

**This is an electronic reprint of the original article.
This reprint *may differ* from the original in pagination and typographic detail.**

Author(s): Nousiainen, Tuula

Title: Miten torjua vihapuhetta sosiaalisen median palveluissa?

Year: 2017

Version:

Please cite the original version:

Nousiainen, T. (2017). Miten torjua vihapuhetta sosiaalisen median palveluissa?. In S. Elo, K. Kaihari, P. Mattila, & L. Nissilä (Eds.), *Rakentavaa vuorovaikutusta : opas demokraattisen osallisuuden vahvistamiseen, vihapuheen ja väkivaltaisen radikalismien ennaltaehkäisyyn* (pp. 29-33). Opetushallitus. Oppaat ja käsikirjat, 1/2017. http://oph.fi/download/182479_rakentavaa_vuorovaikutusta.pdf

All material supplied via JYX is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

MITEN TORJUA VIHAPUHEHTA SOSIAALISEN MEDIAN PALVELUISSA?

Tuula Nousiainen

Viimeaikainen keskustelu vihapuheesta ja sen leviämisestä verkossa on tuskin jäänyt keneltäkään huomaamatta. Se yhdistettynä nuorten erittäin aktiiviseen sosiaalisen median käyttöön on yhtälö, joka tekee mediakasvatuksesta yhä tärkeämpää ja asettaa sille isoja haasteita. Syksyllä 2016 julkaistun SoMe ja nuoret 2016 -tutkimuksen³ perusteella nuoret itsekkin ovat tästä tietoisia. Yli 70 % kyselyyn vastanneista 13–29-vuotiaista nuorista oli nimittäin sitä mieltä, että joko koulussa tai kotona tulisi opettaa sosiaalisen median käyttöä, erityisesti netissä käyttäytymistä ja turvallisuutta.

Tämä artikkeli pyrkii osaltaan vastaamaan tähän tarpeeseen nostamalla esiin käytännön keinoja vihapuheeseen ja ei-toivottuun verkkokäyttäytymiseen reagoimiseen sosiaalisessa mediassa. Sisältö rakentuu kolmen teeman mukaan. Ensimmäisessä osiossa luodaan katsaus työkaluihin, joita sosiaalisen median palvelut tarjoavat asiattomalta käyttäytymiseltä suojautumiseen. Toisessa osassa käsitellään asiattoman sisällön ilmiantamista näissä palveluissa. Ja kolmannessa osassa tarkastellaan lyhyesti sosiaalista mediaa ns. vastapuheen areenana.

Asiattomalta verkkokäyttäytymiseltä suojautuminen

Eri sosiaalisen median välineet tarjoavat tapoja sekä välittömään, henkilökohtaiseen ei-toivotulta sisällöltä suojautumiseen että tällaisten sisältöjen ilmiantamiseen, jotta ne poistuisivat kaikkien käyttäjien näkyviltä. Henkilökohtaisen suojautumisen kannalta on tärkeää tuntea keskeisimmät yksityisyysasetukset sekä käyttäjien estämiseen liittyvät toiminnot.

Sosiaalisen median käyttäjätilin yksityisyysasetuksilla jokainen käyttäjä pystyy rajaamaan sitä, kenelle hänen jakamansa sisällöt näkyvät ja mitä tietoja hänestä voidaan löytää esimerkiksi hakutoiminnon kautta. Palveluun rekisteröidytessä on hyvä tarkistaa, millaiset yksityisyysasetukset palvelussa on oletuksena ja tutustua siihen, miten niitä voidaan muokata. Joissakin palveluissa, kuten Facebookissa, on määritelty erikseen oletusasetukset alaikäisille käyttäjille: Facebookin hakutoiminto ei näytä alaikäisten käyttäjien yhteystietoja, koulua ja syntymäaika julkisissa hakutuloksissa, ja sijaintitietojen jakaminen on oletuksena kytketty pois päältä⁴.

