ETLA

Working Papers | 151

16.10.2026

Miten väestösokit näkyvät *Suomi24*-keskustelupalstalla?

Juha Alho
Helsingin yliopisto
juha.mikko.alho@gmail.com

Julian Lin
Finland Pure-Fountain Ltd.
julian@finlandpurefountain.com

Suositeltava lähdeviittaus:
Alho, Juha & Lin, Julian (16.10.2026). ”Miten väestösokit näkyvät *Suomi24*-keskustelupalstalla?”. ETLA Working Papers nro 151.
https://pub.etla.fi/ETLA-Working-Papers-151.pdf

Julkivapaa perjantaina 16.10.2026 klo 00.01.

Tiivistelmä

Hedelmällisyyden lasku 2010–2024, koronaepidemia 2020–2022 ja nettomuuton jyrkkä kasvu 2022–2024 ovat tuoreita esimerkkejä väestösokeista. Miten sokit on kansan keskuudessa ymmärretty?

Suomi24-keskustelupalsta on toiminut vuoden 2001 alusta. Tarkastelun kohteeksi luotiin neljä aiheryhmää. Viestejä oli runsaat 1,26 miljoonaa, ja ne muodostavat tutkimuksen korpuksen. Aiheryhmissä laskettiin substantiivien ja adjektiivien frekvenssit ja valittiin vajaan 8 000 yleisimmän yhdelmä sanastoksi. Kussakin aihepiirissä havaittujen purskeperiodien kieltä ja sen muutosta tutkittiin kuvaamalla sanat neliulotteiseen avaruuteen. Analyysiä täydennettiin lukemalla otospohjalla yhteensä runsaat 2 000 viestiä. Oman luentamme ohella käytimme Gemini-tekoälyä tuottamaan yhteenvetoja otosviestien sisällöstä.

Eri aihepiirien viestipurskeiden synnyllä on väljä yhteys kiinnostuksen kohteena oleviin väestönmuutosprosesseihin. Syntyvyyden katsotaan johtuvan asenteiden muutoksesta. Kuolleisuuskeskusteluita dominoi koronaepidemia, vaikka se ei merkittävästi vaikuttanut kuolevuuteen. Krooniset sairaudet ja eliniän kasvu saivat vain vähän huomiota. Muuttoliikekeskusteluissa painottui vuoden 2015 pakolaisryntäys. Sen sijaan Ukrainasta tullutta pakolaisvirtaa käsitellään vähän.

Abstract

How Demographic Shocks Show Up in *Suomi24* Discussion Forum?

The collapse of fertility in 2010–2024, corona pandemic in 2020–2024, and the sharp increase of net migration in 2022–2024, are fresh examples of population shocks. How have they been understood by the population at large?

Suomi24 discussion forum has been in operation since 2001. Four topic groups were formed. A corpus of 1.26 million posts was formed. The frequencies of nouns and adjectives were counted in the topic groups. On this basis, a vocabulary of approximately 8,000 lemmas was determined.

In each topic group, the language used and its change, was studied by embedding the lemmas into a four-dimensional Euclidean space. The analysis was complemented by reading, on a sample basis, about 2,000 posts. Gemini AI was also used to provide summaries of the posts we had read.

The bursts in the flow of posts were only indirectly related to the demograhic shocks. Changes in fertility were thought to be caused by changes in attitudes. Corona dominated the discussions on mortality, although its effect on mortality was minor. In discussions on migration, the major burst was related to the refugee crisis of 2015. In contrast, the wave of immigrants provoked by the Ukrainian war was less discussed.

Ph.D. **Juha Alho** on emeritusprofessori Helsingin yliopistossa ja Elinkeinoelämän tutkimuslaitoksen Senior Fellow.

Ph.D. **Julian Lin** on tutkija Finland Pure-Fountain Oy:ssä.

Ph.D. **Juha Alho** is a Professor Emeritus at University of Helsinki and a Senior Fellow at ETLA Economic Research.

Ph.D. **Julian Lin** is a Researcher at Finland Pure-Fountain Ltd.

Kiitokset: Tämä työpaperi on tuotettu osana Suomen Akatemian strategisen tutkimuksen neuvoston rahoittamaa *Väestörakenteen muutosten elämänkaari- ja talousvaikutukset (LIFECON)* -projektia (nro 345218), ja se pohjaa aiemman Citizen Mindscapes -hankkeen (Suomen Akatemian päätös 292906) puitteissa tehdylle tietokantatyölle. Alustavia tuloksia esitettiin ensi kertaa Kielitieteen päivillä Joensuussa toukokuussa 2026. Kiitämme Ismo Söderlingiä maahanmuuttoa koskevista lähteistä.

Tekoäly Geminiä on käytetty kirjoituksessa esitetyllä tavalla yhteenvetojen tekemiseen otosviesteistä. Muuten tekoälyä ei ole käytetty.

Acknowledgements: This working paper is part of *Life course and economic implications of demographic change (LIFECON)* project, funded by the Academy of Finland Strategic Research Council (decision 345218). It uses earlier database work in project Citizen Mindscapes (Finnish Academy decision 292906). Preliminary results were first presented at Kielitieteen päivät -symposium, in Joensuu, May 2026. We thank Ismo Söderling for references related to immigration.

Gemini AI has been used to summarize sample posts, as described in the report. No other use of AI.

Asiasanat: Järjestysluku, Kuolleisuus, Lemma, Muuttoliike, Some, Syntyvyys, Tekoäly, Token

Keywords: Order statistics, Mortality, Migration, Social media, Fertility, AI, Token

JEL: C83, J10, N34, Z13

Tiivistelmä

Väestö on varanto, jonka koko ja ikärakenne muuttuvat syntyvyyden, kuolleisuuden ja muuttoliikkeen myötä. Muutosvirtoja ei julkisuudessa yleensä kuvata tieteellisin käsittein. Tietoon pohjaavan yhteiskunnallisen päätöksenteon oikeuttamisen takia olisi kuitenkin tärkeää, että tasomuutosten vaikutukset tulisivat ymmärretyiksi. Tuoreita esimerkkejä väestösokeista ovat hedelmällisyyden romahdusmainen lasku 2010-2024, koronaepidemia 2020-2022 ja nettomuuton jyrkkä kasvu 2022-2024. Miten sokit on kansan keskuudessa ymmärrretty?

Suomi24-keskustelupalsta on toiminut nykymuotoisena vuoden 2001 alusta. Sen tiedot on päivitetty tutkimuskäyttöön vuoden 2023 loppuun asti. Kuka tahansa voi osallistua, mutta kirjoittajien henkilöllisyys on pääsääntöisesti tuntematon. Esitystapaa ja käsitteitä valitessaan kirjoittajat joutuvat kuitenkin ottamaan huomioon lukijoiden kielitajun. Jos viestien käsitteistö ja kieli muuttuvat sokkien tapahtuessa, se voi kertoa siitä, miten sokit on kansan keskuudessa tulkittu, ja mitä niiden seurauksista pidetään tärkeinä. Toki on selvää, ettei keskustelupalstaa perustettu väestön muutosvirtojen keskustelupaikaksi.

Suomen väestörekisterijärjestelmä tuottaa muutosprosesseista tiedot kuukausittain. Tästä syystä myös *Suomi24*-aineisto on ryhmitelty kuukausittain. Kielipankki ylläpitää aineistoa CSC:n supertietokoneilla, missä sitä voi käyttää Korp-käyttöliittymän kautta, mutta se voidaan myös viedä mm. **R**-ohjelmiston tiedostomuotoon. Aineistolle on tehty automatisoitu lauseenjäsennys (Turku Dependency Treebank), joka tuottaa mm. sanan perusmuodon ja sanaluokan.

Tarkastelun kohteeksi luotiin neljä aiheryhmää sanan tai sanan osien perusteella. Viestejä oli runsaat 1.26 miljoonaa vuosina 2001-2023, ja ne muodostavat tutkimuksen korpuksen. Aiheryhmissä laskettiin substantiivien ja adjektiivien frekvenssit ja valittiin vajaan 8 000 yleisimmän yhdelmä sanastoksi.

Väestöaineistossa tiedettiin olevan merkittäviä sokkeja, ja viesteissä havaittiin purskeisuutta. Tämän takia aineisto jaettiin kuuteen periodiin. Frekvenssijakaumat olivat silti huomattavan vinoja, minkä takia vertailuja varten esiintymisfrekvenssit korvattiin niiden järjestysluvuilla. Sanojen järjestyslukuja aiheryhmissä verrattiin niiden järjestyslukuihin koko aineistossa. Näin sekä ns. stop-sanojen että harvinaisten sanojen vaikutus vertailuihin voitiin minimoida, ja ryhmiä parhaiten erottelevat sanat saatiin määritettyä.

Kussakin aihepiirissä havaittujen purskeperiodien kieltä ja sen muutosta tutkittiin kuvaamalla sanat neliulotteiseen avaruuteen. Tällä keinolla voitiin graafisesti tutkia sanojen frekvenssien muutoksia, ja löytää aihepiirejä eri periodeilla erottelevia sanoja. Kuvaus suoritettiin laskemalla singulaariarvohajoitelma viestien järjestysluvuista muodostetulle matriisille. Tuloksia kuvattiin dynaamisella grafiikalla, sanapilvillä ja niiden muutoksilla. Tuloksena voitiin osoittaa, että sanojen frekvenssit muuttuvat ymmärrettävällä tavalla periodista ja aihepiiristä toiseen.

Tilastollista sanapussimalliin perustuvaa analyysiä täydennettiin kahdella tavalla. Ensinnäkin luimme otospohjalla yhteensä runsaat 2 000 viestiä. Tämän pohjalta voitiin arvioida kirjoittajakunnan rakennetta ja argumentaatiotapaa. Aristoteleen retoriikan kolmijakoa käytettiin jäsentämään muuten melko työlästä luentaa. Oman luentamme ohella käytimme Gemini-tekoälyä tuottamaan yhteenvetoja otosviestien sisällöstä. Tekoäly osoittautui nopeaksi, mutta sen tuloksiin liittyy epäselvyyksiä.

Keskeisin empiirinen tulos tutkimuksesta on, että eri aihepiirien viestipurskeiden synnyllä on vain väljä yhteys kiinnostuksen kohteena oleviin väestönmuutosprosesseihin. Syntyvyyden katsotaan johtuvan asenteiden muutoksesta, mutta keskusteluissa pääpaino on esim. naisten aseman ja perheen roolin muuttumisessa. Vaikutus syntyvyyteen tulee mukaan ”oheistuotteena”. Odotetusti kuolleisuuskeskusteluita dominoi koronaepidemia, vaikka se ei merkittävästi vaikuttanut kuolevuuteen. Pääosa keskusteluista käsittelikin kuolleisuuden sijaan esim. rokotuksia ja niiden sivuvaikutuksia. Krooniset sairaudet ja eliniän kasvu saivat vain vähän huomiota. Muuttoliikekeskusteluissa painottui vuoden 2015 pakolaisryntäys, vaikka sen vaikutus nettomuuttoon oli melko pieni. Sen sijaan Ukrainasta

tullutta pakolaisvirtaa käsitellään vähemmän, vaikka se näyttäisi jättävän isomman jäljen väestöhistoriaamme.

Keskusteluihin osallistujat muodostavat varsin heterogeenisen joukon. Osa näkee paljon vaivaa pysyäkseen perillä ilmiöiden syistä. He vetoavat argumenteissaan tietoon. Tätä useammat pohjaavat argumenttinsa omiin uskomuksiinsa ja asenteihinsa, ei niinkään selkeästi osoitettavaan tietoon. Kolmas pienempi ryhmä käyttää keskusteluita tunneperäisten herjaavien kommenttien esittämiseen, joissa voimasanoilla korvataan väitteiden perustelut.

Osana sosiaalista mediaa *Suomi24* muodostaa uuden tietolähteen kansalaisten muuttuvien kiinnostuksenkohteiden seuraamiseen. Tekstimuotoisen datan tilastolliset ja tietotekniset käsittelymenetelmät ovat kehittyneet siihen pisteeseen, että kohtuullisella vaivalla voidaan selvittää, millaisia muutoksia tapahtuu, lähes reaaliajassa. Raportin lopussa pohditaan näin saatavan tiedon luonnetta verrattuna perinteisiin tilastollisiin luotauksiin ja tekoälyn mahdollisuuksia tehostaa tutkimuksen suorittamista.

Abstract

Population is a stock, whose size and age structure are determined by births, deaths, and migration. The flows are typically not described in scientific terms, in public discourse. However, for knowledge-based decision making to be socially acceptable, it would be important that level changes in flows would be correctly understood. The collapse of fertility in 2010–2024, corona pandemic in 2020–2024, and the sharp increase of net migration in 2022–2024, are fresh examples of population shocks. How have they been understood by the population at large?

Suomi24 discussion forum has been in operation, in current form, since 2001. The data have been updated to the end of 2023, for research use. Anyone can participate, but the identities of the participants are mostly unknown. When deciding on their linguistic expression and concepts, the paticipants are compelled to take into account the linguistic competencies of their readers. Changes in the concepts and laguage used, when shocks occur, may be indicative of how the demographic changes have been understood, and which of their consequences are deemed relevant. To be sure, the discussion forum has not been established for this very purpose.

Finland's population registry provides monthly data on the demographic flows. Therefore, we chose to consider *Suomi24* posts on a monthly basis, as well. Language Bank maintains the discussion forum data in a super computer of CSC, where it can be used via Korp interface. But, it can also be uploaded into R-environment. The data have been preprocessed by an automated parser (Turku Dependency Treebank) that lemmatizes the text, and determines the part of speech, for example.

For our study, four topic groups were formed, based on text tokens. A corpus of 1.26 million posts was formed. The frequencies of nouns and adjectives were counted in the topic groups. On this basis, a vocabulary of approximately 8,000 lemmas was determined.

The demographic data were known to include shocks, and the flow of monthly posts turned out to contain bursts. On this basis, the observation period was divided into six sub-periods. Yet, the frquency distributions of the lemmas turned out to be quite skew. Therefore, the observed frequencies were replaced by order statistics. These were compared in topic groups as compared to the whole data set. This permitted the elimination of the so-called stop-words, and rare words. The lemmas that best distinguish topic groups could be determined.

In each topic group, the language used and its change between periods, was studied by embedding the lemmas into a four-dimensional Euclidean space. This enabled the graphical study of the changes in the use of lemmas, and find lemmas that distinguish topics. The embedding was based on the singular value decomposition of a matrix formed from the order statistics. The results were described using dynamic graphics, word clouds, and their change. As a result, we could show that the linguistic usages differed between topic areas, and changed from period to period, in an interpretable manner.

The statistical bag-of-words model was complemented, in two ways. First, we read, on a sample basis, about 2,000 posts. Based on the reading we could characterize the structure of the authors as a group, and the manner of argumentation. The trichotomy of Aristotle's rhetoric was used to structure the reading. In addition we used Gemini AI to provide summaries of the posts we had read. The AI turned out to be fast, but the results contained unclear features that merit further study.

The central empirical finding was that the bursts in the flow of posts were only indirectly related to the demograhic shocks that were of interest. Changes in fertility were thought to be caused by changes in attitudes. But in the discussions, the emphasis was in changes in women's role, and that of the family. The effect on births is seen as a "by-product". As expected, corona dominated the discussions on mortality, although its effect on mortality was minor. Instead, the discussions revolved, e.g., around vaccinations, and their side effects. Chronic diseases and increasing life expectancy were hardly noted. In discussions on migration, the major burst was related to the refugee crisis of 2015, although its effect on net migration turned out to be small. In contrast, the wave of immigrants provoked by the Ukrainian war was less discussed, although it is likely to leave a more enduring effect on our demographic history.

The participants were rather heterogeneous a group. One part expends much effort to stay abreast with the causal developments. They typically argue based on perceived facts. A larger group base their arguments on beliefs and attitudes, not on facts. A third, smaller group, uses the forum as a place to rant and slur. Expletives replace reasoning.

As part of social media, *Suomi24* provides a new source of information that reflects the public's current interests. The statistical and information technology methods have evolved to a point where it is possible, with a reasonable effort, to monitor changes in views, nearly in real time. The report concludes with a discussion on how the results are related to those from traditional surveys, and on the possibilities of using AI in future research.

1 Johdanto

1.1 Väestönmuutoksen ongelmat

Väestö on *varanto* (engl. *stock*), jonka kokoon *muutosvirrat* (*flows*) vaikuttavat. Kuolleisuus vähentää varantoa ja syntyvyys lisää sitä. Muuttoliike tekee kumpaakin. Vuoden 2026 alun väestön koko oli Suomessa 5 656 900. Vuoden 2025 alussa varanto oli 5 636 000. Kasvua oli 0.34 prosenttia.

Suomen väestötiedot koskevat vakinaisesti maassa asuvaa (*de jure*) väestöä. Käsitteen katsotaan olevan yhteydessä valtion käsitteen syntyyn (Lénoir, 1995). Tosiasiallisesti maassa olevasta (*defacto*) väestöstä ei julkaista Suomessa virallisia arvioita (ks. Alho, 2026).

Muuttoliikkeen tilastointi on monissa maissa ollut ongelma erityisesti sen jälkeen kun rajavalvontaa on Schengen-alueella vähennetty. Suomen lainsäädännön mukaan maahanmuuttajan, joka aikoo olla maassa vähintään vuoden, on ilmoittauduttava viranomaisille, samoin jos on jo ollut maassa 3 kuukautta. Oikeudesta maassaoloon säädetään erikseen esimerkiksi turvapaikanhakijoiden osalta.

Vuosituhannen vaihteen tienoilla Suomessa oli noin 60 000 syntymää ja 50 000 kuolemaa vuodessa. Muuttovoitto oli noin 5 000. Syntyneiden enemmyys johtui nuoresta ikärakenteesta, sillä jo vuodesta 1969 alkaen hedelmällisyys oli ollut alle uusiutumistason. Ensimmäinen otsikon tarkoittama väestösokki oli hedelmällisyyden romahdusmainen lasku 2010-2024. Kuolemat alkoivat ylittää syntymät vuoden 2015 tienoilla. Toinen sokki oli koronaepidemia vuosina 2020-2022. Kolmas sokki oli nettomuuton jyrkkä kasvu vuosina 2022-2024.

Miten sokit on suuren yleisön keskuudessa ymmärrretty? *Suomi24* tarjoaa harvinaisen mahdollisuuden päästä tutkimaan sitä, millaisin käsittein keskusteluihin osallistuvat mieltävät väestöilmiöt yleensä ja sokit erityisesti.

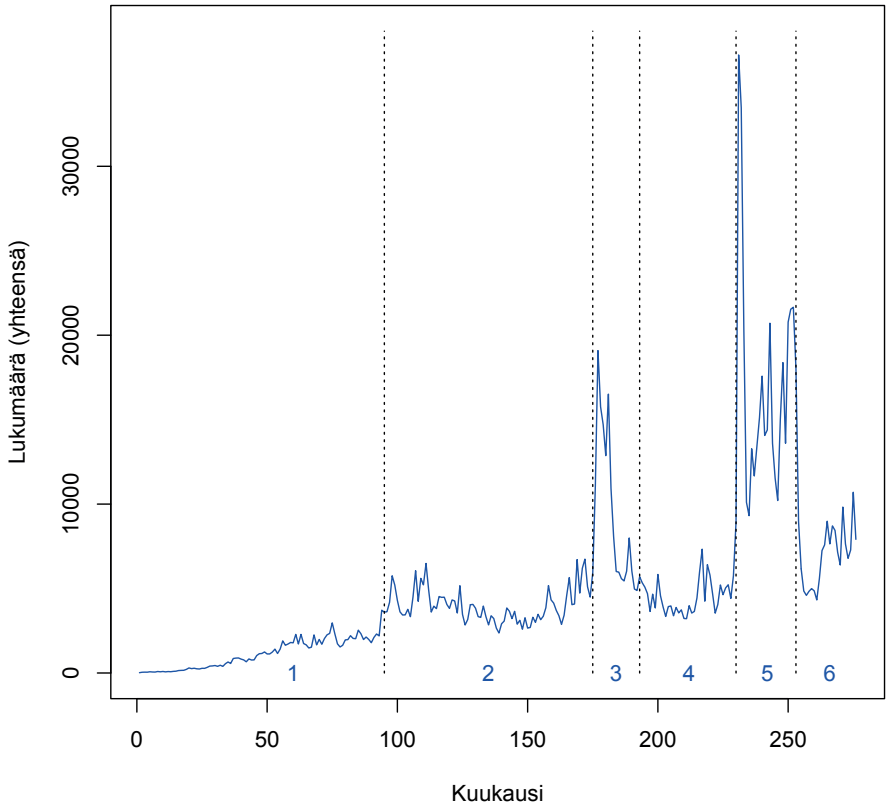
1.2 *Suomi24* tietolähteenä

Suomi24-keskustelupalsta on toiminut nykymuotoisena vuoden 2001 alusta. Kuka tahansa voi osallistua, mutta kirjoittajien henkilöllisyys on pääsääntöisesti tuntematon. Esitystapaa ja käsitteitä valitessaan kirjoittajat joutuvat kuitenkin ottamaan huomioon lukijoiden tietotason, asenteet ja kielitajun. Jos viestien käsitteistö ja kieli muuttuvat sokkien tapahtuessa, se voi kertoa siitä, miten sokit on tulkittu, ja mitä niiden syistä ja seurauksista ajatellaan.

Toisin kuin lehtien mielipidepalstat, *Suomi24* tarjoaa lähes vapaan kanavan mielipiteiden ilmaisemiseen. Tämän nähdään johtavan asiattomuuksiin ja loanheittoon. Toisaalta jo se, että yksilö haluaa viestiä väestoilmiöistä, on osoitus yhteiskunnallisesta tiedostavuudesta.

Kun ilmaisukanavaksi tulee valituksi *Suomi24*, aiottu kohdeyleisö on lähipiiriä laajempi. Halutaan tehdä vaikutus aloittamalla tai osallistumalla dialogiin. Osallistuminen onkin esimerkki ns. *puheaktista* (Austin, 1962; Hudson, 1980).

Onnistuakseen vaikuttamisen tarkoituksessa kirjoittaja joutuu pohtimaan, millaiset argumentit voisivat olla uskottavia. Hyväksyttävät faktaperustelut, vetoaminen lukijakunnan arvoihin sekä mobilisoinnin edellyttämä tunteisiin vetoaminen ovat käyttökelpoisia ohjenuoria viestintään. Lukija havaitsee, että tässä vetoamme Aristoteleen retoriikan kolmijakoon *logos*, *ethos* ja *pathos*, joita soveltaen voidaan vaikuttaa erityisesti henkilöihin, ”jotka eivät kykene näkemään monimutkaisia yhteyksiä eivätkä seuraamaaan pitkiä päättelyketjuja” (ks. teoksen suomennos Aristoteles, 1997, I kirja, luku 2, erityisesti sivut 11 ja 13).



Kuva 1: *Suomi24*-keskustelupalstalle lähetetyt korpukseen kuuluvat viestit kuukausittain vuosina 2001-2023.

Viestien lähettäminen voi olla myös viihdettä ilman tarkoitusta vaikuttaa mihinkään spesifiin asiaan. Tällöin vitsit, herjat ja yksinäisen kirjoittajan purnauksen tai jaarittelun tarve voivat olla viestin takana.

Kuvassa 1 näytetään, miten valittuun korpukseen sisältyvien viestien määrät ovat kehittyneet vuosina 2001-2023. Kuukausia oli $23 \times 12 = 276$ kpl. Kuvasta näkyy viestien tavaton heilunta (*volatility*) sekä se, miten alkuvuosista alkoi viestimäärien kasvu vuoteen 2008 (jonka tammikuu on kuukausi 85) asti. Sitten määrät lievästi vähenivät, kunnes 2010-luvulla kasvoivat jälleen. Alkuvuosina korpukseen kuuluvat viestit olivat osapuilleen vakio kaikista viesteistä, mutta noin vuoden 2008 jälkeen ne kasvoivat ensin hiljalleen, myöhemmin purskeisesti.

1.3 Tutkimuksen lähestymistapa

Tekstimuotoisten aineistojen tilastollinen analyysi on kehittynyt tietokoneiden laskentakapasiteetin kasvun myötä erityisesti 2000-luvulla. Tässä tutkimuksessa käytetään hyväksi **R**-ohjelmiston tekstinkäsittelyproseduureja, joita vielä vuosikymmen sitten ei ollut olemassa.

Lähtökohtana on ajatus, että tekstilajit eroavat niissä käytettyjen sanojen frekvenssien osalta. Kyseessä on ns. sanapussimalli (*bag of words*). Matemaattisesti frekvenssieroja kuvataan ”upottamalla” sanat pisteiksi pieniulotteiseen (alla neliulotteiseen) euklidiseen avaruuteen, jos-

sa keskenään yhteisvaihtelevat sanat muodostavat ryhmiä. Tarkoitusperiltään analyysi vastaa frekvenssiaineistojen korrespondenssianalyysiä (Greenacre, 1984, 1993).

Vuositasolla väestön muutosvirroissa havaitaan merkittävästi suurempia vaihteluita kuin niiden kuukausittainen kausivaihtelu. Prosessit ovat tilastollisessa mielessä epästationaarisia. *Suomi24*:n viestien kuukausivaihtelu on vielä väestönmuutosdataakin suurempaa. Keskeinen havainto on, että selkeää yhteisvaihtelua ei ole. Päinvastoin muutosvirrat ja viestivirrat poikkeavat hätkähdyttävällä tavalla toisistaan eri periodeilla. Periodeja löydettiin kuusi viestivirroissa havaittujen purskeiden ja muutosvirtojen sokkien perusteella.

Olemme tässä raportissa pyrkineet kirjoittamaan tavalla, joka on riittävän yksityiskohtainen, jotta tarvittavat tilastomatemaattiset ja ohjelmointityökalut hallitseva henkilö voisi toistaa analyysit muuhunkin tekstiaineistoon. Teknisimmät kohdat on merkitty symbolilla ” * ” lukujen otsikoissa. Vain tuloksista kiinnostunut henkilö voi jättää ne väliin.

