 |
| <i>osutama</i> | 1,00 | <i>lemmik</i> | 1,00 | <i>korruptsioon</i> | 1,00 | <i>kunstnik</i> | 0,80 |
| <i>porgand</i> | 1,00 | <i>mõnus</i> | 1,00 | <i>kurbus</i> | 1,00 | <i>naljakas</i> | 0,80 |
| <i>samm</i> | 1,00 | <i>pilvitu</i> | 1,00 | <i>maailmasõda</i> | 1,00 | <i>odav</i> | 0,80 |
| <i>sügis</i> | 1,00 | <i>puhkus</i> | 1,00 | <i>mõrv</i> | 1,00 | <i>pikaajaline</i> | 0,80 |
| <i>tellimus</i> | 1,00 | <i>rõõm</i> | 1,00 | <i>peavalu</i> | 1,00 | <i>pikem</i> | 0,80 |
| <i>tähendama</i> | 1,00 | <i>tervislik</i> | 1,00 | <i>pori</i> | 1,00 | <i>rebima</i> | 0,80 |
| <i>vahemaa</i> | 1,00 | <i>unistus</i> | 1,00 | <i>segane</i> | 1,00 | <i>saun</i> | 0,80 |
| <i>võrdlema</i> | 1,00 | <i>õnnelik</i> | 1,00 | <i>süü</i> | 1,00 | <i>spetsiaalne</i> | 0,80 |
| <i>ööpäev</i> | 1,00 | <i>isamaa</i> | 1,00 | <i>töötus</i> | 1,00 | <i>suurendama</i> | 0,80 |
| <i>sõlmima</i> | 1,00 | <i>eestimaalane</i> | 0,60 | <i>vaev</i> | 1,00 | <i>suvila</i> | 0,80 |
| <i>meelitama</i> | 0,60 | <i>jumal</i> | 0,60 | <i>varas</i> | 1,00 | <i>sõrmus</i> | 0,80 |
| <i>peremees</i> | 0,60 | <i>kodune</i> | 0,60 | <i>vihane</i> | 1,00 | <i>sünnikoht</i> | 0,80 |
| <i>põhjustama</i> | 0,60 | <i>kohvik</i> | 0,60 | <i>hall</i> | 0,60 | <i>tunnistus</i> | 0,80 |
| <i>raha</i> | 0,60 | <i>kuld</i> | 0,60 | <i>hapu</i> | 0,60 | <i>töömees</i> | 0,80 |
| <i>saatus</i> | 0,60 | <i>lihtne</i> | 0,60 | <i>jahe</i> | 0,60 | <i>valmis</i> | 0,80 |
| <i>teater</i> | 0,60 | <i>mõistus</i> | 0,60 | <i>külmetama</i> | 0,60 | <i>kulutama</i> | 0,55 |
| <i>tigu</i> | 0,60 | <i>neitsi</i> | 0,60 | <i>langus</i> | 0,60 | <i>perekond</i> | 0,55 |
| <i>tsirkus</i> | 0,60 | <i>parandus</i> | 0,60 | <i>loosung</i> | 0,60 | <i>romaan</i> | 0,55 |
| <i>tuli</i> | 0,60 | <i>populaarsus</i> | 0,60 | <i>neegeer</i> | 0,60 | <i>signa</i> | 0,55 |
| <i>tädi</i> | 0,60 | <i>sirge</i> | 0,60 | <i>paks</i> | 0,60 | <i>viin</i> | 0,55 |
| <i>unenägu</i> | 0,60 | <i>suutma</i> | 0,60 | <i>süütama</i> | 0,60 | <i>ülemus</i> | 0,55 |
| <i>vari</i> | 0,60 | <i>tulevane</i> | 0,60 | <i>tume</i> | 0,60 | <i>tagajärg</i> | 0,55 |
| <i>võistlus</i> | 0,60 | <i>usk</i> | 0,60 | <i>vallutama</i> | 0,60 | <i>sünnitama</i> | 0,50 |
| <i>õrn</i> | 0,60 | <i>vabandama</i> | 0,60 | <i>võim</i> | 0,60 | <i>appi</i> | 0,43 |
| <i>ärimees</i> | 0,60 | <i>vähendalik</i> | 0,60 | <i>õhuke</i> | 0,60 | <i>hoiatama</i> | 0,43 |
| <i>viitsima</i> | 0,55 | <i>üllatama</i> | 0,60 | <i>õppemaks</i> | 0,60 | <i>hoiduma</i> | 0,38 |