Käyttäjä voi suojautua asiattomalta sisällöltä myös estämällä yksittäisiä käyttäjiä ottamasta häneen yhteyttä ja näkemästä hänen päivityksiään. Käyttäjä voi estää myös tiettyjen käyttäjien tuottamien tai jakamien sisältöjen tuleminen hänen itsensä nähtäville. Koska eri sosiaalisen median palvelut perustuvat erityyppiseen ja erilaajuisille yleisöille jakamiseen, myös yksityisyysasetusten muokkausmahdollisuudet ja käyttäjien estämisen tavat vaihtelevat.

3 1 ebrand Suomi Oy & Oulun kaupungin sivistys- ja kulttuuripalvelut. [2016]. SoMe ja nuoret 2016. Suomessa asuvien 13–29-vuotiaiden nuorten sosiaalisen median palveluiden käyttäminen ja läsnäolo. <http://www.ebrand.fi/somejanuoret2016>

4 2 Facebook. Alaikäiset ja yksityisyys. <https://www.facebook.com/help/473865172623776>

Esimerkiksi Facebookissa⁵ suurin osa vuorovaikutuksesta perustuu rajatulle ja käyttäjän itsensä hyväksymälle yleisölle jakamiseen. Jos epäasiallista sisältöä jakanut henkilö on käyttäjän kaverilistalla, hän voi poistaa tämän kavereistaan sekä estää tätä lähettämästä hänelle yksityisviestejä. Helpoin tapa on käyttää esto-toimintoa; tällöin estetty käyttäjä poistuu kaverilistalta ja suora vuorovaikutus (viestit, kaveripyynnöt, tapahtumakutsut, julkaisuihin merkitseminen) on estetty.

Toimintoa voidaan käyttää myös muiden kuin omalla kaverilistalla olevien käyttäjien estämiseen. Jos käyttäjä ei halua poistaa jotakuta kokonaan kaverilistaltaan, hän voi lisätä tämän rajoitettujen käyttäjien listalle, jolloin tämä näkee hänen julkaisunsa vain, jos ne on määritelty julkisiksi. Käyttäjä voi myös säätää sitä, missä määrin tietyn käyttäjän tai sivun jakamia sisältöjä tulee näkyviin hänen omaan uutisvirtaansa joko lopettamalla kokonaan kyseisen henkilön tai sivun seuraamisen tai määrittämällä haluavansa nähdä jatkossa vähemmän tämän tahon jakamia sisältöjä. Kaikki nämä toiminnot on kerätty myös kootusti *pidä tauko* -toiminnon alle, jolloin tietyn henkilön osalta voidaan kerralla määrittää sekä hänen päivitystensä näkyminen itselle, omien päivitysten näkyminen hänelle että sellaisten aiempien julkaisujen näkyminen muille, joissa kumpikin on mukana.

Myös Twitterissä ja Instagramissa on mahdollista määrittää käyttäjätili täysin yksityiseksi. Näiden palveluiden, samoin kuin esimerkiksi YouTuben, perusidea kuitenkin pohjautuu Facebookia vahvemmin julkiseen jakamiseen, joten yleisön rajaamisen ja yksityisyysasetusten hienosäätämisen mahdollisuuksia on tarjolla hieman rajatummin. Yksittäisten käyttäjätilien estäminen on mahdollista kaikissa näissä palveluissa. Twitterissä⁶ estetty käyttäjä ei näe hänet estäneen henkilön twiittejä (suoraan profiilissa eikä hauissa), ei voi lähettää tälle yksityisviestejä eikä seurata tätä. On kuitenkin tärkeää huomioida, että jos tili on julkinen, twiitit näkyvät kaikille – myös estetyille henkilölle silloin, kun tämä ei ole kirjautuneena estetylle tililleen. Instagramissa⁷ esto toimii pitkälti vastaavasti, mutta on hyvä muistaa, että selaimen kautta Instagraminkin julkisia tilejä ja sisältöjä on mahdollista katsella kirjautumatta sisään, vaikka kommenttien lähettäminen ei kirjautumattomana onnistukaan. Myös Snapchatissa⁸ käyttäjiä voi poistaa kavereista sekä estää ottamasta yhteyttä. YouTubeessa⁹ on mahdollista estää käyttäjää kommentoimasta omia videoita ja kanavaa sekä lähettämästä yksityisviestejä.