1.4 Tutkimusraportin sisältö

Luvussa 2 Kuvaamme ensinnä käytettyjen väestötietojen valintaa ja käsittelytapaa. Tämän jälkeen käymme läpi sen, miten *Suomi24*-aineisto on muokattu analyysiä varten. Tuloksena on tilastollinen tutkimusaineisto eli korpus, ja rajattu sanasto. Luvussa 3 näytetään, miten syntyvyys, kuolleisuus ja muuttoliike ovat todellisuudessa käyttäytyneet suhteessa niitä koskevien s-, k- ja m-viestien frekvensseihin. Tämän perusteella suoritetaan vuosilta 2001-2023 kertyneen kuukausittaisen aineiston periodisointi kuuteen osaan. Luvussa 4 esitetään, miten sanastosta on etsitty kutakin kolmea muutosprosessia parhaiten erottelevat lemmat. Keskeistä roolia tässä näyttelee aineiston upotus järjestyslukumuunnoksen avulla. Tulokset esitetään sanapilvien avulla. Luvussa 5 kuvataan kunkin muutosprosessin osalta erottelevien lemmojen muutoksia valittujen vertailuperiodien välillä. Kun tähän asti analyysit ovat perustuneet tilastolliseen tutkimuskorpukseen, niin luvussa 6 tarkastellaan otospohjalta viestien sisällön muutoksia sellaisena kuin kirjoittajat ovat ne tuottaneet. Tekstien luennan tueksi määritetään rakennekehikko ja tuloksia verrataan tekoälyllä saataviin. Luvussa 7 esitetään yhteenveto empiirisistä tuloksista ja pohdimme *Suomi24*-tyyppisten sosiaalisen median aineistojen käytön mahdollisuuksia väestöilmiöiden seurannassa verrattuna kyselytutkimuksiin ja niiden uudempiin versioihin.

2 Aineistot

2.1 Väestötiedot

Kuukausittaiset väestön muutostiedot elävänä syntyneistä, kuolemista ja nettomaahanmuutosta saatiin Tilastokeskuksen nettisivuilta . Tiedot ovat vapaasti kaikkien saatavilla ja ne löytyvät googlaamalla *StatFin*.

Muutostiedot ladattiin excel-muodossa kannettavalle tietokoneelle, jossa ne luettiin **R**-ohjelmiston käyttämään tiedostomuotoon. Kaikki laskelmat (ml. kuvat) on tehty **R**:llä.

Etukäteen oli tiedossa, että muutostiedoissa olisi kausivaihtelua (esimerkkejä löytyy raportista Alho, 2026). Tästä syystä kuukausittaiset tiedot muunnettiin päiväkohtaisiksi, jotta kuukausien pituuden vaihtelu voitiin poistaa. Ajatuksena oli puhdistaa keskimääräinen kausivaihtelu havaintosarjoista. Näin voitaisiin selvittää esimerkiksi kanonisia korrelaatioita (Rao, 1973) tai kointegroituvuutta (Engle ja Granger, 1987) soveltamalla, ovatko poikkeamat assosioituneita *Suomi24*-aineiston viestien lukumäärien tai kielen kanssa.

Nopeasti kävi ilmi, että kanoniseen analyysiin tai kointegraatioanalyysiin sisältyvät ristikorrelaatioiden vakioisuusoletukset eivät aineistossa täyty. Väestönmuutos-sarjoissa on sokkeja ja *Suomi24*-sarjoissa on purskeisuutta (ks. kuva 1), joiden takia sarjojen yhteydet ovat erilaisia eri aikoina.

Tästä syystä aineiston hienosäädöstä luovuttiin. Esimerkiksi tekstiaineiston kuukausifrekvenssejä ei vakioitu eikä esimerkiksi ikävakiointia tehty väestönmuutostiedoille.

2.2 *Suomi24*-aineiston luonti

2.2.1 Keskustelupalsta tietokantana

Suomi24-keskustelupalsta on ollut nykymuotoisena toiminnassa 1.1.2001 alkaen. Se näyttäisi olevan maailman vanhin edelleen toimiva keskustelualusta (*discussion forum*). Esimerkiksi *reddit* on perustettu vasta kesäkuussa 2005. Vanhempiakin on olemassa (esim. *delphi forums*, jotka ovat toimineett Yhdysvalloissa ainakin vuodesta 1994), mutta niiden toiminnan rakenteessa näyttäisi tapahtuneen aikojen kuluessa isoja muutoksia.

Toisaalta keskustelupalstoista on nykyisin tuoreempia versioita, joita tutkijat voivat vapaasti analysoida. Tästä on esimerkkinä itävaltalaisen *Der Standard* -lehden ylläpitämän saksankielisen keskustelupalstan aineisto (Fraxanet, Gomez, Kaltenbrunner ja Pellert, 2026). Osallistummekin tekemillämme tutkimusmenetelmävalinnoilla alan kansainväliseen keskusteluun.

Keskustelupalstan omistajat ovat antaneet aineiston Kielipankin käytöön. Kielipankki tarjoaa yliopistotutkijoille aineiston käyttöön hakutyökaluja mm. Korp-käyttöliittymän kautta, mutta tutkijat voivat myös ladata aineiston ns. *vrt*-muodossa CSC:n supertietokoneille.[1] Tässä tiedostomuodossa viestit on esitetty sana (ml. välimerkit) kerrallaan, alekkain. Kustakin sanasta on sen rivillä automaattisen lauseenjäsentimen (Turku Dependency Treebank; ks. www.bionlp.utu.fi/fintreebank.html) antamat kieliopilliset tiedot, esimerkiksi sanan perusmuoto ja sanaluokka. vrt-tiedostot muunnettiin csv-muotoiseksi ja siitä edelleen **R**-ohjelmiston tiedostomuotoon. Kiinnostuksen kohteena olevat osat lähtötiedostoista ladattiin kannettavalle tietokoneelle.

Suomi24:n viestit ovat aineistossa sellaisena kuin kirjoittaja ne on tuottanut. Sanoissa on kirjoitusvirheitä ja murteellisuuksia. Ne voivat olla vieraskielisiä tai koostua esimerkiksi välimerkeistä tai muista merkistön elementeistä (esim. emojit). Lauseet eivät useinkaan noudata oikeakielisyyssääntöjä, joten niillä ei ole edes periaatteessa yksikäsitteistä kieliopillista tulkintaa. Niitä voineekin pitää merkitykseltään *sumeina* (ks. esim. Zadeh, 1965). Niinpä lauseenjäsennin on eräissä tapauksissa tulkinnut sanojen perusmuodot (lemmat) toisin kuin odottaisi. Tästä on esimerkkejä mm. luvun 4.3 sanapilvissä.

Kuhunkin viestiin liittyy myös metatieto, joka sisältää mm. viestin aikaleiman päivän ja tunnin tarkuudella, sekä järjestelmän tuottaman hierarkkisen aiheluokan.

2.2.2 Korpuksen määrittely: sanat, lemmat, tokenit ja järjestysluvut

Eksploratiivisen data-analyysin (Tukey, 1977) hengessä tarkastelun kohteeksi valittiin joukko *Suomi24*-palstalle lähetettyjä viestejä niiden sisältämien sanojen tai sanan osien, eli *tokenien* perusteella. Tokenit valittiin alun perin haravoimalla substantiivien frekvenssiluetteloita. Lyhyyden vuoksi sanomme, että jos viestissä esiintyy sana, joka alkaa tokenilla *väestö*, kyseessä on *v-viesti*. Viesti, joissa esiintyy sana, joka alkaa tokenilla *syntyv*, *hedelmällisyys*, *kokonaishedelmällisyys*, *lapsiluku* on *s-viesti*; viesti jossa jokin sana alkaa tokenilla *kuolleisuus*,

[1] CSC = Tieteen tietotekniikan keskus, Espoo.

kuolevuus, *elinajanodote*, *epidemia*, *pandemia*, *korona* tai *covid* ovat *k-viestejä*; vastaavasti tokenit *maahanmu*, *maastamuutt*, *muutto* ja *mamu* määrittävät *m-viestit*. Kriteeritokeneita oli siis 16 kappaletta.

Tarkasteluvuosina 2001-2023 *Suomi24*:ään tuli noin 110 miljoonaa viestiä. Näistä 1.33 miljoonaa oli v-, s-, k- tai m-viestejä. Viestien aihepiirien rajauksella päädyttiin lopulta 1.26 miljoonaan viestiin. Rajauksella pyrittiin vähentämään todennäköisyyttä sille että token esiintyisi keskusteluissa, joilla ei tosiasiallisesti ole tekemistä väestöilmiöiden kanssa. Nämä muodostivat tutkimuksemme tilastollisen korpuksen.

Sanojen sijaluokkien ja omistusliitteiden takia sanat eivät sellaisenaan ole tarkoituksenmukainen havaintoluokka. Sen sijaan tarkastelut perustettiin sanojen perusmuotoihin eli *lemmoihin*. Kussakin aiheryhmässä laskettiin substantiivien ja adjektiivien frekvenssit ja valittiin vajaan 8 000 yleisimmän yhdelmä *sanastoksi*. Frekvenssijakaumat ovat huomattavan vinoja. Vertailuja varten frekvenssit korvattiin järjestysluvuilla (*order statistics*). Viestiyhmien erityispiirteitä kuvattiin vertaamalla lemmojen järjestyslukujen erotusta niiden järjestyslukuihin koko aineistossa. Näin sekä stop-sanojen että harvinaisten sanojen vaikutusta voitiin pienentää, ja ryhmiä parhaiten erottelevat lemmat saatiin määritettyä.

Tokenit *kokonaishedelmällisyysluku* ja *elinajanodote* kuuluvat väestötieteen terminologiaan. Oli odotettavissa, että ne olisivat harvinaisia, ja näin myös oli. *Kokonaishedelmällisyysluku* esiintyi vain 57 viestissä ja *elinajanodote* 1 678 viestissä. Nämä tokenit otettiin mukaan, jotta saataisiin esimerkkejä viesteistä, jotka mahdollisimman suurella todennäköisyydellä olisivat asiantuntevien kirjoittajien tuottamia.

3 Väestönmuutosten ja viestifrekvenssien vastaavuus

Ennakko-odotuksemme oli, että väestönmuutokset saattaisivat aiheuttaa niitä koskevia keskusteluita joko samanaikaisesti tai ainakin jälkikäteen. Pari vuosikymmentä sitten tiedonlouhinnan (*data mining*) käynnistyessä uutisoitiin näyttävästi, että google-kyselyt saattoivat ennakoida sairausaaltoja, joten sekin mahdollisuus oli mielessä, että keskustelut voisivat jopa ennakoida joitain väestönmuutoksia. (Väitteet google-hakujen ennakointiominaisuudesta eivät olleet kaikin osin perusteltuja, ks. Lazer, Kennedy, King ja Vespignani (2014).) Näin ei kuitenkaan *Suomi24*:n osalta ollut.

Tämän tutkimuksen keskeiseksi tehtäväksi muodostuikin sen selvittäminen, mistä viestiaineiston purskeissa on kysymys. Tarkastelun pääpaino on s-, k- ja m-viesteissä. v-viestien rooli on toimia vertailuryhmänä.

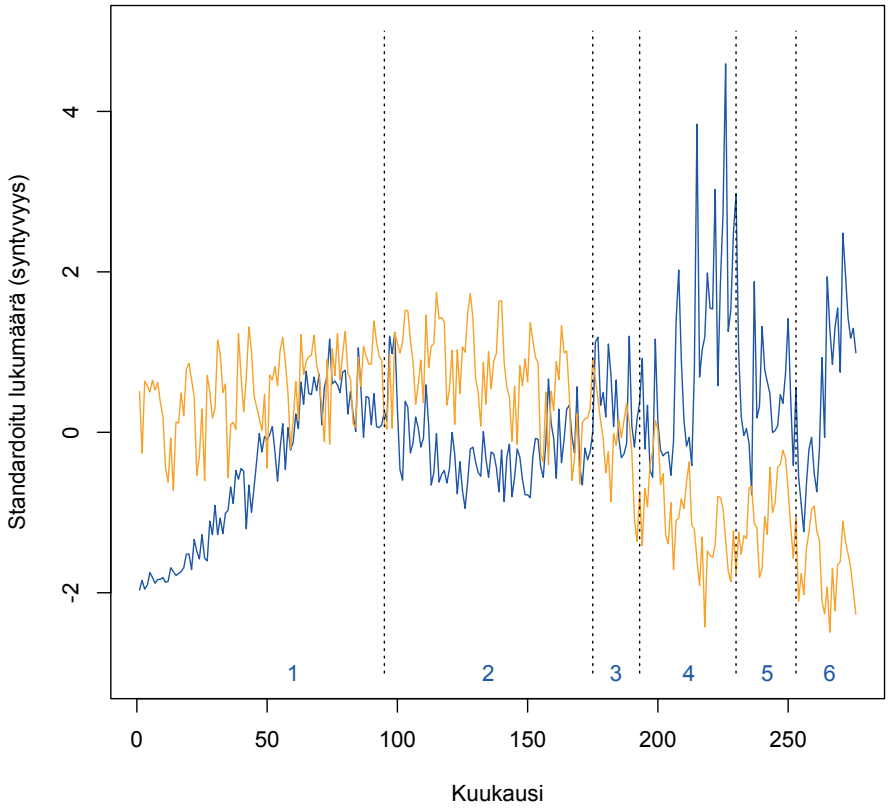
Kuvassa 1 näkyvä havaintokuukausien jako kuuteen periodiin perustellaan kohdassa 3.4.

3.1 Syntyvyys

Syntymiä koskevien s-viestien ja havaittujen syntymien lukumäärät on kuvaa 2 varten normalisoitu vähentämällä keskiarvo ja jakamalla keskihajonnalla.

Syntymien lukumääriä kuvaavasta oranssista käyrästä käy selvästi ilmi syntyminen kausivaihtelu kalenterivuoden sisällä. Korkeimmat arvot havaitaan loppukesällä ja pienimmät vuodenvaihteessa (ks. Alho, 2026, kuva 19 (a)).

Väestönmuutosten taloudellisen merkityksen näkökulmasta kuvan 2 oranssi käyrä on keskeinen. Vuoden 2010 jälkeen (vuoden 2011 tammikuu on kuukausi 121) kokonaishedelmällisyys laski



Kuva 2: *Suomi24*-keskustelupalstalle lähetetyt syntymiä koskevat viestit (sininen käyrä) ja syntymät (oranssi) kuukausittain vuosina 2001-2023.

arvosta 1.8 lasta naista kohti arvoon 1.3. Väestön sisäinen uusiutuminen edellyttäisi noin arvoa 2.07. Hedelmällisyys on siis vajaat 2/3 uusiutumistasosta. Pitkällä tähtäimellä tämä tarkoittaisi (ilman muuttoliikettä) väestön koon pienenemistä 1/3:lla sukupolven aikana. Sukupolven pituus on nykyisin hieman yli 30 vuotta.

Viestifrekvenssejä kuvaavan sinisen käyrän alku vastaa sekä korpuksen viestien kokonaismäärän kasvua (ks. kuva 1) että kaikkien *Suomi24*:n viestien kokonaismäärän kasvua. Vuoden 2010 jälkeen s-viestien määrä on kasvanut nopeammin kuin koko *Suomi24*:n viestien lukumäärä. Aihe on siis tullut suuremman huomion kohteeksi. Heilunta on kuitenkin erittäin suurta.

Korkeimmat frekvenssit havaitaan vuosien 2018-2019 tienoilla. On tärkeää selvittää, miksi niin on.

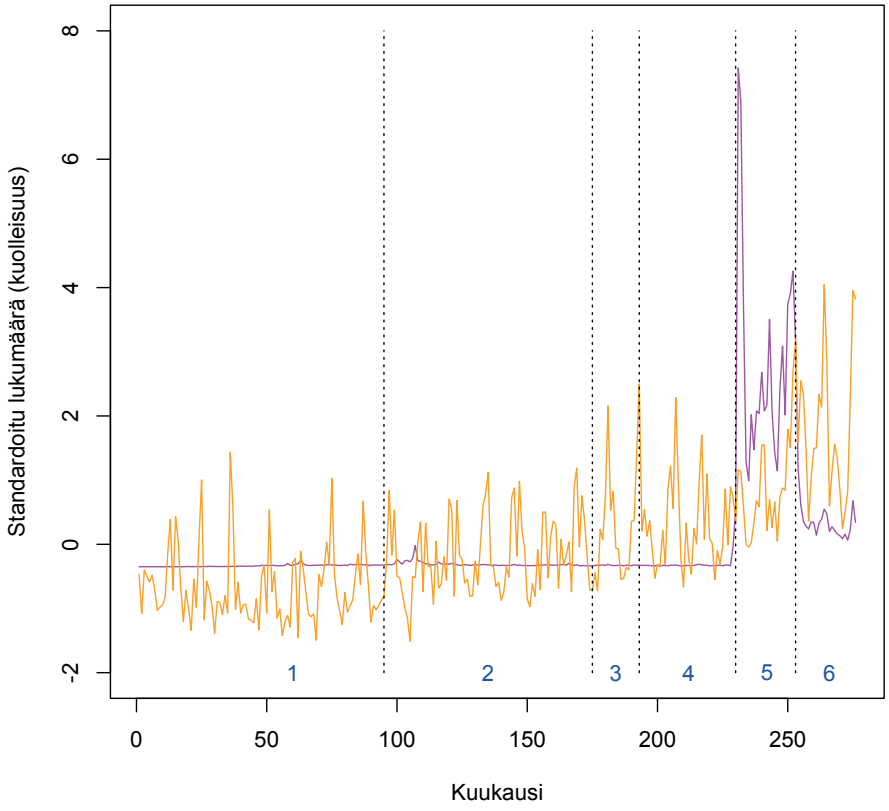
3.2 Kuolleisuus

Kuvan 3 oranssista käyrästä nahdään kuolemantapausten kausivaihtelu. Kuolemien määrät ovat suurimmillaan talvella ja pienimmillään kesällä. Vuosien välinen vaihtelu on kuitenkin suurta. Tämän tiedetään johtuvan vuotuisen influenssakauden ajankohdasta, joka vaihtelee loka-toukokuun välillä.(ks. Alho, 2026, kuva 21).

Kuolemien käyrän loiva nousu johtuu siitä, että ikäryhmittäisen kuolevuuuden alenemisen takia yhä useammat ovat selvinneet ikiin, joissa kuolevuus on korkeaa. Tätä kehitystä korosti se, että vuosina 1945-1950 syntyneet suuret ikäluokat olivat vuonna 2020 siirtyneet kaikki yli 70-vuotiaiden ikäryhmään, mikä kasvatti kuolemien absoluuttista lukumäärää.

Suomi24:n viestejä kuvaava aniliini käyrä poikkeaa radikaalisti kuvasta 2. Nähdään, että ennen vuoden 2020 alkua kuolevuus ei juurikaan kiinnostanut keskustelupalstaan osallistuneita. (Pieni poikkeus ovat vuodet 2009-2010. Tätä nyppylää aikasarjassa setvitään tarkemmin kohdassa 6.2.) Jyrkän nousun syynä vuoden 2020 helmikuussa (kuukausi 230) oli koronaepidemia.

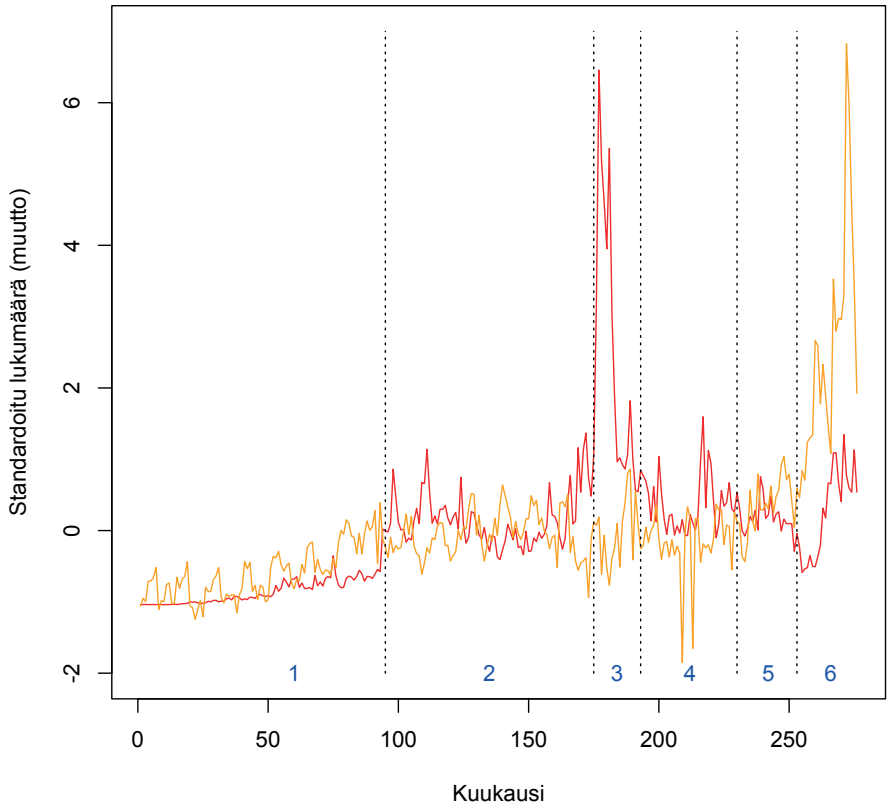
Keskustelut olivat kiivaimmillaan kuukausina 230-252 (vuoden 2020 helmikuusta vuoden 2021 joulukuuhun). Oranssista käyrästä käy ilmi, että kuolleiden lukumäärä tällä periodilla ei ollut erityisen korkea verrattuna sitä seuraavaan aikaan, jolloin keskustelut olivat merkittävästi laantuneet. Kun keskusteluilla ei ollut selvää yhteyttä kuolemien määrään, onkin kiinnostavaa selvittää, mitä kuukausien 230-252 keskustelut koskivat.



Kuva 3: *Suomi24*-keskustelupalstalle lähetetyt kuolemia koskevat viestit (aniliini käyrä) ja kuolemat (oranssi) kuukausittain vuosina 2001-2023.

3.3 Muuttoliike

Kuva 4 poikkeaa täysin kahdesta edellisestä. Oranssista käyrästä käy ilmi muuttovoiton vähittäinen kasvu ja sen ennennäkemätön räjähdys vuonna 2022. Keskeisimpänä syynä oli Venäjän



Kuva 4: *Suomi24*-keskustelupalstalle lähetetyt muuttoliikettä koskevat viestit (punainen käyrä) ja nettomuutto (oranssi) kuukausittain vuosina 2001-2023.

hyökkäys Ukrainaan. Sota aiheutti pakolaisaallon, johon EU reagoi tarjoamalla ns. tilapäistä suojelua pakolaisille. Toisena syynä nettomaahanmuuton kasvuun vuodesta 2021 alkaen oli suomalaisten yritysten ja oppilaitosten aktiivinen rekrytointi erityisesti Kaukoidästä.

Kuvan 4 punainen käyrä poikkeaa oranssista odottamattomalla tavalla. Runsain keskustelu käytiin vuoden 2015 heinäkuusta (kuukaudesta 175) alkaen. Vuosina 2014-2015 Eurooppaan saapui vajaat 2 miljoonaa pakolaista Lähi-Idästä. Heidät päästettiin Balkanin läpi Keski-Eurooppaan ja lopulta osa kulkeutui Suomeen asti. (Tästä tarkemmin kohdassa 7.1.3.) Oranssi käyrä osoittaa, että hieman koronaepidemiaa muistuttavalla tavalla Suomeen muuttaneita oli 2015 jopa hieman vähemmän kuin 2014.

Samoin kuin kuolleisuuden kohdalla on syytä pyrkiä täsmentämään, miksi keskustelu painottuu ajankohtaan, jolloin maahan tuli runsaasti turvapaikanhakijoita, mutta ei aikaan, jolloin maahan tuli räjähdyksenomaisesti maahanmuuttajia.

3.4 Aineiston periodisointi

Edellä on osoitettu että väestön kolmea muutosprosessia koskevat sarjaparit ovat keskenään erilaisia ja kaikki voimakkaan epästationaarisia. Tästä syystä esimerkiksi klassiset *itseään kiihdyttävät* mallit, joissa pienet efektit kumuloituvat ja aiheuttavat lopulta purskeen, eivät näytä aineistoon sopivilta (ks. esim. Hawkes, 1971; Zhuang, Ogata ja Vere-Jones, 2002; Mohler, Short, Brantingham, Schoenberg and Tita, 2011). Tästä syystä valitsimme lähestyä tutkimusaihettamme periodisoimalla havaintojakso niin, että kunkin sarjan sokit/purskeet erottuvat muista periodeista:
1. kuukaudet 1-94 (tammikuu 2001 - lokakuu 2008, yhteensä 94 kk)
2. kuukaudet 95-174 (marraskuu 2008 - kesäkuu 2015, yhteensä 80 kk)
3. kuukaudet 175-192 (heinäkuu 2015 - joulukuu 2016), yhteensä 18 kk)
4. kuukaudet 193-229 (tammikuu 2017 - tammikuu 2020, yhteensä 37 kk)
5. kuukaudet 230-252 (helmikuu 2020 - joulukuu 2021, yhteensä 23 kk)
6. kuukaudet 253-276 (tammikuu 2022 - joukuu 2023, yhteensä 24 kk)

Periodia 1 voi kuvata *Suomi24*:n kypsymisen periodiksi. Yhdessä periodin 2 kanssa se muodostaa kielen muutosten tausta-aineiston. Purskeiden perusteella muotoiltiin kolme tutkimustehtävää. Ne perustuvat kunkin muutosprosessin vertailupareille seuraavasti: I. syntyvyydelle periodit 4 ja 6; II. kuolleisuudelle periodit 5 ja 6; III. muuttoliikkeelle periodit 3 ja 6.

Tokenien *kokonaishedelmällisyys* ja *elonajanodote* perusteella poimitut viestit toimivat vertailuryhmänä. v-viestit toimivat niinikään vertailuryhmänä, joka auttaa etsimään sanoja, jotka mahdollisimman hyvin erottelevat s-, k- ja m-viestejä toisistaan.

4 Erottelevien sanojen etsintä

4.1 Erottelun logiikka*

4.1.1 Viestin rakenne*

Jättämällä tämän raportin kannalta tarpeettomat yksityiskohdat sivuun viestin rakenne on seuraava. Olkoon H kaikkien vuosina 2001-2023 lähetettyjen viestien indeksien h joukko. H:ssa on siis noin 110×10^6 elementtiä. Valittujen 16 tokenin perusteella näistä valittiin tarkastelun

kohteeksi osajoukko $H' \subset H$. Osajoukossa on 1.33×10^6 elementtiä. Ottamalla huomioon viestien aihepiiriluokka pienehkö määrä poistettiin. Tuloksena saatiin osajoukko $H'' \subset H'$, jossa on 1.26×10^6 elementtiä. Tämä muodostaa tutkimuksen *korpuksen*.

Kukin viesti h kuvataan vektorina $\mathbf{v}_h = (v_{h1}, \mathbf{v}_{h2}, \mathbf{v}_{h3}, \mathbf{v}_{h4})$. Vektorin ensimmäinen komponentti sisältää metatiedot. Aineistossa on aikaleima viestin saapumisajasta tunnin tarkkuudella ja varsin tarkka aihepiirin luokitus. Käsillä olevan kuvauksen kannalta vähempi riittää, ja v_{h1} sisältää vain tiedon siitä kuukaudesta, jona viesti saapui. Näin ollen $v_{h1} \in \{1, \ldots, 276\}$ kaikille $h \in H''$. Tämän muuttujan arvojen avulla voidaan löytää esimerkiksi kunkin kuuden periodin viestit.

