Tartu Ülikool
Humanitaarteaduste ja kunstide valdkond
Ajaloo ja arheoloogia instituut
Eesti ajaloo osakond

Tormi Lust

Tubakareklaamid Eesti ajalehtedes aastatel 1920-1940: korpusepõhine analüüs

Bakalaureusetöö

Juhendajad: Peeter Tinits, PhD
Aigi-Rahi Tamm, PhD

Tartu
2025

Sisukord

Sissejuhatus

Tubaka tarvitamine on Euroopas tuntud juba sajandeid.[1] Eestis hakkas tubakakasvatus ja -tööstus hoogsamalt arenema 1920. aastate alguses. 19. sajandi lõpul tegutsesid siin üksikud tubakatöökojad, peamiselt Tartus, Pärnus ja Tallinnas, kus sigareid ja paberosse valmistati valdavalt käsitsi. Kuigi mitmed töökojad suleti 1860.–1870. aastate tubakaaktsiisi seaduste karmistumise tõttu, säilis huvi odava ja kergesti kättesaadava tubaka vastu.[2] Kõrgemad tolli- ja aktsiisimäärad tõid kaasa nii riigi tulude kasvu kui ka salajasi paberosside valmistamiskohti, mida kohalik ajakirjandus nimetas salajasteks tubakavabrikuteks.[3]

Kui varasemalt toodi Eestisse tubakat ja selle saadusi peamiselt Peterburi, Moskva ning osalt ka Riia tubakavabrikutest, siis suurem muutus kaasnes Esimese maailmasõja ja Vabadussõja ajal, mil katkes ühendus Venemaaga, kust tubakasaaduste sissevedu lõppes ning valitsus piiras tubaka sissevedu mittevajaliku kaubana. See tõi kaasa tubakapuuduse, mistõttu hakkasid kodanikud tubakat ise kasvatama ning mitmed ärimehed nägid selles võimalust rajada kaasaegsed tubakavabrikud.[4] 1920. aastate alguses asutati mitmeid ettevõtteid, näiteks OÜ „Tubak“,[5] AS „Laferme“ ja AS „Astoria“, mis importisid toortubakat peamiselt Bulgaariast, Kreekast, Türgist ning Ameerika Ühendriikidest.[6] 1930. aastal oli paberossikauplustes müügil korraga kuni 50 erinevat sorti.[7] Konkurents tubakatööstuses kujunes tihedaks ja järk-järgult läks turg suurtootjate kontrolli alla. Väiksemad vabrikud kas ühinesid või osteti suuremate tegijate poolt ära, paljud pankrotistusid.[8]

1930. aastate lõpus jäi turule peamiselt neli suuremat tubakatehast, mis investeerisid modernsetesse masinatesse[9] ja panustasid turundusse, näiteks reklaamides „nikotiinivabu“ paberosse.[10] Kui arstid avaldasid 1930. aastatel kirjutisi tubaka kahjulikkusest,[11] siis leidsid

[1] Andrzej Grzybowski, “The History of Antitobacco Actions in the Last 500 Years. Part. II. Medical Actions,” *Przeglad Lekarski* 63, nr. 10 (2006), lk 1131 – 1134.
[2] Otto Karma, *Tööstuslikult revolutsioonilt sotsialistlikule revolutsioonile Eestis* (Tallinn, 1963), lk 175–176.
[3] Marge Kannel, “Tubakatööstus Eesti Vabariigis 1920–1940” lõputöö, juhendaja Jaak Valge, Tartu Ülikool, Filosoofiateaduskond, Ajaloo osakond, 2005, lk 9 – 10.
[4] Kannel, “Tubakatööstus Eesti Vabariigis 1920–1940,” lk 12.
[5] 1919. aastal rajati AS “Tubak”, mille asutajateks olid Mihkel Pung, Gustaw Pihlakas ja Karl Lemberg. 1920. aastal rajati OÜ “Tubak”, mille asutajateks olid Eduard Alwer, Eduard Saarepera, Eduard Pool, Ado Tõllasepp ja G. Pihlakas. – *Riigi Teataja* 1919, nr. 82 – 83, lk 553 – 555; *Riigi Teataja* 1920, nr 69 – 70, lk 651 – 657.
[6] Kannel, “Tubakatööstus Eesti Vabariigis 1920–1940,” lk 14.
[7] Postimees, 09.12.1930, lk 2.
[8] Samas.
[9] Päewaleht 08.10.1930, lk 4.
[10] Päewaleht 16.12.1930, lk 32.
[11] Apollon Spiridis. “Tubakasuitsu mõju organismi ja närvisüsteemi.” Uurimustöö. Tartu Ülikool, 1930.

tootjad lahendusi, kirjutades, et nende paberossid on “kahekordse vatiga”[12], nende “paberossidel puudub nikotiinimürk”[13] või, et nende tubakad on “pneumaatilise seadmega tolmust ja kahjulikest lisanditest puhastatud”.[14] Tööstusharu areng katkes 1940. aastal kui Nõukogude Liidu okupatsioon tõi kaasa kõigi tubakaettevõtete natsionaliseerimise, varade riigistamise ning konkurentsipõhise majanduskeskkonna asendumise keskselt juhitud tootmisega.

Senised uuringud Eesti tubaka ajaloost on keskendunud peamiselt selle tööstuslikule arengule ja vastavale seadusandlusele.[15] Digitaalajastu on avanud ajaloolastele ja kultuuriteadlastele enneolematud võimalused.[16] Eesti mäluasutuste poolt läbi viidud ulatuslikud digiteerimisprojektid, sealhulgas Eesti ajalehtede digitaalarhiivi (DIGAR) loomine, on muutnud arhiivimaterjalid kergesti kättesaadavaks. See külluslik andmemaht nõuab uute digitaalhumanitaaria metoodikate arendamist, mis võimaldab poolautomaatselt tekste organiseerida, mõista, otsida ja kokku võtta. Lisaks pakub digitaalne formaat võimalusi kvantitatiivseks ja kvalitatiivseks analüüsiks, mis varem olid raskesti teostatavad.[17]

Rahvusraamatukogu digitaalarhiivis DIGAR digiteeritud ajalehtede alastes uurimustes on keskendutud artiklite sisu analüüsile,[18] kuid seal on eraldi segmenteeritud ka kuulutused. See tähendab, et kuulutuste täisteksti külge on lisatud omadused ehk metaandmed (millisele leheküljele see kuulub, kas tekst on pärit artiklist või kuulutusest, mitu sõna on segmendis kokku). Kuulutused annavad infot ettevõtete turundusstrateegiate kohta (kuidas suitsumarke nimetati, kui tihti neid ajalehes esines), sisaldavad infot tubaka reklaamimise, tarbijakäitumise ja sotsiaalsete normide kajastumise kohta. DIGAR-is digiteeritud ajalehtede reklaamidel on suur uurimispotentsiaal.

---

[12] Päevaleht, 26.01.1939, lk 5.
[13] Päevaleht, 16.12.1930, lk 20.
[14] Sakala, 13.01.1936, lk 2.
[15] Kannel, “Tubakatööstus Eesti Vabariigis 1920–1940”; Ronald Juurmaa, “Eesti tubakatööstus 1919–1940 tubakavabrik ‘Laferme’ näitel,” magistritöö, juhendajad Jaak Valge ja Maie Pihlamägi, Tartu Ülikool, Filosoofiateaduskond, Ajaloo ja arheoloogia instituut, 2012.
[16] M. J. H. F. Wevers, “Consuming America: A Data-Driven Analysis of the United States as a Reference Culture in Dutch Public Discourse on Consumer Goods, 1890–1990,” doktoritöö, Utrecht University, 15.09.2017, lk 41 – 42.
[17] Wevers, “Consuming America,” lk 44 – 45.
[18] Peeter Tinits, “Elekter, aur ja hobujõud 20. saj. vahetusel,” (16.11.2020), https://data.digar.ee/samples/elekter_aur_hobu.html (viidatud 14.05.2025).