Asiattoman verkkokäyttäytymisen ilmiantaminen

Edellä on esitelty yksinkertaisimmat tavat, joilla käyttäjä voi välittömästi vaikuttaa siihen, missä määrin hän itse suoraan joutuu tekemisiin ei-toivotun sisällön kanssa omilla käyttäjätileillään. Suuri osa sosiaalisen median asiattomasta sisällöstä ja vihapuheesta esiintyy kuitenkin muualla kuin suoraan yksittäiseen käyttäjään itseensä kohdistuvana, kuten toisten henkilöiden Twitter-päivitysten, käyttäjän seuraamien Facebook-sivujen ja -ryhmien sekä esimerkiksi YouTube-videoiden kommentteissa. Edellä kuvatut henkilökohtaisen suojautumisen toiminnot ovat peruskäyttäjille yleensä ainakin yleisellä tasolla tuttuja: tiedostetaan omien yksityisyysasetusten tärkeys sekä tiedetään keskeisimmät tavat, joilla voidaan torjua ei-toivotun sisällön esiintymistä omassa viestivirrassa.

5 Facebook. Mitä estäminen tarkoittaa? Mitä tapahtuu, kun estän jonkun? <https://www.facebook.com/help/131930530214371>; Facebook. Jonkun poistaminen kavereista tai estäminen <https://www.facebook.com/help/1000976436606344>

6 Twitter. Blocking accounts on Twitter. <https://support.twitter.com/articles/117063>

7 Instagram. Henkilöiden estäminen. <https://help.instagram.com/426700567389543>

8 Snapchat. Kaverien poistaminen ja estäminen. <https://support.snapchat.com/fi-FI/a/block-friends>

9 YouTube. Käyttäjien estäminen. <https://support.google.com/youtube/answer/56113>

Moni tietää myös, että asiatonta sisältöä – huolimatta siitä, missä se esiintyy – on mahdollista ilmiantaa palvelun ylläpidolle parilla klikkauksella. Ilmiantotoiminto saattaa kuitenkin näyttäytyä käyttäjälle ”mustana laatikkona”, eikä hän välttämättä ole tietoinen siitä, millainen prosessi sisällön ilmiannon yhteydessä todellisuudessa käynnistyy. Suurilla sosiaalisen median palveluilla on käytössä laajat resurssit: ilmiannettuja sisältöjä työkseen läpikäyvissä, ympärivuorokautisesti toimivissa tiimeissä työskentelee satoja eri kieli- ja kulttuuritaustoista tulevia ihmisiä eri puolilla maailmaa. Pelkän kielitaidon lisäksi on osattava sijoittaa ilmiannettu sisältö kontekstiinsa ja arvioida siinä valossa, onko se palvelun sääntöjen vastainen. Kansallisten kontekstien ymmärtäminen on olennaisessa roolissa siinä, että ilmiantojen käsittelijät tunnistavat, onko jokin kommentti vihapuheeksi luokiteltava tai muuten sävyiltään asiaton. Ilmiannettujen sisältöjen arvioinnin ja niihin tarvittaessa puuttumisen perustana ovat sosiaalisen median palveluiden määrittelemät säännöt ja yhteisönormit. Vihapuheella viitataan yhteisönormeissa halventavaan sisältöön, joka kohdistuu yksilöihin tai ryhmiin mm. rodun, etnisyyden, kansallisen alkuperän, uskonnon, seksuaalisen suuntautumisen, sukupuolen, sukupuoli-identiteetin, iän, vamman tai sairauden perusteella¹⁰.