Toinen komponentti $\mathbf{v}_{h2}$ sisältää viestin tekstin sellaisena kuin kirjoittaja on sen lähettänyt. Kyseessä on siis merkkijono, jonka pituus voi vaihdella muutamasta merkistä/sanasta useisiin tekstisivuihin. Tätä komponenttia käytetään, kun luvussa 6 halutaan lukea alkuperäinen viesti.

Kolmas komponentti on muotoa $\mathbf{v}_{h3} = (v_{h31}, v_{h32}, v_{h33}, v_{h34})$, missä $v_{h31} = 1$, jos h on v-viesti, muuten $v_{h31} = 0$. Vastaavasti $v_{h32} = 1$, jos h on s-viesti. Vastaavasti kaksi viimeistä komponenttia indikoivat k- ja m-viestit. Näillä indikaattoreilla voidaan löytää v-, s-, k- ja m-viestit.

Neljäs komponentti $\mathbf{v}_{h4}$ sisältää tiedon viestin sanojen perusmuodoista, lemmoista. Koska kiinnostuksen kohteena oli vain viestijoukko H'', lähtökohtana on H'':n viesteissä esiintyneiden lemmojen joukko. Koska tutkimuksen rajattuna tarkoituksena oli päästä käsiksi väestöilmiöitä koskeviin keskusteluihin, ei tutkia viestejä yleensä, lemmoista otettiin mukaan vain substantiivit ja adjektiivit.

Esimerkki. Oletetaan, että viestin tekstin $\mathbf{v}_{h2}$ jossain kohdassa lukee *koronasta tulee meille vaikee*. Tätä katkelmaa vastaavat lemmat *korona*, *tulla*, *me* ja *vaikea*. Näistä ensimmäinen on substantiivi ja viimeinen on puhekielimäinen adjektiivi. Tieto näistä tulee mukaan viestiä h koskevaan vektoriin $\mathbf{v}_{h4}$. Välissä olevat verbi ja pronomini jätetään pois.

4.1.2 Aineiston piirteet ja dimension pienentäminen

Aineiston empiiriset piirteet ovat seuraavat: H'':ssa oli kaikkiaan 1 280 830 viestiä. Näistä 234 274 oli v-viestejä, 45 046 s-viestejä, 389 564 k-viestejä ja 656 446 m-viestejä. Näiden summa on 1 359 435, mikä osoittaa, että eri tokenien määrittämät viestien osajoukot eivät ole toisensa poissulkevia. Nähdään, että k- ja m-viestejä on kertaluokkaa enemmän kuin s-viestejä.

Tekstitiedon tilastollisen analyysin keskeisin ongelma juontuu komponentin $\mathbf{v}_{h4}$ dimensiosta. Vaikka H'' sisältää vain runsaan prosentin H:n sisältämistä viesteistä ja huomiota kiinnitetään vain substantiivi- ja adjektiivilemmoihin (ei siis tehdä eroa esimerkiksi sijamuotojen välillä), niin H':ssa on (toistot mukaan lukien) 49.9×10^6 lemmaa eli keskimäärin 50 substantiivia tai adjektiiviä viestiä kohti. Erillisiä lemmoja on kaikenkaikkiaan 1.82×10^6 kappaletta. Tämä olisi siis komponentin $\mathbf{v}_{h4}$ dimensio ja samaa suuruusluokkaa kuin havaintoyksikköjen lukumäärä, joka on H'':n elementtien lukumäärä eli 1.26×10^6. Dimensiota täytyy tavalla tai toisella pienentää, kun halutaan löytää lemmoja, jotka erottelevat erityyppisiä viestejä.

Lähtökohtana dimension pienentämiselle tarkastellaan kunkin lemman esiintymisfrekvenssejä. Osoittautuu, että frekvenssijakaumat ovat erittäin *vinoja*. Esimerkiksi niiden lemmojen lukumäärä, jotka esiintyvät vain kerran, on 1.21×10^6, ja niiden, jotka esiintyvät kahdesti, on 0.60×10^6. Vain muutaman kerran esiintyviä lemmoja ei voida hyödyllisesti käyttää kuvaamaan kielen muutosta ajassa.

Toisaalta hyvin yleisesti esiintyville lemmoille on tyypillistä, että ne esiintyvät hyvin monlaisissa viesteissä. Tässä yhteydessä puhutaan ns. *stop*-sanoista, joilla ei ole tästä syystä erottelukykyä. Kymmenen yleisintä lemmaa aineistossamme olivat *Suomi*, *muu*, *ihminen*, *vuosi*, *maa*, *asia*,

aika, *suomalainen*, *maahanmuuttaja* ja *oma*. Ensimmäisen ja viimeisen esiintymisfrekvenssit olivat 640 119 ja 291 706. Jakauman vinoutta kuvaa se, että sadannen frekvenssi oli 49 068, tuhannennen 6 887 ja kymmenennentuhannen 405. Näissä frekvensseissä toistot viestin sisällä on laskettu mukaan. Erottelukykyisiä lemmoja, mikäli sellaisia ylipäätään on, täytyy etsiä välimaastosta.

4.2 Lemmojen upotus järjestyslukujen pohjalta*

Käytämme pääkomponenttianalyysiä ”upottamaan” lemmat pisteinä tasoon. Koska tutkimuksen tarkoituksena on selvittää lemmojen käytön mahdollisia muutoksia ajassa, lemmojen kokonaisfrekvenssien alarajaksi valittiin 553. Tämä vastaa sitä, että lemma voisi esiintyä keskimäärin useammin kuin kahdessa viestissä kuukaudessa ($2 \times 276 = 552$). Tällaisia lemmoja oli 7 957 kappaletta. Nämä substantiivit tai adjektiivit muodostivat erottelevia termejä koskevan tutkimuksen varsinaisen sanaston. Tämä merkitsee myös sitä, että viestien ominaisuuksia kuvaavan osavektorin $\mathbf{v}_{h4}$ pituudeksi tuli 7 957. Tämä on siis noin 0.4 % kaikista H':n erillisistä lemmoista.

Lähtökohtana seuraavassa käytetään hyväksi tietoa valittujen 7 957 lemman *järjestysluvusta*: pienimmän esiintymisfrekvenssin lemman järjestysluku on 1, seuraavan 2 ja esimerkiksi suurimman frekvenssin lemman järjestysluku on 7 957. Kyseessä on siis käänteinen suuruusjärjestys. Ajatuksena on tarkastella lemmojen järjestyslukuja esimerkiksi s-viestien joukossa eri periodeilla ja etsiä sellaisia, joiden järjestysluku on odottamattoman korkea verrattuna niiden järjestyslukuun koko aineistossa.

4.2.1 Upotuksen matematiikka*

Matemaattisesti lemmojen upotus tapahtui seuraavasti. Olkoon W valittujen 7 957 lemman w joukko, muotoa $W = \{1, \ldots, 7957\}$. Koska $w \in W$, sovimme, että lemman nimi w on samalla sen järjestysluku (*hashing*). Ts., esimerkiksi lemma $w = 7957$ vastaa sanaa *Suomi*, koska se on aineistossa eniten esiintyvä substantiivi tai adjektiivi.

Kutakin periodia $J = 1, \ldots, 6$ kohti määritellään 7957×4 -matriisi $\mathbf{W}_j = (W_{wij})$, missä W_{wij} on lemman w esiintymisten lukumäärän järjestysluku periodilla j niissä viesteissä, joiden aihepiiri on i. Tässä i:n arvot, $i = 1, 2, 3, 4$, vastaavat v-, s-, k- ja m-viestejä. Määritellään keskistetyt matriisit $\bar{\mathbf{S}}_j$, jonka alkiot ovat muotoa $\bar{S}_{wij} = W_{wij} - w$, ts. elementtien järjestysluvuista on vähennetty niiden järjestysluku koko aineistossa. Esimerkiksi jos $\bar{S}_{w,2,5} > 0$, niin lemma w esiintyy s-viesteissä $i = 2$ periodilla $j = 5$ useammin suhteessa muihin lemmoihin kuin kaikilla periodeilla yhteensä.

Määritellään keskistettyjen matriisien keskiarvo, $\bar{\mathbf{W}} = (\mathbf{S}_1 + \ldots + \mathbf{S}_6)/6$, ja lasketaan tämän matriisin singulaariarvokehitelmä $\bar{\mathbf{W}} = \mathbf{UDV}^T$, missä $\mathbf{D}$ on 4×4 -diagonaalimatriisi, jonka diagonaalilla $\mathbf{d}$ ovat singulaariarvot, $\mathbf{U}$ on 7957×4 -matriisi, jonka sarakkeissa ovat vastaavat singulaarivektorit, jotka toteuttavat ehdon $\mathbf{U}^T\mathbf{U} = \mathbf{I}_4$ ($= 4 \times 4$ -identiteettimatriisi). Vastaavasti $\mathbf{V}$ sisältää oikeat singulaarivektorit ja toteuttaa niin ikään ehdon $\mathbf{V}^T\mathbf{V} = \mathbf{I}_4$. Tätä hajoitelmaa kutsutaan myös matriisin *pääkomponenttiesitykseksi*. Matriisin $\bar{\mathbf{W}}^T\bar{\mathbf{W}}$ ominaisarvot ovat vektorin $\mathbf{d}$ neliöitä, ja matriisi $\mathbf{V}$ koostuu niitä vastaavista ominaisvektoreista. Poislukien tapaukset, joissa ominaisarvot eivät ole erillisiä, $\mathbf{V}$:n sarakkeet ovat etumerkkiä vaille yksikäsitteisiä. Valitsemalla etumerkit ne määrittelevät koordinaatiston, jossa graafinen kuvaus tapahtuu (vrt. Greenacre, 1993).

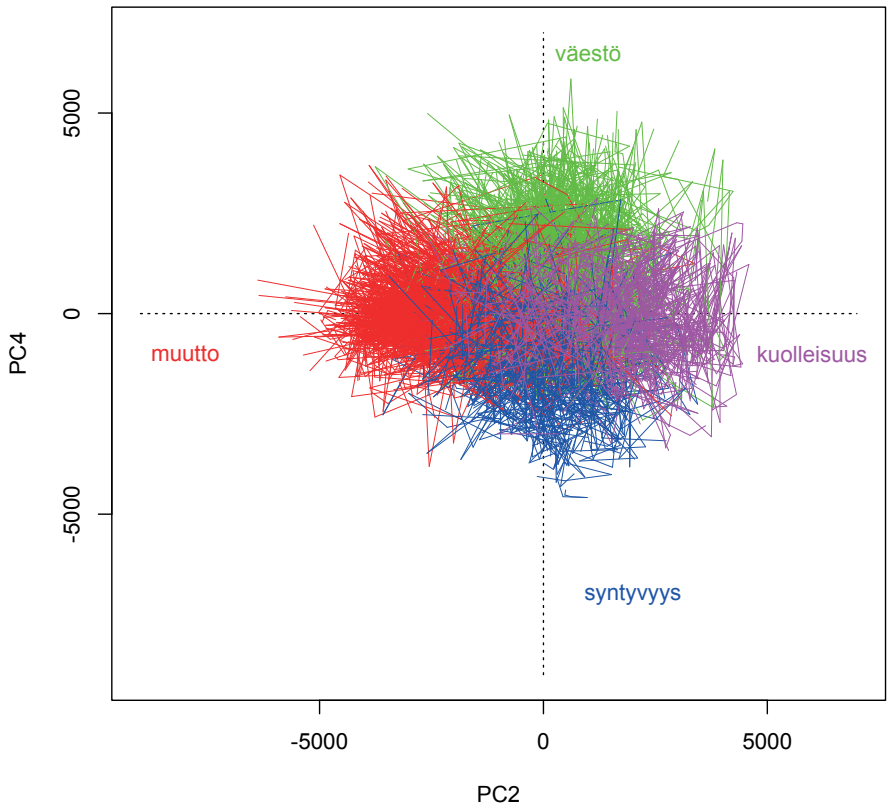
Tulokset ovat seuraavat. Ensinnäkin $\mathbf{d} = (24.5, 14.4, 9.8, 8.6)$. Tästä voidaan laskea, että ensimmäinen pääkomponentti selittää 61 % vaihtelusta, toinen 21 %, kolmas 10 % ja neljäs 8 %.

Matriisin sarakkeet muodostavat siis neliulotteisen avaruuden ortonormaalin kannnan ja sen rivit vastaavat v-, s-, k- ja m-viestejä (ks. esim. Afifi ja Azen, 1979, s. 318). Toiseksi saamme tuloksen

$$\mathbf{V} = \begin{bmatrix} -0.45 & 0.06 & -0.31 & 0.84 \\ -0.63 & -0.08 & -0.55 & -0.54 \\ -0.56 & 0.47 & 0.67 & -0.09 \\ -0.28 & -0.88 & 0.39 & 0.05 \end{bmatrix}.$$

Nähdään, että ensimmäinen komponentti/sarake kuvaa lemmojen yhteisvaihtelua periodista toiseen. Toinen komponentti on kontrasti kuolleisuuden ja muuttoliikkeen välillä, kolmas erottaa lähinnä väestöä ja syntyvyyttä koskevat viestit kuolleisuutta ja muuttoliikkeettä koskevista ja neljäs on kontrasti väestö- ja syntyvyysviestien välillä. Toinen pääkomponentti on selvästi erotteleva, samoin neljäs. Koska neljännen selitysaste on lähes sama kuin kolmannen, valitaan se selkeämmän tulkinnan takia graafisen kuvauksen toiseksi dimensioksi.

Matriisi **UD**, tässä tapauksessa sen toinen ja neljäs sarake, antaa kunkin lemman koordinaatit edellä valitussa koordinaatistossa. Mutta kuten edeltä käy ilmi, esiintymisfrekvensseissä on huomattavan paljon ”hälyä”, jonka alle kiinnostavat kielen erot saattaisivat jäädä. Tästä syystä pyrimme varmistamaan, että keskustelun aihepiirejä mahdollisimman hyvin kuvaavat lemmat saataisiin esiin.



Kuva 5: Erottelevien sanojen muodostaminen v-käyrien (vihreä), m-käyrien (punainen), s-käyrien (sininen) ja k-käyrien (aniliini) paikat valitussa upotuksessa.

4.2.2 Erottelevien lemmojen määrittely*

Otetaan lähtökohdaksi edellä määritellyt keskistetyt 7957×4 -matriisit $\bar{\mathbf{S}}_j, j = 1, \ldots, 6$. Valitaan kunkin matriisin kustakin neljästä sarakkeesta 2 % suurimmista arvoista, eli 160 lemmaa.. Periaatteessa näin voisi tulla valituksi 24 x 2 % = 48 % lemmoista, tai toisin ilmaisten $24 \times 160 = 3840$ eri lemmaa. Päällekkäisyyksiä on kuitenkin runsaasti. Olkoon B_i niiden lemmojen joukko, jotka tulivat valituksi sarakkeen $i = 1, 2, 3, 4$ perusteella. Näissä on seuraavat lukumäärät lemmoja: 681, 537, 592 ja 650. Poistetaan päällekkäisyydet niin, että jäljellä jäävät ne lemmat, jotka esiintyvät vain aihepiirissä i. Olkoot nämä osajoukot $D_i, i = 1, 2, 3, 4$. Tällöin pätee esimerkiksi, että $D_2 = ((B_2 \backslash B_1) \backslash B_3) \backslash B_4$, missä on käytetty joukko-opin vähennyslaskumerkkiä "\". Näin saaduissa joukoissa D_i on 261, 218, 252 ja 421 lemmaa.

4.3 Erottelevimpien lemmojen kuvaus sanapilvillä

Edellä määriteltyjä puhdistettuja lemmajoukkoja $D_i, i = 1, 2, 3, 4$, voidaan nyt tarkastella graafisesti. Kuvassa 5 muutoksia ajassa on pyritty havainnollistamaan pirtämällä kuvaan kunkin lemman polku alkaen lemman paikasta periodilla $j = 1$ viidellä askeleella sen paikkaan periodilla $j = 6$. Edellä esitettyä formalismia käyttäen nähdään, että $\bar{\mathbf{W}}\mathbf{V} = \mathbf{UD}$, mistä käy ilmi miten koordinaatit ovat yhteydessä lemmojen keskimääräisiä järjestyslukuja kuvaavaan matriisiin $\bar{\mathbf{W}}$ ja pääkomponentit määrittelevään kerroinmatriisiin $\mathbf{V}$. Kuvaan tarvittavat periodispesifit luvut saatiin tätä vastaaavasti matriisien $\mathbf{W_j V}, j = 1, \ldots, 6$ toisen ja neljännen sarakkeen muodostamista lukupareista.

Kuvasta käy ilmi, miten eri keskustelunaiheisiin liittyvien polkujen paikat eroavat toisistaan. Kaikkiaan kuvassa on $281 + 218 + 252 + 421 = 1\,172$ käyrää. Lukua voi suhteuttaa siihen, että rajattu sanastomme sisältää 7 957 lemmaa, joten hyvin eritteleviksi määrittelemämme käyrät ovat tästä 15 %.

Vaikka kuvassa 5 eriväriset käyräparvet leikkaavat, tutkimuksen otsikon mukaisia muutoksia on selvästi olemassa. Upotus tarjoaa yleiskuvan lemmojen eroista. Lisäksi on tarpeen nähdä esimerkkejä lemmoista, jotka ovat erityisen spesifejä keskustelunaiheille $i = 1, 2, 3, 4$.

Vaikka sanapilvellä ei ole yleisesti hyväksyttyä yksikäsitteistä tilastotieteellistä määritelmää, 2-ulotteisena se tarjoaa hyödyllisen tavan käyttää näköaistia lemmojen mahdollisten yhteisesiintymisten havaitsemiseen. Seuraavassa on käytetty **R**-ohjelmiston *wordcloud*-pakettia. Lemman w koko määräytyy sen järjestysluvun poikkeavuuden $\bar{W}_{wi} - w$ (missä keskiarvo otetaan yli periodien) mukaan, ei siis sen esiintymisfrekvenssin mukaan. Sanat sijoitellaan järjestyslukujen poikkeavuuden mukaisessa järjestyksessä aloittaen keskeltä. Ohjelma värittää lemmat valitusta paletista (tässä *brewer.pal 8 Dark2*) poikkeavuusluvun mukaisesti. Tasahavaintoja kuitenkin esiintyy eikä tulos ole muutenkaan yksikäsitteinen. Hyötynä on silti, että kuvio "kutsuu katsojaa" pohtimaan eri frekvensseillä visuaalisesti lähellä esiintyvien lemmojen mahdollisia yhteyksiä.

4.3.1 Syntyvyys-lemmat

Kuvaan 6 on valittu harkinnanvaraista alarajaa käyttäen (kaikkiaan 218:sta hyvin erottelevasta lemmasta) 54 lemmaa, joiden keskimääräinen poikkeama s-viestien joukossa on suurin verrattuna niiden järjestyslukuun koko aineistossa. Kiinnostavasti pilven keskellä *syntyvyysluku* on ainoa vähänkään hedelmällisyyden tieteelliseen kuvaamiseen liittyvä lemma, joskaan sitä ei varsinaisessa väestötieteessä käytetä.

Kuvasta voi erottaa lemmaryhmiä, jotka viittaavat (a) raskauteen tai raskaaksi tulemisen edellytyksiin: *vaihdevuodet*, *ehkäisyväline*, *e-pilleri*, *ehkäisykielto*, *syntyvaihe*, *hedelmätön*, *hedelmällisyyshoito*, *hedelmöityshoito*, *sterilisaatio*, *kalenteri*; (b) uskontoon: *Baal*, *SRK:n*[2], *Baabelin*, *lestadiolaisuus*, *epäjumala*, *vapahtaja*, *päivämies*, *herätysliike*; (c) yhdessäoloon: *perhe-elämä*, *kolmekymppinen*, *kuningatar*, *kotiäiti*, *nelikymppinen*, *aviopari*, *synnyttäjä*, *uros*, *pukki*, *syntymäpäivä*, *kani*, *seurustelu*, *elättäjä*, *poikanen*, *naaras*, *ikäero*, *härkä*, *muksu*, *hella*, tai (d) fysiologiaan: *testosteroni*, *hormoni*, *vuoto*, *ovulaatio*, *alapää*, *kives*, *munasarja*, *sukupuolielin*, *tissi*, *lima*. Näiden lisäksi kuvassa ovat mukana *naisellinen*, *seksikäs* ja *viehättävä*, jotka voivat liittyä sosiaalisiin suhteisiin. Yleisessä käytössä olevat lemmat *kynttilä* ja *alaston* saanevat merkityksensä asiayhteydestä. Liitteessä 1 kuvataan erikseen lemmojen *Baal* ja *Baabelin* käyttöä keskusteluissa.

Kun edellä luvussa 3.1 korostettiin maatalan hedelmällisyyden keskeistä merkitystä väestön uusiutumisen näkökulmasta, jo tämä deskriptiivinen analyysi viittaa siihen, että s-viestien kirjoittajien mielessä ovat olleet lasten saannin vaikeudet (kohta (a)), uskonnollis-moraaliset kysymykset (kohta (b)), perhe-elämän normit (kohta (c)), ja sukupuolisen kanssakäymisen fysiologia (kohta (d)). Lähtökohtana näyttäisi olevan yksilö psyykkis-fyysisenä kokonaisuutena.

Kuva 6: Erottelevimmat syntyvyys-lemmat.

4.3.2 Kuolleisuus-lemmat

Kuvaan 7 on valittu 52 eniten k-viestejä erottelevaa lemmaa yhteensä 252:sta hyvin erottelevasta lemmasta. Lemmojen joukosta voi löytää esimerkiksi seuraavia ryhmiä:

[2]Suomen rauhanyhdistysten keskusyhdistys.

(a) Kroonisiin tauteihin liittyviä lemmoja ovat *sydänsairaus*, *diabeetikko*, *sepelvaltimotauti*, *sydänlihastulehdus*, *verisuonitauti*, *verenkierto*, *kirurginen*, *alkoholinkäyttö*, *eliö*, *suolisto*, *tupakoitsija*, *munuainen*, *elohopea* ja *vajaatoiminta*. (b) Tarttuviin tauteihin liittynevät *synteettinen*, *lääkefirma*, *malaria*, *ebola*, *Wuhanin*, *Ebola*, *AIDS*, *suojautua*, *sars*, *virologi*, *influenssavirus*, *espanjantauti*, *labra*, *käsihygienia*, *saippua* ja *jätevesi*, *narkolepsia* ja *mRNA*. (c) Terveydenhuoltoon ja sen hallintoon voisivat liittyä *lääkärilehti*, *kuolleisuusprosentti*, *hoitomuoto*, *Journal*, *Health*, *taudinaiheuttaja*, *vaste*, *löydös*, *yliopistollinen*, *kognitiivinen*, *terveysvaikutus*, *terveysongelma*, *magneettinen*, *terveysriksi*, *kg*, ja *mg*.

Kuvan yläreunasta löytyy *lapsikuolleisuus*, mikä voi liittyä siihen, että lapsikuolevuus on Suomessa tunnetusti maailman alhaisimpien joukossa. Lemma *tuhkarokko* voi liittyä samaan huoleen lasten terveydestä, mutta lemma voi myös olla yhteydessä tartuntatauteihin ja rokotuksiin. Lemmasta *Heikkilä* on lisätietoa liitteessä 1.

Erotukseksi s-viesteistä ja m-viesteistä (luku 4.3.3) k-lemmoissa korostuu lääketieteellinen terminologia. Julkisuudessa toistuvasti esiintyvistä alan edustajista on 1960-luvulta alkaen tullut julkkiksia, jotka ovat yrittäneet kertoa selkokielellä suurelle yleisölle terveyden ja sairauden ilmiöistä. Tarkastelujaksollamme tartuntatauteihin on koronan myötä liittynyt paljon huolta ja kirjoittelua (kohta (b)). Asiantuntijoiden haastatteluita välitettiin mediassa epidemian vuosina 2020-2021 lähes päivittäin. Toisaalta kuten kuvasta 3 kävi ilmi, kuolevuus on kehittynyt suhteellisen tasaisesti vuosina 2001-2023: kuolemantapausten lukumäärä on kasvanut väestön ikärakenteen vanhenemisen takia, vaikka ikäryhmittäin tarkasteltuna kuolevuus on laskenut.

Kuva 7: Erottelevimmat kuolleisuus-lemmat

4.3.3 Muuttoliike-lemmat

Kuvaan 8 on valittu 59 eniten m-viestejä erottelevaa lemmaa yhteensä 421:stä. hyvin erottelevasta. Havaitaan ensinnäkin, että 19 valituista liittyy maahanmuuttoon mutta vain 1 maastamuuttooan. Toisin kuin s- tai k-viestien kohdalla, m-viestit näyttävät käsittelevän muita kuin kirjoittajan omia viiteryhmiiä, ts. maahanmuuttajia ja heidän vaikutuksiaan yhteiskuntaan.

Jäljellä olevista voidaan löytää kolme ryhmää. (a) Maahanmuuton hyötyihin voivat liittyä *huippuosaaja*, *rikastuttaja*, *lastenvaunut*, *vetovoimatekijä*, *matalapalkka-ala*, *kulttuuritausta*, *pitseria* ja *monikultturia*. (b) Kotouttamiseen voivat liittyä *kotoutus*, *kielikurssi*, *haalia*, *kotoutu*, *kielikoulutus*, *kotouttamista*, *kotouttamisesta*. ja *puukottaja*. (c) Maahanmuuton haittoihin voivat liittyä seuraavat: *turvapaikkaturisti*, *taikaseinä*, *kunniaväkivalta*, *hyssyttely*, *joukkotappelu*, *sosiaalietu*, *hyysäys*, *ruokajono*, *elintasopako*, *merkkivaate*, *sossuraha* ja *paskaduuni*. (Näiden ohella esiintyy kaksi poliitikkoa (*Sebastian*, *Rydman*). Jäljempänä on useita vastaavia esimerkkejä, joita emme erityisesti kommentoi.) Liitteessä 1 on lisätietoa lemman *Dublin* käytöstä.

Kielteistä näkemystä kuvaavat lemmat (c) ovat usein murteellisia ja sävyltään pilkallisia, kun taas myönteiseen asenteeseen tai kotoutumiseen liittyvät lemmat (a) ja (b) vaikuttavat neutraaleilta.

Kuva 8: Erottelevimmat muuttoliike-lemmat

Kuva 9: Syntyvyys-lemmat, jotka erottelevat parhaiten periodin 4 periodista 6 (vasen), ja lemmat, jotka erottelevat parhaiten periodin 6 periodista 4 (oikea).