Andmestik ja metoodika

Käesolevas töös analüüsitakse viie suurema päevalehe – Päevaleht, Postimees, Sakala, Kaja (1919 kuni 17.09.1935) ja selle järglase Uus Eesti (alates 18.09.1935)[19] põhjal kui palju varieerus reklaamide hulk ajavahemikul 1920 – 1940, missuguseid sõnu kasutati tubakatoodete iseloomustamiseks ja reklaamimiseks. Selleks kasutatakse töös Rahvusraamatukogu digilabori abivahendeid, mis võimaldab andmeid koguda ja ajalehekuulutuste sõnavaral põhineva tekstikorpuse luua. Loodud korpuse ja sõnade analüüsi töövoog võimaldab analüüsida ajalehekuulutusi masinloetaval kujul.

Ajaleht, ühe mahukama ja järjepidevama digiteeritud allikaliigina, pakub regulaarse ilmumissageduse tõttu rikkalikult kvantitatiivseks analüüsiks sobivat materjali. Sellest tulenevalt on käesoleva uurimuse fookuses just ajalehtedes ilmunud tubakareklaamid, kõrvale on jäetud ajaleheartiklid ja teised reklaamikanalid. Uurimisperioodiks on aastad 1920 – 1940, mis hõlmab Eesti Vabariiki alates iseseisvuse algusest kuni Nõukogude Liidu okupatsioonini kui ettevõtted natsionaliseeriti.

1920. – 1930. aastate Eesti tubakareklaame on analüüsitud vaid üksikjuhtumitena. Silmapaistvaim on Merle Talviku dissertatsioon, kus ta kirjeldab 1930. aastate ajakirjagraafikat ja sealhulgas tutvustab tubakareklaamide kunstnikke.[20] Sel perioodil tegutsenud tubakavabrikuid on kaardistanud Ronald Juurmaa[21] ja Marge Kannel.[22]

Antud töö eesmärgiks on avada ajaloo, ajakirjanduse ja teiste humanitaaria valdkonnaga seotud uurijatele võimalusi, mida saaks rakendada ajalehtede uurimiseks digimeetodeid kasutades. Selleks on loodud andmetöötluse ja -analüüsi töövoog, kus on rakendatud DIGARi digilabori võimalusi, Tartu Ülikooli pakutavat veebipõhist interaktiivset arenduskeskkonda JupyterLab;[23] keelelisi andmeid on töödeldud läbi RStudio, mis on R programmeerimiskeele[24] arenduskeskkond,[25] et analüüsida andmestikku nende sageduse baasil.[26] Tegu on praktilise, rakenduslikku laadi uurimustööga. Analüüsi läbiviimiseks

[19] Epp Lauk, *Peatükke Eesti ajakirjanduse ajaloost* (Tartu: TÜ Kirjastus, 2000), lk 14–15.
[20] Merle Talvik, "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia," doktoritöö, Tallinna Ülikool, Kunstide Instituut, 2010.
[21] Juurmaa, "Eesti tubakatööstus 1919–1940 tubakavabrik 'Laferme' näitel."
[22] Kannel, "Tubakatööstus Eesti Vabariigis 1920–1940."
[23] HPC Public Documentation, "Jupyter," 24.11.2020; muudetud 15.10.2024, https://docs.hpc.ut.ee/public/services/jupyter.hpc.ut.ee/
[24] The Comprehensive R Archive Network (CRAN), viidatud 14.05.2025, https://cran.r-project.org/
[25] RStudio kasutusjuhend, avaldatud 05.05.2025, https://docs.posit.co/ide/user/
[26] Digilabori kohta lähemalt 2.1. alapeatükis ning analüüsi käsitlen 3. peatükis.

töötasin välja töövoo lähtekoodi, mis on koos erinevate versioonide ajaloo ja juhendfailidega avalikult kättesaadavad GitHubi repositooriumis, tagades töö reprodutseeritavuse.[27] Konkreetsemalt on metoodilisi lähtekohti avatud vastavatest alapeatükkides.

Rahvusvaheliselt on näidanud Melvin Wevers[28] Hollandi ajalehereklaamide kvantitatiivsest analüüsist avanevaid metoodilisi võimalusi, millega saab teostada reklaamianalüüsi ka Eesti digiteeritud ajalehe-andmebaasides. Wevers on oma töös rakendanud arvutuslikke meetodeid, et analüüsida suuremahulist digiteeritud ajaleheandmestikku ning uurinud reklaami ja ühiskonna seoseid kvantitatiivselt.[29]

Toetusin käesolevas töös ajastu konteksti tutvustamisel Ronald Juurmaa ning Marge Kannela uuringutele, reklaamipraktikast kirjutades toetusin Merle Talvikule.[30] Oluliseks eeskujuks antud töös on Melvin Wevers'i uurimused, kuna need näitavad, kuidas digitaalseid materjale on võimalik töödelda masinloetaval kujul ning rakendada sarnaseid automaatse tekstianalüüsi põhimõtteid.

Käesoleval bakalaureusetööl on **neli peamist uurimisküsimust**. (1) Millised metoodilised väljakutsed tekivad DIGARis leiduvate ajalehekuulutuste masstöötlusel (tekstituvastuse kvaliteet, sisutõlge)? (2) Milline sõnakasutus joonistub välja tubakareklaamide massanalüüsimisel? (3) Kuidas on reklaamide sõnastus ettevõtete lõikes erinenud? (4) Kas reklaami analüüsimisel esilekerkivad eripärad on seletatavad ajalooliste protsessidega?

Töö koosneb kolmest peatükist. Esimeses antakse ülevaade Eesti tubakatööstuse ajaloost ning 1920. – 1930. aastate reklaamipraktikatest. Teises peatükis kirjeldatakse tubakareklaamide korpuse moodustamist ning selleks kasutatud vahendeid. Kolmandas peatükis hinnatakse tehtud töö kvaliteeti ja analüüsitakse korpust, tubakareklaamide sõnakasutust, mille pinnalt joonistuvad välja teatud eripärad.

---

[27] "Tubakakorpus." Tormi Lust, (GitHub repositoorium. Viimati muudetud 17. mai 2025) https://github.com/tormil/tubakakorpus/

[28] Melvin Wevers, "Mining Historical Advertisements in Digitised Newspapers," *De Gruyter eBooks*, 2022, 227–52, https://doi.org/10.1515/9783110729214-011; Melvin Wevers, Jianbo Gao ja Kristoffer Laigaard Nielbo, "Tracking the Consumption Junction: Temporal Dependencies between Articles and Advertisements in Dutch Newspapers," *Digital Humanities Quarterly* 14 (2019), kättesaadav 27.11.2024: https://digitalhumanities.org/dhq/vol/14/2/000445/000445.html; Melvin Wevers ja Jesper Verhoef, "Coca-Cola: An Icon of the American Way of Life. An Iterative Text Mining Workflow for Analyzing Advertisements in Dutch Twentieth-Century Newspapers," *Digital Humanities Quarterly* 11, nr. 4 (2017).

[29] Wevers, Gao ja Nielbo, "Tracking the Consumption Junction."

[30] Juurmaa, "Eesti tubakatööstus 1919–1940 tubakavabrik 'Laferme' näitel."; Kannel, "Tubakatööstus Eesti Vabariigis 1920–1940."; Talvik, "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia."

Tulevasteks uurimissuundadeks antud teemal on võimalikud mitmed lähenemised: lähemalt saab uurida tubakareklaamide kunstnike töid, tubakaettevõtete ajalugu nii antud kui ka järgneval perioodil, kaasata võiks ka teisi reklaamikanaleid, näiteks plakateid. Kvantitatiivse andmeanalüüsi kaudu annab parandada ajalehtede metaandmete segmenteerimist, tehisintellekti vahenditega tekstituvastuse kvaliteeti, et uurida artikleid ja reklaame, kasutada teisi analüüsimeetodeid, nagu meelestatusanalüüs ja teemade modelleerimine. Lisaks eelnimetatule võib pelgalt sõnasageduste uurimisest kaugemale minna ning hinnata artiklite ja reklaamide vahelisi seoseid kausaalsusanalüüsimeetoditega nagu *Granger Causality Test*.