kontrolli saab nende põhjal siiski teha. Kolmekümnel juhul (88%) langesid loetelukatsetes moodustunud positiivse ja negatiivse hinnangu kategooriad kokku tundetoonihinnangutega käesolevas uurimuses kirjeldatud domineeriva arvamuse meetodil. Seega võib väita, et nelja eksperdi arvamused langesid nende kolmekümne nelja sõna osas valdavalt kokku 200 (kahe uurimuse peale kokku) tavakeelekasutajast katsealuse arvamusega, mida on põhjust pidada representatiivseks. Pole põhjust eeldada, et ekspertide arvamust peaks vähem usaldama ka ülejäänud märksõnastiku sõnade tundetooni hindamisel.

Mingis väga piiratud osas saab põhisõnastiku hinnatud tundetooni võrrelda ka emotsioonisõnade semantika uuringu tulemustega, milles Vainik (2004) lasi 100 (võrdselt eri soost ja eri vanuses) katseisikul sõnu hinnata kokku seitsmel polaarsest omadustest koosneval skaalal. Meetod on väga lähedane semantilise diferentsiaali meetodiga, mida muidu peetakse kõige usaldusväärsemaks sõnade hinnangulisuse mõõtmise meetodiks (vt eestpoolt). Võrrelda saab nendest tulemustest hinnanguid skaalal meeldiv-ebameeldiv, mida võib samastada emotsioonide negatiivse ja positiivse tundetooniga. Siin oli põhisõnastikuga kokkulangevaid sõnu 14 ja nende hinnangud langesid 100% ulatuses kokku ekspertide hinnangutega käesolevas uurimuses, kaasa arvatud sõnale *uhkus* iseloomulik ambivalent. Kuigi üldistustega peab olema ettevaatlik, kinnitab siiski ka see võrdlus, et käesolevas uurimuses kasutatud meetodikat ja eksperte võib pigem usaldada kui mitte.

4. Kokkuvõtte ja üldine arutelu

Kasutatud meetodi positiivse küljena võib mainida selle suhteliselt vähest ressursinõudlikkust – kulus *ca* 4 inimtöönädalat. Teiseks äratub usaldust suur kokkulangevus tulemuste osas, mis olid võrreldavad varasemate uuringutega. Kolmas hea näitaja on see, et antud meetodiga saadud määrangutulemusi on provisoorselt juba rakendatud tekstilõikude emotsionaalse valentsi automaatseks hindamiseks ja sealjuures küllaltki edukalt (Pajupuu jt, ilmumas). Osutub, et määrangud valitud granulaarsuse tasemel – neutraalsus *vs.* tundetoon ning viimase eristused positiivne, negatiivne ning vastuoluline – on piisavad keeletehnoloogilise rakenduse jaoks, mis kalkuleerib pikemate tekstilõikude emotsionaalset valentsi.

Keeletehnoloogilise lahenduse jaoks oleks veelgi selgem, kui sõnade emotsionaalsed valentsid oleks polaarsemad, s.t sõna tundetoon jaguneks selgelt positiivseks ja negatiivseks ega oleks vastuoluline, kuna viimasega on tekstilõigu emotsionaalse valentsi kalkuleerimisel vähem peale hakata. Seega võidakse tunda vajadust sõnavara “ülepolariseerida” – omistada hinnanguid “positiivne” ja “negatiivne” enam, kui see on reaalse elu hinnangute ambivalentidele mõeldes õigustatud. Ülepolariseerimist peaks autori arvates siiski kindlasti vältima, seda eriti keeleõppijast sõnastikukasutaja huve silmas pidades – talle võib eesti keel tunduda liiga komplitseeritud, kui sellega kaasnevad tugevad polaarsed hinnangud sõnade küljes, mille sisu sellist kalduvust ei eelda.