- Facebookin yhteisönormit: <https://www.facebook.com/communitystandards>
- YouTuben yhteisösäännöt: <https://www.youtube.com/yt/policyandsafety/communityguidelines.html>
- Twitterin säännöt: <https://support.twitter.com/articles/18311>
- Instagramin yhteisösäännöt: <https://help.instagram.com/477434105621119>
- Snapchat-yhteisön säännöt: <https://support.snapchat.com/fi-FI/article/guidelines>

Esimerkiksi YouTube¹¹ ja Facebook¹² kuvaavat ilmiantoprosesseitaan omilla ohjesivustoillaan. Ilmiantaa voi sekä yksittäisiä julkaisuja ja kommentteja että kokonaisia käyttäjätilejä. Ilmiannon yhteydessä kysytään syytä – usein sekä valmiiksi annettuina vastausvaihtoehtoina että vapaamuotoisena kuvauksena. Varsinkin siinä tapauksessa, että ilmiannettavan sisällön asiattomuuden ymmärtäminen vaatii olennaisesti tapauskohtaisen kontekstin tuntemista, ilmiannon yhteydessä on tärkeää antaa mahdollisimman tarkka kuvaus ja selkeät perustelut, miksi sisällön suhteen tulisi ryhtyä toimenpiteisiin.

Esimerkiksi päällepäin neutraalilta näyttävä kommentti sosiaalisessa mediassa voi todellisessa kontekstissaan olla osa laajempaa häirintää, nettikiusaamista tai vihakäyttäytymistä, ja vaikka sisältöjä käsittelevissä tiimeissä osataan kymmeniä eri kieliä ja tunnetaan lukuisten maiden kulttuurisia konteksteja, tällaiset paikalliset ja tilannekohtaiset taustat jäävät piiloon, ellei asiattoman kommentin ilmiannon yhteydessä ole avattu syytä ilmiantoon. Toisaalta taas ihmiset toisinaan myös jakavat jonkun muun tuottamaa vihapuhetta mm. siinä tarkoituksessa, että sitä esittämällä halutaan lisätä tietoisuutta vihapuheesta. Tällaisissa tapauksissa sisältöä ei toki ole tarkoituksenmukaista poistaa, mutta tällöin julkaisun tarkoitusperä on ilmaistava selkeästi¹³.

Ilmiannetut sisällöt ja käyttäjätilit käydään läpi tapaus kerrallaan, ja tarvittaessa ilmiantojen käsittelijät ovat myös yhteydessä käyttäjään, jota ilmianto koskee. Ilmiantajan henkilöllisyys pidetään aina luottamuksellisena. Mikäli sisältö todetaan yhteisönormien vastaiseksi, ilmianto voi johtaa sisällön poistamiseen tai siitä vastuussa olevan käyttäjätilin sulkemiseen; seuraukset riippuvat mm. rikkomuksen vakavuudesta ja kyseessä olevan käyttäjän aiem-

10 Facebook. Toisten kunnioittamiseen kannustaminen: Vihapuhe. <https://www.facebook.com/communitystandards#hate-speech>; YouTube. Vihamielinen sisältö. <https://support.google.com/youtube/answer/2801939>; Twitter. Vihamielistä käytöstä koskeva käytäntö. <https://support.twitter.com/articles/20175050>

11 YouTube. Ilmoitus- ja valvontakeskus. <https://www.youtube.com/yt/policyandsafety/reporting.html>

12 Facebook. Ilmianna jotakin. <https://www.facebook.com/help/263149623790594>

13 Facebook. Toisten kunnioittamiseen kannustaminen: Vihapuhe. <https://www.facebook.com/communitystandards#hate-speech>

mista yhteisönormien rikkomuksista. Ilmiantojen määrä ei suoraan vaikuta siihen, poistetaanko sisältö, mutta se voi auttaa käsittelijöitä priorisoimaan ilmiantojen käsittelyjärjestystä. Palveluilla on myös tiettyjä automatisoituja ilmiantoihin perustuvia toimintoja; esimerkiksi YouTube:ssa riittävän monen käyttäjän roskapostiksi tai häirinnäksi merkitsemät kommentit piilotetaan automaattisesti näkyviltä klikkauksen taakse.