1. Tubakatööstuse jõudmine Eestisse ja selle areng 1940. aastani

1.1. Tubaka jõudmine Eestisse ja Venemaa tubakatööstus

Kõige varasem teadaolev märk tubakatootmisest Eestis oli krahv Gustav Diedrich Rehbinder'i tubakalõikamise töökoda, mille ta asutas 1780. aastal. Teiseks oli Õisu mõisahärra Friedrich Wilhelm von Sivers'i[31] tubakalõikuse ja sigaritegemise töökoda, mis tegutses umbes 10 aastat.[32]

1810. aastal, pärast kontinentaalblokaadist tingitud majandusraskusi, tegi Vene keisririigi valitsus uute ettevõtete asutajatele mitmeid soodustusi.[33] Tänu sellele rajati Tallinna viis tubakatööstuse ettevõtet, kus lõigati tubakat ja valmistati käsitsi sigareid. Tolle aja suurim vabrik kuulus G. H. Burrot'ile, kus töötas 1823. aastal kokku 20 töölist.[34]

19. sajandi teisel poolel Vene impeeriumis toimunud tootmise mehhaniseerimisega ulatus tubakamanufaktuuride koguhulk 1860. aastal 551-ni, kuhu Eestist kuulusid kaks ettevõtet Tartus, üks Pärnus ja üks Tallinnas.[35] 1861. aastal kehtestatud aktsiisimäärus tõi kaasa paljude seniste tubakamanufaktuuride sulgemise ja nende koguarv Venemaal langes 304-le. Eestisse jäi vaid kaks tegutsevat ettevõtet – Toepfferi sigarivabrik Tartus ja tubakaettevõte Pärnus.[36] 1868. aastal lisandus Tartusse Fleischhauer-Cordsi sigarivabrik.[37]

1877. a aktsiisiseaduse kehtestamise järel tõusis tubakavabrikute hulk Venemaal 461-ni, kuid turgu hakkasid hõivama suurettevõtted.[38] 1878. a tööndusloenduse andmeil valmistati Tartus sigareid, paberosse ja töödeldi tubakat käsitsi. Tööliste arv Tartu tubakavabrikutes kokku oli 104 ning 1879. aasta toodangu väärtuseks oli 110 000 rubla.[39] 1890. aastatel rajati tubakavabrik Tapale.[40] Samal ajal jäi Tartusse tegutsema vaid Heinrich Wöhrmanni vabrik, kus kasutati juba aurumasinat ning mille 128 töölist andsid 1890. aastal toodangut 56 000

[31] Kannel, "Tubakatööstus Eesti Vabariigis 1920–1940," lk 8.
[32] A. Nõulik ja M. Ets (toim.), *Tubakaraamat* (Tallinn: Eesti Entsüklopeediakirjastus, 2002), lk 98.
[33] Otto Karma, *Tööstuslikult revolutsioonilt sotsialistlikule revolutsioonile Eestis* (Tallinn: Eesti NSV Teaduste Akadeemia, 1963), lk 29.
[34] Juurmaa, "Eesti tubakatööstus 1919–1940 tubakavabrik 'Laferme' näitel," lk 12.
[35] Karma, *Tööstuslikult revolutsioonilt sotsialistlikule revolutsioonile Eestis*, lk 175.
[36] Eesti Postimees, 31.07.1868.
[37] Eesti Postimees, 25.12.1868.
[38] Karma, *Tööstuslikult revolutsioonilt sotsialistlikule revolutsioonile Eestis,* lk 175–176.
[39] Karma, *Tööstuslikult revolutsioonilt sotsialistlikule revolutsioonile Eestis,* lk 175–176.
[40] Kannel, "Tubakatööstus Eesti Vabariigis 1920–1940," lk 9.

rubla väärtuses.[41] 1892. aastal sai Wöhrmann tehases elektrivalgustuse kasutuselevõtu eest kiita Postimehes, artikkel algas kõlava lausega „Esimene elektrivalgus Tartus“.[42]

20. sajandi alguses toodi Eestisse tubakasaadusi ja -tooteid sisse peamiselt Peterburi, Moskva ja Riia tubakavabrikutest. Enamus Venemaa tubakatöötlemisest koondus Ukrainasse, paberosside valmistamine aga Peterburi, mis andis 1880. aastal 67% kogu Venemaa toodangust. Üle 80% Venemaal toodetud sigaritest valmisid samal ajal Riias ja Peterburis.[43] Seaduslikult tegutsenud ettevõtete kõrval leidus ka mitmed salajasi tubakatootjaid, enamasti väikeste töökodade kujul.[44]

Peterburi tubakavabrikandid olid juba tol ajal mõistnud kaubamärgi olulisust ning sellega üritati meelitada Eesti tubakatarvitajaid nende tooteid ostma,[45] andes 1886. aasta Postimehe sõnul paberossidele nimed „Vanemuine“ ja „Kalewipoeg“.[46] Samamoodi toodeti 1907. aastal paberosse „Kalewipoeg” Saatschi ja Mangubi vabrikust Peterburis.[47] Samuti olid müügil paberossid nimega „Mõistatused“, kus paberossidele lisati meelelahutuseks mõeldud mõistatuste leht.[48]

Pilt 1. Saatchi ja Mangubi eestikeelne reklaam 1907. aastal.[49]


!!! Wennad !!! Soowitan suitsetada ainult
Paberossisid
„Kalewipoeg"
25 tükki 15 kop., 10 tükki 6 kop.
Saatschi & Mangubi wabrik Peterburis
Saada igal pool
Paberossid „Kalewipoeg" on walmistatud niisugusest tubakast, mis terwisele mitte kahjulik ei ole, ja maitse on neil kõige parem. Kes neid paberossisid tarwitab, saab jõuu pooleft kangel., kui Kalewipoeg.
Asjatundja suitsetaja.

[41] Samas, lk 9.
[42] Postimees, 25.11.1892, lk 3.
[43] Karma, *Tööstuslikult revolutsioonilt sotsialistlikule revolutsioonile Eestis,* lk 176.
[44] Kannel, “Tubakatööstus Eesti Vabariigis 1920–1940,” lk 9 – 10.
[45] Samas, lk 10.
[46] Postimees, 18.01.1886, lk 2.
[47] vt Pilt 1.
[48] Postimees, 18.01.1886, lk 2.
[49] Postimees, 04.04.1907, lk 3.