5 Muutokset kielessä

Luvun 4.3 sanapilvistä kävi ilmi, että vaikka v-, s-, k- ja m-viesteille on yhteistä se, että ne kaikki liittyvät väestön muutosvirtoihin, löytyy runsaasti lemmoja, jotka erottelevat aihepiirejä toisistaan intuitiivisesti ymmärrettävällä tavalla. Kuvasta 5 käy ilmi, että v-, s-, k- ja m-viestejä erottelevia lemmoja on sadoittain, ja ne muuttuvat ajanoloon. Seuraavassa hyödynnetään luvussa 3.4 luotua jakoa periodeihin (tehtävät I, II ja III) sekä luvussa 4.2 kehitettyä erottelevien lemmojen luokitusta kielen muutoksen analyysiin.

Sanaston osajoukossa, joka oli pohjana kuvassa 5, oli 1 172 lemmaa. Kukin oli valitun kriteerin perusteella *ainakin jollain periodilla ainakin jonkin aihepiirin kannalta* erottelevia. Tästä heuristisesti valitusta joukosta voidaan kunkin tehtävän I, II ja III kohdalla selvittää, mitkä lemmat ovat erityisesti poikenneet keskimääräisestä jommalla kummalla vertailuperiodilla. Tämän vertailun tulokset esitetään kohdissa 5.1 - 5.4. Saadut tulokset ovat pohjana viestien luennalle, jonka tulokset esitellään luvussa 6.

5.1 Erottelevien sanojen muutokset

Muutokset viesteissä käytetyssä sanastossa haijastavat muutoksia ilmiöissä. On selvää, että kun koronavirus ensi kertaa aiheutti pandemian, se tuotti joukon uusia termejä kaikenlaiseen julkiseen keskusteluun, ml. *Suomi24*:ssä käytävään. Samalla kuitenkin syntyi uusia näkökulmia yhteiskunnallisiin asioihin ja niiden tärkeysjärjestykseen. Keskustelupalstaan osallistujat halusivat viestittää näkemyksensä viiteryhmilleen ja muille. Tämä näkyy muutoksena erottelevissa lemmoissa. Kuitenkaan, kuten alta käy ilmi, muutosten merkitystä ei pelkkien lemmojen perusteella ole aivan helppo kuvailla.

5.2 Syntyvyys-lemmojen muutokset

Kuvasta 2 kävi ilmi, että syntyvyys oli laskenut voimakkaasti periodin 2 jälkipuoliskolta alkaen mutta s-viestien määrä kasvoi jyrkästi vasta periodin 4 lopulla. Viestien määrä laski läpi periodin 5 noustakseen uudelleen periodin 6 loppua kohti.

Kuvan 9 kaksi paneelia on tehty seuraavasti: Erottelevien s-lemmojen joukossa (537 kpl) on tarkasteltu järjestyslukupoikkeamien erotusta periodien 4 ja 6 välillä. Olkoon w po. joukkoon kuuluva lemma. Sen järjestysluvun poikkeavuus periodilla j on $W_{w,2,j} - w$, kun esiintyvyyttä tarkastellaan s-viestien joukossa, jolloin $i = 2$. Näiden erotus periodeilla $j = 4$ ja $j = 6$ on $W_{w,2,4} - W_{w,2,6}$. Olkoon erotusten kerkiarvo m ja keskihajonta S. Periodia $j = 4$ kuvaamaan on etsitty ne lemmat w, jotka toteuttavat ehdon $W_{w,2,4} - W_{w,2,6} > m + S$. Ehdon täyttäviä lemmoja oli 87 kpl. Nämä on kuvattu vasempaan sanapilveen. Oikea sanapilvi on saatu soveltamalla samaa menettelyä mutta vaihtamalla periodien roolit. Tuloksena oli 81 lemmaa.

Kuvaa 9 tulkittaessa on syytä pitää mielessä, että vierekkäiset sanapilvet on konstruoitu niin, että ne ovat toisensa poissulkevia. Periodeilla on tietysti runsaasti yhteisiä lemmoja, jotka ovat tässä kuvattuja yleisempiä, ja joista on esimerkkejä kuvassa 6. Lemmojen koot määräytyvät niiden poikkeavuuden mukaan valituissa osajoukoissa.

Sanapilviparia koskeva vertailu voidaan perustaa kuvasta 6 löydettyihin ryhmiin. Uskonnollis-moraalisiin lemmoihin (kuten luvun 4.3.1 kohdassa (b)) kuuluvat voidaan periodilla 4 laskea *SRK:n*, *homoliitto*, *Abraham*, *eaa.*, *luomakunta*, *Vatikaani*, *Jahve*, *perusarvo*, *kreationismi*, *sekulaari*, *insesti*, *vanhoillinenlestadiolaisuus*, *islaminusko*, *avioliittolaki*, *uudestisyntyminen*, *autuas* ja *vihkiä*. Periodilla 6 tätä ryhmää ei näyttäisi olevan ollenkaan, jollei sitten tulkita, että lemmat *palvoja* ja *kuolematon* sekä lyhenteet *eKr.*, *eKr* ja *jKr* liittyisivät uskonnollisiin pohdintoihin.

Raskauteen, yhdessäoloon ja fysiologiaan liittyviä lemmoja (luvun 4.3.1 kohdat (a), (c) ja (d)) on vain muutamia kummassakin sanapilvessä. Ne eivät siis erottele näitä kahta havaintoperiodia.

Sen sijaan oikeanpuoleisesta sanapilvestä voidaan erottaa luvun 4.3.2 ryhmä (c) terveydenhuolto ja sen hallinto. Tähän voivat liittyä *kannustin*, *maksimaalinen*, *masennuslääke*, *hyvinvointialue*, *viisikko*, *hoitojono*, *hierakia*, *omavarainen*, *työttömyysturpa*, *protestanttinen*, *ikäjakauma*, *mielenterveysriski*, *taloustieteilijä*, *ajallinen*, *kuvaaja*, *pistos*, *puoluehallitus*, *haja-asutusalue*, *itsetuhoinen* ja *Gatesin*. Periodilla 4 tätä aihepiiriä voivat edustaa *genomi*, *lääkäriliitto*, *autoimmunisairas*, *pakkokeino*, *kansalaisaloite*, *edelläkävijä*, *peruskoulutus* ja *suuruusluokka*.

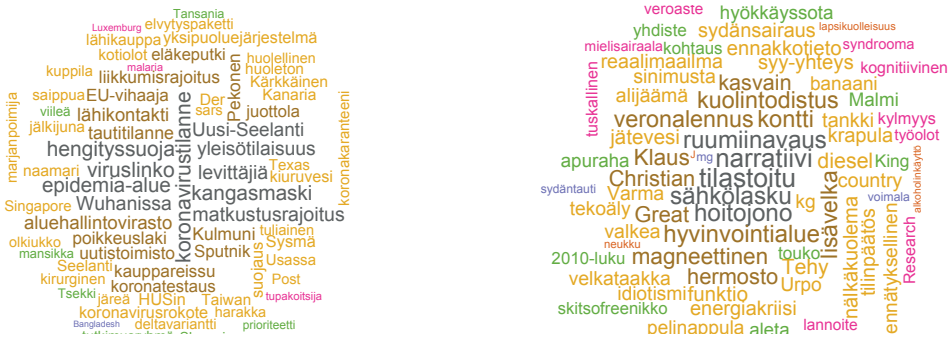
Liitteessä 1 on lisätietoja lemmojen *Abraham* ja *Moktar.Gubta* käytöstä.

Kun periodia 4 parhaiten erotteleva aihepiiri on uskonnollis-moraaliset kysymykset, niin periodilla 6 erottelevaksi aihepiiriksi on vaihtunut terveydenhuolto ja sen hallinto. Luvussa 6 pyritään tarkentamaan tätä eroa: sitä, miksi periodilla 4 oli ylipäätään purske sekä sitä, missä määrin keskustelut liittyvät syntyvyyden romahdukseen.

5.3 Kuolleisuuslemmojen muutokset

Kuvasta 3 kävi ilmi, että kuolleisuus-teema ei vuosiin juurikaan kiinnostanut *Suomi24*:n kirjoittajia, kunnes viestien määrä räjähdysmäisesti kasvoi periodilla $j = 5$, selvästikin koronaepidemian takia. Periodilla $j = 6$ k-viestien määrä oli merkittävästi pienempi, mutta kuitenkin selvästi suurempi kuin periodeilla $j = 1, 2, 3, 4$. Erottelevia k-lemmoja oli 592 kpl. Kuvat 10 ja 11 on tehty samalla tavalla kuin kuva 9. Vasemmassa sanapilvessä on ollut pohjana 68 lemmaa, oikeassa 72.

Analogisesti syntyvyyslemmojen tapaan käytämme kuolleisuuslemmojen luokkia luvusta 4.3.2. Korona-epidemiaan liittyvät sanat periodilta 5 menevät joko luvun 4.3.2 ryhmään (b) tarttuvat taudit tai (c) terveydenhuolto ja sen hallinto. Niihin voidaan vasemmasta sanapilvestä 10 lukea *koronavirustilanne*, *hengityssuoja*, *yleisötilaisuus*, *viruslinko*, *epidemia-alue*, *Wuhanissa*, *Uusi-Seelanti*, *yleisötilasisuus*, *levittäjä*, *kangasmaski*, *matkustusrajoitus*, *poikkeuslaki*, *kauppareissu*, *koronatestaus*, *lähikontakti*, *tautitilanne*, *koronakaranteeni*, *HUS:in*, *koronavirusrokote* ja *deltavariantti*. Lemma *lapsikuolleisuus* liittynee myös kohtaan (c).



Kuva 10: Kuolleisuus-lemmat, jotka erottelevat parhaiten periodin 5 periodista 6 (vasen), ja lemmat, jotka erottelevat parhaiten periodin 6 periodista 5 (oikea).

Periodia 6 erottelevasta sanapilvestä tartuntatauteihin liittyviä lemmoja ei löydy. Sen sijaan terveydenhuoltoon ja sen hallintoon liittyviä kohtaan (c) liittyvät *tilastoitu*, *hoitojono*, *ruumiinavaus*, *kuolintodistus*, *syy-yhteys*, *reaalimaailma*, *alijäämä*, *Tehy*, *ennätyksellinen* ja *tilinpäätös*. Kroonisia sairauksia erottelevaan kohtaan voivat liittyä *sydänsairaus*, *mielisairaala* ja *krapula*.

Näistä tarkasteluista nähdään, että koronaepidemia toi mukanaan sekä siihen itseensä että sen hallintayrityksiin liittyvän terminologian. Kiinnostus kroonisiin sairauksiin sekä terveydenhuolto ja sen hallinto, joka näkyy selvästi kuvassa 7, alkoivat erottua periodilla 6.

Liitteessä 1 on lisätietoja lemmojen *Klaus*, *Christian* ja *King* käytöstä.

5.4 Muuttoliike-lemmojen muutokset

Kuvasta 4 kävi ilmi, että periodilla $j = 3$ oli iso m-viestien purske. Purske liittyy selvästi Euroopan pakolaiskriisiin vuosina 2014-2015. Esimerkiksi Suomeen tuli vuonna 2015 noin 32 000 turvapaikanhakijaa. Suomen nettomaahanmuutossa ei tuolloin kuitenkaan tapahtunut merkitttäviä muutoksia. Sitä vastoin kun periodina $j = 6$ tapahtui merkittävä nettomuuttosokki, m-viestien määrä oli suurin piirtein samalla tasolla kuin periodeilla 4 ja 5.

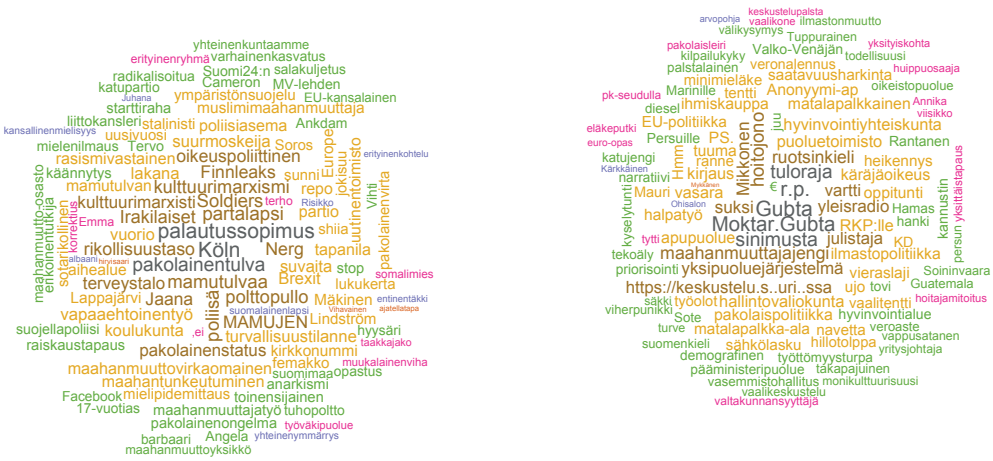
Erottelevia m-lemmoja oli 650 kpl. Kuva 11 pohjautuu vasemmassa sanapilvessä 109 lemmaan, oikeassa 110:een. Noin 60 erottelevinta on mahtunut kuviin mukaan.

Luvussa 4.3.3 muodostettiin kolme m-lemmoja erottelevaa ryhmää: (a) maahanmuuton myönteiset vaikutukset, (b) kotoutus ja (c) maahanmuuton kielteiset vaikutukset. Nämä löytyvät vasemmasta sanapilvestä mutta niitä keskeisempänä periodia 3 erottelevassa ryhmänä sanapilvessä näkyy rajan yli pyrkiviin pakolaisiin reagointi. Tähän liittyviä lemmoja lienevät *palautussopimus*, *pakolainentulva*, *MAMUJEN*, *mamutulvaa*, *mamutulvan*, *pakolainenstatus*, *maahantunkeutuminen*, *pakolainenongelma*, *pakolainenvirta*, *katupartio*, *käännytys*, *turvallisuustilanne* ja *taakkajako*. Näihin lemmoihin näyttäisi liittyvän negatiivinen sävy. Lemmat *kulttuurimarxismi* ja *kulttuurimarxisti* viitannevat kriittiseen sävyyn Frankfurtin sosiologikoulukunnan edustajiin. (*Kulttuurimarxisti* esiintyy myös kuvan 9 vasemmassa paneelissa, joka liittyy periodiin 4.)

Oikeanpuoleisessa sanapilvessä pakolaisongelmaan reagointi ei samalla tavalla näy, vaikka lemma *maahanmuuttajajengi* onkin keskeinen. Tähän näkökulmaan liittyvät kenties myös *hillotolppa* ja *vieraslaji*.

Vaikka maahanmuuton vaikutuksia kuvaavat ryhmät (a), (b) ja (c) näkyvät periodia 6 erottelevassa oikeassa pilvessä, niiden käsittely käytännön hallintotyössä näyttää korostuvan. Esimerkkejä tähän aihepiiriin luettavista lemmoista näyttäisivät olevan *hoitojono*, *tuloraja*, *puoluetoimisto*, *RKP:lle*, *hallintovaliokunta*, *pakolaispolitiikka*, *hyvinvointialue*, *matalapalkka-ala*, *matalapalkkainen*, *saatavuusharkinta*, *ihmiskauppa*, *hyvinvointiyhteiskunta*, *veronalennus*, *käräjäoikeus*, *työolot*, *veroaste*, *vappusatanen*, *halpatyö*, *eläkeputki*, *työttömyysturva*, *apupuolue*, *minimieläke*, *kyselytunti*, *vasemmistohallitus*, *vaalikeskustelu* ja *valtakunnansyyttäjä*. Jopa lemma *demografinen* on päässyt mukaan!

Liitteessä 1 on lisätietoja lemmojen *MV-lehden*, *Ankdam*, *Soldiers*, *Köln*, *Gubta* ja *Moktar.Gubta* käytöstä.



Kuva 11: Muuttoliike-lemmat, jotka erottelevat parhaiten periodin 3 periodista 6 (vasen), ja lemmat, jotka erottelevat parhaiten periodin 6 periodista 3 (oikea).

6 Muutokset aiheissa ja käytetyssä kielessä

Kun edellä on tarkasteltu käytettyjen sanojen muutoksia tehtävissä I, II ja III, niin nyt pyrimme vastaamaan kysymykseen, miksi viestit on ylipäätään kirjoitettu. Mikä on niiden yhteys väestösokkeihin, joita koskeva tilastoaineisto on kiistatonta? Vai tulevatko väestönmuutokset puheeksi enemmänkin kiinnostavampien aiheiden oheistuotteena? Lopulta pitää ottaa huomioon, että *Suomi24*-keskusteluilla on myös ajanvieterooli.

Vastataksemme näihin kysymyksiin luimme kunkin tutkimustehtävän I, II ja III kohdalla kahdelta periodilta 196 viestin satunnaisotoksen.[3] Kustakin viestistä selvitettiin ko. tutkimustehtävän kannalta tarpeelliset tiedot.

[3]Tällöin 10 %-yksikön ero tulisi havaituksi todennäköisyydellä $>$ 50 % testillä, jonka merkitsevyystaso on 5 %.

Edeltävien tarkasteluiden pohjalta kehitimme luokituksen, jonka avulla eri aiheita koskevia keskusteluita voitiin verrata. Kunkin otoksen 196 viestin osalta pyrimme määrittämään, liittyvätkö ne yhteen tai useampiin seuraavista viidestä luokasta:
(1) Käsitteleekö viesti po. demografisen muutosvirran syytekijöitä tai seurauksia?
(2) Esitteleekö viesti kirjoittajansa uskomuksia tai asenteita?
(3) Liittyykö viesti talouden tai hallinnon demografiseen rooliin?
(4) Sisältääkö kirjoitus lähdeviitteitä argumentaation tueksi?
(5) Onko kyseessä herja, lyhyt heitto tai pitempi purnaus?

Kysymykset eivät ole toisensa poissulkevia, eivätkä kata kaikkia mahdollisuuksia. Niinpä esimerkiksi viesti, joka sisältää vain listan tilastotietoja ei kuulu mihinkään viidestä luokasta. Kriteerinä kohtaan (1) kuulumiselle pidimme sitä, että viestissä ilmaistiin arvio syysuhteesta. Luokkaan (5) kuuluvalla ”herjalla” tai ”heitolla” tarkoitamme loukkaavia tai ivallisia puhekielimäisyyksiä sisältäviä viestejä.[4] Se voi sisältää väitteitä syytekijöistä ja kuulua myös luokkaan (1). Luokkien (2) ja (5) ero tehtiin usein viestin lopussa olevien huudahdusten, huutomerkkien yms. perusteella. Luokka (3) pyrittiin rajoittamaan koskemaan viestejä, joissa valtion tai julkisen sektorin päätöksenteko on eksplisiittisesti puheena. Kohdan (4) osalta päädyttiin siihen, että täsmällinen viite oli viestissä annettu. Koko joukko viittauksia, joissa lähes sama informaatio annettiin laveammalla tavalla, jäi näin pois.

Vapaamuotoisten viestien luokittelu muistuttaa avovastausten käsittelyä kyselytutkimuksissa (ks. esim. The SAGE Encyclopedia of Communication Methods.) Kysymysten (1)–(5) merkitys täsmentyi vasta luennan edetessä. Niinpä luokittelu tehtiin kahteen kertaan. Koska luenta perustui satunnaisotoksiin, ilmoitamme luokkien prosenttiosuudet ja osuuksien otantavirhettä kuvaavat keskivirheet suuntaa-antavina.

Avovastausten hyötynä on pidetty sitä, että vastaaja voi tuoda esiin tärkeinä pitämiään näkökulmia, joita kyselyn tekijä ei ole tullut ajatelleeksi. Ilmeisenä haittana on kuitenkin se, että tulosten analyysi on työlästä ja tulokset voivat riippua tutkijan ennakkokäsityksistä. Oman luentamme ohella käytimme tästä syystä Gemini Notebook -tekoälyä otosten sisältämien viestien sisällönanalyysiin. Järjestelmälle voi antaa mm. tekstimuotoista aineistoa ja pyytää sitä tekemään aineistosta yhteenvedon. Pyrimme näin saamaan riippumattoman kuvauksen aineistojen tärkeimmistä piirteistä.

Seuraavassa kirjoitamme *kursiivilla* viesteistä poimittuja sanoja tai lauseita. Lainauksia on selkeytetty lyhentämällä ja kevyellä editoinnilla (sisältöä muuttamatta mutta tekstiyhteyteen sovittamalla), koska ne on joka tapauksessa valittu edustamaan useammassa viestissä esiintulleita näkökulmia.

6.1 Syntyvyys-keskusteluiden muutos periodilta 4 periodille 6

Syntyvyyden romahduksen sokkiluonne perustuu sen pitkän ajan kumuloituvaan vaikutukseen. Sitä vastoin s-viesteissä periodilla 4 havaittavan purskeen kesto oli vain muutamia kuukausia ja siitä oli vähäisempiä merkkejä myös perioidilla 6. Periodilla 4 oli kaikkiaan 8 882 s-viestiä. Periodilla 6 niitä oli merkittävästi vähemmän eli 5 021. Lähtökohtana viestien luennalle oli havainto, että viestien ja syntymien lukumäärät eivät vastanneet millään yksinkertaisella tavalla toisiaan.

Periodi 4. Otoksen viesteistä noin $32_{\pm 3.4}$ % (Keskivirhe alaindeksissä kuvastaa otosvirhettä mutta ei sitä epävarmutta, joka liittyy viestien luokitteluun kirjoittajien toimesta.) käsitteli

[4] Ivallisuus ei ole vain *Suomi24*-ilmiö. The New York Times kertoi 22.08.26, että presidentti Trump kutsuu Kalifornian kuvernööriä *Newsome*a somessa väännöksellä *Newscum*.

kausaalikysymyksiä (1). Syntyvyyden matalan tason katsottiin johtuvan asenteiden muutoksista. Selvästi suurempi osa po. viesteistä käsittelikin ensisijassa yleisiä asenteiden muutoksia ja syntyvyyden tasoa niiden oheistuotteena. Useissa viesteissä ilmaistiin käsitys, että matala syntyvyys johtaa lopulta väestön vähenemiseen ja tarpeeseen lisätä maahanmuuttoa. Toisaalta useassa viestissä tuotiin esiin maapallon liikakansoitus hieman 1960-luvun tapaan, ja pohdittiin keinoja sen vähentämiseen.

Noin $48_{\pm 3.6}$ % viesteistä käsitteli ensi sijassa kirjoittajan omia uskomuksia ja asenteita (2). Niissä näkyy laestadiolaisten keskinäistä keskustelua, jota odotimme näkevämme luvun 5.2 tulosten takia. Tämän ohella otoksessa oli kärjekästä samaa sukupuolta olevien avioliittojen ja valtionkirkon vastustusta (*Papit haluaa, että heterot eroaa kirkosta*) sekä esimerkiksi tyytymättömyyttä miehen asemaan pariutumismarkkinoilla. Kantaväestön matalan ja muslimi- ja somaliväestön korkean syntyvyyden ero tuodaan esiin (*Suomessa sikiää vain lestadiolaiset ja matut*), ja nähdään sen johtavan kantaväestön häviämiseen. Ilmiöiden syinä mainitaan liikasuvaitsevaisuus suhteessa muihin kulttuureihin ja mm. sosiaalisen median negatiivinen vaikutus perhe-elämään.

Noin $10_{\pm 2.1}$ %:ssa viesteistä oli selkeä taloudellisiin oloihin liittyvä painotus ja panimme ne luokkaan (3). Lapsiperheköyhyyttä pidetään yhtenä matalan syntyvyyden syynä, ja hyvinvointipalvelujenkin katsotaan tuottaneen elintavan, johon lapset eivät kuulu: *valkoihoisten syntyvyys laskee, kiitos siitä hallitukselle, joka leikkaa valkonaamojen lapsilisiä, päivähoitoa! Miten perheen elättää jos ei saa työstään palkkaa jolla pärjää.* Toisaalta eräässä viestissä pidetään yksinhuoltajien taloudellista tukea liian runsaana.

Eksplisiittisiä lähteitä (4) oli $9_{\pm 2.1}$ %:ssa otoksen viesteistä. Viesteistä näkyy selvästi, että osa kirjoittajista pyrkii (Aristoteleen oppien mukaisesti) perustelemaan väitteensä tällä keinolla. Kun Raamatun jakeisiin viitataan uskonnollissävytteisissä viesteissä, niin taloudellisesti painottuneissa tietolähteinä mainitaan sekä media (YLE, Helsingin Sanomat, Suomen Kuvalehti) että viranomaiset (Tilastokeskus, Eläketurvakeskus). Eräässä kirjoituksessa viitataan *NewScientist* -lehteen, mikä osoittaa, että kirjoittajajoukko on heterogeeninen.

Tyylillisesti heikkotasoisia herjoja ja heittoja (5) on otoksen viesteistä noin $22_{\pm 3.0}$ %. Näiden kirjoittajat ovat usein itsetietoisia ja katsovat, että heillä on varaa nokkeluuksiin muiden kustannuksella.

Periodi 6. Kausaaliajatteluun kuuluvaa luokkaa (1) edusti periodin 6 s-viesteissä noin $47_{\pm 3.5}$ % viesteistä, kun periodilla 4 tähän luokkaan saatoimme luokitella $32_{\pm 3.4}$ %. Vaikutelma on, että matalan syntyvyyden yhteys maahanmuuttoon ja kotoutumisen ongelmiin on kirjoittajille kirkastunut, aiempaa paremmin. Eräs pohtii: *Suljetun Suomen ominaisuus on nolla mamua. Tarkoittaisi että kaikki mamut pitäisi pakkopalauttaa alkuperämaahansa. Tämä rikkoisi näiden ihmisten perusoikeuksia ja olisi Suomelle haitallista, koska työvoimaa menetettäisiin esimerkiksi terveydenhuollossa. Työvoimapulaa ei voitaisi helpottaa tulevaisuudessa, koska uusia mamuja ei luonnollisesti otettaisi.*

Kirjoittajan omien mielipiteiden ja asenteiden (2) esittelyä painotettiin periodilla 6 noin $60_{\pm 3.5}$ %:ssa viesteistä eli selvästi enemmän kuin periodilla 4. Laestadiolaisuuden piirissä käyty keskustelu ei periodilla 6 enää otoksessa näy kuin muutamana viestinä. Vaikka joitain viittauksia Raamatun teksteihin vieläkin näkyy, uskonnon sijaan viesteissä korostuu näkemys pariutumisen ongelmista, joiden nähdään johtuvan erityisesti naisen roolin muutoksesta (*peräkammarinpojat eivät kelpaa nykynaisille*; *suomalaiset naiset on vemakoita*; *kaikki seksuaalisuus on nykyään häirintää*) ja avioliittoinstituution heikentymisestä (*erotaan heti kun tulee vaikeuksia*). *Vihervasemmistolaisen ajattelun* leviämisen (*homoja löytyy joka käänteeseen ja paikkaan*; *Seta on syyllinen*) ja sukupuolen käsitteen epämääräistymisen ajatellaan johtavan syntyvyyden laskuun. Ongelmana pidetään sitä, että Suomen syntyvyys on alempi kuin Ruotsissa ja Virossa

ja on jo Etelä-Euroopan matalalla tasolla. Keskusteluista löytyy kuitenkin edelleen monissa viesteissä esitetty näkemys, että maailman tulevaisuuden kannalta matalampi syntyvyys on välttämätöntä.