Problemaatiline on ka see, kas kirjeldatud domineeriva arvamuse meetodit oleks mõttekas rakendada märksõnastiku “mittetäistähenduslikele” sõnadele – ase- ja sidesõnadele. Kui näiteks sõna *ja* kohta võib eeldada teatud kalduvust positiivsuse poole (peaaegu homonüüm jaatussõnaga *jah*), siis seda, kas *sest* on pigem posi-

tiivne või negatiivne, oleks ekspertidel arvatavasti väga raske oma kultuurilisele ja keelelisele kompetentsile toetudes öelda. Lahendust tuleks siin otsida teistsugustest meetoditest. Näiteks võiks arvesse võtta selle sõnavara distributsiooni emotsioonide (ja nende valentsi) lõikes, mis esineb eesti keele emotsionaalse kõne korpus⁵ piisava kindlusega tuvastatud lausetes. Senine analüüs on näidanud, et näiteks kurbust väljendavates lausetes esineb tihti just oma “väikeste sõnade”, s.t ase- ja määrsõnade kasutamise muster (Vainik 2010).

Veel oleks arvatavasti keeletehnoloogilisel rakendusel (mitte aga niivõrd keeleõppijal) kasu, kui põhisõnastikus oleks lisaks tundetoonile määratud potentsiaalse afektikalduvuse klass. Afektikalduvuse all mõeldakse siin keeleüksuste võimet aidata kaasa emotsioonide (ja üldisemalt afektiivsuse) kirjeldamisele, väljendamisega, mõjutamisele ja tekstis tuvastamisele. Klass tähendab seda, et eri tüüpi sõnad annavad lause üldisesse afektiivsusesse erineva panuse olenevalt sellest, kas tegu on nn sisusõnadega, mis emotsioone kas nimetavad või väljendavad, või on tegu nn funktsionisõnadega, mille rolliks on vahendada lause üldist eksspressiivsust. Seda, et emotsioonide väljendumine keeles hõlmab mitmeid keeletasandeid ja et ka sõnavara eri gruppidel on täita eri rollid, on märgitud nt tekstilingvistikas (Biber, Finegan 1989, Ochs, Schieffelin 1989).

Esialgu võiks lähtuda viiest klassist, mis emotsioonide väljendamisse või/ja kirjeldamisse põhimõtteliselt eri viisil oma panuse annavad: a) nn võtmesõnad ehk emotsionaalse denotatsiooniga sõnad (otsene emotsionisõnavara nagu *viha, rõõm, kurbus* jne, emotsioonide väljendustegevused nagu *nutt, naer, kallistus* jne); b) nn leksikaalne külgetõmme ehk afektiivselt laetud sõnavara, sh nii tänu denotatiivsele kui konnotatiivsele hinnangule (nt omadussõnad *hea, halb, suur, kõhn, paks*, isikuomadused *laisk, tige, kade, töökas*, käitumisaktide nimetused nagu *valetamine, petmine, abistamine* jne), aga ka emotsioone tekitavad nähtused (*päike, loodus, sõda, infarkt*); c) eksspressiivne sõnavara (hüüdsõnad, nt *ah, ahah, ai, oh, ohoo, oi*), slängiväljendid, vormelid (*aitäh, appi*) jm; d) sõnavara, kus põimuvad ülalmainitud dimensioonid (nt piltlikud emotsioonikirjeldused, mis denotatsiooniga osutavad emotsiooni läbielamise protsessile, on aga ka eksspressiivsed (s.t kujundlikud, väljendusjõulised) ja lisaks tihti hinnangulised (nt *vihast vahutama, moka tõllakil*). Sedasorti leksikat põhisõnavara sõnastik praegu aga ei sisalda. Eraldi afektikallaku klass e) on intensiivsust, evidentsiaalsust, tõenäosusmodaalsust, deiktilisust jmt kõneleja ja kõnesündmuse sisu vahekorda esile toovad sõnad (*arvatavasti, väga, nii, võib-olla* jne). Lisas 1 esitatud väljale “afektikalduvuse klass” on mõeldud sisestada just sedasorti informatsiooni.