Venemaa tubakavabrikud eksportisid oma tooteid ka välismaale. 1897. aasta teatas ajaleht Sakala, et Venemaa paberosside eksport Saksamaale kasvas märkimisväärselt: 1887. aastal saadeti sinna 7 miljonit paberossi, 1889. aastal 20 miljonit, 1892. aastal 37 miljonit ja 1895. aastal juba 40 miljonit paberossi.[50] Eestis muutus tubakatööstuse areng märkimisväärseks siiski alles Eesti Vabariigi ajal.[51]

1.2. Eesti tubakatööstuse kujunemine kuni 1940. aastani

Enne Esimest maailmasõda jõudis Eestisse tubakat eelkõige Peterburi, Moskva ja osaliselt ka Riia tubakavabrikutest. Iseseisvumise, Saksa okupatsiooni ja Vabadussõja puhkemisega katkes senine tarneahel, välisriikidest tubaka toomine muutus väga keeruliseks ning valitsus kehtestas kokkuhoiu eesmärgil tubaka kui mittevajaliku kauba sisseveokeelu.[52] Sellest tulenevalt hakkasid inimesed ise tubakat kasvatama. Tubakapuudusest johtuvalt müüdi tänavatel heina- ja lehepuruga täidetud ‘paberosse’. 1919. a jaanuaris andis Ajutine Valitsus välja määruse, mis keelas tänavatel paberosside müügi, seda võis korraldada vaid müügiõiguse loaga kauplustes.[53] Aeg, mil tubaka sissevedu oli keelatud, pani aluse Eesti tubakatööstuse arengule. Paljud ärimehed hakkasid kasutama kodumaist tubakat, et valmistada käsitsi paberosse, mis küll ei vastanud tööstusliku toodangu kvaliteedile, aga osutus siiski tulusaks ettevõtmiseks. Aja möödudes vähenes käsitsi valmistatud paberosside tootmine. Enamik väiketöökodasid suleti või muudeti suuremateks tootmisüksusteks, näiteks August Reieri tubakavabrik Tallinnas, mis suutis ajaga kaasas käia ja areneda kaasaegseks tööstusettevõtteks.[54]

Kuna Eesti majandus oli Esimese maailmasõja, selle järgnenud pöördeliste muutuste ning Saksa okupatsiooni tõttu raskelt kannatanud, järgnesid Vabadussõjale tööstuse kohanemise ja ümberkorraldamise aastad (1920 – 1924).[55] 1920. aastal võttis Asutav Kogu vastu seaduse, mis nõudis alkoholitööstuste ja tubakavabrikute asutamiseks rahandusministri eriluba.[56] Sama aasta aprillis kehtestati tubakale ja tubakasaadustele aktsiisi-, patendi- ja tollimaks. Aktsiisimaksu võeti valmistehtud tubaka ja tubakasaaduste pealt; patendimaksu

[50] Sakala, 24.04.1897, lk 4.
[51] Juurmaa, “Eesti tubakatööstus 1919–1940 tubakavabrik ‘Laferme’ näitel,” lk 13.
[52] Kannel, “Tubakatööstus Eesti Vabariigis 1920–1940,” lk 11.
[53] Postimees 28.01.1919, lk 2.
[54] Kannel, “Tubakatööstus Eesti Vabariigis 1920–1940,” lk 11.
[55] Juurmaa, “Eesti tubakatööstus 1919–1940 tubakavabrik ‘Laferme’ näitel,” lk 8.
[56] Tubakamaksu seadus, 10.04.1920. – Riigi Teataja 1920, nr 65/66, lk 517 – 522; Seadus napside ja likööride valmistamise ja müügi kohta, 03.09.1920. – Riigi Teataja 1920, nr 145 – 146, lk 1153 – 1154.

tubakavabrikutelt, paberossitehastelt, tubakaladudelt ja tubakakauplustelt; tollimaksu välismaalt sissetoodava tubaka ja tubakasaaduste pealt.“[57]

Aktsiisimaks moodustas 5 – 40% müügihinnast, tollimaks töötlemata importtubakalt 800 marka puudast[58] ning töödeldud importtubakalt 320 – 800 marka naelast[59] ja patendimaks 100–500 marka aastas.[60]

1921. aastal oli tubakatööstusettevõtete põhikapitali kogusumma 38 miljonit marka, 1922. aastal aga juba 80 miljonit marka,[61] millest osa koosnes tõenäoliselt pangalaenudest. Tubaka valmissaaduste sissevedu Eestisse vähenes järjekindlalt aasta–aastalt, töötlemata tubaka import aga kasvas, andes tunnistust Eesti tubakatööstuse arengust. Kui 1920. aastal veeti sisse 1411 puuda, siis 1923. aastal vaid 135 puuda valmissaadusi. Töötlemata tubakat veeti sisse 1920. aastal 1748 puuda, kuid 1923. aastal juba 50700 puuda.[62] Sissetoodud töötlemata tubakast ¾ pärines Kreekast, Bulgaariast ja Türgist (nn orienttubakas), ülejäänud ¼ aga USA-st (Virginiast), NSV Liidust, Hiinast ja Hollandist.[63]

Aastal 1926 kaaluti Eesti poliitilistes ringkondades riikliku monopoli kehtestamist teatud kaupadele, sealhulgas tubakale, tuletikkudele ja õllele. Kehtinud aktsiisid kavatseti kaotada ja asendada osalise müügimonopoliga. Sel moel loodeti suurendada riigieelarve tulusid 80 miljoni marga võrra. Ka rahaminister pooldas monopolide kehtestamist ning rõhutas, et seeläbi paraneb toodete kvaliteet ja suureneb läbimüük. Monopoliseaduse eelnõu koostamisel arvestati ka hindade ja aktsiisimäärade kavaga.[64] 1930. aastal kaaluti seda taas, kuid majandusministeerium lükkas riikliku monopoli tagasi, kaaludes teisi variante.[65]

1930. aastate majanduskriisist oli mõjutatud ka Eesti tubakatööstus. Aastatel 1932 – 1935 oli Eesti tubakatoodang languses, kuid 1936. aastaks oli toodang uuesti kasvavas trendis. 1931. aastal oli paberosside toodang 1,12 miljardit, kuid 1932. aastal oli paberosside toodang kukkunud 643 miljonini, kuhu kanti see ka mõneks ajaks jäi. 1938. aastaks oli tootmine taas tõusnud peaaegu miljardi paberossini (kokku toodeti 951 miljonit paberossi).[66]

---

[57] Tubakamaksu seadus, 10.04.1920. – Riigi Teataja 1920, nr 65/66, lk 517 – 522.
[58] üks puud = 16,38 kg
[59] üks nael = 409,512 g
[60] Kannel, “Tubakatööstus Eesti Vabariigis 1920–1940,” lk 12–13.
[61] Samas, lk 13.
[62] J. Michelson, “Tubakatööstus,” *Eesti: maa. Rahvas. Kultuur,* toim H. Kruus (Tartu, 1926), lk 605.
[63] Kannel, “Tubakatööstus Eesti Vabariigis 1920–1940,” lk 14.
[64] Juurmaa, “Eesti tubakatööstus 1919–1940 tubakavabrik ‘Laferme’ näitel,” lk 22.
[65] Samas, lk 23 – 24.
[66] Samas, lk 17.