Viestejä, joissa on selvä talousnäkökulma (3), oli periodin 6 viesteistä noin $21_{\pm 2.9}$ %, eli enemmän kuin periodilla 4. Viesteissä nähdään mm., että vika on hallinnossa (*naispäättäjät ovat ajaneet naiset ansiotyöhön*; *kotimaisen syntyvyyden vei ihan kotimainen politiikka, tulevaisuus on epävarma ja rahat tiukalla*) ja tukipolitiikassa (*somaleille turha antaa rahaa, työllistyvät heikosti*; *maailman suurimmat tuet lapsiperheille ei saa syntyvyyttä kasvuun*). Ilmastonmuutoksesta kannetaan huolta (*vastaus ilmasto ongelmiin on syntyvyydensäännöstely*).

Toisaalta osa kirjoittajista näkee maahanmuuton kasvun välttämättömäksi, vaikka sen mahdollisuuksiin suhtaudutaan epäilevästi (*Suomen sijainti on syrjäinen, ilmasto kylmä, kieli outo. Lisäksi yli 1 000 km rajaa Venäjän kanssa*; kiinnostavasti vedotaan mm. Pentti Linkolaan: *teollisuusmaiden kuormitusta kasvattavat kaikkein kiivaimmin köyhistä maista saapuneet siirtolaiset, joiden syntyvyys noudattaa heidän lähtökulttuuriensa tasoa tai sosiaalihuollon korkeamman tason ansiosta ylittääkin sen.*

Keskustelusta löytyy myös huomautuksia siitä, että hedelmällisyys on ollut alle uusiutumistason jo 1960- 1970-lukujen vaihteesta, siis *kultaiselta Kekkosen ajalta, jolloin alle työikäisten ja eläkeläisten ikäluokat olivat pieniä.* Periodista 4 poiketen periodin 6 otokseen on sattunut myös normaalin asiaproosan muodossa esitettyjä arvioita esimerkiksi vihreiden puolueohjelman syntyvyyttä sivuavista näkemyksistä sekä Pohjois-Karjalan ja Uudenmaan kokonaishedelmällisyyden eroista. (Kirjoituksessa siis esiintyy muuten harvinainen *kokonaishedelmällisyys*-sana, joka oli otettu mukaan aineiston valinnan kriteeriksi juuri sen selvittämiseksi, voisiko se erottaa asiatyylillä kirjoitettuja viestejä muista. Tässä tapauksessa näin tosiaan oli.)

Lähdeviittauksia (4) periodilla 6 oli $12_{\pm 2.3}$ %, eli kutakuinkin sama määrä kuin periodilla 4. Vaikutelma on, että viittaukset nettiin olisivat lisääntyneet, mutta nytkin otokseen sattui mukaan yksi viesti, jossa viitattiin tutkimuskirjallisuuteen (*Annals of Medicine*).

Heikkotasoisten herjojen ja heittojen (5) osuus oli arviolta $16_{\pm 2.6}$ %, eli vähemmän kuin periodilla 4.

Yhteenveto tuloksista. Syntyvyyden pienenemisen keskeisimpänä syynä nähdään *Suomi24*:ssä asenteiden muutos Suomen kantaväestössä. Tälle löydetään selitykseksi mm. naisten työssäolo, feminismi ja perheinstituution heikentyminen. Ongelmaksi nousee tällöin kantaväestön absoluuttinen väheneminen ja korvautuminen maahanmuuttajilla, joilla on paljon lapsia, mutta joita kuitenkin tarvitaan työvoimaksi. Periodilla 4 havaittu s-viestien purske seuraa periodilla 3 tapahtunutta turvapaikanhakijoiden aaltoa Eurooppaan ja Suomeen.

Huippukuukauden viestit. Tarkastelimme viestejä tarkemmiin, kun periodin 4 viestipurske oli huipussaan kuukaudella 226 (siis vuoden 2019 lokakuussa), jolloin viestejä tuli 542 kpl. Viesteistä käy ilmi, että viestipurskeen aiheutti Tilastokeskuksen väestöennuste, joka julkaistiin 30.09.2019. Ennusteessa kokonaishedelmällisyyden tasoksi oli valittu tuolloinen arvo 1.35, joka oli tilastohistorian alhaisin. Seurauksena tulevat syntymät olivat noin 10 000 alemmat kuin vuoden 2015 ennusteessa. Eräs realisti kirjoitti: *Syntyvyys laskenut vuosikymmeniä, alkoi 1995-2005. Vauhti kiihtyi 2010 tasaantuen lähivuosina.* Syyllisiä alettiin kuitenkin etsiä: *Mamutulva alkoi neljä ja suomalaisten syntyvyyden notkahdus kolme vuotta sitten. Yhteys täytyy olla!* Toinen säesti: *Päivän uutisissa rummutetaan syntyvyyden laskua ja sitä, miten rahoituspohja romahtaa. Vauvakato on pelkkä keppihevonen maahanmuuton lisäämiseksi.* Ja edelleen *selvästi mamutulva ahdistaa suomalaisia niin, että se vie lisääntymishalut* kirjoittaa kolmas. Käy siis ilmi, että kirjoittajien mielissä yhteys pakolaiskriisin ja syntyvyyskeskustelun välillä oli olemassa. Kaikki eivät ole kuitenkaan samaa mieltä: *Mamuilla ei ole tähän osaa ei arpaa muualla*

kuin EU-vihaajien valeissa. Toisaalta useat kirjoittajista katsovat, että koko maailman tasolla väestön kasvu on yhteydessä ilmastonmuutokseen, ja sitä pitäisi pyrkiä hillitsemään.

Akateemista kirjoitustapaa ei pieniä poikkeuksia lukuun ottamatta löydy. Noin 1/5 viesteistä on heikkolaatuisia herjoja ja heittoja. Ne voivat vaikuttaa muidenkin viestien kieleen. Olkoonkin että viesteissä on merkkejä siitä, ettei omanarvontuntoisiin kirjoittajiin vetoa löperö argumentaatio. Näiden kirjoittajien joukko ei rajoitu ”lähdeviittaajien” kymmenykseen. Olisiko tämä ”harkitsevien” joukko sekin noin 1/5?

Litteessä 2 on esitetty, sellaisenaan, Gemini Notebookin tekemät yhteenvedot periodien 4 ja 6 s-viestiotoksista. Nähdään, että vaikka koosteet on laadittu hieman eri logiikalla ja kirjoitettu eri tyylillä kuin omamme, tulokset ovat olennaisilta osin samat. Vertailu osoittaa, että koska tekoäly tuottaa tulokset vähällä vaivalla ja nopeasti, se on s-viestien osalta kelpo työkalu tutkimuskäyttöön.

6.2 Kuolleisuus-keskusteluiden muutos periodilta 5 periodille 6

Tutkimustehtävä II voisi liittyä kuolevuuden trendinomaiseen laskuun, mutta tosiasiassa se liittyy lähinnä korona-epidemian aiheuttamaan purskeeseen k-viesteissä periodilla 5. Luvussa 5.3 koronan vaikutus korostui periodia parhaiten erottelevissa lemmoissa, mutta hieman samaan tapaan kuin syntyvyyden muutosten kohdalla periodia 6 erottelivat parhaiten terveydenhuoltoon ja erityisesti sen hallintoon liittyvät lemmat. Löytyykö tälle vastinetta k-viesteissä?

k-viestien lukumäärä periodilla 5 oli 285 887. Periodilla 6 se oli enää 77 966, mutta tämäkin luku on noin kymmenkertainen s-viestien määrään verrattuna. Juuri tämä tavaton suuruusluokkaero oli sanafrekvenssien jakaumien vinouden ohella syynä järjestysluku-muunnoksen käyttöön tilastollisessa analyysissä.

Periodi 5. Ensinnä on sanottava, että kroonisiin sairauksiin liittyviä viestejä oli niin vähän, ettei niitä 196 viestin otokseemme sisältynyt ainuttakaan. Tämä on hyvin linjassa kuvan 3 kanssa, mistä näkyy, että kiinnostus kuolleisuuteen on ylipäätään ollut viestipalstalla vähäistä.

Koronan aiheuttamaan kuolemanriskiin tai muihin demografisiin kausaalikysymyksiin (1) liitettäviä viestejä oli periodin 5 otoksessa vain noin $9_{\pm 2.1}$ %. Viesteissä tunnistettiin vanhusten ja jo valmiiksi sairauden heikentämien erityinen riski kuolla koronaan, samoin yksikäsitteisen kuolemansyyn määrittämisen vaikeus. Sekä ulkomaisia että kotimaisia tartunta- ja kuolleisuuslukuja selostettiin useissa viesteissä. Rokotettujen ja rokottamattomien kuolemantapausten määriä vertaillaan ottamatta huomioon riskipopulaatioiden kokoa. Kiinnostavana yksityiskohtana voi mainita, että eräässä viestissä pohdittiin Ruotsin omaksumaa, Suomea liberaalimpaa epidemian hallintapolitiikkaa, siinä määrin yksityiskohtaisesti, että mainituksi tulivat sekä vastuuvirkamies Tegnell että hänen taustavoimakseen nimetty prof. T. Britton[5]. Tämä on osoitus siitä, että *Suomi24*-kirjoittajien joukko on heterogeeninen.

Kirjoittajan omia käsityksiä ja asenteita (2) korostettin $55_{\pm 3.5}$ %:ssa viesteistä. Viesteissä oli kyse tartunnalta suojautumisesta, esimerkiksi maskien käytöstä julkisissa tiloissa, ja sairastumiseen liittyvistä vaivoista (*morbidity*), ei kuolemanriskistä (*mortality*). Vähemmistössä otoksen viesteistä vähätellään tartunnan riskejä, mutta muut kirjoittajat pitivät sitä tietämättömyyden merkkinä. Tilastojen luotettavuutta ja koronapassin hyödyllisyyttä epäiltiin. Käsidesin käytön ja rokottautumisen merkitystä korostetaan useissa viesteissä, mutta rokotusten haittavaikutukset ja tehottomuus huolettavat useita. *Nuorison bilettämisen* arvioidaan levittävän tautia. *Rokottamattomat koronalingot pitäisi laittaa eristykseen* vaaditaan eräässä viestissä.

[5]Tukholman yliopisto. Kenties kansainvälisesti arvostetuin epidemioiden matematiikan tutkija Pohjoismaissa.

Yhteiskunnan ja terveydenhuollon hallinnon roolia koronan hallinnassa (3) korostettiin $10_{\pm 2.1}$ %:ssa viesteistä. Edellä mainitut rajoitustoimet hyväksyttiin pääosin, vaikka matkustukselle aiheutuvia ongelmia tuotiin monessa viestissä esiin. Esimerkiksi Uudenmaan eristäminen mainittiin otoksen viesteissä vain ohimennen. Eräässä viestissä todettiin, että yksittäinen skenaariolaskelma ei ole pätevä perustelu vaalien siirtämiseen. Marinin hallituksen toimia arvosteltiin aluksi liian hitaiksi ja myöhemmin liian järeiksi (*on sotatilalait*). Samoin Ruotsin erilaista koronapolitiikkaa sekä moitittiin että kiitettiin. *Koronanyrkin* puutetta moititaan ja Uudesta Seelannista kehotetaan ottamaan oppia. Vaikutelmaksi jää, että osa keskustelijoista on nähnyt varsin paljon vaivaa ollakseen perillä siitä, miten eri maiden terveydenhuoltojärjestelmät ovat koronaan reagoineet.

Lähdeviitteitä (4) oli $10_{\pm 2.2}$ %:ssa viesteistä. Tämä on käytännössä sama kuin edellä s-viestejä sisältävissä otoksissa, mutta kansainvälisiä lähteitä on huomattavan paljon. Nämä ovat pääosin englanninkielisiä, mutta italialainen La Repubblica on myös mukana. Myös esimerkiksi youtube-videoihin viitattaan. Iltapäivälehdet, YLE ja Helsingin Sanomat ovat mukana. Kuten s-viestien kohdalla yksittäisiä viitteitä tutkimuskirjallisuuteen tuli otokseen mukaan.

Herjoja, heittoja ja nyt käsiteltävässä aihepiirissä myös pitempiä emotionaalisia purnauksia (5) on periodilla 5 noin $33_{\pm 3.4}$ %. Osuus on kaksinkertainen s-keskusteluihin verrattuna. Sävy on usein pilkallinen, kun kohteena on pääministeri, vihreät ja vasemmistopuolueet. Herjaavien viestien isosta määrä näkyy siinäkin, että periodin 5 k-viesteistä huomattavan moni oli erityisen lyhyitä. Dialogi lienee ollut nopeaa!

Periodi 6. Tällä periodilla otoksessa oli kaksi viestiä, joiden voi tulkita koskevan kroonisia tauteja. Kuolemanriskiin liitettäviä viestejä (1) oli periodin 6 otoksessa noin $9_{\pm 2.0}$ % eli saman verran kuin periodilla 5. Eräissä viesteissä todettiin (aivan oikein), että kuolemantapauksia tuli vuonna 2022 enemmän kuin odotettiin. Keskustelun pääpaino oli kuitenkin rokotteissa ja niiden aiheuttamissa sivuvaikutuksissa. Erityistä hämmennystä kirjoittajissa aiheutti rokotuksen ottaneiden ja ei-ottaneiden kuolemien lukumääristä julkaistut tiedot. Riskiväestöjen koosta kun ei ollut samalla olemassa tietoja. Erityisesti mRNA-rokotteen arveltiin useassa viestissä aiheuttaneen kuolemantapauksia.

Kirjoittajan omia käsityksiä ja asenteita (2) korostettin $45_{\pm 3.6}$ %:ssa viesteissä. Tämä on vähemmän kuin periodilla 5. Viesteissä näkyy yksilön valintojen vaikutus ja merkitys esimerkiksi julkisissa tiloissa käyttäytymisessä. Aiemmista poiketen näiden viestien joukossa on useampia henkilökohtaiseen kokemukseen liittyviä sekä myös eräs viesti, joka oli selväesti naisen naisille kirjoittama. Henkilökohtaisen puolustatumisen strategiat ovat esillä. Selvästikin koronan aiheuttama terveysriksi on pääosin otettu vakavasti.

Yhteiskunnan ja terveydenhuollon hallinnon vastuun näkökulma (3) oli noin $13_{\pm 2.4}$ %:ssa viesteistä. Tämäkin on hienokseltaan enemmän kuin periodilla 5. Aihepiirin viesteissä korostuivat hallituksen rokotuspolitiikkaan liittyvät toimet. Viestien joukossa oli myös muutamia sellaisia, joilla ei ollut muuta kuin satunnaista yhteyttä kuoleman riskiin ja terveyteen (esimerkiksi valtion velanotto) sekä puoluepoliittiset kannanotot. Marinin hallituksen velanottoa kritisoidaan mutta toisissa viesteissä yritetään ymmärtää. Samoin pääministeriä itseään. Alatyylisissä viesteissä maalitauluna on enimmäkseen pääministeri ja vihervasemmisto, SDP erityisesti. Rintama ei kuitenkaan ole yhtenäinen. *Persuja* sekä kehutaan että moititaan muutamassa otoksen viesteistä. *Kepua* sekä moititaan että myötäillään.

Lähdeviitteitä (4) oli $13_{\pm 2.4}$:ssa viesteistä. Periodin 5 havainnot pätevät sellaisinaan periodin 6 viittauksiin. Keskustelupalstalle ajautui nytkin kirjoittamaan joukko henkilöitä, joilla on erityinen tarve esittää tietämystään. Tällä kertaa mukaan tuli viesti, jossa oli sivun verran yksityiskohtia eri tyyppisten kasvomaskien suodatuskyvystä, ja toinen, jossa oli kuusi lähdeviittausta ukrainalaisiin lähteisiin. Hieman yllättäen näihin korona-epidemiaa koskeviin viesteihin tuli

mukaan kaksi, joissa vedottiin Raamattuun.

Herjoja, heittoja ja pitempiä purkauksia (5) viesteistä oli noin $28_{\pm 3.2}$ %. Otoksen perusteella nämä viestit näyttävät edelleen lyhentyneen, mikä viittaa siihen, että ne on kirjoitettu hetken innostuksesta. Vaikka joukkoon mahtuu myös asialliseen sävyyn kirjoitettuja heittoja, puolet viesteistä on pääosin riitelyä. Aivan tällaisia ei syntyvyyttä koskevissa keskusteluissa ollut, vaikka niissäkin heikkotasoisia viestejä esiintyi.

Esimerkki aikaisemmasta tasonmuutoksesta k-viesteissä. Kuvasta 3 kävi ilmi, että k-viesteissä oli (koronaan verrattuna) pieni nyppylä kuukauden 107 (eli marraskuun 2009) tienoilla. Lähemmässä tarkastelussa kävi ilmi, että tämä purske aiheutui vuonna 2009 alkaneesta sikainfluenssaepidemiasta, jota vastaan oli kehitetty Pandemrix-rokote. Keskusteluissa epäiltiin rokotuksella olevan sivuvaikutuksia. Vuoden 2010 keväällä tutkijat löysivätkin mahdollisen yhteyden rokotteen ja narkolepsian välillä. Sairastuneille päätettiin maksaa korvauksia, vaikka biologista vaikutusmekanismia ei ole pystytty yrityksistä huolimatta osoittamaan.

Yhteenveto. Kun tehtävässä II haettiin sitä tapaa, millä *Suomi24*-keskustelut vastaavat kuolleisuuden muutoksia, niin lyhyt vastaus on, että yhteys on yllättävä ja kiinnostava, eikä se vastaa ennakko-odotuksia. Tutkijan mielessä kuolleisuus liittyy elinikään ja terveisiin elinvuosiin. Nämä aiheet eivät esiinny periodien 4 ja 6 otoksissamme, vaikka *elinajanodote*-lemman perusteella valituissa viesteissä ne löytyvät. Kuolevuuden aleneminen liittyy kiinteästi väestön ikärakenteen vanhenemiseen. Tämäkään ei näy viesteissä. Sen sijaan kirjoittajat ovat kiinnostuneet, reaaliajassa, korona-epidemian tuottamista ongelmista yksilölle ja yhteiskunnalle.

Korona lienee ollut ensimmäinen pandemia, jota on voitu seurata ympäri maailmaa lähes reaaliajassa radiossa, TV:ssä ja netissä. Tietoa ja taitoa puuttui medialta, mutta ei yleisölläkään ollut valmiuksia sulattaa informaatiovirtaa. Terveydenhuollon hallinnossa jouduttiin tekemään päätöksiä suurennuslasin alla.

Huippukuukauden viestit. Periodin 5 viestipurske oli huipussaan kuukaudella 231, siis vuoden 2020 maaliskuussa. Ensimmäinen koronadiagnoosi oli tehty 29.01.2020 Rovaniemellä Wuhanista kotoisin olleelle matkailijalle. Korpuksessamme mukana olevia k-viestejä saapui maaliskuussa 31 289 kpl, ts. noin 60 kertaa niin paljon kuin periodin 4 s-viestipurskeen suurin arvo, ja noin 1 000 viestiä per päivä. Tutustuimme tämän kuukauden viesteihin otospohjalla. Alatyylisten ja loukkaavien viestien osuus oli suurempi kuin periodeille 5 ja 6 yleensä. Käsitykset epidemiasta olivat ristiriitaisia. Yksi arveli, että *korona ei ole juuri kausi-influenssaa vaarallisempi.* Toinen pelkäsi tartuntaa: *Kummallista, kun Italiaaan pitää mennä hakeman Korona.* Kolmas näki tulevaisuuteen: *Ollaan vielä pulassa koronan kanssa, iäkkäät ihmiset ja vähän tehohoitoa.* Aineistosta paljastui myös maan sisäisiä ristiriitoja: *Etelä-Suomi toi Koronan Lappiin* ja *Hämeessä on voimakas mökkiläisten vastainen ilmapiiri!*

Suomi24:n kuolevuuskeskustelijoiden joukossa oli kriitikkoja, jotka haastoivat *valtamedian* kautta tulevaa informaatiota hakemalla eksoottisiakin tietolähteitä. Vaikka keskustelut olivat usein alatyylisiä ja kirjoituksina heikkotasoisia, nämä piirteet karsimalla kuva korona-epidemiasta oli silti pääosin sama, joka tuli *valtamedian* kautta.

Väestötieteellisessä mielessä tärkeimmät asiat jäivät k-viesteissä katveeseen. Kuolevuus ei kasvanut ns. koronavuosina 2020-2021, ja vaikka se kasvoikin vuonna 2022, niin kuolevuuden aleneminen jatkui vuonna 2023. Samalla kävi ilmi, ettei tämä tulevaan väestönrakenteeseen eniten liittyvä aspekti saanut huomiota keskustelupalstalla.

Liitteen 2 kuolleisuutta koskevat Geminin tekoäly-yhteenvedot poikkeavat varsin paljon edellä esitetystä. Geminiä ei ohjeistettu tarkastelemaan viestejä nimenomaan kuolleisuutta koskevina. Niinpä se onkin painottanut viestien niitä puolia, joissa korostuvat yhteydet esimerkiksi salaliittoteorioihin. s-viestien kohdalla Gemini tarkasteli viestejä samasta näkökulmasta kuin

mekin. Sikäli kuin kuolleisuutta käsitellään, johtopäätökset ovat kuitenkin samansuuntaiset kuin omamme.

6.3 Muuttoliike-keskusteluiden muutos periodilta 3 periodille 6

Probleema III liittyy paradoksiin, joka syntyy siitä, että periodilla 3 oli merkittävä m-viestien purske mutta ei juuri muutosta nettomuutossa. Sen sijaan periodilla 6 oli merkittävä sokki maahanmuutossa mutta m-viesteissä ei vastaavaa muutosta näy. Periodilla 3 *Suomi24*-keskustelupalstalle tuli 142 027 m-viestiä ja periodilla 6 niitä oli 71 095. Periodin 3 luku on puolet koronaperiodin 5 k-viesteistä ja suunilleen sama kuin k-viestin määrä periodilla 6. Kuten luvussa 6.2 todettiin, jälkimmäinen luku on kymmenkertainen s-viestien lukumääriin verrattuna. Toteamme, että eniten väestönmuutoksiin liittyviä viestejä syntyi koronan takia (vrt. kuva 1).

Luvussa 5.4 havaittiin, että periodilla 3 havaittujen maahanmuuttokriittisten lemmojen sijaan periodilla 6 viestitys on muuttunut asiapitoisempaan ja käytännönläheisempään suuntaan. Liittyykö tämä periodin 6 muuttoliikesokkiin?

Periodi 3. Muuttoliikkeen syihin (1) tai sen kausaaliyhteyksiin, esimerkiksi valtiontalouteen, liittyviä viestejä oli vain $3_{\pm 1.2}$ % viesteistä. Määrä s- ja k-viesteihin verrattuna oli pieni. Otoksen viesteistä käy ilmi, että Suomi tarvitsee työllistyvää maahanmuuttoa, mutta syitä siihen, miksi maahantulijat ovat halunneet lähteä kotimaastaan, ei pohdita. *Mamujen* mahdollisuuksia kotoutua Suomeen epäillään. Suomen vetovoimatekijänä pidettiin liian anteliaita yhteiskunnan tukia (mm. lapsilisät ja vastaanottorahat) turvapaikanhakijoille.

Kirjoittajan omia käsityksiä ja asenteita (2) painottavia viestejä oli $40_{\pm 23.5}$ %:ssa m-viestejä periodilla 3. Kirjoittajia harmittaa se, että *suvakit syyllistävät kansaa rasismilla, ärsyttävät ivallisiin kannanottoihin*, vaikka *ihan maltillisia olemme rehellistä työtätekevää maahanmuuttajaa kohtaan.* Tätä useammat näkevät maahanmuuttajat sosiaalitukien väärinkäyttäjinä ja rikollisina (tappelut, ryöstöt, raiskaukset; eräässä viestissä on oikeusministeriön sivuilta kootut tiedot 64 maahanmuuttajien rikostuomiosta ajalta 05.08.2014–06.08.2015). Erityisesti somalit, muslimit ja arabit nähdään ongelmina. Näiden viestien kieli on usein karkeaa ja alatyylistä. Osa kuitenkin huomauttaa suomalaisten itse tekevän runsaasti rikoksia ja olevan syypäitä omiin ongelmiinsa. Lisäksi eräässä viestissä todetaan ilmiön monimuotoisuus: *monet pakolaisista puhuivat sujuvaa englantia, olihan joukossa paljon korkeakoulutettuja ihmisiä kuten lääkäreitä, arkkitehtejä ja hammaslääkäreitä. Siis juuri sellaisia ihmisiä, jollaisia Suomeen tarvittaisiin paikkaamaan työvoimavajetta.*

Valtionhallinnon ja valtiontalouden rooli (3) oli keskeisen $22_{2.9}$ %:ssa viestejä. Pääosassa viestejä arvioidaan turvapaikanhakijakriisin olevan itseaiheutettu. *Varsinaisia syyllisiä ovat lainsäätäjät*, joita ovat mm. *vihermystöliiton valtionvirassa makoilevat luonnonsuojelijat*, *kukkahattutädit* ja *suvakit.* Maahanmuuttajille (*jotka eivät tee päivääkään työtä suomalaisen yhteiskunnan hyväksi*) menevien kotiutus-, työttömyys- ja asuntotukien katsotaan olevan *kantaväeltä* pois. Ääri-islamista ja sharia-laista varoitetaan. Toisaalla todetaan, että maahanmuutto voisi hidastaa maaseudun tyhjentymisen seurauksena syntyvää asuntojen hintojen laskua.

Viitteitä (4) oli käytetty argumentaation tukena $15_{\pm 2.5}$ %:ssa viestejä, eli kutakuinkin sama määrä kuin s-viesteissä, mutta hieman vähemmän kuin k-viesteissä. Lähteiden joukossa ovat mm. Kaleva, Iltasanomat, Iltalehti, mtv, Kotimaa, yle, Uusi Suomi, Helsingin Sanomat, MV-lehti, Talouselämä, wikipedia, Eurostat ja Oikeuspoliittinen tutkimuslaitos. Yleensä viitataan nettisivuille.

Erilaisia herjoja, heittoja ja äriseviä purkauksia (5) oli otoksen m-viesteistä $46_{\pm 3.5}$ %. Tämä on noin 10 %-yksikköä enemmän kuin k-viesteissä ja 30 %-yksikköä enemmän kuin s-viesteissä.