Sosiaalinen media ja vastapuhe

Vaikka sosiaalinen media tarjoaakin usein valitettavan helpon alustan vihapuheen levittämiselle ja suuri osa verkkovihapuheeseen reagoinnista keskittyy siltä suojautumiseen, suurimmilla sosiaalisen median palveluilla on toisaalta parhaimmillaan myös paljon vaikutusvaltaa vihamielisen retoriikan vastustamisessa niin sanotun vastapuheen avulla. Vastapuheella (counter-speech) tarkoitetaan sitä, että käyttäjät vastaavat vihapuheeseen ja haastavat sitä monin eri tavoin samoilla areenoilla, joilla vihapuhettakin esiintyy.¹⁴

Erityisesti Twitterissä esiintyvään vihapuheeseen ja siihen vastaamiseen liittyen on julkaistu ytimekäs *Stopping Hate: How to Counter Hate Speech on Twitter?* -opus¹⁵. Oppaassa jaetaan käytännönläheisiä ohjeita ja vinkkejä vastapuheen esittämiseen vaikuttavasti ja turvallisesti aina tavoitteiden pohtimisesta argumenttien valmisteluun, viestien muotoiluun sekä tehokkaisiin hashtagien käyttötapoihin. Opas myös kehottaa huolehtimaan turvallisuudesta ja toimimaan vastapuheen esittämisessä mahdollisuuksien mukaan osana jotakin laajempaa yhteisöä tai ryhmää kuin toimimaan täysin yksin.

YouTube puolestaan on käynnistänyt vuoden 2016 lopulla *Creators for Change*¹⁶ -nimisen aloitteen, joka pyrkii saamaan aikaan positiivista vaikutusta yhteiskunnassa. Kampanjassa joukko eri puolilta maailmaa tulevia ja eri kulttuuritaustoja edustavia suosittuja tubettajia tuottaa ja jakaa omilla kanavillaan videoita, joissa he lähestyttävällä tavalla nostavat esiin erilaisia yhteiskunnallisia näkökohtia, pyrkien lisäämään tietoisuutta ja edistämään keskustelua näistä ilmiöistä. Teemat liittyvät mm. vihapuheen, muukalaisvastaisuuden ja ekstremismin vastustamiseen sekä yleisemmin empatian ja suvaitsevaisuuden edistämiseen. *Creators for Change* -ohjelmalla on tarkoituksena myös käynnistää ja tukea paikallisia kampanjoita yhteistyössä eri organisaatioiden kanssa.

Lopuksi

Sosiaalisen median palveluihin kohdistuu yhä enemmän painetta ottaa vastuuta vihapuheen kitkemisestä. Kesällä 2016 Facebook, Twitter, Google ja Microsoft sopivat EU:n kanssa käytännöt, joissa ne lupaavat käsitellä alle vuorokaudessa suurimman osan niille ilmiannetusta vihapuheesta.¹⁷ Google, Facebook ja Twitter tekivät vastaavanlaisen sopimuksen jo aiemmin

14 Bartlett, J. & Krasodowski-Jones, A. (2015). Counter-speech examining content that challenges extremism online. London: Demos. <https://www.demos.co.uk/wp-content/uploads/2015/10/Counter-speech.pdf>

15 Iganski, P. & Sweiry, A. (2016). Stopping Hate: How to Counter Hate Speech on Twitter? Media Diversity Institute. <http://www.stoppinghate.getthetrollsout.org>

16 YouTube. Creators for Change. <https://www.youtube.com/yt/creators-for-change>; YouTube blog. Introducing YouTube Creators for Change. <https://youtube.googleblog.com/2016/09/introducing-youtube-creators-for-change.html>

17 The Guardian. (2016). Facebook, YouTube, Twitter and Microsoft sign EU hate speech code. 31.5.2016. <https://www.theguardian.com/technology/2016/may/31/facebook-youtube-twitter-microsoft-eu-hate-speech-code>

Saksassa, mutta seurantatietojen mukaan yhtiöt eivät toistaiseksi ole päässeet tavoitteissa tyydyttävälle tasolle, ja niitä voivatkin uhata sanktiot, ellei tilanne parane.¹⁸