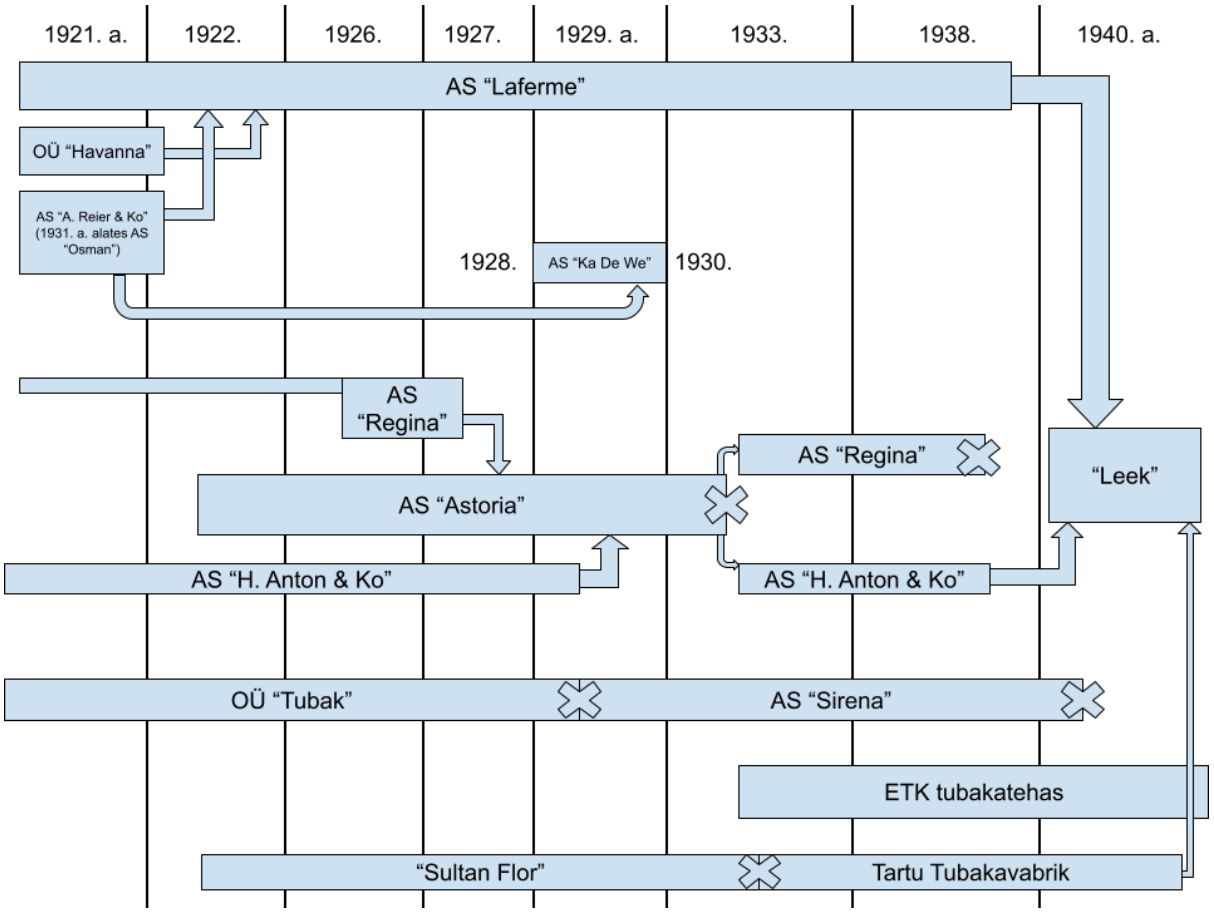
1931. aasta Postimehes kirjutatakse, et sel ajal tubakasaaduste väljavedu praktiliselt puudus ja kogu toodang läks Eesti turu varustamiseks. Autori sõnul 1920. aastate lõpus tarvitati aastas keskmiselt ligi 1 miljardit paberossi (ulatudes 1.6 miljardini 1929. aastal)[67] ning 250 000 kg pakktubakat.[68] 1932. aastaks langes tarvitamine miljardilt 900 miljoni paberossile aastas.[69] Lisad 1 – 3 demonstreerivad tubakaettevõtetega seotud majandustegevust aastatel 1919 – 1938.

1930. aastal olid Eestis tegevad veel kaheksa tubaka tootmise ettevõtet.[70] Suurettevõteteks loeti 20 või enama töötajaga ettevõtteid. 1930. aastate alguses hakkas ettevõtete arv taas kasvama, kuid stabiliseerus 1930. aastate lõpuks nelja suurema juures.[71] Mõned tehased üritasid turule naasta, kuid enamasti ei suudetud konkurentsis püsida.[72]

1940. aastal alanud Nõukogude okupatsioon tõi kaasa märkimisväärseid muutusi nii tubakatööstuses kui ka majanduse korralduses tervikuna. Vabaturu reklaamivabadus ja konkurents lõppesid.[73] Natsionaliseerimise protsessi käigus integreeriti kõik väiksemad tubakaettevõtted suurettevõtete Laferme ja ETK tehase koosseisu, mille käigus toimus ka juhtkonna vahetus.[74]

[67] Kaja, 06.03.1935, lk 4.
[68] Postimees, 29.08.1931, lk 5.
[69] Kaja, 06.03.1935, lk 4.
[70] "Tubakawabrikute arw oli kõige suurem 1921. a. – 13, mille arv langes hilisematel aastatel. 1930. a töötasid 8 wabrikut, sellest 2 wabrikut tööliste arwuga kuni 50 ja 6 varbikut üle 500 töölisega." – Samas, lk 5.
[71] Juurmaa, "Eesti tubakatööstus 1919–1940 tubakavabrik 'Laferme' näitel," lk 17. Tabel 1.
[72] Samas," lk 16 – 17.
[73] Juurmaa, "Eesti tubakatööstus 1919–1940 tubakavabrik 'Laferme' näitel," lk 10 – 11.
[74] Samas, lk 54 – 55.

Joonis 1. Suuremate tubakaettevõtete elukäik aastatel 1921 – 1940.*[75]



* Nool tähistab ettevõtete ühinemist või ettevõtte äraostmist selle ettevõtte poolt, kuhu nool näitab. Rist tähistab pankrotistumist või enampakkumisele minemist. Samal joonel asuvad ettevõtted tegutsesid samades ruumides. OÜ “Tubak” jäi kaubamärgina alles.[76] Samamoodi “Havanna” ja “A. Reier & Ko”.[77]

[75] Allikad: Juurmaa, “Eesti tubakatööstus 1919–1940 tubakavabrik ‘Laferme’ näitel,” lk 14 – 16, 25 – 29, 36, 46, 54 – 55; Kannel, “Tubakatööstus Eesti Vabariigis 1920–1940,” 11, lk 16 – 17.; Michelson, “Tubakatööstus,” lk 605 – 606; Kodumaa Saaduste Propaganda Keskkorraldus, *Eesti tööstus ja kaubandus.* (Tallinn, 1934), lk 98; Päevaleht 20.01.1940, lk 6; Waba Maa 06.12.1922, lk 1; Postimees 14.12.1933, lk 5; Tallinna Teataja 28.05.1921, lk 2; Lõuna–Eesti 25.11.1922, lk 4; Päevaleht, 23.10.1928, lk 5. Esmaspäev 02.06.1930, lk 1; Postimees 24.11.1934, lk 1.

[76] Juurmaa, “Eesti tubakatööstus 1919–1940 tubakavabrik ‘Laferme’ näitel,” lk 14 – 15.

[77] Kodumaa Saaduste Propaganda Keskkorraldus, *Eesti tööstus ja kaubandus*, lk 98.

1.3. Reklaamipraktika üldine areng ajavahemikul 1920 – 1940

Tubakatootmine koondus aastatel 1920 – 1940 järk-järgult suurettevõtetesse. Suurematel tubakatootjatel tekkis järjest selgem vajadus oma kaubamärke laiaulatuslikult ja süstemaatiliselt tutvustada, sest suure tootmismahu jätkusuutlikkus eeldas stabiilset ja laiapõhjalist turunõudlust.[78] Sel ajajärgul kujunesid välja üha professionaalsemad reklaamipraktikad, kus kasutati kutsuvaid kaubamärkide nimetusi, oma kunstnikke ning trükipressi võimalusi ja ajalehtede laia kandepinnaga kuulutuste veerge.[79] Tubakatoodete reklaamidel ajakirjades ja -lehtedes olid teatud eelised teiste kanalite ees: suur šrift, visuaalsed ja ikonograafilised kuulutused, mis andsid võimaluse püüda tähelepanu ning suhelda laiema publikuga.[80] 1930. aastatel said reklaamid tooteturunduse tähtsaks osaks[81] ning sel ajal rajati ka esimesed reklaamiagentuurid.[82]

1920. aastatel olid ettevõtjad reklaami kujundamise osas veel võrdlemisi ükskõiksed. Suust suhu reklaami peeti vaat et kõige väärtuslikumaks, ent 1930. aastateks oli see hoiak muutumas ja üha enam hakati tähelepanu pöörama erinevate reklaamikanalite sihipärasele rakendamisele.[83] 1920. aastatel oli ka graafilise disaini pool nõrk, kuid 1930. aastatel läksid reklaamid palju loomingulisemaks ja kunstilisemaks.[84] 1920. aastate alguseks olid reklaamid veel väga tekstikesksed, kuid graafiliste kunstnike tulekuga hakati tarbija tähelepanu püüdmiseks kasutama pilte.[85]