Kirjoittelu on pääosin alatyylistä ja tietoisen ilkeää. Vaikka *Suomi24*-keskustelupalstaa moderoidaan ja viestejä poistetaan, silmiin osuu pari viestiä, jotka saattavat olla osa kriminalisoitavaa keskusteluketjua, ts. rikoksen suunnittelua. *Persuille* irvistellään epäjohdonmukaisuudesta, kun päästävät maahan *mamuja*. Tunnettuja poliitikkoja halveksutaan ja samaan keskusteluketjuun osallistuville nälvitään. Jotkut arvelevat, että ilmaisun teho kasvaa, kun voimasanoja yhdistää: *Olet tyhmä pelle ja älytön apina*.

Periodin 6 erityinen piirre oli, että Venäjän hyökkäyssota Ukrainaan tuotti pakolaisaallon, jollaisesta periodilla 3 ei ollut tietoakaan (olkoonkin, että Venäjä valtasi Krimin niemimaan 2014 eikä peitellyt laajentumishalujaan). Ukrainalaisten pakolaisten suuri määrä periodilla 6 näkyy tähän nähden *Suomi24*-keskusteluissa vain vähän. Olihan se Suomen väestöhistoriassa ennätyksellisen suuri. Keskustelua käytiin edelleen pakolaisten vastaanoton ongelmista ja erityisesti humanitaarisen pakolaisuuden väitetyistä seurauksista.

Muuttoliikkeen kausaaliproblematiikkaan (1) liitettäviä oli periodilla 6 selvästi enemmän kuin periodilla 3, nyt $21_{\pm 2.9}$ % viesteistä. Kuten periodilla 3 keskustelijoita ei kiinnostanut, mikä on saanut muuttajat lähtemään kotimaastaan. Se, että Suomi tarvitsee työvoimaa, näyttää tulleen hyväksytyksi, mutta avokätisinä pidettyjen tukien arveltiin vaikuttaneen tulijoiden määrään. Outojakin näkemyksiä esitetään: *Ilmeisesti migrin hallinnossa naisia, jotka suosivat miespuolisia maahantunkeutujia ja katkeria naisille, jotka oppivat suomen ja työllistyvät vanhuksien hoitajina*. Samoin: *Työperäinen maahanmuutto pahentaa ongelmaa. Suomi on Euroopan perälkorpea. Ei tule korkean tuottavuuden alojen osaajat, vaan sosiaalitukielätit ja B-luokan osaajat.* Mutta toisaalta: *Vaihtoehtona suomalaiset mummot ja papat kuolevat nälkään sekä hoivan puutteeseen. Mamut kuskaavat ruokaa ja tavaroita kohtuuhintaan kotiovelle.*

Kirjoittajan uskomukset ja asenteet (2) painottuivat $39_{\pm 3.5}$ %:ssa viestejä, ts. osuus oli käytännössä sama kuin periodilla 3. Kannanotot ovat aiempaa enemmän yksilöön käyviä ja karkeita, mutteivät yksimielisiä: *muuttuisikohan myönteisyys mamuihin, jos mamut raiskaisi oman tytön?*, mutta toisaalta *persuilu perustuu vähämielisten provokaatioon työläisiä ja maahanmuuttajia vastaan*. Eräs tiivistys kulttuurieroista menee näin: *Pakolaiset polkee hintojs, suomalaisia ei palkata. Siksi ei synny lapsia. Mamuja tuetaan isoilla rahoilla. Heillä on lapsia ja äidit kotona. Mamumiehillä useita vaimoja. Suomalaiset joutuu sitten elättämään*. Työllisyyspalveluiden arvellaan toisaalta toimivan huonosti maahanmuuttajille. Kirjoittajien heterogeenisyys paljastuu jälleen: *On muuten hyvä, että maahanmuuttajatkin kirjoittelevat tänne!*

Valtionhallinnon ja valtiontalouden rooli (3) oli keskeinen $31_{3.3}$ %:ssa viestejä, eli jälleen käytännössä sama kuin periodilla 3. Sävy on kuitenkin toisenlainen, lähempänä normaalia poliittista debattia. Puolueiden kannanottoja maahanmuuttoon pohditaan aiempaa analyyttisemmin. Esimerkiksi perussuomalaisten näkemyksiä sekä kannatetaan että vastustetaan, mutta kiintoisasti sekä liian maahanmuuttovastaisena että liian vähän maahanmuuttovastaisena. EU:n turvapaikkalinjauksia kritisoidaan, mutta aivan asiallisella kielellä. Viesteistä näkyy perehtyminen työvoiman määrän roolista taloudellisen kasvun osatekijänä. Hallituksen nähdään edustavan eri intressejä: *Persuilla ja kokoomuksella iso ristiriita, kun EK tahtoo lisää ulkomaalaistyövoimaa ja persut haraavat vastaan*. Eräs kompromissinhakija kirjoittaa: *paljon ei-kantasuomalaisilla vaatimuksia, jotka eivät suomalaisten eduksi. Ensimmäisen polven maahanmuuttajille pitäisi antaa vain oleskeluoikeus ja toisellekin polvelle vain rajatut oikeudet.*

Viitteitä (4) oli käytetty argumentaation tukena $11_{\pm 2.2}$ %:ssa viestejä, eli hieman vähemmän kuin periodilla 3. Nyt viitteisiin on sattunut mukaan myös Raamattu, joka oli s-viesteillä periodilla 4 tärkeä lähde. Mediaan viitataan periodin 3 tapaan: Iltalehti, Iltasanomat, svt, yle, Helsingin Sanomat, Kansan Uutiset, abcnews, mtv Uutiset, sisäministeriö, valtiovarainministeriö ja Golman Sachs tulevat tietolähteinä myös esiin. Nyt kaikki otoksen viitteet olivat nettisivuille.

Erilaisia herjoja, heittoja ja purkauksia (5) oli otoksen m-viesteistä edelleen runsaasti, eli $34_{\pm 3.4}$

%. Näissä viesteissä perussuomalaisten kriitikkoja oli runsaasti: *persut toi suomeen 32000 mamua 2015 mamuvastaisena puolueena. nyt nämä rellestää kaduilla, kouluissa on oppiminen heikkoa heidän takiaan.* Kuten aina, puolustajiakin löytyi: *onko perussuomalaiset pitäneet maahanmuuttoa Ukrainasta negatiivisena asiana?* Eräs kirjoittaja tiivistää kantansa: *Persut ovat siis ilmastonmuutosvastaisuuden, maahanmuuttovastaisuuden, yrittäjävastaisuuden lisäksi myös kulttuurivastainen puolue.* Viitteitä Ruotsin maahanmuuttopolitiikkaan esiintyi sekä periodilla 3 että periodilla 6: *Ruotsissa on paljon mamukriittisiä, ei siis pelkästään Ruotsi demokraatit.* Erikoislaatuisia käsityksiä löytyy jälleen: *Vaikka osa islamilaisista maahanmuuttajista onkin suht koht normaaleja ihmisiä, niin joukkoon mahtuu silti niin monia ääri-ihmisiä, että tilaisuuden tullen saavat paljon ruumiita aikaan länsimaissa.*

Yhteenveto. Tehtävässä III oli tarkoituksena selvittää, miksi m-viestejä tuli erityisen paljon periodilla 3, muttei lainkaan samassa määrin periodilla 6, jolla kuitenkin oli historiallisen suuri maahanmuuttoryntäys maahamme. Vastaus on selvä. *Suomi24*-kirjoittajat eivät olleet ensi sijassa kinnostuneita nettomaahanmuutosta, vaan vuonna 2014 alkaneeseen pakolaiskriisiin liittyvistä oikeudellisista ja taloudellisista ongelmista.

Periodin 5 korona-pursketta hoiti mediassa THL:n asiantuntijajoukko, joiden kielestä tuli pian kansankieltä ja joiden mielipiteistä osattiin keskustella. Pakolaisaallon kohdalla ei ollut näin. Kun TV:n asiantuntijat vetosivat Suomen kansainvälisiin sopimuksiin, kyseessä olivat abstraktit asiat, joista ei ollut omakohtaisia kokemuksia. Ei nähty kuinka pakolaisten kotouttaminen voisi toimia. Näin siitä huolimatta, että meillä olisi pitänyt olla suhteellisen tuoreessa muistissa, miten Ruotsiin muutto vuoden 1970 tienoilla toteutui. Kirjoitettiinhan Ruotsin lehdissä meitä harmittavasti ”en Finne igen”, kun joku maanmiehemme oli syyllistynyt rötöksiin. Lopultahan ruotsinsuomalaiset ovat sopeutuneet hyvin, eikä heitä syytetä Ruotsin nykyisistä maahanmuutto-ongelmista.

Yhteistä kummankin periodin m-viesteille on, että Suomi nähdään (lähinnä muslimi-) maahanmuuton kohteena, johon valtaapitävät (EK, hallitus, Sitra ja muut *hyysärit*) ovat osaltaan myötävaikuttaneet. Maahanmuuton hyödyllisyys maan taloudelle myönnetään, mutta harvat viestit ottavat tämän puolen puheeksi. Pakolaisuuteen työntävät voimat eivät ole keskusteluissa läsnä. Vaikka periodilla 6 oli jo tiedossa, että periodilla 3 maahan tulleet olivat pääosin lähteneet maasta, tämä ei näkynyt viesteissä. Ukrainalaisten tilapäistä suojelua ei viesteissä myöskään juuri käsitellä.

Huippukuukauden viestit. Eniten m-viestejä tuli kuukaudella 177 eli syyskuussa 2015. Viestejä tuli 17 130 kpl. Määrä on runsaat puolet k-viestien huippuarvosta, mutta 30-kertainen verrattuna s-viestien huippuun. Maahanmuuttoviraston mukaan syyskuussa Suomeen tuli 10 837 (koko vuonna 32 476) turvapaikanhakijaa, näistä irakilaisia 9 101 (koko vuonna 20 485). Tulijoiden lukumäärät eivät näytä olleen kirjoittajien tiedossa, ei liioin irakilaisten suuri osuus. Eräs järkeili kuitenkin, että *Suomeen tulee 300 pakolaista päivässä, 10 000 kuukaudessa ja 40 000 vuoden loppuun mennessä*, ja saa suuruusluokan oikein. Viestien kieli on tavallista kärjekkäämpää, mutta monet olisivat varmaan samaa mieltä seuraavan kanssa: *Pikakäsittely ja -käännytys ainoa keino! Irakilaiset ovat kuulleet, että Suomessa myönnetään helpommin kansalaisuus kuin Ruotsissa. Kyse on turvapaikkashoppaamisesta, ei mistään hätäpakolaisista.*

Kaiken kaikkiaan useat pitävät hallitusta syypäänä maahanmuuttajien määrän kasvuun, eikä tulijoita enimmäkseen pidetä oikeina turvapaikanhakijoina.

Liitteen 2 Gemini-yhteenvedot asettuvat syntyvyyttä ja kuolecvuutta koskevien yhteenvetojen väliin suhteessa siihen, missä määrin niissä korostuu tutkimuskohteeksemme valitsemamme väestönmuutosprosessi. Jälleen, siltä osin kuin kyse on muuttoliikkeestä, johtopäätökset ovat samansuuntaisia kuin omamme. Metodisessa mielessä on syytä huomata, että koska Geminiä ei erityisesti ohjeistettu ottamaan väestöntutkimuksellista näkökulmaa, se käsittelee kaikkia

viestejä samanarvoisina. Niinpä yhteenvedon kohdassa Gemini kuvaa pitkähköä viestiä, jossa käsitellään Israelin historiaa, mutta joka ei kuulu tutkimuskohteemme piiriin lainkaan.

6.4 Ovatko keskustelut *kokonaishedelmällisyydestä* tai *elinajanodotteesta* erilaisia?

Teknisten termien käyttö voi johtua kirjoittajan perehtyneisyydestä väestötieteen käsitteisiin ja ajatustapohiin, erotukseksi muista *Suomi24*-keskustelijoista. Selvitimme, näkyykö tämä s- ja k-viesteissä, joissa nämä lemmat esiintyvät. Vastaus on myönteinen.

Kokonaishedelmällisyys esiintyy vain 57 viestissä. Luimme ne kaikki. Viestien sävy poikkesi selvästi kohdassa 6.1 käsitellyistä s-viesteistä. Herjoja ei ollut lainkaan. Lähteitä käytettiin enemmän ja viestien sisällöissä korostui syntyvyyttä koskevien tilastolukujen erittely. Yhteys väestön huoltosuhteeseen tuodaan esiin. Alan tutkijoita siteerattiin muutamassa viestissä. Yksi kirjoittaa: *Tilastokeskuksen väestöennuste on vetänyt monen poliitikon nenän kalpeaksi.* Toinen säestää: *Suomessa on ennen kokematon vauva-ahdistus, suomen kansan kohtalon ajat käsillä.* Kolmas jatkaa: *tolkun ihminen ymmärtää, että maassa tarvitaan lisää työntekijöitä.*

Elinajanodote esiintyi 1 678 viestissä. Luimme niistä 196 kpl. Tähän joukkoon mahtui jo muutama herjakin, mutta ne olivat poikkeuksia, ja kirjoitusten sävy oli pääpiirteittäin asiallinen. Eräs huoli oli, että vaikka eliniän nähdään olevan kasvussa, se voi merkitä lisää sairausvuosia. Eräät epäilevät eliniän kääntyneen koronan myötä ensi kertaa laskuun. Eläkkeiden elinaikakerrointa käsitellään myös. Samoin eksoottisia taudinaiheuttajia.

Kummassakin aineistossa tehtiin sekä kansainvälisiä vertailuja että vertailuja historiallisten aikaperiodien välillä. Suomen ohella myös Venäjän demografia tuntui erityisesti kiinnostavan kirjoittajia. Putinin sotapolitiikan puolustajia ei aineistossa ollut.

Lintuinfluenssa 2005. Viestejä tarkasteltaessa kävi ilmi, että edellisiin tarkasteluihin liittyen *Suomi24*:n osallistujien joukko on hyvin todennäköisesti rakenteeltaan muuttunut. Tarjoamme esimerkiksi tästä vuoden 2005 lintuinfluenssaa koskevan keskustelun. Keskustelu sijoittui kuukausille 58–64, eli marraskuusta 2004 maaliskuuhun 2005. Keskustelun tekee erityislaatuiseksi se, että se näyttää lähes kauttaaltaan asiantuntevien henkilöiden kirjoittamaa. On mahdollista, että tämän tyyppinen kirjoittajakunta on edelleen keskustelupalstalla mukana, mutta sen kontribuutio ei aineistossamme näy.

7 Johtopäätökset

7.1 Väestönmuutokset

On kiinnostavaa, että syntyvyyden tutkijoiden näkemyksiä ei kirjoituksissa juuri siteerata. Esimerkiksi *väestöliitto* ja *perhebarometri* eivät esiinny otoksissamme kertaakaan. Sanastossamme, jossa on 7 957 lemmaa, *väestöliitto* on sentään 1 008:ksi yleisin (järjestysluku 6950) ja esiintyy siten (viestin sisäiset toistot mukaan lukien) noin 6 880 kertaa, ks. luku 4.1.2. Esimerkiksi *perhebarometri* ei päässyt mukaan sanastoomme.

Vertailun vuoksi lemma *tilastokeskus* on 1 037:nneksi yleisin. *Eläketurvakeskus* tai *ETK* eivät ole mukana sanastossa vaikkakin *ETK* on mukana joissakin viesteissä, ja näin voisi periaatteessa *perhebarometrikin* olla. Esimerkiksi *ETLA, VATT, PTT, TTL* ja *EVA* eivät myöskään esiinny sanastossamme. Näistä poiketen *KELA* on sanastossamme sijalla 502. Korona-keskusteluissa

siteeratut tutkijat ovat olleet esillä lähinnä viran puolesta. Muuttoliiketutkijoita ei osunut aineistoihimme.

Sosiaalinen media, eli some, on ollut lisääntyvässä määrin mukana lähes kaikissa yhteiskunnallisissa prosesseissa havaintojaksollamme 2001-2023. Vain muutamassa viestissä viitataan mahdollisuuteen, että somen käyttö voisi olla keskeinen syy hedelmällisyyden romahdukseen. Lemma *somessa* onkin vasta 4 598:s sanastossamme. (Selvästikään lauseenjäsennin ei ole oikein tunnistanut sanan perusmuotoa.)

Ajatus somen keskeisestä merkityksestä lieneekin epäuskottava. Kun perheen perustaminen ja lasten hankinta on pitkäkestoista ja kallista investoinnin luonteista toimintaa, niin somen käytölle on tyypillistä hetkessä eläminen ja halpa hinta. Sosiologisessa ajattelussa kohorttikokemuksilla on ajateltu olevan tärkeä rooli asennemuutoksissa (Ryder, 1965). Tätä näkemystä perustellaan yksityiskohtaisemmin kirjoituksessa Alho (2022). Aivan eri tason asioista näyttäisi olevan kysymys, vaikka some on tarjonnut uusia kanavia asenteisiin vaikuttamiseen.

7.1.1 Syntyvyyssokkien syyt ja seuraukset

Olisi odottanut, että ne muutamat väestöpolitiikan elementit, jotka maamme yhteiskuntapolitiikassa vielä ovat, ts. lapsilisät, päivähoito, kotihoidon tuki, lapsiperheiden asumistuki ja muihin tukiin liittyvät lapsikorotukset ,olisivat kiinnostaneet *Suomi24*:n kirjoittajia. Eivät kiinnostaneet.

Kun tuet mainitaan, niiden tehoa epäillään. Olisiko tähän osasyynä myös se, että kirjoittajat ovat (arviomme mukaan) pääasiassa miehiä? Tai se, että ajatukseen syntyvyyteen vaikuttamiseen valtionhallinnon keinoin suhtaudutaan skeptisesti. Kuten edeltä käy ilmi viesteistä löytyy juonne, jossa valtiovaltaa pidetään suorastaan syyllisenä matalaan hedelmällisyyteen.

Myöskään eräiden tutkijoiden esittämä ajatus siitä, että Suomen hedelmällisyyden lasku vuoden 2010 jälkeen johtuisi lasten hankinnan lykkäämisestä (ns. *postponement*), ei ole mukana yhdessäkään otostemme viesteistä. Tämä arvio osoittautuikin jälkikäteen tarkasteltuna virheelliseksi. Lapsia näyttää tulevan entistä vähemmän.

Syntyvyyden romahduksen syynä nähdään *Suomi24*:ssä asenteiden muutos, jota selittää mm. naisten työssäolo, feminismi ja perheinstituution heikentyminen, mistä aiheutuu kantaväestön väheneminen ja korvautuminen maahanmuuttajilla.

7.1.2 Kuolleisuussokkien syyt ja seuraukset

Kuolevuutta väestötasolla määrittävät lähinnä krooniset sairaudet ja niiden vaatima hoito. Nämä aiheet eivät saaneet paljoakaan huomiota keskustelupalstalla. Ei liioin elinajanodotteen systemaattinen kasvu tai se, että kuolevuuden aleneminen johtaa väestön ikärakenteen vanhenemiseen.

Kaiken kaikkiaan pääosa k-viesteistä liittyy korona-epidemian tuottamiin ongelmiin. Kuolleisuus tulee puheeksi enemmänkin sivuasiana, silloin kun se mainitaan. Jotkut tosin huomasivat, että koronan vaikutus kokonaiskuolevuuteen oli pieni vuosina 2020-2021. Koronakeskusteluun osallistuneita asiantuntijoita kritisoitiin melko vähän. Sen sijaan ravinnon ja elämäntapojen terveysvaikutuksia käsittelevissa viesteissä kriittiset äänenpainot dominoivat.

Yllättävää on, että koronan torjunnan varsin huomattavat kustannukset mainitaan vain muutamissa viesteissä. Yksilön vastuu esimerkiksi maskien käytössä ja julkisissa tiloissa käyttäytymisessä on tuntunut tärkeämmältä.

Koronarokotukset ja niiden sivuvaikutukset olivat suuren kiinnostuksen kohteena. Vaikka monet puolustavat rokotteiden käyttöä ja tehoa, suurempi osa viesteistä on skeptisiä. Rokotettujen ja rokottamattomien koronakuolemien lukumääriä vertaillaan ilman tietoa riskipopulaatioiden koosta. Eräissä viesteissä epäillään rokotteiden vaikuttavan jopa hedelmällisyyteen.

7.1.3 Muuttoliikesokkien syyt ja seuraukset

Maastamuutto näyttää olevan sekä *Suomi24*:ssä että mediassa yleensäkin katveessa. Tämä on suomalaistaustaisten osalta ymmärrettävää, sillä sen vaihtelut ovat olleet vähäisiä (ks. esim. Alho, 2026, kuvio 17).

Nyt kun maahanmuuton taso on merkittävästi korkeampi kuin vielä 2000-luvun alussa, luulisi kotoutumisen olevan kiinnostava aihe. Niissä viesteissä, joissa aihe tulee esiin, kirjoittajat suhtautuvat skeptisesti kotoutumisen onnistumiseen. Syinä mainitaan sää, kieli ja kulttuuri. Myrskylä ja Kainulainen (2025) esittävät tiiviin analyysin maahanmuuton roolista suomalaisessa hyvinvointivaltiossa. Keskustelulle kotoutuksesta olisi runsaasti aihetta!

Kuten Leppäkorpi (2025) osoittaa hallituksen esitysten perusteella, maamme maahanmuuttopolitiikassa on pyritty antamaan ulospäin kuva maasta vastaanottavana muttei liian houkuttavana. *Suomi24*:n kirjoittajille tämä ei ole tullut mieleen. Edelleen Helsingin Sanomien kuukausiliitteessä julkaistiin 02.03.2019 Migrin selvitys vuoden 2015 turvapaikanhakijoista. Tulijoita oli noin 32 000. Neljän vuoden kuluttua heistä 14 000 oli maassa luvallisesti, 12 500 oli poistunut maasta ja noin 4 000:n olinpaikkaa ei tiedetty. Osalla päätösprosessi oli vielä kesken. Kuten kuvasta 4 käy ilmi, *de jure* väestöön pakolaisaallon lopullinen vaikutus näyttää pienemmältä kuin 14 000. Virallisen tilaston mukaan vuosina 2015-2021 nettomuutto Irakista Suomeen oli 10 800.

Suomi24:n kirjoittajien mielissä näytti painottuvan käsitys, että Suomeen houkutellaan avokätisin tuin muuttajia, jotka eivät saa työtä. Toisin kuin syntyvyyden kohdalla m-viesteissä kotouttamisraha, asumistuki ym. sosiaalituet ovat tärkeä purnauksen aihe.

Viesteihin liittyy erityinen rasismiin vivahtava epäilys muslimien maahanmuuttoa kohtaan. Jyrkkyydessään ilmiö on yllättävä, sillä vastaavaa ei esiinny esimerkiksi Intiasta tai Vietnamista tulevia buddhalaisia kohtaan.

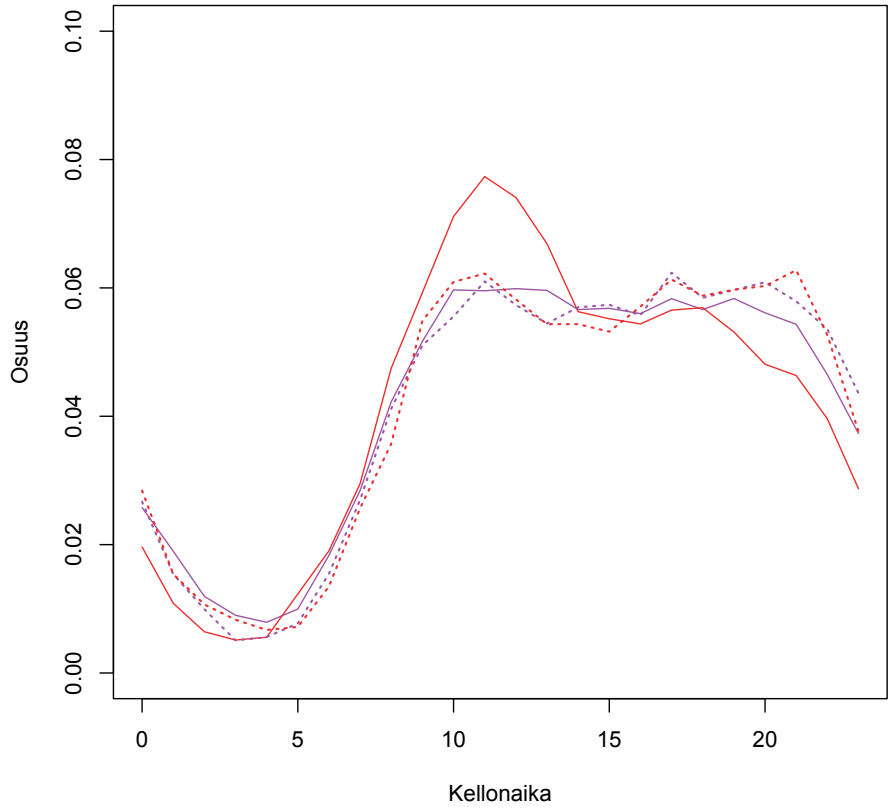
Työhön johtavaan maahanmuuttoon suhtaudutaan enimmäkseen myönteisesti, mutta esimerkiksi yksilön roolia kotouttamisessa ei tunnisteta. Tämä poikkeaa koronaan suhtautumisesta.

Muuttoliikkeen vaikutus väestön etniseen rakenteeseen tunnistetaan selvästi, vaikka syntyvyyden ja kuolleisuuden vaikutusta ikärakenteeseen ei näitä aiheita koskevissa viesteissä tunnisteta. Ukrainasta tullutta pakolaisvirtaa käsitellään melko vähän, vaikka se jättää luultavasti isomman jäljen väestöhistoriaamme kuin Lähi-idästä tullut pakolaisvirta.

7.2 Metodologian käyttökelpoisuus

7.2.1 Ketkä kirjoittavat?

Koska kirjoittajat eivät yleensä esiinny omalla nimellään, on vaikea arvioida, mikä kirjoittajien sukupuolijakauma on. Esimerkiksi periodien 4 ja 6 s-viestien otoksiin sattui vain yksi viesti, jossa näyttäisi olevan naisen näkökulma. Samaan tapaan vain yhdessä viestissä kirjoittaja näyttäisi olevan nuori aikuinen. Joissain viesteissä puhutaan eläkeläisten asioista, mutta ei selviä, onko kyse kirjoittajan omista kokemuksista. Arvelemme, että kirjoittajajoukko koostuu



Kuva 12: Viestien lähetyksen jakauma kellonajanmukaan k-viesteille periodilla 6 (jatkuva, aniliini) ja kuukaudella 231 (pisteviiva, aniliini) ja m-viesteille periodilla 6 (jatkuva, punainen) ja kuukaudella 177 (pisteviiva, punainen).

lähinnä työikäisistä miehistä. Eräistä k-viesteistä käy ilmi, että ne ovat osa naisten keskustelua. m-viesteistä käy ilmi, että myös sekä naiset että maahanmuuttajat osallistuvat, tosin vähemmistönä.

