

TARTU ÜLIKOOL
Arvutiteaduse instituut
Informaatika õppekava

Sabiina Haiba
KUIDAS ON NETIKEEL MUUTUNUD
AASTATEL 2001–2008?
Bakalaureusetöö (9 EAP)

Juhendaja: vanemteadur Heiki-Jaan Kaalep

Tartu 2016

Kuidas on netikeel muutunud aastatel 2001–2008?

Lühikokkuvõte:

Käesoleva töö eesmärk on koostada sagedusloendeid perioodi 2001–2008 netikeele korpustest ning leida erinevused keelekasutuses. Korpuseks võeti Planet Foorumitest pärinevad tekstid. Töö käigus loodi arvutiprogramm, mille abil võeti korpuse tekstidest vajaminev informatsioon ning moodustati sagedusloendid. Sagedusloendite võrdluse teel uuriti muutusi täheortograafias, lühendite ning emotikonide kasutuses.

Töö käigus selgitati välja, et üldisemate järelduste tegemiseks eesti netikeele kohta oleks vaja võtta korpuseks laiema kasutajaskonnaga suhtlusplatvorm. Käesolevas töös tehtud järeldused on Planet Foorumite põhised. Uurimusest selgus, et foorumi kasutajad on hakanud järjest rohkem kirjutama korrektsemas eesti keeles ning mõnevõrra enam järgima keelereegleid.

Võtmesõnad:

Netikeel, emotikon, foorum, korpus

CERCS: P175, Informaatika, süsteemiteooria

How has the Internet language changed from 2001 to 2008?

Abstract:

The purpose of this research is to find changes in the use of language on the Internet between the years 2001 – 2008. This is accomplished by analyzing posts from the Planet Forums. A program was written to parse the forum posts, extract all the necessary information and analyze language use with word frequency lists. These frequency lists are used to study changes in spelling, abbreviations and emoticon usage.

It became apparent during the research that a larger sample size would be required in order to make generalizations about the Estonian language on the Internet. The generalizations made in this paper are based only on the Planet Forums. It was found that the users of the forums have gradually shifted towards writing more correct Estonian.

Keywords:

Internet language, emoticon, forum, corpus

CERCS: P175, Informatics, systems theory

SISUKORD

SISSEJUHATUS	4
1. NETIKEEL	5
2. KORPUS	6
3. UURIMISMEETOD	8
4. KASUTAJATE JA POSTITUSTE DÜNAAMIKA	11
5. KEELETUNNUSTE DÜNAAMIKA	16
5.1. SUURE JA VÄIKESE ALGUSTÄHE KASUTUS	16
5.2. EMOTIKONIDE KASUTUS	18
5.3. TÄHEKORDUSED	21
5.4. VÕÕRTÄHTEDE KASUTUS	23
5.5. LÜHENDITE KASUTUS	25
6. UURIMUSE JÄRELDUSED	28
7. EDASISED TÖÖSUUNAD	30
8. KOKKUVÕTE	31
VIIDATUD ALLIKAD	32
LISA I	34
LITSENTS	34

Sissejuhatus

Eestikeelset internetikeelt on väga vähe uuritud. Tiit Hennoste ja Merike Tederi [1] väitel on sellel kolm põhjust: netikeel muutub kiiresti, materjali on liiga palju ja suur hulk tekste ei ole avalikud. Väidet, et netikeel muutub kiiresti, peetakse justkui iseenesestmõistetavaks, kuigi statistikat selle kohta ei ole. Käesolevas töös püütaksegi välja selgitada, kuivõrd see väide tegelikult paika peab, arvestades kolme aspekti: täheortograafiat, metagraafikat ning lühendeid.

Käesoleva bakalaureusetöö eesmärk on koostada sagedusloendid ühe kindla ajavahemiku tekstikorpusest ning leida selle põhjal erinevused netikeele kasutuses. Sagedusloendite võrdluse teel uuritakse muutusi täheortograafias ning metagraafikas ning tehakse kindlaks, kas varem kasutatud lühendid on kadumas või on tekkinud uusi.

Bakalaureusetöö koosneb kuuest peatükist ning kokkuvõttest. Esimeses peatükis antakse ülevaade internetikeelest: mõistest, funktsionaalsusest, võrdlusest tavakeelega ning netikeele omapäradest, näiteks emotikonide olemasolust. Teises peatükis kirjeldatakse uuritavat korpust, milleks on eestikeelne internetikeskkond Planet Foorumid [2]. Töö empiiriline osa koosneb programmi kirjutamisest, korpuste ehk elektrooniliste tekstikogude andmete töötlemisest ning andmete põhjal sagedusloendite koostamisest. Sellest annab ülevaate kolmas peatükk. Neljas peatükk sisaldab sagedusloenditel põhinevat ülevaadet Planet Foorumite kasutajate ning postituste dünaamikast, mille põhjal vastatakse kahele küsimusele: kas selle korpuse järgi saab hinnata keelekasutuse muutumist üldiselt ning kas korpus esindab paljude või ainult väheste inimeste keelekasutust. Viiendas peatükis kirjeldatakse foorumi kindlate keeletunnuste (metagraafika, lühendite, täheortograafia) dünaamikat. Kuuendas peatükis tehakse kahe eelneva põhjal uurimise järeldused. Kuna teema on palju laiem, on see analüüs loodetavasti heaks aluseks edasistele eestikeelset netisuhtlust käsitlevatele uurimustele. Sellest, millistele probleemidele ja kuidas võiks edaspidi lahendusi leida, räägitakse seitsmendas peatükis „Edasised töösuunad“.

1. NETIKEEL

Internetikeskkondades kasutatakse keelt, mida on hakatud nimetama netikeeleks ja mis ei vasta üldjuhul õigekeelsusreeglitele ning sarnaneb pigem kõnekeelele. Netikeele kohta puuduvad kindlad kirjapandud juhised, mistõttu igaüks võib kirjutada, nagu soovib, ning teatud tavad kujunevad üksteise eeskuju järgides. Nendest tavadest kujunevad omamoodi reeglid, mis võimaldavad internetis suhtlejal üksteisest küllaltki hästi aru saada isegi siis, kui nad räägivad erinevaid keeli.

Anni Oja [3] sõnul on oluline mõista, et internetisuhtluse puhul on ennekõike tähtis kiirus. Ta lisab, et kiiresti kirjutamiseks optimeeritakse keelt ning selle tõttu ei pöörata tähelepanu kirjavahemärkidele, suurele ja väiksele algustähele ning kasutatakse mitmesuguseid fraaside lühendusi (nt *LOL* – *Laughing Out Loud*, *mb* – *my bad*). Lühendite kasutamine muudab internetis suhtlemist mugavamaks ja kiiremaks ning mitmed sellised lühendid on nii laialt levinud, et on lisatud isegi sõnastikesse. Näiteks lühend *LOL* on olemas „The Oxford English Dictionary’s“ [4]. Tiit Hennoste jt [1] on toonud mitmeid näiteid, kuidas lühendatakse tavalisi eestikeelseid sõnu, jättes ära vokaalid (nt *vbnđst* – vabandust) või lihtsustades kaashäälikuühendeid (nt *ks – x*, *ts – c*). Töös vaadeldakse selliste lühendite kasutamise arengut: kas leidub lühendeid, mida on kasutatud kogu vaadeldava perioodi jooksul ning kas lühendite puhul on märgata muudatusi nende pikkuses ja kirjapildis.

Emotikonid pakuvad keeleuurijatele suurt huvi, sest peaaegu kõik tänapäeval internetis suhtlejad kasutavad neid oma tekstides igapäevaselt [6]. Kirjavahemärkidest koosnevad piltmärgid ehk emotikonid annavad internetisuhtluses edasi emotsioone, mida suulises kõnes väljendatakse näoilmetega (nt kurva meele puhul „:(“). Lisaks sellele väljendatakse emotikonidega tihti ka sarkasmi [7] (nt „Sa ei meeldi mulle üldse :)“). Peale emotsioonide väljendamise on veel emotikone tegevuste ning asjade kujutamiseks (nt „:*“ on suudlus ning „<3“ on süda). Häälightsusi ja hääle tugevust näidatakse suurtähelise kirjaga, mitme hüüumärgi või tähekordustega (nt *OOOOOMMMGGGG!!!!*). Töös uuritaksegi muu hulgas, kui palju on aastate jooksul muutunud emotikonide kasutus: kas mõned emotikonid on kasutuselt kadunud ja kas juurde on tulnud uusi ning kuidas on muutunud nende esinemiste arv internetisuhtluses. Tähekorduste puhul uuritakse, kas seda viisi emotsioonide näitamiseks kasutatakse rohkem või vähem.

2. KORPUS

Käesoleva bakalaureusetöö jaoks vajaliku keelematerjali leidmine oli äärmiselt tähtis, sest sellest peab saama piisavalt põhjaliku ning ülevaatliku andmekogu, et selle põhjal järeldusi teha. Lisaks peavad analüüsiks olema kättesaadavad nii vanemad kui ka uuemad tekstid. Samuti peab postituste juures olema postitaja kasutajanimi, selleks et uurida ühe autori sõnakasutuse või mõne keeletunnuse muutumist eri aegade tekstides. Selleks et võrrelda tekstide muutumist ajas, peavad nad olema dateeritud.

Siinses töös kaaluti järgmisi internetipõhiseid suhtlusplatvorme.

1. Foorumid – veebikeskkonnad, milles kasutajad saavad vahetada infot ning arutleda mingitel kindlatel teemadel. Foorumis toimub suhtlus asünkroonselt, kasutajad saavad hiljem varasemaid postitusi lugeda. Üldiselt on kasutajate valida, millist keelt nad kasutavad, kuid mõnes foorumis on sätestatud kindlad reeglid, mille järgimist kontrollivad moderaatorid.
2. Jututoad – suhtlusplatvormid, milles saavad omavahel sünkroonselt ehk samal ajal suhelda mitmed kasutajad. Nagu foorumiteski, on tihti ka jututubades moderaatorid.
3. Kommentaariumid – veebikeskkonnad, mis annavad kasutajatele võimaluse reageerida uudistele või artiklitele ning jagada teistega oma arvamust. Enamikus kommentaariumides on võimalik jääda anonüümseks. Anni Oja [3] väitel on see üks põhjusi, miks ollakse seal postitades julgemad.

Jututoad ei sobi korpuseks seetõttu, et suur osa sealsest suhtlusest toimub privaatsest, ning kommentaariumid seetõttu, et võimaldavad kasutajatel jääda anonüümseks, mistõttu ei ole võimalik ühe kindla kasutaja kohta statistikat luua. Käesoleva uuringu jaoks on parimaks korpuse valikuks foorum, kus enne postitamist nõutakse konto loomist. Foorumi puhul pääseb ligi ka vanematele postitusele ning iga postitus on märgistatud aastaarvuga, mil postitus on tehtud. Lisaks on iga postituse juures selle kirjutaja kasutajanimi. Kuna varasemalt on juba kogutud foorumi Planet Foorumid 2001.–2008. aasta tekstid korpuseks¹, mis koosneb 8,7 miljonist sõnast (kirjavahemärke ei ole arvestatud), siis kasutatakse seda.

Planet Foorumid sobib analüüsi üldkogumiks hästi, sest see ei spetsialiseeru kindlale alale, vaid sisaldab teemasid seinast seinale, k.a näiteks haridus ja sport. Lisaks on seal mitmeid alafoorumeid, igapäevase jutu ajamiseks on seal näiteks alafoorum „Päeva(vargad) ja (t)öö inimesed“. Foorumi kasutajate vanuse kohta info puudub, kuna kasutajatelt pole nõutud nende tegeliku vanuse avaldamist ning arvatavasti ei oleks see info ka usaldusväärne, sest internetis suheldes valetatakse tihti enda vanuse kohta. Sama kehtib ka soo kohta.

Muidugi ei ole kogu eestikeelne netisuhtlus koondunud foorumisse, kuid sealt saadava informatsiooni põhjal on foorum käesoleva bakalaureusetöö eesmärgi täitmiseks parim. Planet Foorumites on hulk kasutajaid, kes on olnud aktiivsed kogu vaadeldava perioodi jooksul (2001–2008), seega saab nende postitatu põhjal analüüsida, kas ja kui palju on nende


























¹ Internetilehelt

http://www.cl.ut.ee/korpused/segakorpus/uusmeedia/foorumid_uudisgrupid_kommentaariid.php?lang=e_t laadida alla failikaust „Foorumid, uudisgrupid ja kommentaarid eemaldatud kordustega“. Korpuse failide leidmiseks avada failikaust „xml“ ja seejärel kaust „foorumid“ [5].

kirjakeeles muudatusi toimunud. Korpus sisaldab ka emotikone. Postitusele võib lisada emotikone foorumi poolt etteantud järjendist. Sellised emotikonid on korpuses asendatud seda kirjeldava nimetusega. Tabelites 1 ja 2 on toodud kõik valitud foorumis olevad emotikonid, millel on pildiga vaste.