Afektikalduvuse näol on tegu hajusate piiridega ja koostiselt heterogeense kategooriaga, mida määratlev printsiip on funktsionaalne. See on hõlmavam, kui käesolevas artiklis käsitletud sõnavara tundetoonilisus ehk emotsionaalne valents. Afektikalduvuse arvessevõtt tundub perspektiivikas selles mõttes, et kuigi 75% jooksvast ajalehetekstist koosneb umbes 4000 sagedasemast sõnast (Pajupuu jt 2009), ei ole praegu veel selge, kui suur protsent eesti keele lausete n-ö “jooksvatest sõnadest” on täistähenduslikud ja kui suur on nende seas afektirehkendust võimaldavate mitteneutraalsete sõnade arv. Afektikalduvuse klasside arvessevõtt aitaks tuvastada ka neid lauseid, milles väljendub ainult emotsionaalne eksspressiivsus ja mitte ilmingimata hinnangulisus.

⁵ <http://peeter.eki.ee/5000/> (25.04.2011).

Problemaatiline on afektikalduvuse märkimise puhul aga asjaolu, et informatsiooni selle kohta saadaks sõnade tegelikust kasutamisest eesti emotsionaalse kõne korpuses ja seega on need andmed metoodiliselt teist moodi kogutud ning nad erinevad ka teoreetilise lähte-eelduse poolest selle kohta, kas keelt tuleb uurida reaalses kasutuses või siis pigem potentsiaalina inimeste peades (vt ptk 1.1). Selleks, et erinevalt kogutud ja määratud informatsioon oleks andmebaasis ka erinevalt talletatud, on põhisõnastiku “emotsioonigrupi” toimetamisvaates (vt lisa 1) ette nähtud lahter “määramise alus”. Sinna tuleb siis märkida alus, mille põhjal on otsus langetatud. Seega, kuigi on tehtud valik üht teatud tüüpi informatsiooni eelistamise ees, ei ole välistatud ka võimalus edaspidi toetuda teistsugustele meetoditele (vt ptk 1.2).

Lõpetuseks võiks öelda, et sissejuhatuses mainitud kahe praktilise ülesande lahendamine samaaegselt on kahtlemata võimalik, kuid tundetoonide määramist ja märkimist see ikkagi mõnevõrra komplitseerib. Teha tuleb valikuid, mis ei tundu kummagi ülesande seisukohast just ideaalsed. Lahenduseks on siin see, et sõnastiku andmebaasi väljadel sisaldub rohkem ja detailsemat infot, kui seda kuvatakse keeleõppijast kasutajale.

Viidatud kirjandus

- Aavik, Toivo; Allik, Jüri 2002. The structure of Estonian personal values: A lexical approach. – *European Journal of Personality*, 16 (3), 221–235. <http://dx.doi.org/10.1002/per.439>
- Aitchison, Jean 2003. *Words in the Mind: An Introduction to the Mental Lexicon* (3rd ed.). Blackwell Publishing.
- Altrov, Rene; Pajupuu, Hille 2008. The Estonian emotional speech corpus: Release 1. – *The Third Baltic Conference on Human Language Technologies*. Vilnius: Vytauto Didžiojo Universitetas; Lietuvių kalbos institutas, 9–15.
- Atkins, B. T. Sue; Rundell, Michael 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Barsalou, Laurence W. 2003. Situated simulation in the human conceptual system. – *Language and Cognitive Process*, 18 (5–6), 513–562. <http://dx.doi.org/10.1080/01690960344000026>
- Bestgen, Yves 1994. Can emotional valence in stories be determined from words? – *Cognition and Emotion*, 8 (1), 21–36. <http://dx.doi.org/10.1080/02699939408408926>
- Biber, Douglas; Finegan, Edward 1989. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. – *Text*, 9 (1), 93–124. <http://dx.doi.org/10.1515/text.1.1989.9.1.93>
- Bower, Gordon H.; Forgas, Joseph P. 2000. Affect, memory and social cognition. – Eric Eich, John F. Kihlström, Gordon Bower, Joseph P. Forgas, Paula M. Niedenthal (Eds.). *Cognition and Emotion*. Oxford: Oxford University Press, 87–168.
- Colombetti, Giovanna 2005. Appraising valence. – *Journal of Consciousness Studies*, 12 (8–10), 103–126.
- Cruse, Alan D. 2000. *Meaning in Language. An Introduction to Semantics and Pragmatics*. Oxford: Oxford University Press.
- Evans, Vyvyan; Green, Melanie 2006. *Cognitive Linguistics: An Introduction*. London: Lawrence Erlbaum.
- Feldman Barrett, Linda 2004. Feelings or words? Understanding the content in self-report ratings of experienced emotion. – *Journal of Personality and Social Psychology*, 87 (2), 266–281. <http://dx.doi.org/10.1037/0022-3514.87.2.266>