Vasemmistolaisten näkemysten puolustajia on selvä vähemmistö, mutta ilmastonmuutoksen torjuntaa pitävät monet tärkeänä ja *perskeko*-päättäjät saavat kritiikistä osansa. Vaikutelmaksi jää, että iso osa *Suomi24*:ään kirjoittavista koostuu *valtamediaan* kriittisesti suhtatuvista tyoikäisistä miehistä. Osa heistä näyttäisi olevan valmiita näkemään vaivaa tietojen kaivamiseen epätavallisistakin tietolähteistä. Kirjoittajat eivät useinkaan näytä olevan mukana järjestötoiminnassa. Sen sijaan *Suomi24* näyttää olevan ”heidän” kanavansa. Viesteissä puhutaan esimerkiksi palstan *älyköistä*, jotka ovat ansainneet ryhmän arvostuksen.

Esimerkkinä epäsuorista keinoista ymmärtää kirjoittajakuntaa kuvassa 12 on verrattu k- ja m-viestien lähettämisen kellonaikoja periodilla 6 (jatkuvat käyrät) kellonaikoihin kuukausilla 231 ja 177 (pisteviivat), jolloin viestivirroissa havaittiin maksimiarvot. Nähdään, että purskeaikoina myöhäisiltoina ja yöllä lähetetttyjen viestien osuus on selvästi suurempi kuin keskipäivällä, jolloin kirjoittelu on muutoin tiheintä. Ovatkohan ärsykkeet saaneet tunneperäisesti suuntautuneet ihmiset liikkeelle? On hämmästyttävää, miten lähellä koronaan ja turvapaikanhakijoihin pääsiassa liittyvät pisteviivat ovat toisistaan.

Vastaavia eroja kirjoittajien käyttäytymisessä on havaittu mm. emotikonien (emojien) käytössä eri aihealueilla (Alho, 2020; Kruse, 2026).

7.2.2 Täydentävä vaihtoehto rekisteri- ja kyselytutkimuksille

Väestösokit ovat keskeinen yhteiskunnalliseen päätöksentekoon vaikuttava ilmiö, joita koskevia näkemyksiä ja asenteita Suomen tietorekisterit eivät systemaattisesti kata. Näkemykset ilmastonmuutoksesta tai nuorison pahoinvoinnista ovat esimerkkejä, joita koskevia mielipiteitä tunnetaan samalla tavalla huonosti.

Kyselytutkimukset tarjoavat tietolähteen, mutta perinteiset kyselytutkimukset ovat hitaita suorittaa ja kalliita. Niiden vahvana puolena on todennäkösyysteoria, jolle niiden tulosten edustavuus voidaan perustaa (ks. esim. Cochran, 1977).

Suomi24:n kirjoittajakunta on itsevalikoitunut, eivätkä esitetyt mielipiteet edusta sellaisenaan minkään helposti identifioitavan väestöryhmän näkemyksiä. Sosiaalisesta mediasta on kuitenkin 2000-luvulla tullut instituutio, joka synnyttää *sosiaalisia faktoja*. Durkheim (1937) viittaa käsitteellä yhteisön arvoihin, normeihin tai instituutioihin, joilla on pakottavaa voimaa yksilöön nähden. Bourdieu (1984) laventaa käsitteen kattamaan yhteisön alaryhmien makuja ja käyttäytymistapoja. Becker (1996) päätyy samantapaisiin pohdintoihin mikrotaloustieteen näkökulmasta tarkastelemalla korreloituneita valintoja. *Suomi24*:n kirjoittajien mielipiteet voivat vaikuttaa niidenkin kansalaisten mielipiteisiin, jotka eivät suoraan osallistu keskusteluihin, mutta jotka joutuvat ottamaan huomioon, vaikkapa omissa arkipäivän keskusteluissaan, esitetyt mielipiteet.

Lisäksi on huomattava, että perinteisten kyselytutkimustenkin edustavuusominaisuudet pätevät vain silloin kun vastauskatoa ei ole. On kehitetty monia menetelmiä (mm. kalibrointi tunnettuihin marginaaleihin, vastaustodennäköisyyksien mallinnus ja monipaikkaus; ks. esim. Little ja Rubin (1987); Särndal, Swensson ja Wretman (1992)) vastauskadon vaikutuksen pienentämiseksi, mutta niillä on omat oletuksensa ja heikkoutensa erityisesti silloin kun vastaajien osuus jää pieneksi.

Nettipaneelit ovat lupaava tapa nopeaan ja halpaan tiedonkeruuseen, mutta niidenkin edustavuuteen liittyy epäilyksiä itsevalikoitumisen takia. Tästä aihepiiristä enemmän lähteissä Bethlehem (2010), Baker et al. (2013) ja Japec et al. (2015). Niinpä tämänkin tietolähteen oheen on syytä pohtia vaihtoehtoja.

On myös syytä huomata, että kyselytutkimuksissa vastaajan identiteetti tulee tavallisesti tutkimuksen tekijöiden tietoon. Tämän tiedetään johtavan sellaisten mielipiteiden sensurointiin, joita vastaaja arvelee paheksuttavan. Tästä ovat esimerkkinä esimerkiksi maahanmuuttoon liittyvät rasistiset asenteet (ks. esim. N.R.C., 2004). *Suomi24*-aineistossa tällaista harhaa ei näyttäisi esiintyvän ainakaan siinä mielessä, ettei rasistisävyisiä mielipitä esiintyisi.

Tutkimusmetodisesti *Suomi24*:n tai muun sosiaalisen median tuottamien tekstimuotoisten aineistojen analyysi muistuttaa kyselytutkimusten avovastausten analyysiä. Keskeisin ero on, että kyselyssä vastaajaa ohjataan spesifiin aihepiiriin. Sosiaalisen median aineistot syntyvät tutkijasta riippumattomista syistä, ja tutkijan ensimmäinen tehtävä on löytää tekstivirrasta osajoukko, jossa kiinnostuksen kohteena olevat ilmiöt voivat tulla puheeksi.

Väestösokkeja koskevassa tutkimuksessamme etsittiin ensinnä sanat, joiden tiedettiin liittyvän väestönmuutoksiin. Keskusteluiden aihepiiriluokitusta käytettiin korpuksen rajaamiseen edelleen. Lukemissamme viesteissä oli silti muutamia, jotka eivät kuuluneet mihinkään luokkaan. Näille oli tyypillistä, etteivät ne tosiasiassa käsitelleet tämän väestötutkimuksen aihepiiriä lainkaan. Onneksi näitä näytti olevan vain vähän.

Tärkeämpi havainto oli, että usein viestien tosiasiallinen aihe näytti olevan muu kuin jokin kolmesta väestönmuutosprosessista. Viestit käsittelivät silloin myös po. muutosprosesseja, mutta ”oheistuotteena”. Tästä on esimerkkinä korona-epidemia, jolla oli välttämätön yhteys kuolleisuuteen, vaikka onkin käynyt ilmi, ettei sen enempää korona kuin kirjoittajia erityisesti kiinnostaneet rokotteiden haittavaikutukset ole olleet kuolinsyinä kovin tärkeitä. Toinen esimerkki oli Euroopan pakolaiskriisi, jolla oli välttämätön yhteys muuttoliikkeeseen, vaikkei se lopulta juurikaan vaikuttanut Suomen maahanmuuttotilastoihin. Näitä piirteitä emme osanneet etukäteen arvata, ja tämä onkin tutkimuksen keskeinen metodologinen löydös.

Metodisesti käsillä olevan tutkimuksen kontribuutio on järjestyslukumuunnoksen käyttö lemmojen frekvenssien sijaan. Se näytti eliminoivan automaattisesti stop-sanat. Sanastoon kuulumiselta edellytettiin kuitenkin tiettyä absoluuttista frekvenssiä, jotta keinotekoisia järjestyslukujen erotuksia ei syntyisi sanaston harvinaisimpien lemmojen välillä. Liitteestä 1 käy ilmi, että jotkin varsin harvinaiset lemmat tulivat mukaan sanapilviin. Tästä ovat esimerkkeinä *Dublin* sanapilvessä 7 ja *King* kuvan 10 oikeassa sanapilvessä. Nämä esimerkit viittaavat siihen, että jonkin verran pienempi sanasto olisi kenties johtanut informatiivisempaan kuvaan kielen muutoksista. Tiedossamme ei ole, että järjestyslukumuunnokseen perustuvaa upotusta olisi aiemmin käytetty. Tilastotieteessä järjestyslukumuunnosta on sen sijaan käytetty ainakin vuosisata, ks. esim. Gibbons (1971).

Tässä tutkimuksessa käytimme graafista esitystä (sanapilvet ja upotukseen perustuvat kuviot) tarjoamaan yksinkertaistettuja näkökulmia kieleen ja sen muutokseen. Tekstien lukua tilastolliset analyysit eivät voi kokonaan korvata, mutta toisaalta viestien lukeminen osoittautui työlääksi. Niinpä arvelemme, että molemmilla lähestymistavoilla on jatkossakin roolinsa.

Suomi24:n ohella on monia alustoja, joissa syntyvällä tekstiaineistolla on vastaava institutionaalinen sosiaalisen faktan luonne. Kaikissa näissä vastaava tilastollisen analyysin ja tekstien luennan yhdistelmä voi olla hyödyllinen. Kun luimme runsaat 2 000 viestiä, niin sekin on vain noin 0.16 % korpuksesta, joka puolestaan oli noin 1.1 % koko *Suomi24*-aineistosta. Aivan kaikkea ei ole tarpeen läpikäydä!

Tekoäly voinee tulevaisuudessa toimia luentaosuuden nopeuttajana. Omat kokemuksemme Geministä ovat rohkaisevia. Se on nopea ja halpa. Tuotetun yhteenvedon laatuun vaikuttaa kuitenkin se, että kun sama yhteenvetopyyntö toistetaan, niin tulokset muttuvat. Tämä näyttää johtuvan siitä, että Geminin tulokset riippuvat käytettyjen algoritmien satunnaisista alkuarvoista. Havaitsimme myös, että Geminin antamien vastausten takana on jonkinlainen ”maailmankuva”, jonka pohjalta se rakentaa yhteenvedot. Lisätutkimukset ovat tarpeen maailmankuvan paremmaksi ymmärtämiseksi. Aivan kuten kaksi lukijaa voi poimia annetusta tekstistä eri puolia ja kertoa siitä kolmannelle hyvinkin erilaisella tavalla, niin näyttää tekevän tekoälykin.

Ensi kertaa keskustelupalstan viestejä lukevalle mieleen jäävät alatyyliset ja röyhkeät viestit. Kirjoittajajoukko on kuitenkin heterogeeninen. Varsin monet, kenties 1/5, ovat selvästi omanarvontuntoisia. Syystäkin, sillä pidetäänhän eräitä heistä palstansa *älykköinä*.

Viestit ovat myös erilaisia muodoltaan, eikä niitä voi millään yksinkertaisella tavalla sovittaa Aristoteleen retoriikan kolmijakoon. On kuitenkin selvää, että ainakin *älyköt* ovat ymmärtäneet *logoksen* merkityksen. Runsas puolet kirjoittajista vaikuttaa ymmärtäneen *ethoksen* merkityksen, sillä omien asenteiden ja käsitysten painotus on tavallista.

Lopulta jäljelle jää kenties 1/5 kirjoittajista, joiden viesteissä (usein pahantahtoinen) *pathos* on vahvasti esillä, mutta ymmärrettävää sisältöä on vähemmän. Useista viesteistä käy ilmi, ettei tämä herjoja heittävä joukko ole kaikkien arvostama. Niinpä eräs neuvoi: *Jos haluaa käännyttää toisen puolelleen, ei kannata haukkua häntä vihersuvakkimädättäjäksi.* Aristoteles olisi epäilemättä ollut samaa mieltä.[6]

[6]Tämän raportin viimeistelyvaiheessa tietoomme tuli Eronen-Valli ja Lehti (2026). Luvussa 6 kuvataan so-

Tekoälyn viimeaikainen kehitys näyttää johtavan siihen, että enenevä osa netistä, ml. keskustelupalstat, on osittain tai kokonaan tekoälyn tuottamaa. Tekoälythän on alunperin opetettu sellaisilla tietolähteillä kuten *Suomi24*! Ongelma on tiedostettu (ks. esim. Kirchebauer et al., 2023), mutta tätä kirjoitettaessa on epäselvää, miten tämä tulee vaikuttamaan keskustelupalstojen kieleen ja sen analyysiin.

siaalista mediaa Aristoteleen retoriikan näkökulmasta. Kirjassa käytetään osallistujista hauskaa termiä *reettori*, joka olisi istunut hyvin omaan tekstiimme.

Viitteet

Afifi, A.A. & Azen, S.P. (1979). *Statistical analysis, Second edition.* Academic Press, New York.

Alho, J. (2020). *What authors reveal of themselves in internet discussions?* Paper presented at Joint Statistical Meetings, August 2020. Philadelphia U.S.A.

Alho, J. (2022). Syntyvyys on hankala mallinnettava. *Suomen tilastoseuran vuosikirja 2019-2022: 8-24.*

Alho, J. (2026). DemoGrafiikka - Väestökuvasto. *ETLA Working Papers nro 138.*

Aristoteles (1997). *Teokset IX. Retoriikka.* Gaudeamus, Helsinki.

Austin, J.L. (1962). *How to do things with words.* Clarendon Press.

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J. & Tourangeau, R. (2013). *Report of the AAPOR Task Force on Non-Probability Sampling.*

Becker, G.S. (1996). *Accounting for tastes.* Harvard University Press, Cambridge MA.

Betlehem, J. (2010). Selection bias in web surveys. *International Statistical Review*, 78 161-188.

Bourdieu, P. (1984). *Distinction.* Harvard University Press, Cambridge MA.

Cochran, W.G. (1977). *Sampling techniques, 3rd ed.* Wiley, New York.

Durkheim, É. (1937). Les régles de la méthode sociologique. Presses Universitaires de France, Paris.

Gibbons, J.D. (1971). Nonparametric statistical inference. McGraw-Hill, New York.

Engle, R.F. & Granger, C.W.J. (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica* 55, 251–276.

Eronen-Valli, M. & Lehti, L. (2026). *Digitaalinen retoriikka.* Vastapaino, Tampere.

Fraxanet, E., Vicenç Gómez, V., Kaltenbrunner, A. & Pellert, M. (2026). A decade of news forum interactions: threaded conversations, signed votes, and topical tags. *Nature Scientific Data* 13:1193.

Greenacre, M.J. (1984). *Theory and applications of correspondence analysis.* Academic Press. London.

Greenacre, M.J. (1993). Biplots in correspondence analysis. *Journal of Applied Statistics*, 20, 251-269.

Hawkes, A.G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58, 83-90.

Hudson, R.J. (1980). *Sociolinguistics.* Cambridge University Press.

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, JK., O'Neil, C. ja Usher, A. (2015). Big data in survey research: AAPOR Task Force report. *Public Opinion Quarterly*, 79, 839-880.

Kirchenbauer, J., Geiping, J., Yuxin Wen, Katz, J., Miers, I. ja Goldstein, T. (2023). A watermark for large language models. *Proceedings of the 40th International Conference on Machine Learning, PMLR* 202:17061-17084

Kruse, M. (2026). Does this emoji make me look old? *The New York Times* 05.09.2026.

Lazer, D., Kennedy, R., King, G. & Vespignani, A. (2014). The parable of google flue: traps in big data analysis. *Science* 343, 1203-1205.

Lénoir, R. (1995). L'Invention de la démographie at la formation de l'état. *Actes de la Recherche en Sciences Sociales*, 108:, 36-61.

Leppäkorpi, M. (2025). Jotta Suomi ei olisi muita houkuttelevampi – Vetävät, pitävät ja työntävät tekijät hallituksen esityksissä ulkomaalaislaiksi vuosina 1990–2020. *Sosiologia* 2025:4.

Little, R.J.A. & Rubin D.B. (1987). *Statistical analysis with missing data*, Wiley, New York.

Mohler, G.O., Short, M.B., Brantingham, B.J., Schoenberg, F.B. & Tita, G.E. (2011). Self-Exciting point process modeling of crime. *Journal of the American Statistical Association*, 106, 100-108.

Myrskylä, P. & Kainulainen, S. (2025). Maahanmuuttaneet Suomessa. Hyvinvointivaltion uhka vai ainoa mahdollisuus? *DIAK Puheenvuoro* 25.

N.R.C.(2004). *Measuring racial disscrimination*. National Academies Press.

Rao, C.R. (1973). *Linear statistical inference and its applications, 2nd ed.* Wiley, New York.

Ryder, N.B. (1965). The cohort as a concept in the study of social change. *American Sociological Review* 30, s. 843-861.

Särndal, C.-E., Swensson, B., Wretman, S. (1992). *Model assisted survey sampling*. Springer, New York.

Tukey, J. W. (1977). *Exploratory data analysis*, Addison-Wesley, Reading, MA.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control* 8: 338–353.

Zhuang, J., Ogata, Y. & Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrence. *Journal of the American Statistical Association*, 97, 369-380.

Liite 1. Yllättävien lemmojen taustaa

Abraham (järjestysluku 666)[7]. Nimen takana voidaan erottaa viisi henkilöä. Näistä tärkein on Abraham Lincoln, joka tunnetusti sanoi (osapuilleen), että *joitakin voi huijata kaiken aikaa, mutta kaikkia ei voi huijata kaiken.*

Ankdam (534). Nimimerkki, joka osallistuu mm. maahanmuuttajien suomen- ja ruotsinkielen opetukseen liittyviin keskusteluihin.

Baabelin (379). Viittaa paheellisena pidettyyn Baabelin (Babylonin) kaupunkiin (*Baabelintorni*, *portto*).

Baal (1932). Vanhassa testamentissa esiintyvä kaananilaisten jumala.

Christian (224) Viittaa useimmiten tanskalaiseen jalkapalloilijaan, joka tuupertui maaottelussa Suomea vastaan.

Dublin (38). Viittaa yleisimmin koulujen sulkemiseen Dublinin kaupungissa koronan takia.

Finnleaks (2108). Nettisivusto, jolla näyttää olevan ajatuksellinen yhteys *MV − lehteen*. Julkaisee mm. listoje maahanmuuttajien rikoksista.

Heikkilä (3206). k-viestejä erottelevien lemmojen sanapilvessä 7 esiintyvä *Heikkilä* viittaa useimmiten Antti Heikkilä -nimiseen lääkäriin, jonka kannaotot koskien ruokavalion ja diabeteksen yhteyttä sekä sikainfluenssarokotuksia ovat kiinnostaneet keskustelijoita. Muitakin *Heikkilä*-nimisiä henkilöitä esiintyy k-viesteissä esimerkiksi kroonisia tauteja koskevissa keskusteluissa.

King (12). Useita viittauksen kohteita: King James -Raamattu, Martin Luther King, Burger King. Kirjoituskielenä usein englanti.

Klaus (1975). Viittaa usein ns. Davosin kokousten perustajaan K. Schwab.

Köln (275). Viittaa Kölnissä uutenavuotena 2015/2016 tapahtuneeseen joukkoahdisteluun Kölnin päärautatieasemalla.

Moktar.Gubta (392) tai *Moktar* (5264), *Gubta* (128). Nimimerkki, joka on osallistunut aktiivisesti maahanmuuttoa ja perussuomalaisia koskeviin keskusteluihin. Hänen asenteensa maahanmuuttoon näyttää positiiviseltä ja hän on kriittinen perussuomalaisia kohtaan.

MV-lehti (3881). Nettilehti, jossa mm. erilaisia salaliittoteorioita.

Soldiers. (1217) Englanninkielisissä viesteissä koskien espanjantautia sotilaskuolemat. Myös *of Odin* ja katupartiot.

Liite 2. Gemini-yhteenvedot tehtäville I, II ja III

I. Syntyvyys, periodi 4.

Syntyvyyden lasku Suomessa ja sen syyt

Suomen syntyvyys on laskenut voimakkaasti, ja tilannetta verrataan useissa keskusteluissa jopa 1860-luvun suuriin nälkävuosiin. Kokonaishedelmällisyysluku on pudonnut historiallisen alas 1,43:een, mikä tarkoittaa syntyvyyden olevan noin 32 % väestön uusiutumistason alapuolella. Keskusteluissa esitetään monia rinnakkaisia ja usein ristiriitaisia syitä tälle kehitykselle:
* **Talous ja politiikka**: Taloudellinen epävarmuus, pätkätyöt, työttömyys sekä lapsiperheisiin kohdistuvat leikkaukset (kuten lapsilisien ja koulutusmahdollisuuksien heikentäminen) vähentävät halukkuutta perheen perustamiseen. Myös hyvinvointivaltion koetaan poistaneen ihmisen luonnollisen tarpeen hankkia lapsia oman vanhuuden turvaksi, kun taas korkea verotus

[7] Täten lemma on 7957 - 665 = 7292:nen frekvenssiltään koko aineistossa.

rasittaa työssäkäyviä. * **Yhteiskunnalliset ja kulttuuriset muutokset**: Naisten kouluttautuminen ja uraan keskittyminen vähentävät ja lykkäävät lapsenhankintaa. Osa kirjoittajista syyttää myös digitalisoitumista; älypuhelimet, sosiaalinen media ja videopelit vievät aikaa siten, etteivät pariskunnat enää kohtaa toisiaan. Lisäksi mainitaan "MeToo-kampanja, jonka väitetään tekevän naisten lähestymisestä liian vaikeaa tai pelottavaa miehille, sekä individualistiset ja "itsekkäät"elämänasenteet. * **Biologiset tekijät ja stressi**: Miesten siittiöiden määrän laskun mainitaan uhkaavan väestön säilymistä. Toisaalta yhdysvaltalaistutkimuksen mukaan raskaus ja äitiyteen liittyvä stressi voivat solutasolla lyhentää kromosomien telomeereja, mikä vastaa solujen ennenaikaista ikääntymistä.

Maahanmuutto ja huoltosuhde

Kysymys maahanmuutosta herättää paljon keskustelua syntyvyyden laskun yhteydessä: * **Työperäinen maahanmuutto** ja maahanmuuton lisääminen yleisesti nähdään toisten mielestä välttämättömänä ratkaisuna työvoimapulaan ja eläkemaksujen nousupaineen hillitsemiseen. * Kriitikot puolestaan kantavat huolta siitä, että maahanmuuttajien (kuten muslimien) korkeampi syntyvyys johtaa kantaväestön jäämiseen vähemmistöksi omassa maassaan ("väestönvaihto"). He myös huomauttavat, että turvapaikanhakijoiden kotouttaminen ja tulonsiirrot aiheuttavat suuria julkisia kustannuksia, mikä heikentää kansantaloudellista hyötyä.

Globaali väestönkasvu ja sen rajoittaminen

Maailmanlaajuista väestöräjähdystä pidetään merkittävänä uhkana ympäristölle ja ilmastolle. Ratkaisukeinoiksi globaaliin ylikansoittumiseen tarjotaan naisten koulutuksen ja aseman parantamista, ehkäisyn ja sterilisaation saamisen helpottamista sekä suoraa syntyvyyden säännöstelyä, kuten globaalia yhden lapsen politiikkaa useamman sukupolven ajan.

Vanhoillislestadiolaisuuden hoitokokoukset

Tekstissä käsitellään laajasti myös Suomen Rauhanyhdistysten Keskusyhdistyksen (SRK) piirissä 1970-luvulla pidettyjä niin sanottuja "hoitokokouksia". * Kokousten taustalla olivat 1960-luvun laaja yhteiskunnallinen arvomurros ja uudet ilmiöt, kuten suhtautuminen televisioon ja kunnallisen ehkäisyneuvonnan myötä herännyt keskustelu syntyvyyden säännöstelystä. * Vaikka kokoukset alkoivat sielunhoidollisina keskusteluina yhteisen uskon varjelemiseksi, 1970-luvun puolivälistä alkaen niissä alkoi esiintyä kristillisyydelle vieraita piirteitä. * Väärinkäytöksiin kuuluivat muun muassa oppiharhat (kuten erinimisistä hengistä parannuksen vaatiminen eli "henkioppi"), saneltu julkinen ripittäytyminen, yksilön omantunnon vähättely sekä henkinen ja hengellinen väkivalta. SRK:n johtokunta on myöhemmin pahoitellut näitä tapahtumia avoimesti ja todennut tosiasioiden kohtaamisen olevan tärkeää.

I. Syntyvyys, periodi 6.

*** Suomen syntyvyyden kehitys ja nykytila

Suomen syntyvyys on laskenut **historiallisen alhaiseksi**. Se alitti luonnollisen uusiutumistason jo 1960- ja 1970-lukujen taitteessa . Vaikka se välillä nousi tilapäisesti esimerkiksi vuoden 1973 päivähoitolain voimaantulon jälkeen , **2010- ja 2020-luvuilla syntyvyys on romahtanut nopeasti**. Suomen syntyvyys on nykyisin jäänyt jälkeen muista Pohjoismaista ja pudonnut Etelä-Euroopan tasolle.

*** Syntyvyyden laskuun liittyvät selitykset ja näkemykset

Keskusteluissa esitetään monia toisistaan eroavia ja osin voimakkaan ristiriitaisia näkemyksiä syntyvyyden laskun syistä:

* **Elämäntavat ja yksilölliset valinnat**: Syntyvyyden alenemisen syiksi nimetään **lastensaannin lykkääminen liian myöhäiseksi, ylipaino sekä haitalliset elämäntavat** . Myös miesten

syrjäytymistä ja perinteisen yökerho- ja alkoholikulttuurin vähentymistä pidetään parisuhteiden syntymisen esteinä . Nettideittailussa naisten korkeiden vaatimusten kumppaninetsinnässä katsotaan jättävän monia miehiä parisuhteiden ulkopuolelle. * **Yhteiskunnallinen ja taloudellinen epävarmuus**: Taloudellinen epävarmuus, tiukat rahatilanteet ja työnantajien oikuttelu pätkätöineen tekevät perheen perustamisesta vaikeaa . Valtion rahoituksen painopisteen väitetään siirtyneen nuorilta ja koulutuksesta vanhoille ja eläkkeisiin , ja poliittisia päätöksiä, kuten Kela-korvausten leikkauksia, kritisoidaan . * **Arvojen ja politiikan muutos**: Vihervasemmistolaista politiikkaa, feminismiä ja perhepoliittisia kokeiluja (kuten aiempaa yhteisverotuksen poistamista ja aborttilakeja) syytetään perinteisen perhemallin ja sitoutumisen rapauttamisesta. Myös **sukupuolineutraaliuden edistämistä, Setaa ja epäselvyyksiä sukupuolirooleissa** pidetään tekijöinä, jotka vaikeuttavat parinmuodostusta ja laskevat syntyvyyttä . * **Koronarokotteet ja depopulaatioteoriat**: Useat kirjoittajat uskovat koronarokotteiden (erityisesti mRNA-geeniterapeuttisten ruiskeiden) vahingoittavan hedelmällisyyttä tai aiheuttavan sikiökuolemia. Ruiskeiden väitetään olevan osa globaalin eliitin, WEF:n ja Maailman terveysjärjestön (WHO) suunnitelmallista sterilointia ja väestönvähennystä. Toisissa kommenteissa nämä väitteet tyrmätään perusteettomina "foliohattuiluna"ja viitataan tutkimuskatsauksiin, joiden mukaan rokotteet eivät vaikuta raskauden todennäköisyyteen .