Liikuvad emotikonid			
smash		armastus	
spineyes		rtfm	
lol		cry	
ingel			

Tabel 1. Planet Foorumite liikuvate emotikonide nimetused koos vastavate piltidega

Liikumatud emotikonid			
lambike		h2bi	:o
naerab		kartlik	
evilsmile		happy	
r66mus		uhoh	
teeb_silma		evil	
n2itab_keelt		jahe	
n2itab_p6ialt		nool	
vihane		naeratus	
segaduses		tähelepanu	
kurb		kahtlane	
punastab		kysimark	?
küsimus		pöial üles	
vaatab_yles		õnetu	
pöial alla			

Tabel 2. Planet Foorumite liikumatute emotikonide nimetused koos vastavate piltidega

Lisaks sellele leidub korpuses küllaldaselt emotikone, millel ei ole pildiga vastet (nt „=))))“ , „;o“). Käesolevas töös uuritakse ainult neid emotikone, mis on kogutud korpuses märgistatud emotikoni sildiga. Need, mida korpusse koostaja ei ole tuvastanud, jäetakse välja.

3. UURIMISMEETOD

Suurte koguste andmete töötlemiseks on otstarbekas kasutada arvutiprogrammi [8]. Kuna käesolevas töös on vaja sagedusloendeid moodustada korpustest, milles on 8,6 miljonit sõna, siis on mõttekas seda teha arvutiprogrammi abil. Programmeerimiskeeleks on valitud Python, kuna see ei sea kasutajale piiranguid, mida mitmed muud programmeerimiskeeled teevad.

Kõik töös esitatud joonised ja tabelid on koostatud arvutiprogrammi abil korpuse XML-i failidest saadud andmete põhjal. Käesolevaks tööks loodud arvutiprogramm asub veebilehel <https://bitbucket.org/sabiinah/netikeel/src>. Kõik teegid, mida programmis on kasutatud, on loetletud failis „requirements.txt“.

Selleks, et koostada sagedusloendit, tuleb kõigepealt korpuse tekst arvutiprogrammi abil korrastada ning esitada vajalikul kujul. Algne tekst sisaldab XML-i (*Extensible Markup Language*) märgendeid (vt joonis 1).

```
<div3 type="postitus"> <p <name type="postitaja">Vilbiks</name> </p>
<p><time> 17-08-2008 , 13:30 </time></p>
<p> <s> <gap rend="armastus_emotikon"/>Tre Piu-oled ikka kena kasutaja nime leidnud ? </s> </p>
<p> <s> Endast lähemalt paluks ? </s> </p> </div3>
```

Joonis 1. Korpusest võetud originaalne foorumi postitus

Järgnevalt on toodud korpuses ning joonisel 1 olevate XML-i märgendite tähendused:

nurksulud (<, >) eraldavad elemente;
</> on elemendi lõpu märgend;
<div3> [postitus] </div3>;
<p> [postituse paragrahv] </p>;
<s> [lause] </s>;
<time> [postituse kuupäev ja kellaaeg] </time>;
<name> [kasutajanimi] </name>;
<gap /> märgib emotikoni, juhul kui märgendil on *rend*-atribuut.

Nagu näha, on korpuse tekstis juba eraldi ära märgitud vajaminev informatsioon: aeg, kasutajanimi, emotikonid ning laused ja paragrahvid on üksteisest eraldatud. Eraldi võetakse välja veel iga lause puhul esimene sõna. Selleks, et teksti põhjal loendeid moodustada, muudetakse tekst Pythoni objektiks, nagu võib näha joonisel 2.

```
autor: 'Vilbiks'
aeg: '2008-08-17 13:30:00'
laused: '['Tre Piu-oled ikka kena kasutaja nime leidnud ?', 'Endast lähemalt paluks ?']'
lause esimesed sõnad: '['Tre', 'Endast']'
emotikonid: '['armastus_emotikon']'
```

Joonis 2. Korpusest võetud postitus, mis on muudetud Pythoni objektiks

Järgnevalt on toodud sammud, mida programm korpuse teksti puhastamiseks teeb.

1. XML-i märgendite abil võetakse korpusest välja vajalik info: autor, aeg, laused, lause esimesed sõnad ning emotikonid.
2. Laused, lause esimesed sõnad ning emotikonid esitatakse loendina, kuna Pythoniga on neid lihtne töödelda.

3. XML-i märgendid eemaldatakse.
4. Postitused, millel puudub aasta (selle asemel on näiteks märgend „Eile“), jäetakse välja, sest ilma aastaarvuta postitused on analüüsiks kõlbmatud.

Pärast seda, kui vajalik informatsioon on kategoriseeritud ning järjenditesse lisatud, on lihtne lisada programmi kood (vt joonis 3), mis väljastab vajaminevaid loendeid.

```
def get_emoticon_frequencies(self):
    result = {}
    for year, posts in self.posts_by_year.items():
        result[year] = Counter(emoticon for post in posts for emoticon in post.emoticons)
    return result

def draw_emoticon_graph(self):
    top_emoticons = {}
    for year, emoticons in self.emoticon_frequencies.items():
        for emoticon, count in emoticons.most_common(3):
            top_emoticons[emoticon] = []
    for year, counter in self.emoticon_frequencies.items():
        for emoticon in top_emoticons.keys():
            top_emoticons[emoticon].append(counter[emoticon])
    line_chart = pygal.Line()
    line_chart.title = 'Aastas enimkasutatud emotikonid'
    line_chart.x_labels = map(str, range(2001, 2008))
    for emoticon, series in top_emoticons.items():
        line_chart.add(emoticon, series)
    line_chart.render_to_png("emotikonid.png")
```

Joonis 3. Kolme sagedaima emotikoni leidmise ning graafi moodustamise Pythoni kood

Joonisel 3 on koodiosa, mis leiab kõigepealt funktsiooniga `get_emoticon_frequencies()` iga aasta kohta, milliseid ja kui palju erinevaid emotikone on foorumis kasutatud. Funktsioon `draw_emoticon_graph()` leiab, millised olid igal aastal kolm kõige tihedamini kasutatud emotikoni. Selles meetodis on kasutatud ka PyGali², mis on Pythoni teek graafide joonistamiseks.

Sisukate järelduste tegemiseks on vajalik leida, milliseid aspekte uurimuses arvesse võtta. Seega alustatakse tööd erinevate andmete kogumisest ning sagedusloenditesse korrastamisest, et tekiks ülevaade, milliste andmetega on mõtet tööd jätkata ning millistega mitte. Parim viis selleks on koguda andmed Exceli tabelitesse, sest seal on võimalik neid kergelt sorteerida. Esimese ülevaate annab tabel 3, kus on näidatud, kui palju postitusi on igal aastal foorumis olnud.

Tabel 3. Planet Foorumite postituste arv aastate kaupa

2001	2002	2003	2004	2005	2006	2007	2008	Kokku
18878	37511	30443	49226	58207	40922	43870	38681	317738

Tabelist 3 on näha, et kõige rohkem postitusi tehti aastal 2005 ning kõige vähem aastal 2003.

Exceli tabelleid moodustati aasta kohta järgmiste andmete järgi.

1. Iga kasutaja postituste arv.
2. Emotikonide üldine kasutussagedus (vt tabel 7).

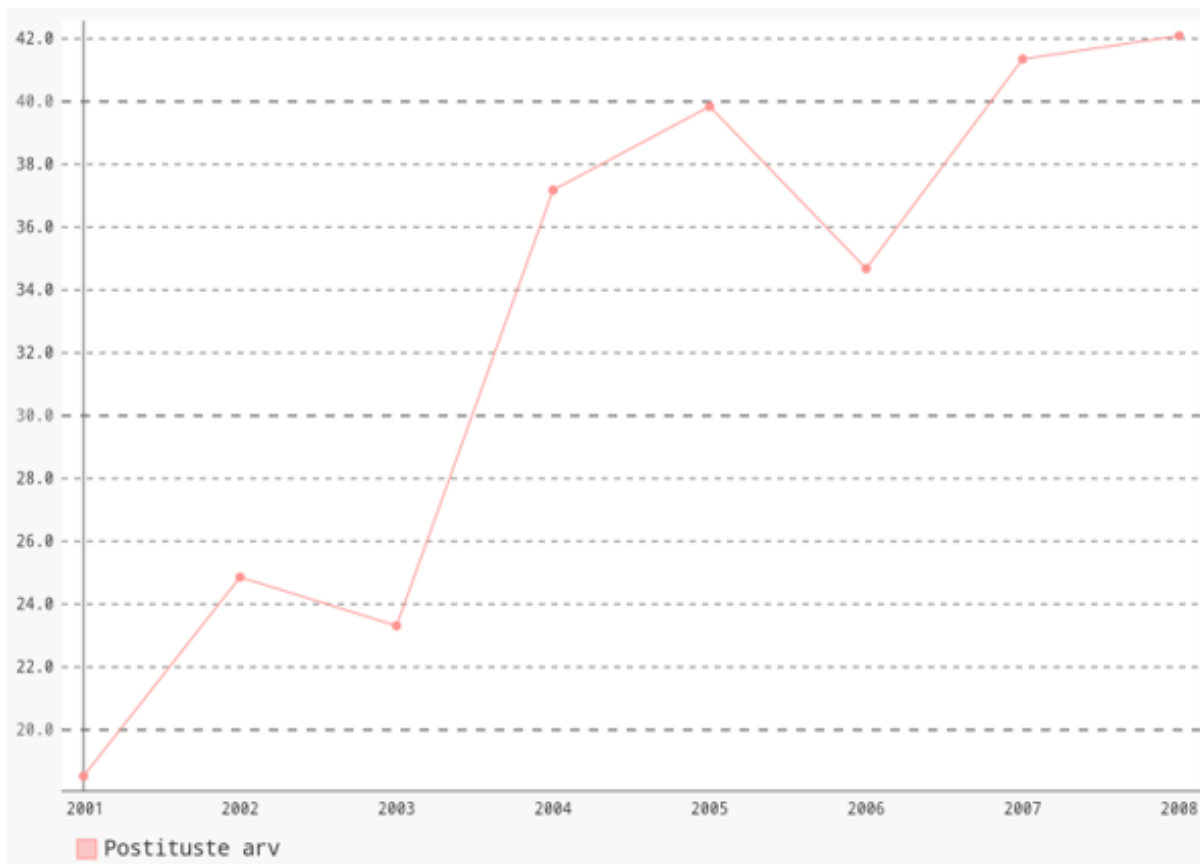
² <http://www.pygal.org/en/latest/> (vaadatud 20.03.2016).

3. Iga erineva emotikoni kasutussagedus.
4. Iga kasutaja kohta, kui palju on ta emotikone kasutanud.
5. Iga kasutaja kohta, kui palju on ta kõiki erinevaid emotikone kasutanud.
6. Suure ja väikese algustähe kasutussagedus (vt tabel 5).
7. Suure ja väikese algustähe kasutussagedus kõigi kasutajate kohta.

Need andmed annavad põgusa ülevaate muudatustest, mis on foorumis toimunud. Näiteks selgub, et on mitmeid kasutajaid, kes on postitanud vaid ühel aastal ning siis on nad teinud vaid ühe postituse.

4. KASUTAJATE JA POSTITUSTE DÜNAAMIKA

Enne keelekasutuse uurimist tuleb vaadelda foorumi üldisemaid näitajaid. Kokku on foorumis 2001.–2008. aastal tehtud 317 738 postitust. Joonisel 4 on näidatud, milline on igal vaadeldaval aastal olnud keskmine postituste arv kasutaja kohta.



Joonis 4. Foorumi keskmine postituste arv kasutaja kohta

Jooniselt 4 on näha, et keskmiste postituste arv on üsnagi pidevalt tõusnud, v.a vahemikes 2002–2003 ning 2005–2006. Kõige suurem hüpe toimus vahemikus 2003–2004. Selleks et selgitada, kas foorumiga liituvate, uute liikmete arv mõjutab keskmist postituste arvu tõusu või langust, on tehtud tabel 4. Kuna andmed on olemas alates aastast 2001, siis on loetud kõik kasutajad, kes on 2001. aastal postitanud, uuteks kasutajateks.