- Feldman Barrett, Linda 2006. Solving the emotion paradox: Categorization and the experience of emotion. – *Personality and Social Psychology Review*, 10 (1), 20–46. http://dx.doi.org/10.1207/s15327957pspr1001_2
- Foolen, Ad 1997. The expressive function of language: Towards a cognitive semantic approach. – Susanne Niemeier, Rene Dirven (Eds.). *The Language of Emotions: Conceptualization, Expression and Theoretical Foundation*. Amsterdam, Philadelphia: John Benjamins Publishing Company, 15–32.
- Gallese, Vittorio; Lakoff, George 2005. The brain's concepts: The role of the sensory-motor system in conceptual knowledge. – *Cognitive Neuropsychology*, 22 (3-4), 455–479. <http://dx.doi.org/10.1080/02643290442000310>
- Grefenstette, Gregory; Qu, Yan; Evans, David A.; Shanahan, James G. 2006. Validating the coverage of lexical resources for affect analysis and automatically classifying new words along semantic axes. – James G. Shanahan, Yan Qu, Janyce M. Wiebe (Eds.). *Computing Attitude and Affect in Text: Theory and Applications*. Dordrecht: Springer, 93–107.
- Heise, David R. 1965. Semantic differential profiles for 1000 most frequent English words. *Psychological Monographs*, 79 (601). American Psychological Association.
- Heise, David R. 1970. The semantic differential and attitude research. – Gene F. Summers (Ed.). *Attitude Measurement*. Chicago: Rand McNally, 235–253.
- Huumo, Tuomas 2008. Joko lingvistin nojatuoli joutaisi kaatopaikalle? Introspektio-lingvistiikan asemasta kognitiivisessa kielitiedessä. – *Emakeele Seltsi aastaraamat*, 53 (2007), 163–180.
- Hübler, Axel 1998. *The Expressivity of Grammar. Grammatical Devices Expressing Emotion Across Time*. Berlin, New York: Mouton de Gruyter.
- Iida, Akemi; Campbell, Nick; Higuchi, Fumito; Yasumura, Michiaki 2003. A corpus-based speech synthesis system with emotion. – *Speech Communication*, 40 (1-2), 161–187. [http://dx.doi.org/10.1016/S0167-6393\(02\)00081-X](http://dx.doi.org/10.1016/S0167-6393(02)00081-X)
- Kallas, Jelena; Tuulik, Maria 2011. Eesti keele põhisõnavara sõnastik: ajalooline kontekst ja koostamispõhimõtted. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 7, 59–76. <http://dx.doi.org/10.5128/ERYa7.04>
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar I. Theoretical Prerequisites*. Stanford: Stanford University Press.
- Le Guen, Olivier 2009. The ethnography of emotions: A field worker's guide. – Asifa Majid (Ed.). *Field Manual*, Vol. 12. Nijmegen: Max Planck Institute for Psycholinguistics, 31–34.
- Ochs, Elinor; Schieffelin, Bambi 1989. Language has a heart. – *Text*, 9 (1), 7–25. <http://dx.doi.org/10.1515/text.1.1989.9.1.7>
- Orav, Heili 2006. *Isiksuseomaduste sõnavara semantika eesti keeles*. *Dissertationes Linguisticae Universitatis Tartuensis*, 6. Tartu: Tartu Ülikooli kirjastus.
- Osgood, Charles; Suci, George. J.; Tannenbaum, Percy. H. 1975 [1957]. *The Measurement of Meaning*. Urbana, Chicago: University of Illinois Press.
- Pajupuu, Hille 1994. A nation's autostereotype. – *Oral Memory & National Identity. Papers of the International Conference held in Tallinn, September 18–19, 1993*. Tallinn, 31–39.
- Pajupuu, Hille; Kerge, Krista; Alp, Pilvi 2009. Sõnavara loomulik rikkus haritud keeleoskaja tekstides. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 5, 187–196. <http://dx.doi.org/10.5128/ERYa5.12>
- Pajupuu, Hille; Kerge, Krista; Altrov, Rene (ilmumas). Detecting emotional valence of text by using a small dictionary. – *Empiricism and Analytical Tools for Applied Linguistics in the 21st Century*. Salamanca: Universidad de Salamanca.
- Pavlenko, Aneta 2008a. Emotion and emotion-laden words in the bilingual lexicon. – *Bilingualism: Language and Cognition*, 11 (2), 147–164. <http://dx.doi.org/10.1017/S1366728908003283>