*** Väestönrakenne, maahanmuutto ja "väestönvaihto"

Suomen ikääntyminen ja huoltosuhteen heikkeneminen herättävät huolta, sillä työssäkäyvien osuus eläkeläisiin nähden laskee jatkuvasti . Ratkaisuiksi esitetään kaksi vastakkaista linjaa: * **Maahanmuuton puoltajat**: Katsovat, että **voimakas työperäinen maahanmuutto** on ainoa keino pelastaa maa autioitumiselta, täyttää työvoimavaje ja turvata hoiva-alan resurssit suuryhteiskunnan ikääntyessä . * **Maahanmuuton vastustajat**: Epäilevät maahanmuuttajien työllistymistä ja katsovat heidän kuormittavan sosiaaliturvajärjestelmää. He pelkäävät suomalaisten muuttuvan vähemmistöksi omassa maassaan (**"väestönvaihto"**) maahanmuuttajien (erityisesti muslimien) korkeamman syntyvyyden vuoksi. Heidän mukaansa työvoimavaje tulisi ratkaista yksinomaan kotimaista syntyvyyttä nostamalla ja poistamalla tuet tai lapsilisät vieraskielisiltä . * Keskusteluissa pohditaan myös nationalistista utopiaa **"suljetusta Suomesta"**, joka olisi täysin omavarainen ja jossa ei olisi maahanmuuttajia . Tämä utopia kuitenkin tyrmätään toisaalla epärealistisena, autoritaarisena ja taloudelle sekä terveydenhuollon työvoimalle äärimmäisen haitallisena hankkeena .

*** Globaali konteksti ja ympäristökysymykset

Syntyvyyden aleneminen nähdään myös globaalina kehityksenä, joka koskee Suomen lisäksi monia Aasian maita, kuten Thaimaata, Vietnamia ja Kiinaa . Toisaalta Afrikan ja Lähi-idän väestönkasvun katsotaan olevan liian suurta, ja kehitysapua ehdotetaan suunnattavaksi sinne **syntyvyyden säännöstelyyn**.

Ympäristönäkökulmasta alhaista syntyvyyttä pidetään toisaalta parhaana ilmastotekona, ja suomalaisten syyllistämistä ilmastopäästöistä pidetään syynä siihen, miksi ihmiset kieltäytyvät tekemästä lapsia . Vihreää puoluetta syytetään ihmisvastaisuudesta, koska se ajaa aborttioikeutta ja avustettua kuolemaa, joiden katsotaan vähentävän tulevaisuuden sukupolvia .

II. Kuolleisuus, periodi 5.

Suomen hallituksen ja viranomaisten toimet sekä poliittinen vastakkainasettelu

Keskusteluissa esiintyy voimakasta kritiikkiä **Sanna Marinin hallitusta ja terveysviranomaisia (kuten THL:ää)** kohtaan. Hallitusta syytetään aluksi hitaasta reagoinnista ja vastuun välttelystä, kunnes muiden maiden esimerkin myötä otettiin käyttöön poikkeusolot. Erityistä närkästystä herättävät Marinin päätös hylätä asiantuntijoista koottu "koronanyrkki", maskisuositusten ja -selvitysten hitaus, rajatarkastusten poistaminen liian aikaisin sekä pääministerin

oma toiminta altistumistilanteessa . Toisaalta hallituksen toimintaa myös puolustetaan, ja sen katsotaan hoitaneen kriisiä mallikkaasti asiantuntijoihin tukeutuen.

Poliittiset jännitteet ovat kiivaita: vasemmistohallituksen toimia syytetään jopa "viherpunakaartin märäksi uneksi", jolla pyritään rajoittamaan sananvapautta ja siirtämään valtaa. Myös puolueiden, kuten Perussuomalaisten ja SDP:n, kannatuksen kehityksestä ja kriisijohtajuudesta väitellään ahkerasti. Osassa kirjoituksista verrataan Suomen mallia Ruotsiin, jossa terveysviranomaiset saivat perustuslain puitteissa päättää toimista ilman maallikkopoliitikkoja, sekä Ruotsin vähäisempiin rajoituksiin, joita sikäläinen johto piti muiden maiden "ihmiskokeiden"vastakohtana korkeammista kuolinluvuista huolimatta. Toiset taas nostavat esiin Uuden-Seelannin onnistuneet, nopeat eristystoimet esimerkillisinä.

Koronapassi, rokotteet ja salaliittoteoriat

Mielipiteet **koronarokotteista ja koronapassista** jakavat kirjoittajia erittäin jyrkästi: * **Skeptisyys ja vastustus:** Osa kirjoittajista pitää koronaa pelkkänä tavallisena flunssana ja rokotetta "kokeellisena litkuterapiana", jonka väitetään muuttavan ihmisen perimää tai aiheuttavan vakavia haittoja ja kuolemia . Koronapassia kritisoidaan tehottomaksi ja sitä pidetään pelkkänä yritystoiminnan ylläpitovälineenä, joka luo väärää turvallisuudentunnetta ja tekee passinhaltijoista "superlevittäjiä". Kirjoituksissa esiintyy myös epäilyksiä siitä, että virus olisi tehty wuhanilaisessa laboratoriossa ja levitetty tahallaan, tai että pandemialla pyritään viemään kansalaisoikeudet ja rakentamaan "Saatanan valtakuntaa". * **Puolustus ja tieteellinen näkökulma:** Toiset kirjoittajat korostavat rokotteiden pelastavan miljoonia henkiä ja muistuttavat, että niiden hyödyt ylittävät moninkertaisesti mahdolliset haitat . Rokotteen ottamista puolustetaan myös kollektiivisen vastustuskyvyn saavuttamiseksi ja terveydenhuollon kuormituksen keventämiseksi . mRNA-rokotteiden todetaan tieteellisesti olevan sellaisia, etteivät ne voi muuttaa solun DNA:ta .

Talousvaikutukset ja koronatuet

Pandemian **taloudelliset seuraukset** herättävät suurta huolta. Valtion ennätysmäistä velkaantumista ja EU:n elpymispaketin kuluja kritisoidaan. Erityistä paheksuntaa herättää Business Finlandin ja ELY-keskusten jakama taloustuki, jota kirjoittajien mukaan on valunut "zombie-yrityksille", nollaliikevaihtoa tekeville tai käytännössä konkurssikypsille yhtiöille. Esimerkkeinä mainitaan poliitikkojen tai heidän yritystensä saamat tuet, kuten Tanja Karpelan koirayrityksen saamat kymmenien tuhansien eurojen avustukset samanaikaisesti kun hän nostaa sopeutumisrahaa . Myös ravintola- ja matkailualan (MaRa) rajoitukset ja saamat tuet herättävät keskustelua.

Arki, hygienia ja asenteet

Kirjoituksissa käsitellään paljon **arkipäiväistä suojautumista**, kuten käsien pesua, desinfiointia ja suojamaskien käyttöä. Sukupolvien välillä nähdään jännitteitä: nuoria syytetään välinpitämättömyydestä ja ohjeiden laiminlyönnistä baareissa ja kotibileissä, kun taas toiset kyseenalaistavat sen, miksi yli 70-vuotiaita syyllistetään ja käsketään pysymään kotona, vaikka tilastollisesti tartuntoja on paljon nuoremmissa ikäluokissa. Osa kokee jatkuvan uutisoinnin ja rajoitukset "koronahysteriana"ja toivoo koko aiheesta puhumisen lopettamista.

II. Kuolleisuus, periodi 6.

Koronapandemia, rokotteet ja suojautuminen

Lähdemateriaalissa käydään erittäin vilkasta ja polarisoitunutta keskustelua koronaviruksesta ja rokotuksista: * **Rokotekritiikki ja salaliittoteoriat**: Monet kirjoittajat kyseenalaistavat koronarokotteiden turvallisuuden ja väittävät mRNA-injektioiden aiheuttavan laajaa ylikuolleisuutta. Rokotteita kuvaillaan "bioaseiksi"tai "troijalaisiksi", jotka heikentävät kehon im-

muunijärjestelmää ja pahentavat itse tautia. Osa kirjoittajista epäilee itse koronaviruksen olemassaoloa tai pitää pandemiaa "plandemiana"ja tavallisena flunssana . Myös kasvomaskien suodatustehoa epäillään, sillä mikään standardimaski ei kirjoittajan mukaan suodata 0,1 mikrometrin kokoista koronavirusta, vaikka FFP-luokituksia ja niiden suodatustehoja esitellään yksityiskohtaisesti. * **Rokotteiden puolustus**: Toiset kirjoittajat puolestaan korostavat, että lääketieteellistä näyttöä koronarokotteiden hyödyistä on massiivisesti . Heidän mukaansa rokotteet suojaavat erittäin tehokkaasti vakavilta tautimuodoilta ja vähentävät merkittävästi sairaalahoidon tarvetta sekä kuolemia. Sairaalaan joutuneista suuren osan kerrotaan olevan rokottamattomia. * **Hoitomuodot**: Kirjoituksissa kiistellään ivermektiinin ja hydroksiklorokiinin kaltaisten lääkkeiden tehosta. Erään vastaajan mukaan tutkimukset ovat kuitenkin osoittaneet, ettei ivermektiinistä ole hyötyä koronapotilaiden hoidossa.

Yhteiskunnalliset salaliittoteoriat ja globalisaatio

* **Mielenosoitukset ja Great Reset**: Kanadan "Freedom Convoy-liikettä ja muita koronamielenosoituksia tarkastellaan kriittisesti salaliittoteorioiden ja niin sanotun "Great Reset-agendan kautta, jota kuvataan pyrkimykseksi sulauttaa julkinen ja yksityinen sektori suuryritysten hallintaan . * **Kansainväliset sopimukset**: Suomen poliitikkojen väitetään syyllistyneen "maanpetokseen"allekirjoittamalla WHO:n kansainvälisen pandemiasopimuksen, jonka katsotaan siirtävän Suomen autonomian terveysasioissa ulkopuolisille tahoille . * **Cabal ja muut teoriat**: Lähteissä viitataan "Cabalin kaatuminen-dokumenttisarjaan, joka käsittelee salaseuroja ja globaalia pahuutta , väitetään Ukrainan sodan olevan vain "hämäystä"todellisten tavoitteiden peittelemiseksi ja vertaillaan uutisia Raamatun ennustuksiin .

Suomen politiikka ja valtiontalous

* **Hallituksen ja velanoton kritiikki**: Marinin hallitusta syytetään huonosta taloudenpidosta, holtittomasta velanotosta, tehottomista koronatuista ja sähkön hintakriisin heikosta hoitamisesta . * **Kriisien ja hallituksen puolustus**: Toisaalta hallituksen toimintaa puolustetaan huomauttamalla, että velkaantuminen tapahtui poikkeuksellisten kriisien, kuten koronapandemian, Ukrainan sodan ja puolustushankintojen vuoksi . Lisäksi korostetaan hallituksen onnistuneen viemään Suomen Natoon, toteuttamaan sote-uudistuksen ja pitämään työllisyyden korkeana vaikeissa oloissa . Myös oppositiopuolue Perussuomalaisten roolista ja politiikasta kiistellään puolesta ja vastaan.

Alueelliset ja muut paikalliset aiheet

* **Väestönkehitys**: Lähteissä seurataan Tampereen ja Turun seutujen väkiluvun kehitystä marraskuussa 2021 ja todetaan Tampereen kasvun olleen marraskuussa hieman rauhoittumaan päin. * **Paikalliset palvelut ja ilmiöt**: Kirjoituksissa arvostellaan Siun soten johtoa, Porissa toimivia "kuolleistaherättäjiä"sekä Muonion koulujen jaksojärjestelmäuudistuksia, joissa sivistyslautakunnan katsotaan sivuuttaneen enemmistön mielipiteen . * **Historia**: Mukana on myös katsaus Puolan ja Ukrainan väliseen sotaan (1918–1919) ja sen seurauksiin ukrainalaisten vankileireillä .

III. Muuttoliike, periodi 3.

*** 1. Maahanmuuton ja kotouttamisen taloudelliset vaikutukset

* **Kansantaloudelliset kulut ja velkaantuminen:** Useissa kirjoituksissa maahanmuutto ja erityisesti turvapaikanhakijoiden vastaanottaminen nähdään merkittävänä taloudellisena taakkana, joka velkaannuttaa valtiota entisestään. Esimerkiksi pakolaisten aiheuttamien vuosittaisten kustannusten arvioidaan olevan yli kaksi miljardia euroa, mikä vastaa lähes kahta prosenttia valtion budjetista. * **Peruspalveluista leikkaaminen:** Monet kommentoijat kritisoivat sitä, että valtio ja kunnat ohjaavat huomattavia resursseja maahanmuuttajien tukiin ja kotouttamiseen samaan aikaan kun kantaväestön omista palveluista – kuten vanhustenhuollosta,

terveydenhuollosta, päivähoidosta ja koulutuksesta – leikataan jatkuvasti. Sosiaalitukien, asuntotukien ja perheenyhdistämisen kuluja pidetään kohtuuttomina suhteessa siihen, että monet maahanmuuttajat jäävät vuosikymmeniksi tukijärjestelmän varaan. * **Aluetaloudellinen hyöty ja työllisyys:** Vastakkaisena näkökulmana esitetään, että vastaanottokeskusten perustaminen voi tuoda merkittävää aluetaloudellista piristystä ja työllisyyttä etenkin maaseudulle, sillä keskukset ostavat laajasti ruoka-, siivous-, sosiaali- ja terveys- sekä turvapalveluita paikallisilta yrityksiltä. Lisäksi osa kirjoittajista uskoo maahanmuuttajien tuovan tarpeellista työvoimaa ja korvaavan vanhenevaa väestöä.

*** 2. Yhteiskuntarauha, rikollisuus ja turvallisuus

* **Rikollisuus ja väkivalta:** Kirjoituksissa ilmenee voimakas huoli yhteiskuntarauhan rikkoutumisesta. Maahanmuuttajien väitetään syyllistyvän kantaväestöä huomattavasti herkemmin rikoksiin, kuten väkivaltaisiin ryöstöihin, jengiytymiseen ja seksuaalirikoksiin. Lähteissä viitataan Oikeuspoliittisen tutkimuslaitoksen selvitykseen, jonka mukaan Afrikassa ja Lähi-idässä syntyneiden miesten rikollisuustaso oli 17-kertainen syntyperäisiin suomalaisiin nähden. Lisäksi teksteihin on koottu laaja luettelo Suomessa annetuista rikostuomioista. * **Katupartiot:** Keskustelu Soldiers of Odin -katupartioliikkeestä ja muista vastaavista ryhmistä jakaa mielipiteitä. Jotkut pitävät niitä tarpeellisina ennaltaehkäisevän turvallisuuden ja järjestyksen ylläpitäjinä, kun taas viranomaiset ja oikeustieteen professorit arvioivat niiden toiminnan olevan ei-toivottavaa, turvattomuutta lisäävää ja poliisilain puitteissa hajotettavaa. * **Salailu ja vähättely:** Mediaa ja viranomaisia syytetään siitä, että maahanmuuttajien tekemiä rikoksia salataan, peitellään tai selitellään "yksittäistapauksina", mikä kirjoittajien mukaan vain lisää ennakkoluuloja ja ärsytystä.

*** 3. Kulttuurierot ja sopeutumisen vaatimukset

* **Maassa maan tavalla:** Monet kirjoittajat vaativat ehdotonta sopeutumista suomalaiseen kulttuuriin ja lakeihin, ja esittävät, että sopeutumattomien tulisi palata takaisin kotimaihinsa. Islaminuskon ja Lähi-idästä tulevan kulttuurin katsotaan olevan ristiriidassa länsimaisen naisten tasa-arvon, seksuaalivähemmistöjen oikeuksien sekä sanan- ja uskonnonvapauden kanssa. * **Uskonnolliset jännitteet:** Teksteissä tuodaan esiin huoli kristittyjen vainoista Lähi-idässä ja Välimerellä, ja pelätään, että pakolaisvirtojen mukana Eurooppaan ja pakolaiskeskuksiin rantautuu vihamielisiä asenteita kristittyjä kohtaan. Äärimmäisenä ehdotuksena esitetään tiukkoja sääntöjä, joilla muslimeilta evättäisiin pysyvät oleskeluluvat ja sharialaki kiellettäisiin. Toisaalta varoitetaan, että monikulttuurisuuden ihannointi sulauttamisen kautta saattaa vähitellen hävittää maiden omat, ainutlaatuiset kulttuurit.

*** 4. Kotouttamisen haasteet ja alaikäiset turvapaikanhakijat

* **Kotouttamisen epäonnistuminen:** Kotouttamistoimia pidetään monissa puheenvuoroissa epärealistisina ja tehottomina. Kirjoittajien mukaan kotouttaminen ei ole onnistunut missään päin maailmaa, ja sen seurauksena Suomeen on muodostumassa pysyvästi sosiaalitukien varassa elävä pakolaispohjainen köyhälistö. Työllistymistä vaikeuttavat puutteellinen kielitaito ja traumat. * **Vapaaehtoistyö ja tuki:** Toisaalta aineistossa esitellään myös myönteisiä esimerkkejä, joissa vapaaehtoiset ja kirjastotyöntekijät tekevät aktiivisesti työtä tulijoiden auttamiseksi, kielenopetuksen tarjoamiseksi ja viranomaispäätösten helpottamiseksi, nähden heidät voimavarana. * **Alaikäiset hakijat ja ikätestit:** Ruotsiin yksin saapuvien alaikäisten turvapaikanhakijoiden määrän voimakkaan kasvun taustalla nähdään ihmissalakuljettajien verkostot sekä niin sanottu "ankkurilapsiteoria", jossa perheen toivotaan seuraavan myöhemmin perässä. Suomessa alaikäisiksi ilmoittautuneille tehdyistä ikätesteistä mainitaan, että testatuista 149 henkilöstä 90 osoittautui todellisuudessa täysi-ikäisiksi.

*** 5. Poliittinen päätöksenteko, media ja instituutiot

* **Poliitikkojen arvostelu:** Kirjoittajat arvostelevat istuvia ja aiempia hallituksia (erityisesti Keskustaa, SDP:tä ja Kokoomusta sekä poliitikkoja kuten Sipilää, Stubbia, Orpoa ja Kataista) holtittomasta maahanmuuttopolitiikasta ja siitä, että Suomen vastaanottamien turvapaikanhakijoiden määrä kasvoi vuonna 2015 peräti 822 prosenttia edellisvuodesta. * **Instituutioiden ja median kritiikki:** Suomen Punaista Ristiä (SPR) ja Maahanmuuttovirastoa (Migri) syytetään "turvapaikkabisneksen"pyörittämisestä valtion kustannuksella ilman riittävää valvontaa. Median (kuten Yle ja Helsingin Sanomat) katsotaan harjoittavan propagandaa ja sensuroivan tai leimaavan rasistiseksi kaiken hallitsematonta maahanmuuttoa vastustavan asiallisenkin keskustelun.

*** 6. Muut erilliset teemat

* **Lähi-idän historialliset konfliktit:** Tekstit sisältävät yksityiskohtaisen kronologian vuoden 1948 tapahtumista Palestiinassa, kuvaten juutalaisten terroristijärjestöjen (kuten Haganah, Stern Gang ja Irgun) tekemiä pommi-iskuja, murhia ja kyliin kohdistuneita hyökkäyksiä arabiväestöä vastaan ennen Israelin valtion perustamista. * **Sateenkaarinuorisotilat ja Seta:** Kirjoituksissa kritisoidaan seksuaalivähemmistöjä edustavien järjestöjen pyrkimyksiä viedä monimuotoisuuskasvatusta päiväkoteihin ja kouluihin. Erityistä huola esitetään Vantaan ja Espoon sateenkaarinuorisoilloista, joissa 13-vuotiaat ja jopa 29-vuotiaat aikuiset tuodaan samaan tilaan vapaa-ajanvieton ja seksuaalisuusteeman ympärille ilman vanhempien riittävää tiedottamista. * **Suomalaisen kulttuurin kritiikki:** Yhdessä kirjoituksessa kuvataan erittäin kriittisesti suomalaisten piirteitä, arkkitehtuuria, huonoja ajotaitoja, autojen palvontaa, heikkoa asiakaspalvelukulttuuria sekä ihmisten katkeruutta ja taipumusta uhriutumiseen.

III. Muuttoliike, periodi 6.

*** Maahanmuutto, kotouttaminen ja taloudelliset vaikutukset

Keskusteluissa esitetään voimakkaita ja keskenään ristiriitaisia näkemyksiä maahanmuuton seurauksista Suomelle: * **Taloudellinen taakka ja sosiaalituet:** Monet kommentoijat kuvaavat maahanmuuttoa kulueränä, joka rasittaa julkista taloutta, luo pitkiä jonoja Kelaan ja vaatii runsasta sosiaalitukea . Lähteissä esitetään esimerkkejä tukijärjestelmän tietoisesta hyödyntämisestä (kuten erillään asuminen tukien maksimoimiseksi) sekä eritellään yksityiskohtaisesti **Janvier Nsabimanan** (alias "January Muhire") oikeustapausta. Nsabimana tuomittiin 70 päivän ehdolliseen vankeuteen väärien henkilötietojen antamisesta ja rekisterimerkintärikoksesta, mutta Maahanmuuttoviraston yli 50 000 euron korvausvaatimus vastaanottokustannuksista hylättiin oikeudessa, koska hän olisi ollut oikeutettu samoihin etuuksiin myös omalla nimellään . * **Työmarkkinavaikutukset:** Osa keskustelijoista näkee maahanmuuton keinona tuoda maahan matalapalkkaista "halpatyövoimaa", mikä heikentää suomalaisten omaa palkkatasoa ja hyödyttää veroparatiisiyrityksiä. Jussi Halla-aho täsmentää perussuomalaisten vastustavan nimenomaan halpatyövoiman maahanmuuttoa, mutta ei työntekoon perustuvaa maahanmuuttoa sinänsä . Toisaalta tuodaan esiin, että **tekoälyn kehitys** voi pian poistaa tarpeen ulkomaiselle työvoimalle korvaamalla suuren osan nykyisistä tehtävistä . Vastakkaisissa puheenvuoroissa maahanmuuttoa ja syntyvyyden kasvua pidetään välttämättömyytenä heikkenevän huoltosuhteen vuoksi , ja maahanmuuttajien työllistymistä palvelualoille (kuten ruokaläheteiksi tai hoiva-alalle) pidetään elintärkeänä yhteiskunnan toimivuudelle . * **Rikollisuus, turvallisuus ja kotoutuminen:** Katkelmissa ilmaistaan huolta **katujengeistä**, joiden jäsenten todetaan poliisin mukaan olevan pääosin ensimmäisen tai toisen polven maahanmuuttajia . Myös seksuaalirikoksia ja rikollisuutta vankiloissa käsitellään laajasti. **Ruotsia** käytetään keskusteluissa toistuvasti varoittavana esimerkkinä epäonnistuneesta integraatiosta ja kasvaneesta jengiväkivallasta.

*** Puoluepoliittinen keskustelu ja hallitusvastuu

Suomalaisen puoluekentän toimet maahanmuuttokysymyksissä herättävät vilkasta ja kriittistä keskustelua: * **Perussuomalaiset:** Kannattajat pitävät puoluetta ainoana toimijana, joka pyrkii parantamaan maata. Kriitikot taas syyttävät heitä tyhjistä vaalilupauksista (kuten polttoaineiden hintojen laskusta tai rajojen sulkemisesta) sekä tekopyhyydestä, sillä Sipilän hallituskaudella vuonna 2015 – perussuomalaisten ollessa hallituksessa – Suomeen saapui ennätysmäärä turvapaikanhakijoita . Riikka Purran vaatimukset ja lausunnot herättävät myös ristiriitaisia reaktioita. * **Muut toimijat:** Kokoomusta kritisoidaan suomalaisten etujen polkemisesta ja veroparatiisiyritysten tukemisesta halpatyövoimalla. Sanna Marinin hallitusta syytetään valtionvelan kasvattamisesta ja löysästä maahanmuuttopolitiikasta , mutta toisaalta puolustetaan sote-uudistuksen ja maksuttoman toisen asteen koulutuksen kaltaisilla saavutuksilla . RKP:n (erityisesti Eva Biaudetin) vastustusta maahanmuuttosäädöksiin kritisoidaan voimakkaasti , kun taas RKP:n virallisena tavoitteena mainitaan humanitaarisen maahanmuuton näkeminen voimavarana . * **EU-tason kehitys:** Lähteissä kuvataan myös Saksan sisäministeri Nancy Faeserin ajamaa aloitetta **"vastaanottohalukkaiden maiden liitosta"**, jossa halukkaat maat ottaisivat vastaan turvapaikanhakijoita ja muut maat osallistuisivat taakanjakoon rahallisesti.

*** Itärajan tilanne ja hybridisota

* **Venäjän hybridioperaatio:** Itärajan turvapaikanhakijavirtaa pidetään Venäjän tietoisena hybridisodankäynnin välineenä, jolla pyritään horjuttamaan Suomen yhtenäisyyttä ja demokratiaa . * **Viranomaisten ja hallituksen toiminta:** Keskusteluissa arvostellaan hallituksen ja sisäministeri Mari Rantasen toimintaa rajan sulkemisessa . Erityisesti viitataan apulaisoikeuskanslerin antamaan huomautukseen siitä, että Sisäministeriön ensimmäiset rajasulkuvaihtoehdot olivat juridisesti puutteellisia ja termistöltään virheellisiä ennen kuin luonnokset muotoiltiin uudelleen yhteistyössä oikeuskanslerin viraston kanssa .

-

ETLA

Elinkeinoelämän tutkimuslaitos

ETLA Economic Research

ISSN-L 2323-2420
ISSN 2323-2420 (print)
ISSN 2323-2439 (pdf)

Kustantaja: Etlatieto Oy

Puh. 09-609 900
www.etla.fi
etunimi.sukunimi@etla.fi

Arkadiankatu 23 B
00100 Helsinki