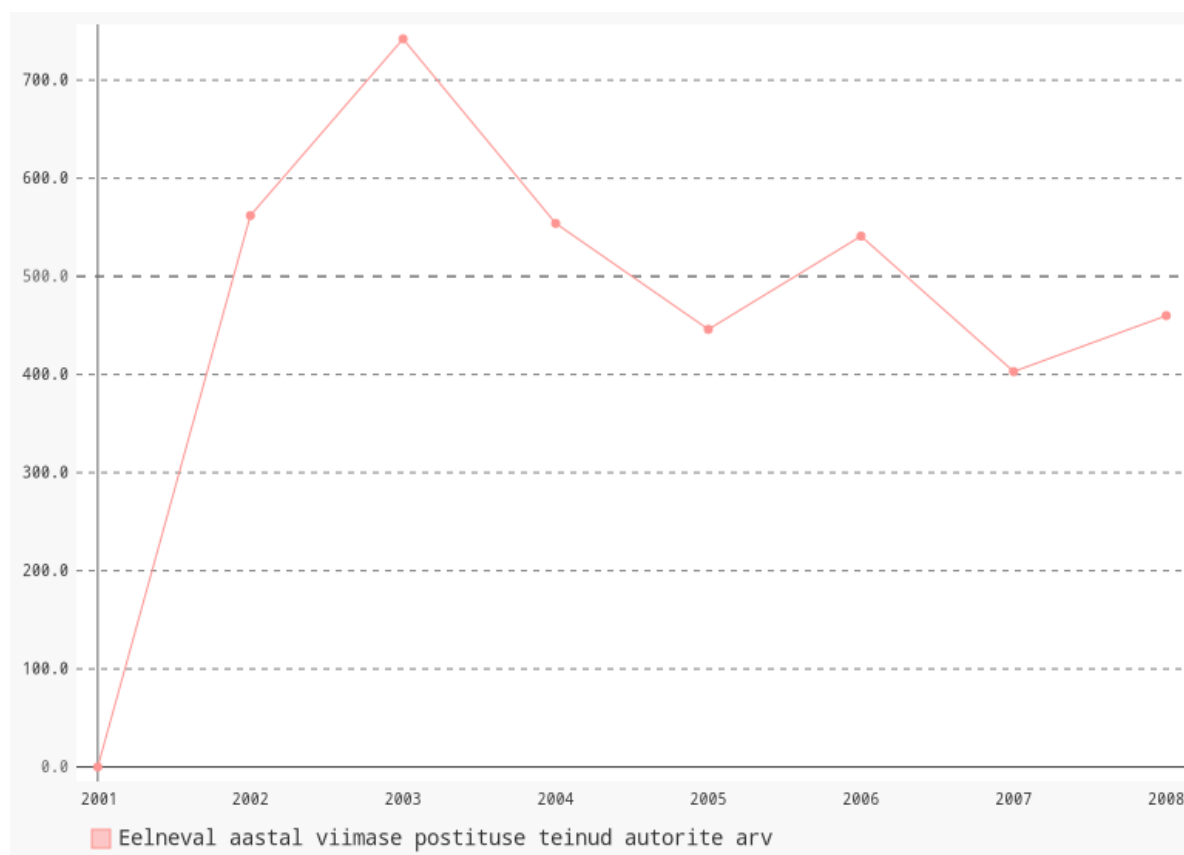
On võimalik, et tegelikult on foorumiga liitunud rohkem kasutajaid, lihtsalt nad ei ole teinud ühtegi postitust. Uute kasutajate liitumise puhul ei saa ka välistada seda, et mõni inimene on näiteks teinud endale mitu erinevat kontot või on lihtsalt vahetanud oma kasutajanime. Samamoodi ei saa ka olla kindel, kas kõik lahkunud kasutajad on ikka lahkunud. Seda aga pole võimalik selle korpuse puhul välja lugeda ning seega vaadeldakse siinses töös erineva kasutajanimega isikuid erinevate inimestena.

Tabelist 4 on näha, et uusi kasutajaid on tulnud foorumisse juurde igal aastal, samas on igal aastal, alates aastast 2002 ka viimase postituse tegijaid. Kõige rohkem on foorumiga liitunud algusaastatel, seejärel on uute kasutajate tulek aeglustunud.

Tabel 4. Uute ja vanade kasutajate osakaal foorumis

Aasta	2001	2002	2003	2004	2005	2006	2007	2008
Olemasolevad kasutajad		427	707	748	870	904	815	681
Uued kasutajad	1019	1082	599	576	591	276	246	238
Lahkunud kasutajad		562	742	554	446	541	403	460
Kasutajaid kokku	1019	1509	1306	1324	1461	1180	1061	919

Tabelis 4 olevad andmed on esitatud graafidena joonistel 5 ja 6.

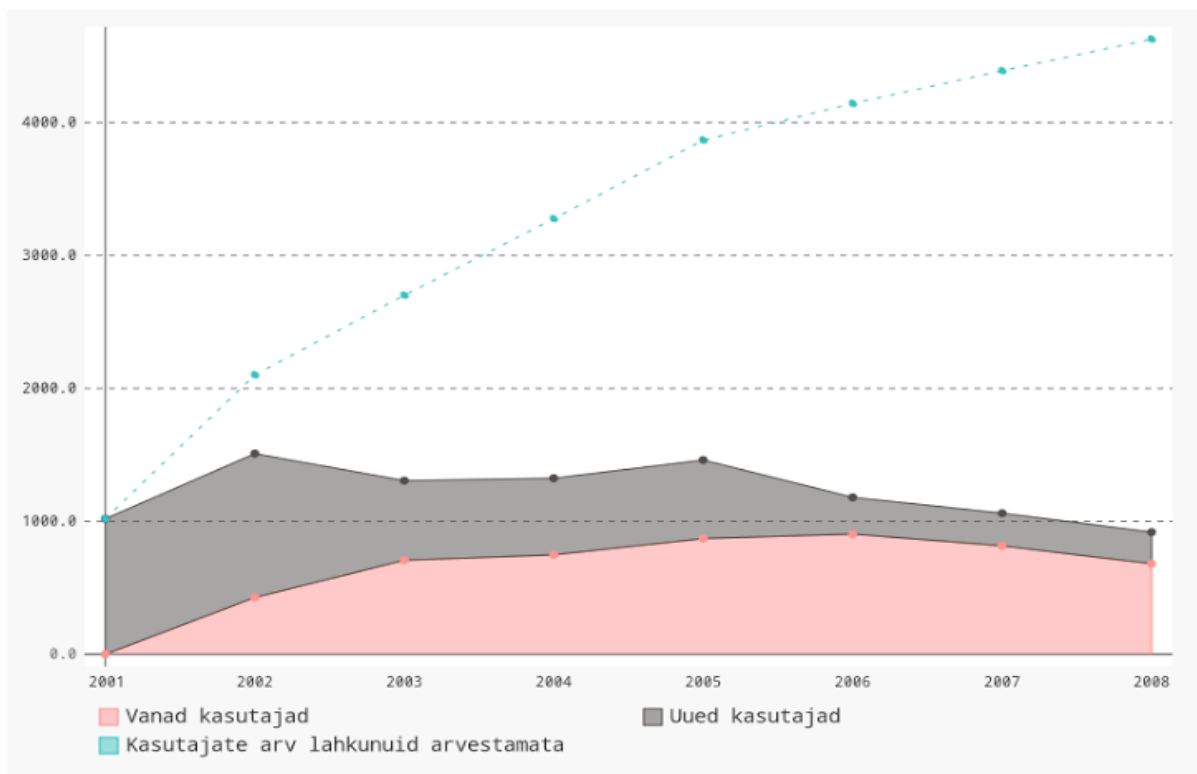


Joonis 6. Foorumist lahkunud kasutajate arv

Foorumist lahkunute arv on iga aastaga kõikunud, küll aga mitte nii palju kui uute liitujate arv, ainult aastal 2003 on rekordiliselt lahkunud 742 inimest. Lahkunuteks loetakse kasutajaid, kes on teinud oma viimase postituse eelneval aastal (aastal 2002 on lahkunud kasutajaid 562, see

tähendab, et kõik need inimesed on teinud oma viimase postituse 2001. aasta jooksul). Igal aastal kokku olevaid kasutajaid on arvatud järgmiselt: (kasutajaid kokku eelmisel aastal – lahkunud) + uued kasutajad. Näiteks aastal 2002 oli 562 lahkunut, 2001. aastal oli kokku 1019 kasutajat, ehk siis 2002. aastal oli kokku $(1019 - 562) + 1082 = 1509$ kasutajat.

Joonisel 6 on hästi näha, et suure osa foorumist moodustavad just vanemad kasutajad ning võrreldes nende hulka uute kasutajatega, on neid alates 2003. aastast alati üle poole rohkem. Eraldi on joonisel toodud veel välja, kui palju oleks foorumis kasutajaid, kui keegi ei oleks lahkunud, ehk sel juhul oleks olnud foorumis aastaks 2008 peaaegu 5000 liiget. Tundub, et foorumis on üha enam tekkinud n-ö oma seltskond, sest uusi liitujaid on alates 2005. aastast vähem.

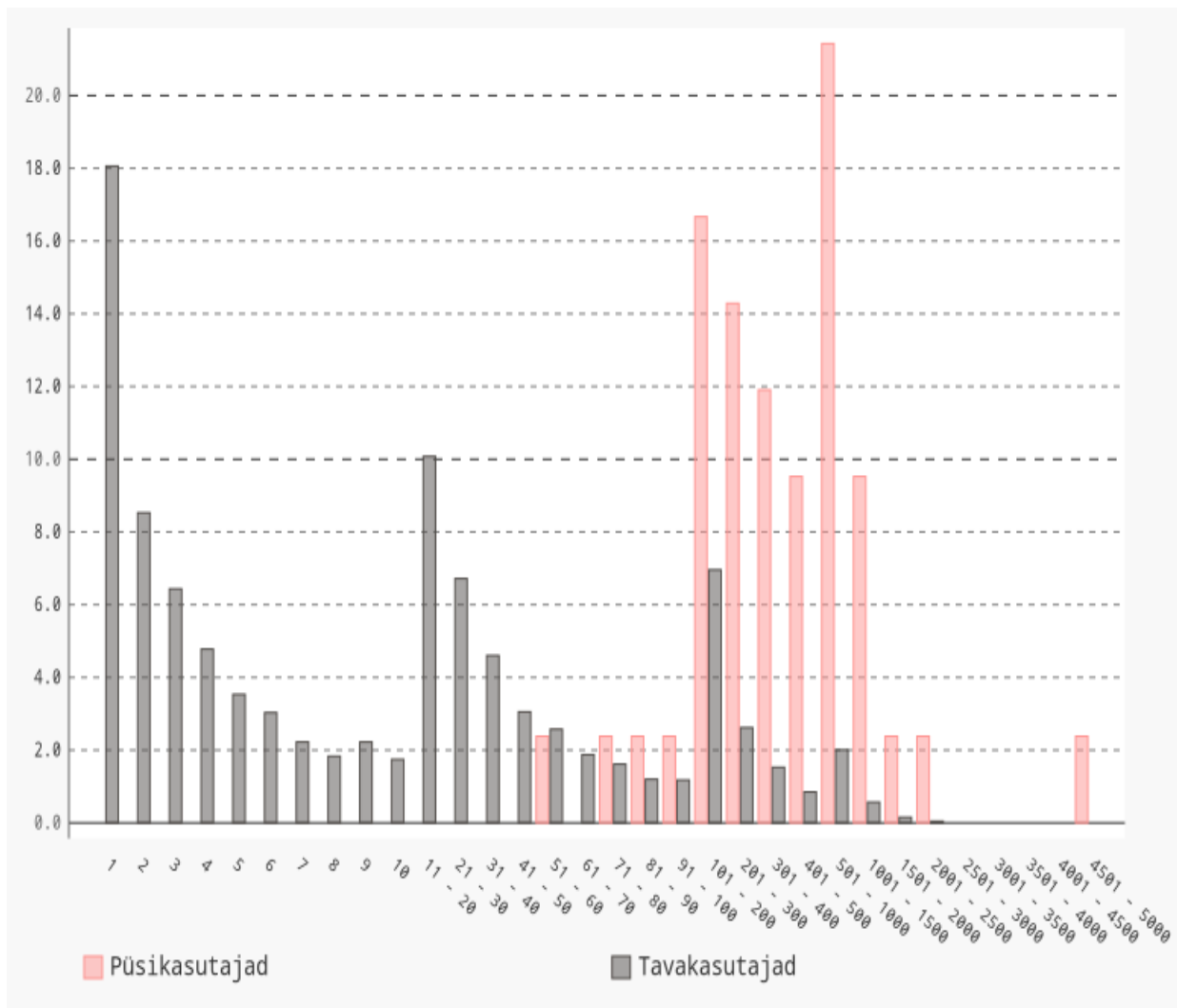


Joonis 6. Uute ning olemasolevate kasutajate osakaal foorumis aastate kaupa

Kõikidest kasutajatest eristatakse püsikasutajaid ehk kasutajaid, kes on perioodil 2001–2008 postitanud igal aastal. Kokku on selliseid postitajaid 42. Kõiki ülejäänud kasutajaid nimetatakse käesolevas töös tavakasutajateks.