- Pavlenko, Aneta 2008b. Structural and conceptual equivalence in the acquisition and use of emotion words in a second language. – *The Mental Lexicon*, 3 (1), 91–120. <http://dx.doi.org/10.1075/ml.3.1.07pav>
- Pennebaker, James W.; Francis, Martha; Booth, Roger J. 2001. *Linguistic Inquiry and Word Count*. Mahwah: Erlbaum Publishers
- Sapir, Edward 2009. *Keel. Sissejuhatus kõne uurimisse*. Tallinn: Eesti Keele Sihtasutus.
- Skrandies, Wolfgang; Chiu, Ming-Jang 2003. Dimensions of affective semantic meaning – behavioral and evoked potential correlates in Chinese subject. – *Neuroscience Letters*, 341 (1), 45–48. [http://dx.doi.org/10.1016/S0304-3940\(03\)00137-X](http://dx.doi.org/10.1016/S0304-3940(03)00137-X)
- Sutrop, Urmas 2001. List task and a cognitive salience index. – *Field Methods*, 13 (3), 263–276. <http://dx.doi.org/10.1177/1525822X0101300303>
- Vainik, Ene 2010. Kuidas õpetada kõnesüntesaatorile empaatiat? Emotsiooni automaatse tuvastuse võimalustest eestikeelses kirjalikus lauses sisalduva info põhjal. – *Eesti Rakenduslingvistika Ühingu aastaraamat*, 6, 327–347. <http://dx.doi.org/10.5128/ERYa6.20>
- Vainik, Ene 2004. *Lexical Knowledge of Emotions: The Structure, Variability and Semantics of the Estonian Emotion Vocabulary*. *Dissertationes Linguisticae Universitatis Tartuensis*, 5. Tartu: Tartu Ülikooli kirjastus.
- Vainik, Ene 2002. Millest on tehtud eestlaste emotsioonisõnavara? – *Keel ja Kirjandus*, 8, 537–553.
- Valitutti, Alessandro; Strapparava, Carlo; Stock, Oliviero 2004. Developing Affective Lexical Resources. – *PsychNology Journal*, 2 (1), 61–83.
- Ververidis, Dimitrios; Kotropoulos, Constantine 2006. Emotional speech recognition: Resources, features, and methods. – *Speech Communication*, 48 (9), 1162–1181. <http://dx.doi.org/10.1016/j.specom.2006.04.003>
- Vogt, Thurid; Andre, Elisabeth; Wagner, Johannes 2008. Automatic recognition of emotions from speech: A review of the literature and recommendations for practical realisation. – Christian Peter, Russell Beale (Eds.). *Affect and Emotion in Human-Computer Interaction*. LNCS, 4868. Heidelberg: Springer, 75–91. http://dx.doi.org/10.1007/978-3-540-85099-1_7
- Voll, Piret 2002. Tähenähtuse konnotatiivne aspekt ja selle kajastamine ükskeelses sõnaraamatus. Käsikirjaline magistritöö Tartu Ülikoolis.
- Whissell, Cynthia M. 1989. *The dictionary of affect in language*. – Robert Plutchik, Henry Kellerman (Eds.). *Emotion: Theory, Research, and Experience*. New York: Academic Press, 113–131.