Ajakirjad ja ajalehed andsid hõlpsaid võimalusi kommertskunstnikele,[86] millest kasvasid välja põhilised tubakareklaamide autorid, keda Merle Talvik kategoriseerib nn autodidaktideks. Rohked tellimused kujundusgraafika alal kutsusid inimesi, kelle haridus või erialane kogemus ei olnud piisav, originaalseid lahendusi pakkuma. Reklaamgraafika on enamjaolt allkirjata, seega autorid jäid suures osas anonüümseks ning olid tihtipeale ka reklaami- või tubakaettevõtete varjatud töötajad.[87]

---

[78] vt Lisa 3.
[79] Talvik, "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia," lk 86.
[80] Samas, lk 29.
[81] Samas, lk 29.
[82] Karin Paulus, *Eesti disaini ja reklaami 100 aastat* (Tallinn: Post Factum, 2018), lk 55.
[83] Samas, lk 55.
[84] Merle Talvik, "Eesti kunstnikud ajakirjandusgraafikas 1930. aastail," *Mäetagused* 33 (2006): lk 18.
[85] Tiina Kaukvere, "Linnamuuseum avas vana reklaami näituse," *Postimees*, 12.04.2014
[86] Talvik, "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia," lk 100.
[87] Talvik, "Eesti kunstnikud ajakirjandusgraafikas 1930. aastail," lk 29.

Sageli iseloomustas selle ajastu reklaame välismaiste eeskujude järeleaimamine, väga tõenäoliselt soovisidki töösturid näha just läänelikke kujundusi.[88] Levinud motiiviks toodete reklaamimisel olid näiteks erinevad naisfiguurid.[89] 1930. aastatel kujutati reklaamides tihti tarbijat ja tootepakendit ning kompositsioonid olid detailiderikkad.[90] Meesfiguure kujutati tüüpiliselt rahulolevana ja suitsetava Hollywoodi staarina.[91] Seda, kuidas lugejad reklaame tegelikult vastu võtsid ja tõlgendasid, me otseselt ei tea.

Tuntumad kunstnikud, kes tegelesid just tubakareklaamidega, olid Paul-Aleksander Pedersen, August Vahtel (Laferme),[92] Karl Vanaveski, Aleksander Laar (ETK) ja Kaarel Joon (1937 aastani Jürgens, A. Reier ja Ko, Regina, Astoria ja Laferme).[93] Lisaks tegelesid selles valdkonnas veel Axel Bernhard Rossman, Günther Reindorff, Paul Luhtein, Johan Naha, Peet Aren ja Jaan Vahtra.[94]

Reklaamikunstnikke on lähemalt uurinud Merle Talvik, käsitledes nende ajakirjades ilmunud autoritöid, kuid nende ajalehtedes ilmunud tööd on kõrvale jäetud, hinnatud on nende loomingut eelkõige kunstilisest aspektist.[95] Siiski on ajalehereklaamidel sotsiaalne, poliitiline ja majanduslik tähendus.[96]

Nimekaimad tarbegraafikud olid August Vahtel, Valter Kõrver ja Karl Vanaveski.[97] Kaarel Joon (1937 aastani Jürgens; 1892 – 1981) oli spetsiaalselt AS "A. Reier & Ko" ning AS "Regina" ettevõtte kunstnik ning tarbegraafik 1922 – 33 aastail. Merle Talvik toetudes Oskar Raunami käsikirjadele, on väitnud, et Jürgens töötas 1933 – 37 aastail tubakaettevõttes AS "Astoria",[98] kuid 1932. a finantspettuse skandaali tõttu pankrotistus "Astoria" 1933. aastal ning selle likvideerimisega alustati ametlikult 1935. aastal.[99] Samuti leidub "Astoria" reklaame tema nimega 1925. aastast.[100] Samuti mainib Raunam, et peale "Astorias" töötamist töötas Jürgens (Joon) AS "Laferme" 1937 – 44.[101] Tubakaettevõttel AS "Astoria" oli ka enda

[88] Talvik, "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia," lk 81.
[89] Samas, lk 122.
[90] Samas, lk 64.
[91] Samas, lk 92.
[92] Talvik, "Eesti kunstnikud ajakirjandusgraafikas 1930. aastail," lk 30 – 31.
[93] Talvik, "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia,", lk 183, 198, 201.
[94] Paulus, *Eesti disaini ja reklaami 100 aastat*, lk 85.
[95] Talvik, "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia," lk 14
[96] Talvik, "Eesti kunstnikud ajakirjandusgraafikas 1930. aastail," lk 7.
[97] Talvik, "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia," lk 92.
[98] Samas, lk 198.
[99] Juurmaa, "Eesti tubakatööstus 1919–1940 tubakavabrik 'Laferme' näitel," lk 15 – 16.
[100] Päevaleht, 01.11.1925, lk 3; Postimees, 06.11.1925, lk 3.
[101] Talvik, "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia," lk 198.

reklaamiagentuur, mis pakkus teistele ettevõtetele etikettide ja pakendite kavandeid.[102] (Vt lisa 4, kus on Karl Jürgensi poolt loodud reklaam AS "Astoria" paberossidele "Kabaree").

Valter Kõrver (1904 – 1941) tegutses reklaamigraafikuna AS "Laferme" juures. Paul-Aleksander Pedersen (sünd. 1906) töötas 1934 – 44 reklaamikunstnikuna AS "Laferme" ettevõttes, kavandades ka tubakapakendeid. August Vahtel (1911 – 1943) töötas koos Paul Pederseniga AS "Laferme" juures. Karl Vanaveski (1909 – 1973) tegi ETK'le tubakareklaame.[103]

1929. aastal loodud[104] reklaamijate organisatsioon Eesti Reklaam-Klubi juhtivate liikmete sekka kuulus ka Laferme juhatuse liige Adam Grünbaum, mis näitas Laferme tahtmist olla sel alal aktiivne ning esirinnas sammuda. Lafermel oli ka omaenda reklaamiosakond, kelle juhiks oli Grünbaum. Laferme tegi koostööd kunstnikega Paul Aleksander Pedersen, August Vahtel, Valter Kõrver, kellest Pedersen kuulus Eesti Reklaami Agentuuri,[105] mis oli organisatsiooni asemel eraldi reklaamijatest koosnev ettevõte.[106]

Konkureeriv turundus kadus koos Nõukogude okupatsiooniga 1940. aastal. Kui Eestis pärast 1940. aastat eraettevõtete reklaam lõppes, siis Lääne turgudel jätkus tubaka agressiivne reklaamimine.[107]

---

[102] Talvik, "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia," lk 198.
[103] Samas, lk 199 – 201.
[104] Eesti Reklaam-klubi, *Eesti Reklaam-klubi põhikiri: asutatud 24. veebr. 1929. a.* (Tallinn: Eesti Reklaam-klubi, 1937).
[105] Paulus, *Eesti disaini ja reklaami 100 aastat,* lk 55 – 57.
[106] Eesti Reklaami Agentuur, *ERA Reklaam kataloog käsiraamat* (Tallinn: ERA, 1937), lk 3.
[107] Hussein A. Samji and Robert K. Jackler, "'Not one single case of throat irritation': misuse of the image of the otolaryngologist in cigarette advertising," The Laryngoscope 118, no. 3 (March 2008): lk 415 – 27; Stanford Research into the Impact of Tobacco Advertising, https://tobacco.stanford.edu/ (viidatud 14.05.2025).