Joonisel 7 on võrreldud püsikasutajate ning tavakasutajate postituste arve. Graaf annab veelgi selgema pildi sellest, kui suur osakaal on foorumis just püsikasutajatel. Ükski püsikasutaja ei ole teinud alla 40 postituse, kuid tavakasutajate hulgas leidub neid hulganisti: üle 18% tavakasutajatest on postitanud vaid ühe korra. Korpuse tekste uurides tundub, et enamik selliseid kasutajaid on postitanud just kindla temaga alafoorumites. Mitmel korral on nad ise

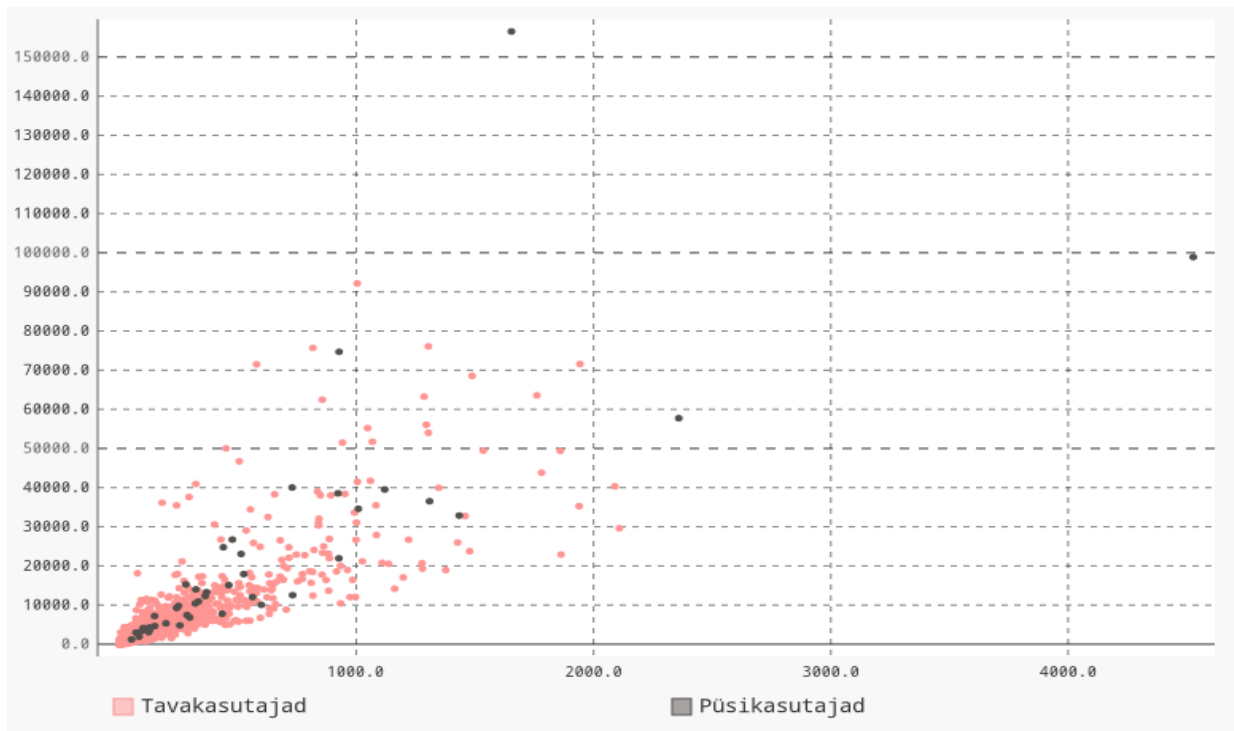
alustanud alaforumites teemat, kus paluvad abi või informatsiooni millegi kindla kohta ning mõnikord on see nende ainukeseks postituseks jäänud.



Joonis 7. Foorumi kasutajate jaotus postituste arvu järgi (protsentides)

Jooniselt 8 on näha, et sagedamini postitavate kasutajate postitused sisaldavad ka rohkem sõnu ehk graafil kujutatud postituste arvu ning pikkuse vaheline suhe on küllaltki lineaarne. On ka erandeid, näiteks leidub kasutaja (kasutajanimega *troskal*), kes on teinud 4527 postitust, kuid keskmine postituse pikkus on vaid ~21,8 sõna, ning leidub kasutaja (kasutajanimega *naine*), kes on teinud 1654 postitust, kuid keskmine postituse pikkus on ~94,6 sõna.

Joonise 8 x-teljel on postituste ning y-teljel sõnade arvud.



Joonis 8. Foorumi kasutajate jaotus postituste ning sõnade arvu järgi

Graafilt tuleb selgelt välja, et väga suur hulk tavakasutajaid on teinud vähe postitusi ning nende postitused on ka võrdlemisi lühikesed.

5. KEELETUNNUSTE DÜNAAMIKA

Selleks et tuvastada, mis on keelekasutuses teistmoodi kui varasemalt, vaadeldakse muudatusi suure ja väikese algustähe, emotikonide, tähekorduste, võõrtähtede ning lühendite kasutuses.

5.1. Suure ja väikese algustähe kasutus

Eesti keeles alustatakse üldiselt uut lauset suure algustähega. Internetis suheldes kiputakse seda reeglit aga eirama. Käesolevas peatükis kirjeldatakse, kui palju on muutunud ajas väikese ja suure algustähe kasutus valitud foorumis ning kas see, kuidas kirjutavad foorumi kõige aktiivsemad postitajad, võib avaldada mõju teistele. Vaadeldakse ainult lause algustähti, välja jäetakse nimede jms kirjutamise uurimine. Kuna info selle kohta, millist lause algustähte foorumis keegi kasutab, on juba olemas, siis jagatakse kasutajad kaheks: suure algustähe kasutajad ning väikese algustähe kasutajad. Jaotus tehakse selle põhjal, kumba viisi on kasutaja rohkem kasutanud.

Tabel 5. Suure ja väikese algustähe kasutussagedus foorumis aastate kaupa

Aasta	2001	2002	2003	2004	2005	2006	2007	2008	Kokku
Suur algustäht	25354	51475	42583	86138	98618	75439	74654	62526	516787
Väike algustäht	35243	72301	69984	83383	89972	55012	54960	39446	500756

Tabel 5 näitab vaid, kui palju on üldse kasutatud väikest või suurt algustähte. Sealt selgub, et enne 2005. aastat oli väikese algustähe kasutus suurem, pärast jällegi suure algustähe kasutus.

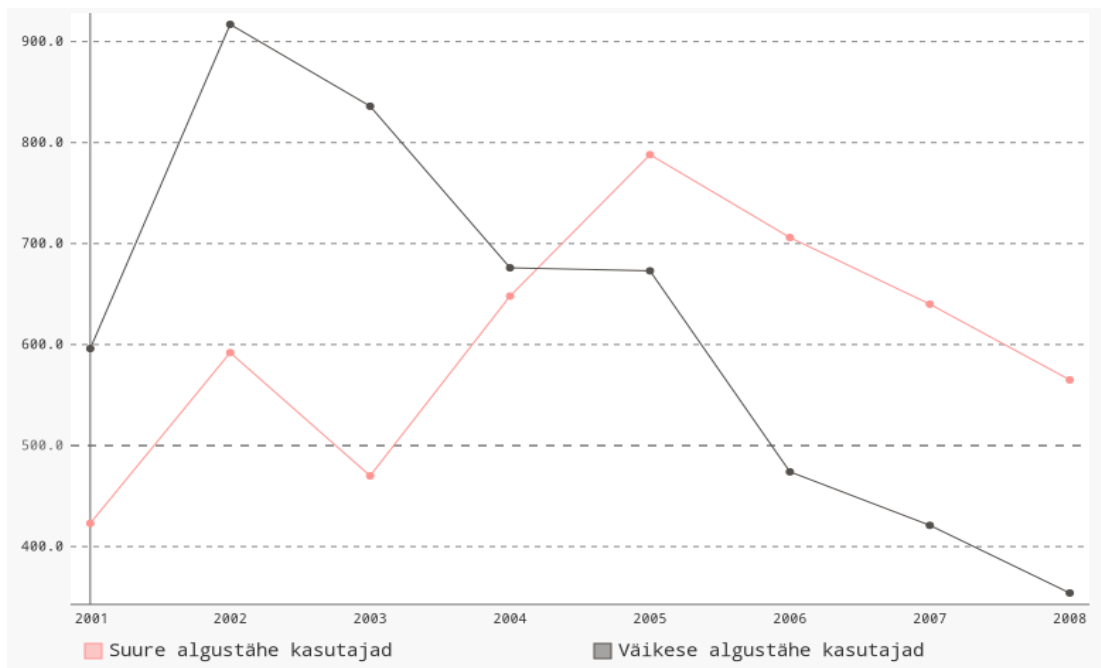
Tabelis 6 on näha, kui palju on väikest algustähte kasutanud püsikasutajad ja kui palju ülejäänud kasutajad. Selgub, et 2008. aastal kasutati suurt algustähte peaaegu poole rohkem kui väikest.

Tabel 6. Foorumi väikese algustähe kasutajate osakaal (ümardatud)

Aasta	2001	2002	2003	2004	2005	2006	2007	2008
Püsikasutajad	67%	57%	62%	38%	31%	24%	24%	24%
Tavakasutajad	58%	61%	64%	51%	47%	41%	40%	39%

Joonis 9 kirjeldab seda, kui palju on foorumis olnud suure või väikese algustähe kasutajaid aastate kaupa. Jooniselt saab välja lugeda, et väikese algustähe kasutus on langenud alates aastast 2002 ning ei ole tõusnud isegi aastal 2005, kui foorumis oli kõige rohkem postitajaid. Kas on võimalik öelda, et teatud kasutajate käitumine mõjutab teisi? Selleks vaadatakse, kuidas

on käitunud püsikasutajad võrreldes tavakasutajatega.



Joonis 9. Foorumis suure ja väikese algustähe kasutajad aastatel 2001–2008

Joonisel 10 on püsikasutajaid võrreldud ülejäänud väikese algustähe kasutajatega.



Joonis 10. Foorumi püsi- ja tavakasutajate võrdlus väikese algustähe kasutuses

Alates aastast 2003 on mõlema hulga puhul väike algustäht oma populaarsust kaotanud. Üldiselt võib jääda mulje, et kui põhilised foorumisse postitajad muudavad oma kirjaviisi, siis

ülejäanud hakkavad seda järgima. On tehtud mitmeid uuringuid sellest, kuidas inimesed omavahel suheldes hakkavad enesele teadmata imiteerima teise inimese kõnepruuki [9]. Võimalik, et seda tehakse ka internetisuhtluses, sest teistega sarnaselt kirjutamine võib luua ühtsustunnet teiste foorumi kasutajatega. Samas on aastal 2002 tavakasutajate puhul väikese algustähe kasutus kasvanud, aga püsikasutajate puhul langenud. Seos erinevate kasutajagruppide vahel pole siin selge.

Küll aga võiks see, et järjest rohkem hakatakse kasutama lause alguse suurtähte, nii nagu eesti keeles grammatiliselt õige on, olla üks näitajaid, et õigekeele tähtsus internetis on kasvanud.

5.2. Emotikonide kasutus

Internetis suheldes väljendatakse oma emotsioone tihti ikoonide abil. Üldise ülevaate emotikonide kasutusest annab tabel 7.

Tabel 7. Foorumi emotikonide kasutamiste arv aastate kaupa

2001	2002	2003	2004	2005	2006	2007	2008	Kokku
5743	6663	5137	7551	7820	3791	4051	3487	44243

Tabelist 7 on näha, et emotikonide kasutuses toimub kõige suurem kõikumine aastate 2005 ning 2006 vahel, kus emotikone hakatakse kasutama vähem. Aastal 2005 on postitustes neid pseudograafilisi märke olnud kõige rohkem.

Edasi vaadatakse, milline seos on postitajate arvu ning emotikonide arvu vahel, andmed selle kohta on toodud ära tabelis 8.

Tabel 8. Foorumi emotikonide arv 100 postituse kohta (ümardatud) ning postitajate arv aastate kaupa

Aasta	2001	2002	2003	2004	2005	2006	2007	2008
Emotikonide arv 100 postituse kohta	30,4	17,8	16,9	15,3	13,4	9,3	9,2	9,0
Postitajate arv	1019	1509	1306	1324	1461	1180	1061	919

Tabelist 8 selgub aga, et emotikonide arv postitustes on vähenenud vaatamata postitajate arvule. Näiteks on olnud aastal 2002 kõige rohkem kasutajaid, kuid emotikonide arv on võrreldes eelneva aastaga langenud. Ka aastal 2005 on olnud üsna palju kasutajaid, kuid emotikonide populaarsus on vähenenud veelgi. Seega ei olene emotikonide kasutus sellest, kui palju on foorumis kasutajaid.

Joonis 11 väljendab, kui palju on aja jooksul emotikonide arv (100 postituse kohta) muutunud ning kui palju on postitajate arv aja jooksul muutunud. Nagu juba tabelist 7 selgus, on ka graafilt näha, et vastastikune seos nende kahe muutuja vahel on väga nõrk: emotikonide kasutus

on jätkuvalt vähenenud, kuid postitajate arv on selles ajavahemikus ka mitu korda tõusnud.



Joonis 11. Foorumi emotikonide kasutus võrreldes kasutajate arvuga

Edasi uuritakse, milline seos on püsikasutajate ja tavakasutajate vahel ning kas ühed mõjutavad teisi. Tabelis 9 on toodud püsikasutajate ja tavakasutajate emotikonide kasutuse võrdlus.