Koska verkkovihapuhe on koko ajan kasvava haaste sosiaalisen median palveluille, ne joutuvat joka tapauksessa jatkuvasti kehittämään uusia tapoja ei-toivotulta sisällöltä suojautumiseen ja sen poistamiseen. Esimerkiksi Twitter tiedotti äskettäin parannetuista sisältöjen suodatusmahdollisuuksista, ilmiantoprosessin kehittämisestä sekä ilmiannettuja sisältöjä läpikäyvien työntekijöidensä lisäkoulutuksesta mm. kulttuurisiin ja historiallisiin konteksteihin liittyen.¹⁹ Kannattaakin säännöllisesti seurata käyttämiensä verkkopalveluiden ohjesivustoja, blogeja ja muita tiedotuskanavia, joissa kerrotaan mahdollisista uusista toiminnallisuuksista ei-toivottuun sisältöön reagoimiseksi.

Vaikka useimmat vihapuheen ja asiattoman sisällön torjumiseen liittyvistä ongelmista ja ratkaisuvaihtoehdoista (kuten estäminen ja ilmiantaminen) ovat yhteisiä kaikille yhteisöpalveluille, eri palveluissa korostuvat kuitenkin erilaiset haasteet. Esimerkiksi Facebookissa, jossa valtaosa käyttäjistä on omilla nimillään, erityisen tärkeiksi nousevat käyttäjän yksityisyyden ja henkilökohtaisten tietojen suojaamiseen liittyvät toimenpiteet etenkin silloin, jos käyttäjä on joutunut ei-toivotun käyttäytymisen kohteeksi. Twitterissä puolestaan päähaasteet liittyvät anonymiteetin tarjoamiin mahdollisuuksiin vihapuheen suoralle, aggressiiviselle ja laajamittaiselle levittämiseksi. Mediakasvatus on tärkeässä roolissa siinä, että käyttäjät – erityisesti nuoret – ymmärtävät erityyppisten ympäristöjen erityispiirteet, osaavat ottaa ne huomioon ja mukauttaa toimintaansa niiden mukaan.

Edellä on kerrottu sosiaalisen median tarjoamista työkaluista vihapuheelta suojautumiseen sekä sen estämiseen ja vastustamiseen verkossa yleisesti, mutta erityisesti paikallisella tasolla ilmenevään vihapuheeseen, nettikiusaamiseen ja muuhun asiattomaan verkkokäyttämiseen puuttumisessa ja sen ennaltaehkäisemisessä keskeisimmässä roolissa ovat kasvattajat ja muut paikalliset toimijat.

Aiheeseen liittyen kannattaa tutustua erityisesti Mediakollektiivin ja Koulukinon tuoreeseen *Disinformaatio, vihapuhe ja mediakasvatuksen keinot* -selvitykseen²⁰, joka on ajankohtainen ja kattava katsaus mediakasvatuksen tarjoamista näkökulmista ja keinoista vihapuheen kitkemiseksi. Julkaisussa pohditaan eri tahojen (mm. koti, koulu, media, viranomaiset, poliittiset toimijat, kulttuuri) rooleja ja vastuuta sekä tarjotaan laaja aineistoluettelo, josta löytyy sekä opetusaineistoja että monentyyppisiä muita aiheeseen liittyviä resursseja ilmiöiden ymmärtämiseksi ja niihin puuttumisen tueksi.

18 The Guardian. (2016). Germany to force Facebook, Google and Twitter to act on hate speech. 17.12.2016. <https://www.theguardian.com/technology/2016/dec/17/german-officials-say-facebook-is-doing-too-little-to-stop-hate-speech>

19 Twitter blog. Progress on addressing online abuse. <https://blog.twitter.com/2016/progress-on-addressing-online-abuse>

20 Töyry, M. (toim.) (2016). Disinformaatio, vihapuhe ja mediakasvatuksen keinot. Osuuskunta Mediakollektiivi ja Koulukino. <http://koulukino.fi/index.php?id=3705>