2. Ajalehtedel põhineva baasandmestiku moodustamine

Eesti tubakatehased reklaamisid oma tooteid erinevaid kanaleid pidi (ajalehereklaamid, raadio, plakatid, neoonvalgustusseaded, tootepaviljonid näitustel),[108] kuid oluline roll kuulus siiski ajalehereklaamile. Ajalehed kujutavad endast kõige ulatuslikumat ja ajaliselt järjepidevamat säilinud allikat ning tänu regulaarsele ilmumissagedusele sobivad need kvantitatiivseks analüüsiks. Digitaalarhiivide olemasolu võimaldab ajalehti käsitleda masinloetaval kujul ning rakendada tekstikaeve meetodeid (st arvutipõhist tekstianalüüsi) tubakareklaamide uurimiseks. Andmeanalüüsi aluseks on valitud sel ajaperioodil ilmunud viie suurema päevalehe kuulutusveerud, tulenevalt nende regulaarsest ilmumisest uuritaval ajavahemikul: Päevaleht, Postimees, Sakala, Kaja (1919 kuni 17.09.1935) ja selle järglane Uus Eesti (alates 18.09.1935).[109] Päevalehtedest loodud tekstikorpus ei esinda kõiki digiteeritud ajalehti aastatel 1920 – 1940, sest DIGARi keskkonnas on segmenteeritud vaid ligikaudu pooled olemasolevatest ajalehtedest (s.t. et tekst lehekülgedel on jaotatud täpsemateks üksusteks, näiteks artikliteks, reklaamideks ja reklaamisektsioonideks.[110] Segmenteerimine käigus on täisteksti külge lisatud ka omadused ehk metaandmed (millisele leheküljele see tekst kuulub, kas tekst on pärit artiklist või kuulutusest, mitu sõna on segmendis kokku). Reklaamid valisin käesoleva töö peamiseks analüüsi objektiks, kuna lähtusin eesmärgist vaadelda andmeid eelkõige ettevõtete põhiselt. Siin töös kasutan mõisteid kuulutus ja reklaam läbivalt samatähenduslikena, viidates mõlemaga ajalehtedes ilmunud tubakatoodete reklaamtekstidele.

Siinse töö metoodika piiritlesin populaarsemate digihumanitaaria ja andmeteaduse võtetega: loomuliku keele töötlus läbi R programmeerimiskeele arenduskeskkonna[111] ning korpuslingvistika meetodid nagu sagedusanalüüs ja kollokatsioonianalüüs. Selles peatükis kirjeldan töös kasutatud allikate andmeid, meetodeid, nendega kaasnenud probleeme ning võimalikke alternatiivseid lahendusi. Tubakareklaamidest koosnev korpus on tehtud iteratsioonides, kuid siin kirjeldan seda lineaarselt. Igast etapist selgusid õppetunnid andmete puhastamise ja analüüsimeetodite kohta. Terve töövoog ehk lähtekood ning arendusprotsessi

[108] Paulus, *Eesti disaini ja reklaami 100 aastat,* lk 56; Eesti Reklaami Agentuur, *ERA Reklaam kataloog käsiraamat*, lk 5 – 8; Päevaleht, 28.06.1931, lk 1.
[109] Lauk, *Peatükke Eesti ajakirjanduse ajaloost*, lk 14 – 15.
[110] Peeter Tinits, "Digiteeritud Eesti ajalehed uurimisallikana," *Acta Historica Tallinnensia* (ilmumas).
[111] RStudio kasutusjuhend.

erinevad iteratiivsed etapid (tööversioonid) on üles laetud autori Github-i repositooriumisse.[112]

2.1. DIGAR ja Rahvusraamatukogu digilabor

Uurimistöö andmestiku allikaks on Eesti Rahvusraamatukogu digitaalarhiiv DIGAR, täpsemalt selle alamosa DIGARi Eesti artiklite portaal (DEA) ning analüüsi läbiviimiseks kasutatav digilabori keskkond ajalehetekstidele ligipääsuks. Ajalehtede digiteerimine annab ajaloolastele innovaatilised tööriistad tekstide analüüsimiseks nii mahuliselt kui ka metoodiliselt, võimaldades uurida kultuuriloolisi nähtusi uuel tasemel.[113] Oma töös kasutan digihumanitaaria lähenemist, mis annab võimaluse mahukaks kvantitatiivseks analüüsiks töödeldes digitaalseid materjale masinloetavale kujule ja rakendan automaatse tekstianalüüsi põhimõtteid.

DIGAR on Rahvusraamatukogu digitaalarhiiv ja kasutajakeskkond, mis pakub juurdepääsu mitmesugustele digiteeritud ja digitaalselt sündinud väljaannetele, sealhulgas e-raamatutele, ajalehtedele, ajakirjadele, kaartidele ja pildimaterjalile.[114] DIGARi Eesti artiklite portaal (DEA) võimaldab juurdepääsu kõigile läbi aegade Eestis või välismaal ilmunud eestikeelsetele ajalehtedele ning osale ajakirjadest, jagades kollektsiooni ajaloolisteks perioodideks, mille kättesaadavus ja otsimisvõimalused on erinevad. Kõige varasem ajalooline periood ehk 1811 – 1944 aastad on osaliselt kättesaadavad, selle perioodi väljaandeid lisatakse aja jooksul nimetuste haaval.[115]

DIGARi põhiliides (veebileht digar.ee) pakub liht- ja detailotsingut, mis toetab märksõnu, fraase ja lihtsamaid loogikaoperaatoreid (AND, OR, NOT) ning sõnatüvede otsingut (*), kuid keerukamate tekstinähtuste ja -mustrite otsimiseks ja analüüsiks on kasutajaliides piiratud. Täieliku regulaaravaldiste süntaksi kasutamise võimalust tavakasutaja otsingus DIGARi otsifunktsioonis ei ole.[116] Selliste keerukamate päringute teostamiseks ja digiteeritud sisu

[112] Lust, "Tubakakorpus," https://github.com/tormil/tubakakorpus/
[113] Wevers, "Consuming America," lk 44 – 45.
[114] "DIGARist," DIGAR – Eesti Rahvusraamatukogu digiarhiiv, https://www.digar.ee/arhiiv/et/info/digarist (viidatud 14.05.2025).
[115] "DIGARi Eesti artiklid," DIGARi Eesti artiklite portaal, Eesti Rahvusraamatukogu, https://dea.digar.ee/?a=p&p=about (viidatud 14.05.2025).
[116] Eesti Rahvusraamatukogu, DIGARi Eesti artiklite portaal, "Abi," https://dea.digar.ee/?a=p&p=help (viidatud 14.05.2025).

programmiliseks töötlemiseks on vaja spetsiifilisemaid vahendeid, mida pakub Rahvusraamatukogu digilabori keskkond.

Andmete digiteerimine võimaldab tekstilist materjali töödelda masinloetaval kujul, avades võimalusi arvutuslikuks uurimiseks, analüüsiks ja visualiseerimiseks, mida nimetatakse ka tekstikaeveks.[117] Sellise andmepõhise uurimise edendamiseks on loodud Rahvusraamatukogu poolt "digilabor",[118] mille eesmärk on muuta andmed digitaalselt paremini kättesaadavaks ja kasutatavamaks. Digilabor ühendab kultuuripärandi andmestikud (sh DIGARist ja DEAst pärinevad materjalid) infotehnoloogiliste vahenditega, pakkudes tööriistu andmestike töötlemiseks ja visualiseerimiseks ning suunates tegevusi humanitaar- ja sotsiaalteadlastele, andmeteadlastele ning üliõpilastele.[119] Materjalidele digilaboris ligipääsuks ja tekstide töötlemiseks kasutatakse näiteks Rahvusraamatukogu poolt JupyterLab'il põhinevat lahendust, mis pakub veebipõhist arenduskeskkonda, rakendades programmeerimiskeeli nagu R või Python.[120] Käesolevas töös kasutatakse RStudiot, mis on arenduskeskkond programmeerimiskeelele R.[121]