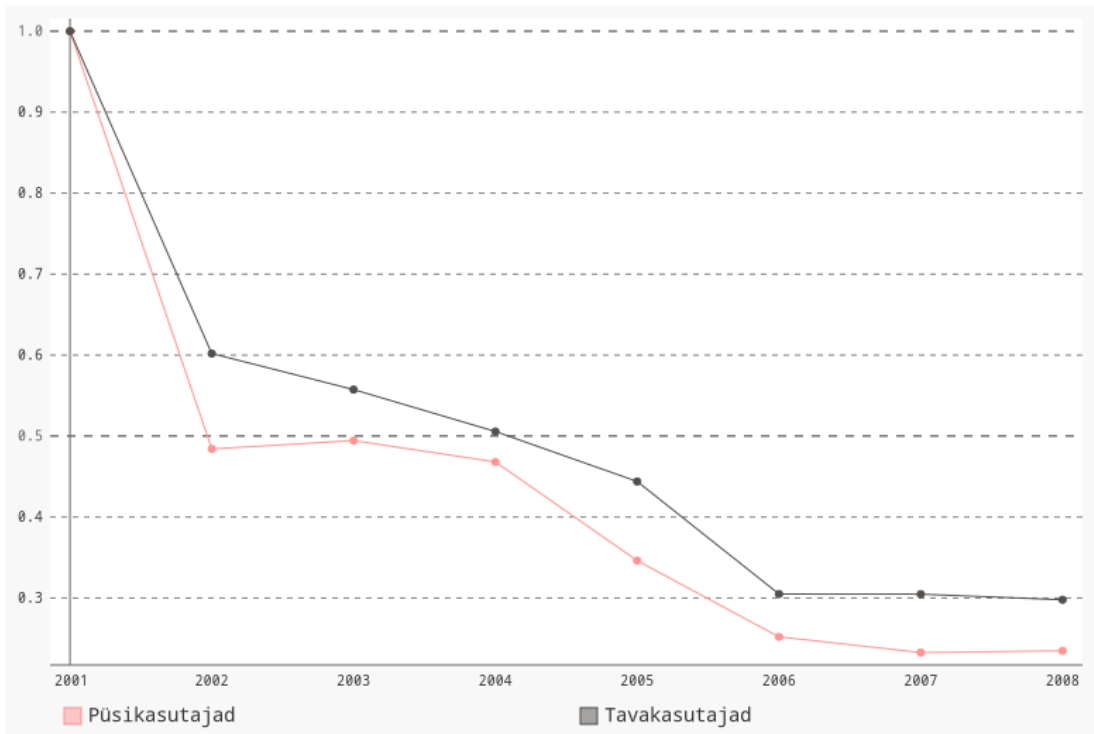
Tabel 9. Foorumi kasutajate emotikonide kasutus saja postituse kohta (ümardatud)

Aasta	2001	2002	2003	2004	2005	2006	2007	2008
Püsikasutajad	20,5	3,6	11,4	8,1	5,4	4,6	6,2	5,9
Tavakasutajad	30,6	18,2	17,1	15,6	13,7	9,5	9,4	9,2

Tabelist on näha, et tavakasutajate puhul on olnud muudatus suurem: perioodil 2001–2008 on emotikonide kasutus vähenenud 21,4% võrra. Arv on iga aastaga ainult langenud. See-eest on püsikasutajate puhul emotikonide arv sel perioodil rohkem kõikunud, kuid aastaid 2001 ja 2008 võrreldes lõpuks siiski langenud 14,6% võrra. Vaadates aga kasutajat, kes on foorumis teinud kokku kõige rohkem postitusi (kasutaja *troskal*), siis selgub hoopiski, et ta on kasutanud aastal 2001 ainult kahte emotikoni ning vahemikus 2002–2007 ka võrdlemisi vähe, aga aastal 2008 koguni 250 emotikoni. Teisena on kõige rohkem postitanud kasutaja nimega *treudenbauer*, kes ei ole püsikasutaja, kuid kellel on 2359 postitust. Ka tema on kõige rohkem emotikone postitanud aastal 2008.

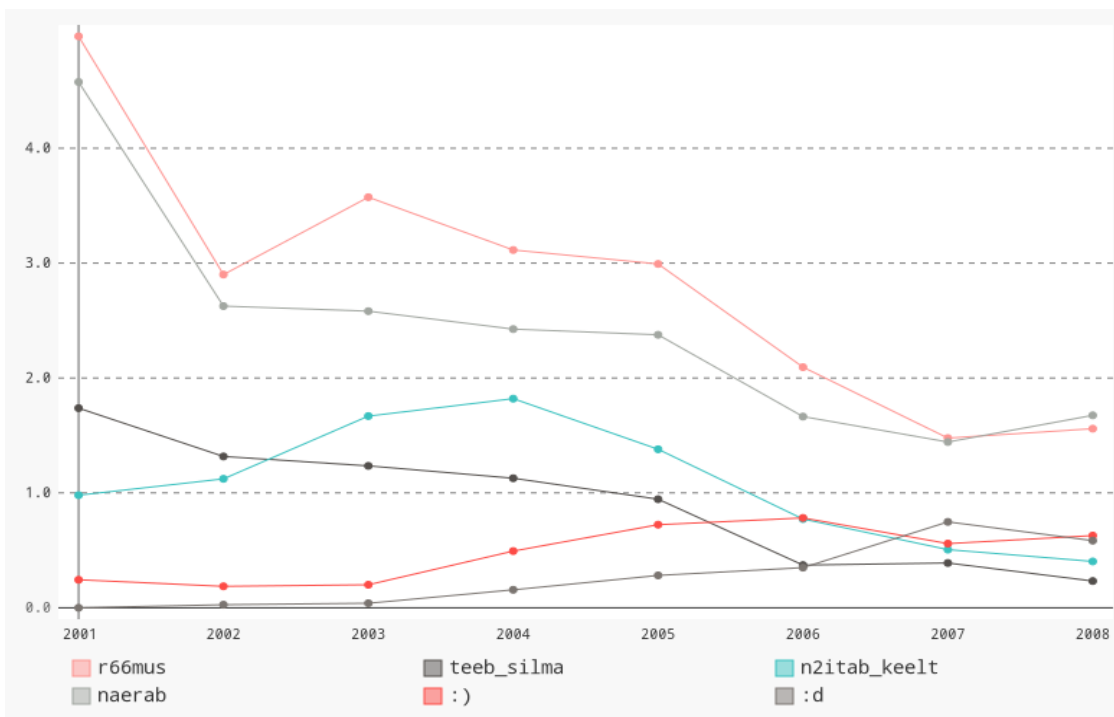
Tabeli 9 andmed on esitatud joonisel 12. Jooniselt tuleb välja, et emotikonide kasutus püsikasutajate ning tavakasutajate vahel on küllaltki sarnane. Mõlema puhul on arv pidevalt langenud, püsikasutajatel vahepeal natuke tõusnud, kuid lõpuks siiski väga palju madalam kui aastal 2001. Selle põhjal väita, et püsikasutajad on tavakasutajate emotikonikasutust mõjutanud

või vastupidi, ei ole võimalik. Saab vaid väita, et emotikonide menüüs on sel perioodil jätkuvalt vähenenud.



Joonis 12. Foorumi püsikasutajate emotikonikasutuse sageduse võrdlus tavakasutajatega

Joonis 13 toob esile enim kasutatud emotikonid.



Joonis 13. Foorumi populaarseimad emotikonid aastas (100 postituse kohta)

Jooniselt 13 on näha, et vaid kahte emotikoni, „:)“ ning „:d“ on kasutatud perioodi lõpus ehk 2008. aastal rohkem kui varem. Ülejäänud emotikonide kasutus on perioodi lõpuks vähenenud.

Kaks kõige populaarsemat emotikoni, *r66mus* ning *naerab*, püsivad kogu perioodi ajal kõige populaarsemana, kuigi mõlema kasutus langeb nähtavalt. Eriti on nende kahe emotikoni kasutus vähenenud vahemikus 2001–2002.

Emotikonid *r66mus* ning „:)“ tunduvad olevat samad, üks on pildiga emotikon ning teine pildita. *R66mus* on aga alates aastast 2003 olnud vähem kasutatud ning emotikoni „:)“ on alates aastast 2003 rohkem kasutatud, kuni aastani 2006. On võimalik, et vahepeal eelistati ilma pildita versiooni. Vahemikus 2007–2008 on aga mõlema kasutus jällegi tõusnud, seega pole võimalik öelda, et üks hakkaks teist asendama.

Tavaliselt seostatakse emotikoni „:d“ („:D“) suure naeratava näoga. See sarnaneb emotikoniga nimega *naerab*. Nende kahe emotikoni puhul on näha sarnaseid muudatusi nagu emotikonidega *r66mus* ja „:)“. Ära võiks veel märkida, et sellel ajal, kui emotikoni *naerab* kasutati kõige vähem, oli emotikoni „:d“ sisaldavaid postitusi just kõige rohkem. Seega võiks arvata, et mõnel perioodil on eelistatud pildita, kuid sama emotsiooni väljendavat emotikoni.

5.3. Tähekordused

Tähekordustega väljendatakse internetis erinevaid emotsioone, nii viha kui ka rõõmu. Eesti keeles leidub sõnu, kus on kahekordne täis- või kaashäälik. Juhul kui on enam kui kaks täishäälikut järjest, on tegu liit- või võõrsõnaga. Kui sõnas on rohkem kui kaks ühesugust kaashäälikut järjest, on tegu liitsõnaga või liitega *-ki*. Seega sagedusloendi moodustamisel leitakse sõnad, kus leidub kolm või rohkem sama täis- või kaashäälikut järjest, jättes välja *ki*-partikliga sõnad ning liit- ja võõrsõnad.

Sagedusloendit tehes selgus, et postitustes leidub väga palju kirjavahemärkide kordusi, näiteks neli või enam punkti, hüüumärki või küsimärki järjest. Korduste alt leiti, et palju on kasutatud tärne sõimusõnade asemel, näiteks *bull******. On ka lihtsalt tärne, näiteks *******, mille puhul pole üldse selge, mis sõnaga on tegu.

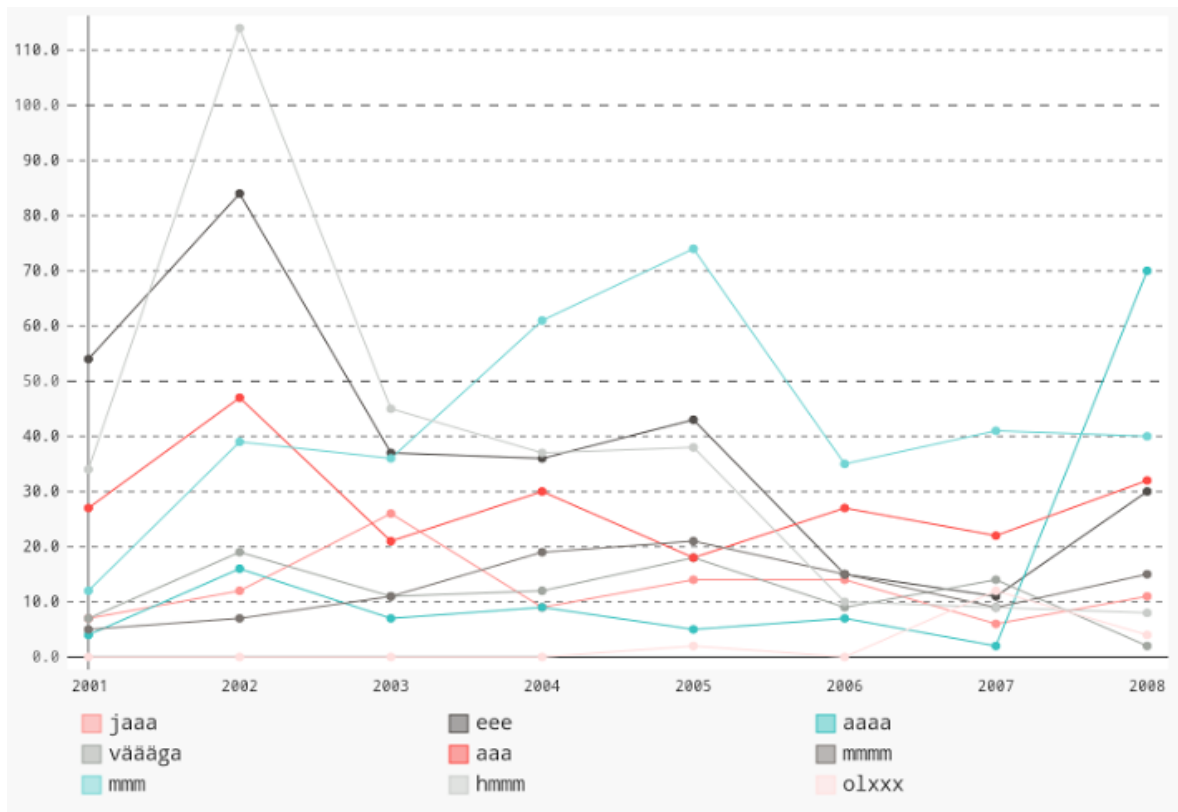
Joonis 14 on saadud, otsides iga aasta (vahemikus 2001–2008) kohta, millised olid tol aastal neli kõige rohkem kasutatud tähekordusi sisaldavat sõna. Välja on jäetud kordused, mis sisaldavad kirjavahemärke või numbreid. Lisaks on jäetud välja *kkk* ehk korduma kipuvad küsimused ning *piii*, mille all on mõeldud protsessorit Pentium III.