Võrgumaterjalid

- Eesti emotsionaalse kõne korpus <http://peeter.eki.ee:5000/> (25.04.2011).
- Firma Connotative Intelligence koduleht <http://www.connotative.com> (25.04.2011).
- Kontentanalüüsi hõlbustava veebivara General Inquirer koduleht <http://www.wjh.harvard.edu/~inquirer/Home.html> (16.04.2010).

Ene Vainiku (Eesti Keele Instituut) uurimisobjekt on eesti keele emotsioonisõnavara, selle struktuur, semantika ja varieerumine. Huvipakkuvateks teemadeks on ka psühholingvistika, kognitiivne semantika ja keele ning emotsioonide interaktsioon laiemalt.
Ene.Vainik@eki.ee

Lisa 1. Vaade EKI eesti keele põhisõnavara sõnastiku toimetamisliidese alale "Emotsioonigrupp"

Kõide: 1. koide A - Y

märksõna üllatus

Rada [üllatus*]: c:A[1]/c:S[1]/c:tp[1]/c:emg[1]

Toimetamisala XML Tabelina Vaatena

| | |
|-----------------------|---------------------|
| sisu | |
| eelnev ühisinfo | |
| tähendusnumbri plokk | 1 |
| grammatikagrupp | |
| semantika grupp | |
| emotsioonigrupp | |
| tundetoon | positiivne |
| määramise alus | 4E1 |
| afektikalduvuse klass | |
| tugevus | 0.75 |
| kuupäev | 2011-11-26T15:19:29 |
| tähendusgrupp | |
| semantilised suhted | |

Vaade

üllatus

♪ <üllatus, üllatuse, üllatust, üllatusse;
mitmus: üllatused, üllatuste, üllatusi 11 S>

DETECTING EMOTIONAL VALENCIES FOR THE ESTONIAN VOCABULARY

Ene Vainik

Institute of the Estonian Language

The need for detection of the emotional valencies of the Estonian vocabulary originates in two practical applications. First, there is a need for a device capable of automatic detection of emotion based on the information present in written Estonian in order to model the naturalness of speech synthesis. Such a device can rely on a support dictionary that covers the most frequent and central part of the vocabulary. Second, in the course of compiling the Estonian Basic Dictionary (around 4000 entries), there is (among other layers of lexicographic description) also the task of informing users of the dictionary about the potential emotional valence of the entries. Both practical approaches are based on an assumption that emotional meanings and/or connotations are not only subjective phenomena but are at least partly shared and conventional. The idea of the study is to use the lexical resources of the Basic Dictionary as the support dictionary for the purposes of automatic emotion detection.

The present article tackles the theoretical background and accounts for the selection of an optimal method for detecting the emotional valencies of words. The theoretical approaches were found to differ mostly in whether they take the emotional properties of words as part of the structure of language (its usage as a fact) or as a part of the mental structure of the speaker (as a potential). In the present study the second approach is followed. The method suggested for the detection of emotional values – the method of predominant opinion – relies on the speaker's evaluations.

The selected method is tested in the case study, and the results and further prospects are discussed in the final part of the article. The procedure of evaluation was carried out in two series (with an interval of two months) and using four independent experts. An agreement rate stronger than 0.6 was set as a threshold for tagging the items in the database as neutral, positive, negative or ambivalent (average in the two series of evaluation tasks). The level of granularity of the evaluations was based on psychological reliability and the rationale of giving enough information for the process of automatic emotion detection but not too much for the human user of the dictionary. The method used in the case study gave quite reliable results as compared with some previous studies into the vocabulary of emotion and personal traits. However, to account better for the potential of words to contribute to the expression, description or infliction of emotion, the evaluative dimension may not suffice. Therefore, a more detailed classification of lexical affect bias is tentatively suggested at the end of the paper. The practical application of the classification should be proved or disproved by further studies

Keywords: emotional valence, method, vocabulary, affect biasedness, basic dictionary, Estonian