Jooniselt on näha, et mitmed tähekordusi sisaldavad sõnad imiteerivad hääliitsusi, näiteks mõtisklemist väljendav *hmmm*, *mmm(m)*, *eee*. Arusaamist väljendab *aaa(a)* ning *eee* oleks justkui mõttepaus vestluses. Vastavalt kontekstile võib neil olla ka teisi tähendusi. Sõna *olxxx* on arvatavasti eestikeelne sõna *olgu*. Leidub ka lihtsalt pikendatud eestikeelseid sõnu, nagu väga – *väääga*. Sõna *jaaa* on rohkem kasutatud kui versiooni sõnast *jah*, mitte ei ole pikendatud sidesõna *ja*.

Kõige huvitavam on *aaaa* kasutus, mis on hakanud alates 2007. aastast äkitselt tõusma. Samas on kolmetähelist *aaa*-d kasutatud järjepidavalt ning üsnagi võrdselt igal aastal. Kas see on täiesti juhuslikult nii?

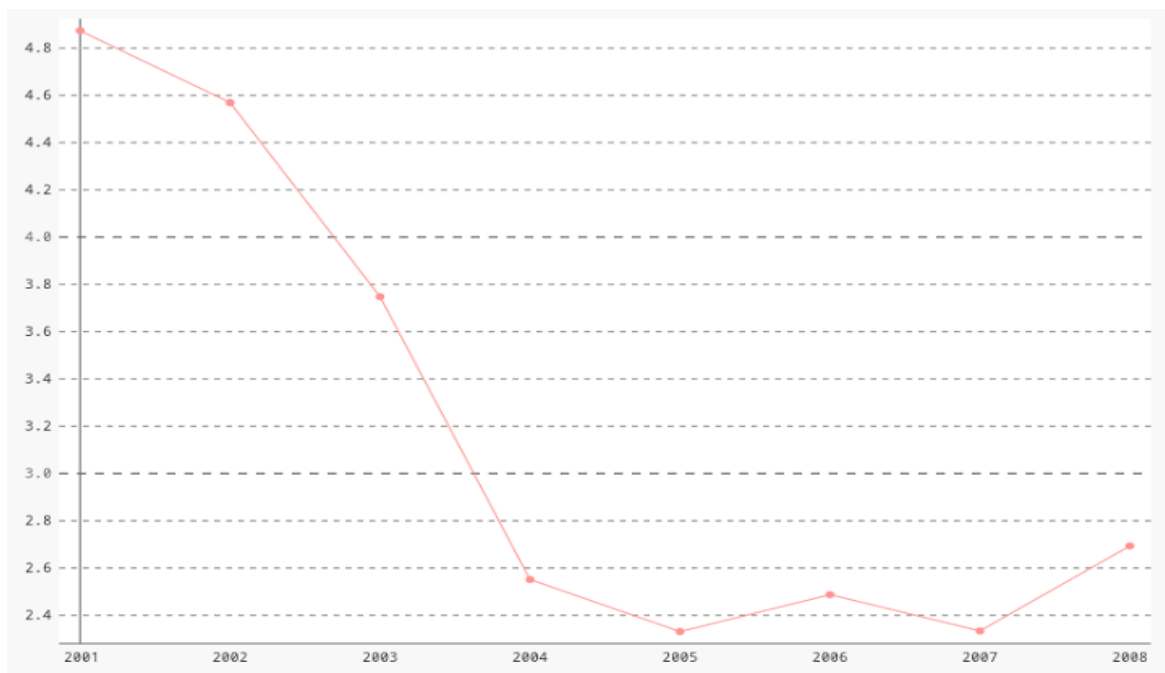
Ka sõna *olxxx* paistab teiste seast välja, kuna varem ei kasutatud seda üldse. Vahemikus 2004–2005 on seda postitustes kasutatud ja siis on see jälle ära kadunud ning 2006–2008 jälle rohkem kasutatud. Tekib küsimus, kas sellel joonisel võib mõni sõna olla selline, mida kasutab vaid

üks inimene oma postitustes. Kontrollides sõna *olxxx*, selgus aga, et seda on kasutanud 15 erinevat kasutajat.



Joonis 14. Kolm enim kasutatavat tähekordust sisaldavat sõna iga aasta kohta

Jooniselt 15 on näha, milline oli tähekordusi sisaldavate sõnade osakaal erinevatel aastatel.



Joonis 15. Tähekordustega sõnade kasutus 100 postituse kohta

Tähekordusi sisaldavate sõnade osakaal on vahemikus 2001–2004 palju vähenenud. Pärast

seda langust on kõikumisi olnud vähem.

5.4. Võõrtähtede kasutus

Võõrtähti (enim kasutatavad on *c*, *q*, *w*, *x*, *y*) eestikeelsetes sõnades ei leidu [10]. Küll aga kasutatakse neid netikeeles tihti eestikeelsetes sõnades olevate häälikuühendite lühendamiseks. Näiteks *ts* asendatakse *c*-ga (*suits* – *suic*), *ks* asendatakse *x*-iga (*miks* – *mix*) või *gu* *q*-ga (*nagu* – *naq*). *W*-d kasutatakse eesti tähestikus oleva *v* asemel ning *y*-d *ü* asemel. See tuleneb arvatavasti sellest, millist klaviatuuri kasutatakse, ingliskeelseks seatud klaviatuuril ei ole näiteks *ü*-tähte. Lisaks asendatakse veel *hv* kahe *f*-iga (*vahva* – *vaffa*) [11]. Sagedusloendi moodustamisel tuleb arvesse võtta, et sinna võivad sattuda ingliskeelsed sõnad. Need ei paku siinse uurimuse puhul huvi ning tuleb välja jätta. Huvi pakuvad vaid sõnad, kus võõrtähti on kasutatud eestikeelsete sõnade lühendamiseks.

Arvutiprogrammil lasti otsida korpusest kaksümmend populaarsemat sõna, mis sisaldavad tähti *c*, *q*, *w*, *x*, *y* või *ff*. Tulemused on toodud tabelis 10.

Tabel 10. Foorumis leitud sõnad, milles on kasutatud võõrtähti

Sõna	Kasutuskordade arv
kyll	2042
nyyd	1442
XP	1313
irw ³	1248
olex	983
ntx	861
yks	783
xp	777
cd	700
yle	679
btw	575
wtf	549
ylitse	543
sry	499
windowsi	490
yhe	439
peax	433
2x	354
offtopic	345
yles	326

Tabelist 10 on näha, et kõige rohkem asendatakse *ü*-d *y*-ga (nt sõnades *kyll*, *yks*, *nyyd*). *Hv* asendamist *ff*-iga ei leidu üldse või lihtsalt ei mahu see kahekümne enim kasutatava hulka. Samuti on häälikuühendi *ts* asendamisega *c*-ga. Küll aga on mitmes sõnas asendatud *ks* *x*-iga (nt *peaks* – *peax*, *oleks* – *olex*, näiteks – *ntx*). Tabelis on sõnu, mis käesolevas töös huvi ei paku: näiteks ingliskeelsed sõnad *offtopic* ja nimetus Windows. Loendisse sattunud ingliskeelseid lühendeid (nt *sry*, *wtf*, *btw*) uuritakse lähemalt järgmises alapeatükis.

³ Väljendab naeru.

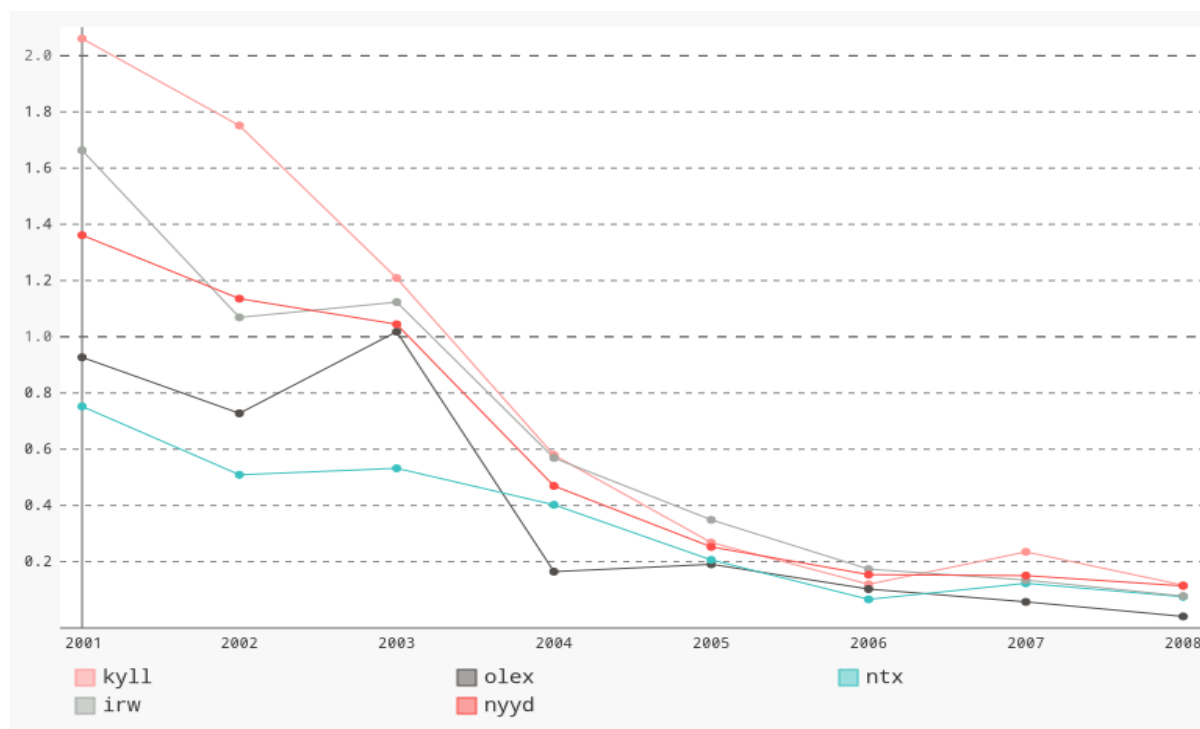
Mõne sõna uurimiseks tuli vaadata korpuse tekste läbi käsitsi. Selle tulemusena selgus, et lühendit *XP* (või *xp*) on kasutatud kõige rohkem arvutiteemalistes alafoorumites (korpuse fail „Arvutid_valiidne.xml“), mõeldes selle all Windows XP-d. Sama on sõna *cd* puhul: kõige rohkem kasutatakse seda arvutitega seoses, mõeldes selle all laserplaati (*compact disc*). *2x* on fraasi „kaks korda“ lühend.

Eemaldades järjest sõnad, mis ei paku võõrtähtede mõttes huvi (erinevad nimetused, lühendid (erandiga *ntx*), ingliskeelsed sõnad), saadi järgnev tabel 11.

Tabel 11. Kümme foorumis enim kasutatavat võõrtähti sisaldavat sõna (eemaldatud on mitte huvipakkuvad sõnad)

Sõna	Kasutuskorrad
kyll	2052
irw	1700
nyyd	1552
ntx	923
yks	839
yle	699
yldse	552
yhe	445
peax	441

Jooniselt 16 on võimalik näha, kuidas on aja jooksul muutunud viie enim kasutatava võõrtähte sisaldava sõna kasutus aja jooksul.



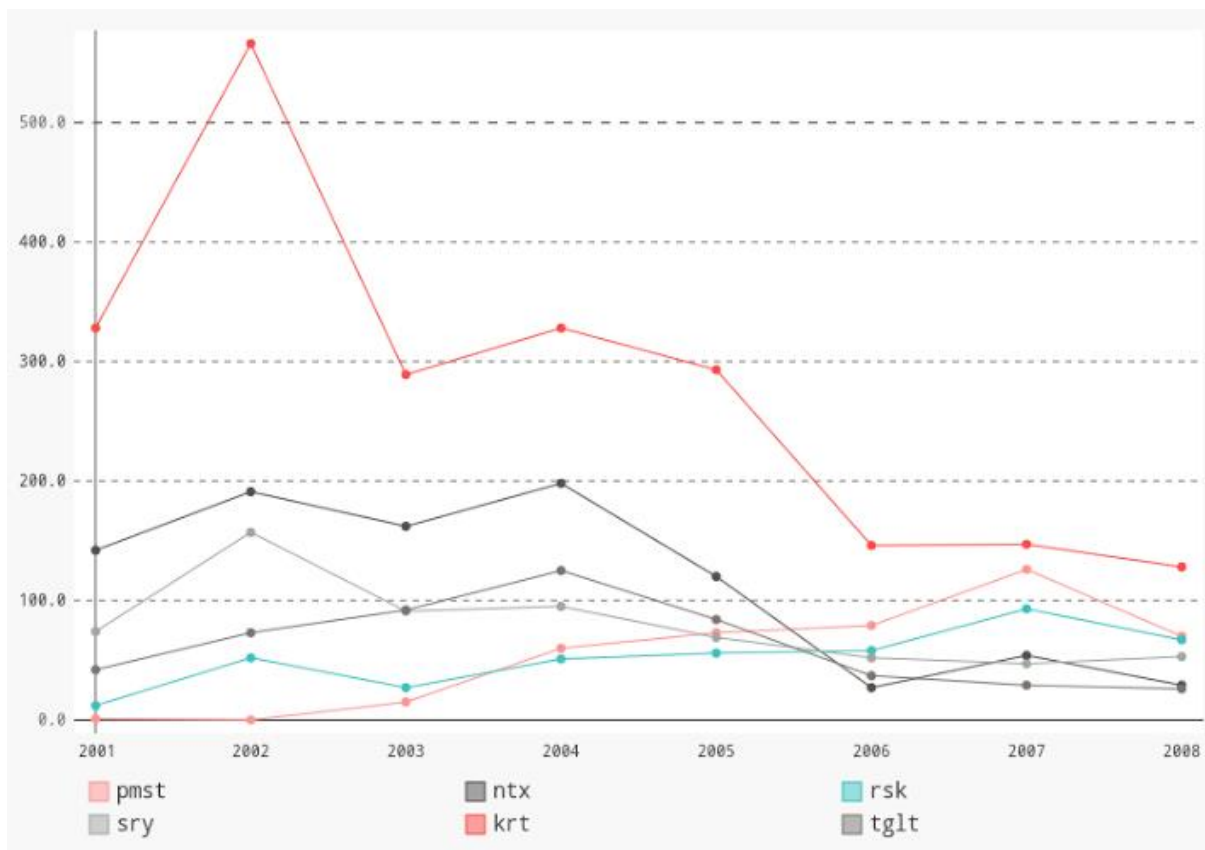
Joonis 16. Võõrtähti sisaldavate sõnade kasutus foorumis (100 postituse kohta)

Nagu näha, on kõiki neid sõnu kasutatud järjest vähem, aastatel 2006–2008 peaaegu enam üldse. Mingil põhjusel on küll sõna *olex* kogunud populaarsust vahemikus 2002–2003, kuid

siis hakanud jällegi järsult ära kaduma. Kuna kõik sellised sõnad nagu „küll“, „oleks“, „nüüd“, „näiteks“ on eesti keeles tavapärased sõnad, siis on neid arvatavasti siiski ka edaspidi vestlustes kasutatud, kuid õige sõnakujuga.

5.5. Lühendite kasutus

Lühendite puhul vaadatakse eraldi eesti- ja ingliskeelseid lühendeid. Selgituseks tuleb mainida, et uuritakse lühendeid, mis ei ole grammatiliselt korrektsed, vaid on internetis suhtlevate inimeste loodud selleks, et lühendada kirjutamisele kuuluvat aega. Nagu töös varem mainitud, on eesti netikeele puhul omane moodustada lühendeid sõnadest vokaalide ärajätmisega. „Eesti keele käsiraamatus“ [12] on kirjas reegel: „Häälikuühendis kirjutatakse iga häälik ühe tähega“, lisatud on, et sellele leidub erandeid. Reeglit arvestades moodustatakse sagedusloend, mis näitab, millised sõnad korpuses sisaldavad rohkem kui kolme kaashäälikut üksteise järel, ning selliste sõnade esinemise sagedust. Kaashäälikuühendi pikkuseks on valitud neli tähte, kuna kolmetähelisi erandeid leidub eesti keeles küllaltki palju. Lisanduvad teadaolevad kahe- ja kolmetähelised lühendid, nagu näiteks: *nv* – nädalavahetus, *krt* – kurat, *ptv* – pole tänu väärt, *ntx* – näiteks, *tre* – tere, lisaks neljätäheline *pmst* – põhimõtteliselt, *sry* (ingliskeelsest sõnast *sorry*, 'vabandust', valitud, kuna eesti kõnekeeles kasutatakse väljendit *sorri*). Lühendite hulka on arvatud ka eitust väljendav *mkm* [13] ja jaatust väljendav *mhm(h)*.



Joonis 17. Iga aasta kohta kolm siis enim kasutatavat eestikeelset lühendit

Joonis 17 on saadud, leides kolm kõige rohkem kasutatavat eestikeelset (v.a *sry*) lühendit igal aastal (aastate lõikes tuli seega esile kuus enim kasutatavat lühendit). Arvutiprogrammi abil lühendite leidmiseks tuli järjest eemaldada sõnu, mis sisaldasid kaashäälikuühendeid, kuid olid

korrektsed eestikeelsed sõnad või lühendid. Jooniselt 17 selgub, et kõik lühendid, mida on kasutatud, on tekkinud vokaalide ärajätmisel. Populaarseim neist püsis igal aastal sõna *krt* – kurat, kuigi selle lühendi kasutus siiski aastatega vähenes. Kaks lühendit, mille esinemissagedus tõusis, olid *pmst* ning *rsk* – raisk. Ülejäänud lühendite kasutus vähenes. Silma jääb lühend *ntx* – näiteks, milles on lühendamiseks ära jäetud *ks* ning selle asemel kirjutatud *x*.

Ingliskeelsete lühendite puhul uuritakse tabelites 12 ja 13 olevate lühendite ilmnemist korpuses. Kuna ingliskeelseid lühendeid on väga palju, on kõiki neid kirja panna võimatu. Selle tõttu on käesoleva töö jaoks tehtud valik populaarseimast [14, 15]. Ingliskeelsete lühendite seletused on esitatud tabelis 12.

Tabel 12. Ingliskeelsete lühendite seletused

	Ingliskeelne tähendus	Tähendus netikeeles
LOL	laughing out loud	naer
LMAO	laughing my ass off	naer
PLZ	please	palun
ROFL	rolling on the floor laughing	naer
WTF	what the fuck	mida kuradit
BRB	be right back	olen kohe tagasi
SRY	sorry	vabandust
IMO	in my opinion	minu arvamus
IMHO	in my honest opinion	minu aus arvamus
BTW	by the way	muide, muuseas
FTW	for the win	(millegi väga hea väljendus)
OMG	oh my god	jumal küll

Ingliskeelsete lühendite puhul on sagedused korrapäratumad. Eestikeelsete lühenditega võrreldes on näha, et ingliskeelsed lühendid on moodustunud pigem fraasis olevate sõnade esitähedest. Ainuke, mille puhul vokaal on ära jäetud, on *sry*. Kõige rohkem on kasutatud lühendit *IMO* ning kokku samatähendusliku *IMHO*-ga on neid kasutatud 1684 korda. Algselt on kasutatud pigem pikemat varianti, kuid selle sagedus on aastate jooksul vähenenud ning lühendi *IMO* sagedus tõusnud. Seega on hakatud eelistama veel lühemat sõna. Lühendi *FTW* puhul tundub, et see on tekkinud alles aastal 2004, kuna varem seda ei leidunud. Aastatel 2007–2008 on seda aga kasutatud 108 korda. Päris ühtset muudatust ingliskeelsete lühendite puhul

ei toimu: mõne kasutus vähenes (nt *BTW*, *IMHO*, *sry*) aastatega ning mõne oma tõusis (nt *WTF*, *LOL*, *IMO*).

Tabel 13. Valik populaarsemaid internetis kasutatavaid ingliskeelseid lühendeid⁴

	2001	2002	2003	2004	2005	2006	2007	2008	Kokku
LOL	35	92	61	39	68	108	229	235	867
LMAO	1	0	10	4	10	1	2	1	29
PLZ	11	27	20	13	13	17	31	22	154
ROFL	2	12	8	10	13	4	5	5	59
WTF	7	46	68	178	167	117	115	125	823
BRB	0	0	0	1	1	0	2	1	4
SRY	74	157	91	95	69	52	47	53	638
IMO	19	68	81	184	214	135	244	216	1161
IMHO	58	87	71	96	78	40	63	30	523
BTW	80	144	119	173	135	57	67	44	819
FTW	0	0	0	3	2	14	64	44	127
OMG	14	10	40	64	76	87	60	42	393

Tundub, et ingliskeelsed lühendid võtavad eestikeelseid üle, sest neid esineb sagedamini ning eestikeelsete kasutus on aastatega vähenenud, sellal kui ingliskeelsete oma on tõusnud.

⁴ Valik on tehtud lehekülgedel <https://www.italki.com/article/154/lol-lmfao-wtf-figuring-out-english-internet-slang-and-texting-language> (vaadatud 05.04.2016) ning <http://www.connexin.net/internet-acronyms.html> leiduvatest lühenditest (vaadatud 05.04.2016).

6. UURIMUSE JÄRELDUSED

Teises peatükis selgitati, miks Planet Foorumid tundub olevat sobiv korpus netikeele muudatuste uurimiseks: palju erinevaid kasutajaid ja teemasid, olemas on püsikasutajad ning informatsioon kasutajate kohta. Selleks, et välja uurida, kui otstarbekas on see foorum keeleliste muudatuste leidmiseks, koostati mitmeid sagedusloendeid ning graafe foorumi kasutajate dünaamikast.

Selgus, et foorumis on olnud palju kasutajaid, kellel on väga vähe postitusi. Seega ei ole nende kohta käiv statistika parim uurimuse läbiviimiseks. Püsikasutajaid ehk kasutajaid, kes postitasid igal aastal 2001–2008, on vaid 42. Nende 42 inimese postituste osakaal on tavakasutajate omast palju suurem. See tähendab, et igas loendis, kus on võetud andmed kõikide kasutajate postituste kohta korraga, domineerivad püsikasutajad. Töös vaadeldi ka, kui tihti vahelduvad foorumis kasutajad, millest selgus, et lõpuaastatel tuli foorumisse juurde järjest vähem inimesi, mis viitab rohkemgi sellele, kui suur tähtsus on neil 42 püsikasutajal foorumi korpusel. Seega järeldused tehtud tööst tulevad vägagi nende 42 kasutaja kesksed.

Kuigi korpus ei olnud ideaalne, uuriti siiski selles esinevaid keelelisi muutusi. Algselt püstitatud küsimusele: „Kas netikeel on muutunud?“ on võimalik selle korpuse põhjal vastata: „Jah, on küll.“ Töö eesmärk oli leida, kuidas on keel internetis muutunud, uuenduste põhjustele ei pööratud tähelepanu. Lisaks ei vaadatud teksti semantikat. Kuigi Planet Foorumite korpus ei osutunud piisavaks, et oleks võimalik teha üldistusi kogu eesti netikeele kohta, on võimalik töösse võetud korpuse puhul öelda, et netikeel on hakanud järjest rohkem kindlaid reegleid järgiva kirjakeelega sarnanema.

See, et netikeel muutus tavapärase ortograafia pärasemaks (vähemalt Planet Foorumites), tuleb välja peaaegu kõigist keelelistest tunnustest, mida uuriti. Esiteks tuli väikese ja suure algustähe kasutajate vaatlemisest välja, et järjest rohkem hakati kasutama just suurt algustähte, nii püsikui ka tavakasutajate puhul. Aastatel 2006–2008 kasutas püsivalt väikest algustähte vaid 24% püsikasutajatest ning tavakasutajate puhul langes väikese algustähe kasutus 58%-lt 39%-le. Seega üle poolte kasutajate alustas oma lauseid korrektselt. Kuna foorum õigesti kirjutamist ei nõua, siis on see kasutajate endi algatatud. Seos nende kahe kasutajate hulga puhul pole aga piisavalt tugev, et öelda, et nad on üksteisest sõltuvad.

Tähekorduste kohta on selle korpuse põhjal keeruline järeldusi teha, üldiselt ei leidunud neid sisaldavaid sõnu väga palju. Põhilisi tähekordusi, mida kasutati, leidis kogu vaadeldava perioodi jooksul. Võib veel öelda, et mõned tähekordused olid omapärasemad ning neid kasutas väiksem ring inimesi. Enamik populaarseid tähekordustega sõnu väljendasid kas emotsioone (*väääga*, *mmmm*) või jäljendasid kõnekeeles leiduvaid häälightsusi (*aaa*, *eee*), kuid üldiselt selliste sõnade sagedus langes (v.a *aaaa*, mis tegi viimasel vaadeldaval vahemikul suure hüppe).

Peatükis „Võõrtähtede kasutus“ tehti kindlaks, kui paljud korpuses esinevad sõnad sisaldasid tähti *c*, *q*, *w*, *x*, *y* ning ei olnud võõrkeelsed sõnad. Enim kasutatud sõnad neist olid *küll* ning selle järel *nyyd*. Võimalik, et *ü* asendamine *y*-ga tuleneb lihtsalt sellest, millist klaviatuuri kasutatakse. Kuid *ks*-i asendamine *x*-iga on võetud kasutusele kiiruse saavutamiseks trükkimisel. Töös vaadeldi viit enim kasutatavat võõrtähte sisaldanud sõna ning moodustati selle põhjal graaf, mis võimaldas näha, kuidas muutus nende sõnade sagedus. Graafilt selgub, et kõiki selliseid sõnu kasutati aastal 2008 väga palju vähem kui aastal 2001. Kuna ükski neist

leitud sõnadest, mis sisaldas võõrtähti ja ei olnud võõrkeelne, ei ole eesti keeles korrektne, siis võiks nende harvenemine taas näidata, et foorumi kasutajad on hakanud rohkem hoolima õigekirjast.

Lühendite sagedusloend näitas aga teistsuguses suunas liikumist, ingliskeelsete lühendite kasutus tõusis. Algselt vaid eestikeelseid lühendeid vaadates tundus, et kuna neid, mis ei ole eesti keeles korrektsed, hakati kasutama vähem, võiks see justkui olla märk õigekeele paranemisest. Samas aga selgus, et ingliskeelseid lühendeid on hakatud kasutama rohkem. Seega võtavad ingliskeelsed lühendid eestikeelseid üle. Varem sai ka mainitud, et näiteks lühend LOL on lisatud sõnaraamatussegi, mis tähendab, et see peab olema internetis väga laialt levinud. Seetõttu pole imestada, et sellised lühendid on jõudnud ka eestikeelsesesse netisuhtlusesse.

Kogudes kokku kõik teadmised, mis on saadud Planet Foorumite kasutajate keelekasutuse kohta, saab öelda, et keel on muutunud korrektsemaks, kuid on näha inglise keele mõju suurenemist. See paneb mõtlema selle üle, mida meedias tihtipeale väidetakse: et netikeel rikub kirjakeelt ning eriti just noored ei oska enam kirjutada. Planet Foorumite põhjal seda küll ei ütleks, pigem on lugu vastupidi: inimesed on ise hakanud rohkem rõhku panema õigekirjale.

7. EDASISED TÖÖSUUNAD

Juba töö alguses, võttes arvesse bakalaureusetöö mahtu, selgus, et kõiki keelelisi aspekte ei ole võimalik uurida. Siinses töös ei uuritud, miks erinevad muudatused ning uuendused on toimunud. Mõne muudatuse puhul oli võimalik teha eeldusi, kuid selleks, et selgitada välja kindlad põhjused, tuleks uurimist jätkata. Seega on see töö justkui alguspunkt neile, kes soovivad uurida netikeelt veel rohkem süvitsi. Nagu eelmisest peatükist välja tuli, jääb Planet Foorumite korpusest uuenduste tagamaade uurimiseks väheks, täpsemaks analüüsiks tuleks võtta juurde veel korpusi. Korpuste seas võiks siiski leiduda ka foorumeid. Foorumite puhul tuleks aga jälgida, et kasutajad vahelduks tihedalt, kuid samas leiduks seal püsikasutajaid. Kui võimalik, siis tasuks lisada korpusteks ka kommentaariumeid ning jututubasid, kus on nõutud registreerimine, sest siis on võimalik kasutajate kaupa vaadata, kuidas on muutunud nende keeleline käitumine.

Väikese ja suure algustähe kasutuse üle tuleks veel arutleda. Siinses töös tuli selgelt välja, et järjest rohkem on hakatud kasutama lause alguse suurtähte. Jäi aga kindlaks tegemata, kas tegu on tõesti sellega, et inimesed panevad keele õigsusele rohkem rõhku, või on see rohkem moeasi. Veel tuleks leida korpus, mille põhjal oleks võimalik teha kindlaks, kui palju teatud kasutajad mõjutavad teisi. Täheortograafia puhul on muudki uurida. Autor leiab, et vaadeldes näiteks tähekorduste esinemist erinevates suhtlusvõrgustikes, kus on esindatud rohkem vanusegrupe, tuleksid tulemused teised kui käesolevas töös. Kuna töö alguses sai mainitud, et tähekordustega väljendatakse tihti emotsioone, aga sellist väljendusviisi kasutatakse Planet Foorumites küllaltki vähe, siis millist viisi eelistatakse? Kas on selleks emotikonid?

Emotikonide kohta tehtud sagedusloenditest on näha, et emotikone on hakatud vähem kasutama. Kuna selles töös olid korpusteks võetud vaid enam-vähem kindla kasutajaskonnaga foorumid, siis ei olnud võimalik teha kindlaks, millest muudatused on tingitud. Seega võiks edasi uurida, kust need suunad pärinevad. Ivar Põllu [7] väidab, et emotikone kasutatakse tihti sarkastiliselt. Oleks huvitav teada, milliseid emotikone selleks kasutatakse ning kuidas on see aja jooksul muutunud. Seda on aga väga raske kindlaks teha sagedusloendite ja arvutiprogrammide abil, kuna oluline on lausete tähendus. Seega pole võimalik kasutada suuri korpusi. Emotikonide puhul võiks huvi pakkuda ka see, kuidas erineb nende kasutus näiteks foorumite ja privaatsete vestluste puhul: kas privaatsetes vestluses tuntakse end vabamalt ning lisatakse emotikone rohkem ja kergekäelisemalt?

Kuna siinses töös olev korpus on saadud foorumi tekstidest aastatest 2001–2008, ei jõutud uurimisega tänapäeva. Kuna tehti selgeks, et netikeel sellel perioodil muutus, võiks eeldada, et järgneva kaheksa aastaga on see veelgi muutunud. Jäi mulje, et vahemikus 2001–2008 muutus netikeel grammatiliselt korrektsemaks, seega oleks huvitav teada, kas see on süvenenud viimase kaheksa aasta jooksul, jäänud samaks või muutunud hoopis vastupidises suunas. Huvi äratab ka see, kuidas on muutunud emotikonide ja lühendite kasutus: kas emotikonid *rõõmus* (sarnaneb „:)“-ga) ja *naerab* (sarnaneb „:D“-ga) on ikka veel enim kasutatavad? Ka täheortograafias jääb nii mõndagi uurida. Käesolevas töös leiti, milliseid eestikeelseid tähti asendati võõrtähtedega Planet Foorumites aastatel 2001–2008, kuid millised asendused on rohkem kasutatud viimase kaheksa aasta jooksul? Lühenditest oli populaarseim inglise keelest pärit *IMO*, mille tulekuga oli vähenenud enam-vähem sama tähendusega lühendi *IMHO* kasutus. Sel perioodil täiesti uuena võeti kasutusele lühend *FTW*. Seega muudatusi on toimunud palju ning vaevalt jäid need seisma aastal 2008. Enamik lühendeid on võetud just inglise keelest, seega võiks arutada ka, kui palju mõjutab inglise keel eesti netikeelt.

8. KOKKUVÕTE

Bakalaureusetöö eesmärk oli koostada sagedusloendeid ühe kindla ajavahemiku tekstikorpusest ning leida selle põhjal erinevused netikeele kasutuses. Teiseks uurimisküsimuseks oli, kas Planet Foorumitest pärinev korpus esindab paljude või väheste inimeste keelekasutust ehk kas selle korpuse põhjal on võimalik teha üldistusi eestikeelse netikeele muutumise kohta.

Töö käigus leiti vastused uurimisküsimustele. Selleks kirjeldati alguses netikeelele iseloomulikke jooni ning valiti, milliseid neist tuleks lähemalt uurida. Teises peatükis selgitati, miks Planet Foorumid tundus selle töö jaoks hea valik. Anti ka ülevaade sellest, milline korpus välja näeb ning millist informatsiooni see sisaldab. Kuna korpuse maht on küllaltki suur, siis on loodud selle töötlemiseks arvutiprogramm, mille abil moodustati erinevaid sagedusloendeid, tabelleid ning graafe.

Esiteks koguti üldist informatsiooni korpuse kohta, et panna paika korpuse dünaamika. Foorumitest võetud tekstid on pärit aastatest 2001–2008 ning selle perioodi sees on kokku tehtud 317 738 postitust. Kasutajate uurimiseks jagati need kahte rühma: püsikasutajad ehk inimesed, kes on teinud igal aastal vähemalt ühe postituse, ning tavakasutajad ehk kõik ülejäänud. Sagedusloendeid vaadates selgus, et püsikasutajatel on olnud palju suurem osakaal, kuna nad on teinud palju rohkem postitusi, tavakasutajate hulgas leidub palju selliseid, kes on teinud vaid paar postitust. Ka uusi liitujaid on tulnud võrdlemisi vähe, eriti viimastel aastatel (2005–2008). See tähendab, et kasutajad on vahetunud harva, seega saab Planet Foorumite korpuse põhjal teha järeldusi vaid teatud kasutajate kohta, mitte kogu eesti netikeele kohta.

Seejärel uuriti korpuse põhjal keelelisi muutusi: väikese ja suure algustähe kasutust, tähekorduste sagedust, võõrtähtede, lühendite ning emotikonide kasutust. Sellest selgus, et netikeel on vahemikus 2001–2008 hakanud järjest rohkem sarnanema ortograafiliselt korrektse kirjakeelega. Esiteks suurenes järjest lause alguses suure tähe kasutus, mis on eesti keeles korrektne. Lisaks leidis perioodi lõpupoole vähem grammatiliselt ebakorrektsid tähekordusi ning vähenes emotikonide kasutus.

Planet Foorumite 2001.–2008. aasta postituste kohta saab öelda, et kasutajad hakkasid sel perioodil rohkem tähelepanu pöörama, kui korrektselt nad kirjutavad. Seda, kas selline muudatus toimus ka muudes suhtlusvõrgustikes, tuleb uurida edaspidi ning suuremate korpuste põhjal.

VIIDATUD ALLIKAD

- [1] **Hennoste, Tiit; Teder, Merike 2013.** Kuule ma eemale nüüd. – Postimees, 7. detš. <http://arvamus.postimees.ee/2622918/tiit-hennoste-kuule-ma-eemale-nuud> (vaadatud 15.12.2015).
- [2] Planet Foorumite koduleht: <http://forum.planet.ee/> (vaadatud 20.03.2015).
- [3] **Oja, Anni 2006.** Eesti keel internetis. – Keel ja arvuti. Tartu: Tartu Ülikooli Kirjastus, lk 259–267.
- [4] The Oxford English Dictionary. <http://www.oxforddictionaries.com/definition/english/lol?q=LOL>.
- [5] Uue meedia korpus: foorumid, uudisgrupid, kommentaarid: http://www.cl.ut.ee/korpused/segakorpus/uusmeedia/foorumid_uudisgrupid_kommentaarid.php?lang=et (vaadatud 15.03.2016).
- [6] **Gear, Clive Thompson 2016.** The Emoji Is the Birth of a New Type of Language (No Joke). <http://www.wired.com/2016/04/the-science-of-emoji/> (vaadatud 05.05.2016).
- [7] **Põllu, Ivar 1998.** See on väga tõsine lugu :-). – Keel ja Kirjandus, nr 10, lk 724.
- [8] **Hovy, Dirk 2012.** Programming in Python for Linguists. A Gentle Introduction, lk 2. http://www.dirkhovy.com/portfolio/papers/download/pfl_handout.pdf (vaadatud 28.02.2016).
- [9] **Hough, Andrew 2010.** Humans 'subconsciously mimic other accents', psychologists claim. <http://www.telegraph.co.uk/news/science/science-news/7931299/Humans-subconsciously-mimic-other-accents-psychologists-claim.html> (vaadatud 24.04.2016).
- [10] Kiri ja tähestik. Eesti keele käsiraamat. <http://www.eki.ee/books/ekk09/index.php?p=2&p1=1> (vaadatud 24.04.2016).
- [11] **Kaalep, Heiki-Jaan; Muischnek, Kadri; Sirel, Raul 2012.** Korpuslingvistiline lähenemine eesti internetikeele automaatsele morfoloogilisele analüüsile, lk 6. http://www.keelevaab.ee/dict/corpus/comments/uue_meedia_morf.pdf (vaadatud 24.04.2016).
- [12] Täheortograafia. Eesti keele käsiraamat. <http://www.eki.ee/books/ekk09/index.php?p=2&p1=6> (vaadatud 23.04.2016).
- [13] AFK, PTV, LOL – mida see kõik tähendab? 2014. <http://digiabi.ee/2014/01/afk-ptv-lol-mida-see-koik-tahendab/> (vaadatud 23.04.2016).
- [14] Internet Slang And Texting Language. <https://www.italki.com/article/154/lol-lmfao-wtf-figuring-out-english-internet-slang-and-texting-language> (vaadatud 05.04.2016).

[15] 50 Popular Internet Acronyms. <http://www.connexin.net/internet-acronyms.html>
(vaadatud 05.04.2016).

LISA I

Litsents

Lihlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, **Sabiina Haiba**,
(*autori nimi*)

1. annan Tartu Ülikoolile tasuta loa (lihlitsentsi) enda loodud teose
Kuidas on netikeel muutunud aastatel 2001–2008?,
(*lõputöö pealkiri*)

mille juhendaja on **Heiki-Jaan Kaalep**,
(*juhendaja nimi*)

- 1.1. reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace'is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;
- 1.2. üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace'i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.
2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.
3. kinnitan, et lihlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartus **16.05.2016**