Üks võimas abivahend digitaalse tekstianalüüsi vallas, mis võimaldab efektiivselt töödelda suuri tekstikorpuseid ja leida keerukaid mustreid, on regulaaravaldised. Regulaaravaldised (inglise keeles *regular expressions* ehk *regex*) kujutavad endast spetsiaalset süntaksit tekstist mustrite otsimiseks ja kirjeldamiseks. Nendega pannakse kirja oodatud sümbolid, nende järjekord, variandid (kas üks või teine) ja kordused. Need võimaldavad kirja panna mitmekesiseid päringuid ja teisendusi, määrates reeglid selle kohta, millised tähed või nende grupid peavad esinema, millises järjekorras, kordustes või alternatiivides. Võrreldes lihtotsinguga on regulaaravaldised tunduvalt paindlikumad ja võimsamad keerukate või varieeruvate mustrite (nt kuupäevade, spetsiifiliste struktuuridega fraaside) leidmiseks, mida lihtsalt sõnadega kirjeldada ei saa. Regulaaravaldiste spetsiifiline ning keerukas süntaks kasutab palju erimärke ehk metamärke, millel on mustri kirjeldamisel eriline tähendus. Regulaaravaldisi toetavad paljud programmeerimiskeeled ja tööriistad ning need on sageli kergesti ülekantavad ühest keskkonnast teise. Regulaaravaldised on peamiselt kasutusel tekstiga töö tegemisel, failide otsimisel, töötlemisel ja eraldamisel. Tänapäeval toetavad

[117] Wevers, "Consuming America," lk 52.
[118] Digilab – RaRa, https://digilab.rara.ee/ (viidatud 14.05.2025).
[119] Digilab – RaRa, "Meist," https://digilab.rara.ee/meist/ (viidatud 14.05.2025).
[120] Peeter Tinits, "Estonian National Library Overviews," OSF-projekt, https://doi.org/10.17605/OSF.IO/3GZXE (viidatud 14.05.2025).
[121] RStudio Team, (GitHub-repositoorium), https://github.com/rstudio/rstudio (viidatud 14.05.2025).

regulaaravaldisi mõningatel viisidel ka tekstiredaktorid, Excel, Google Sheets jne.[122] Regulaaravaldised võimaldavad suurendada päringute katvust digiteeritud tekstides kuna nende abil saab arvesse võtta tekstituvastusest (OCR) tingitud sagedasi vigaseid vorme või õigekirjavariante,[123] näiteks kui AS A. Reier & Ko asemel on hoopis "n/5. B. Reter & Ko",[124] saab teostada päringu "Re?er & Ko", mis võtab arvesse i-tähe õigekirja variatsioone.

Tekstituvastus (OCR ehk inglise keeles Optical Character Recognition) ehk automaatne tekstituvastus on protsess, mille käigus analüüsitakse digiteeritud dokumente spetsiaalse tarkvara abil eesmärgiga tuvastada neis sisalduvad tekstialad, samuti pildid ja tabelid. Kuigi automaatne tekstituvastus (OCR) võimaldab digiteeritud materjalides teostada täistekstiotsinguid, tuleb selle tulemuste puhul arvestada olemusliku ebatäpsusega. Selle täpsuse aste sõltub mitmest asjaolust, sealhulgas väljaande trükikvaliteedist, skaneerimisel kasutatud mikrofilmi või skanneri kvaliteedist, originaaldokumendi seisukorrast (nt paberi kvaliteet, kahjustused), kujunduselementidest (nt väike trükk, erinevad fondid, keerukas veergude paigutus) ning kasutatavast tekstituvastustarkvarast.[125] Võimalik puudus on ka asjaolu, et 100% õige tekstituvastusega (OCR) jäävad välja tubareklaamid, kus ei esine valimisse kaasatud märksõnu.

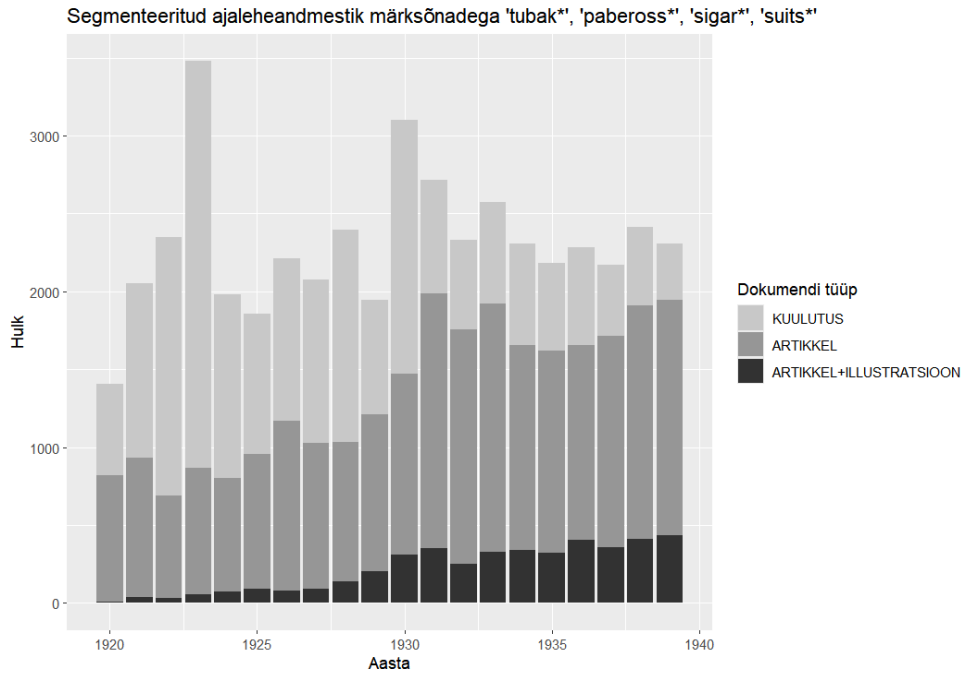
2.2. Reklaamkorpuse loomine

Esimese sammuna kasutasin digilaborist pärit JupyterHub'il põhinevat lahendust, et DIGARi toorandmetele ligi pääseda. Seejärel tutvusin ajalehtede toorandmestikuga ja lõin esimese alaandmestiku sorteerides välja 1920. – 1930. aastate kõik digiteeritud ajalehed. Kasutades digilaborist pärit JupyterHub'il põhinevat lahendust, teostasin regulaaravaldiste otsingu põhimärksõnadega "tubak", "pabeross", "sigar" ning "suits". Sõnavara laiema haarde saavutaseks katsin regulaaravaldistega lisaks põhiterminitele ka nende liitsõnad ja erinevad käändevormid, hõlmates näiteks termineid "suitsetaja" või "sigarett". Põhimärksõnadega jäin võimalikult laia mustriotsingu juurde kuna DIGARi tekstituvastuse ebatäpsusest tingituna leidus reklaamide tekstides vigu, eriti 1930. aastate kunstnike poolt loodud graafiliselt kaunistatud reklaamides. Alustasin korpuse loomist kõige algelisemast regulaaravaldisega otsingust, et hõlmata kõikvõimalikud tubakaga seotud kuulutused ühe katuse alla ehk

[122] Jaak Vilo, loeng „Tekst, infootsingud, masintõlge – ekskurss" (kursus Digitaalne maailmapilt, LTAT.00.020), 7. loeng, Tartu Ülikool, kevadsemester 2025.
[123] Wevers, "Consuming America," lk 85.
[124] Kaja, 08.11.1925, lk 5. OCR-tekst kättesaadav: https://dea.digar.ee/article/kaja/1925/11/08/1/54.1 (viidatud 15.05.2025.)
[125] Eesti Rahvusraamatukogu, DIGARi Eesti artiklite portaal, "Abi."

esialgsesse andmestikku. Laiemat andmestikku on alati võimalik hiljem täpsemate kriteeriumite alusel kitsendada.

Joonis 2. Regulaaravaldisega filtreeritud esialgne ajaleheandmestik aastatel 1920 – 1940.



Kõigepealt vaatasin DIGARi andmete põhimärksõnade mainimist läbi kõikide digiteeritud ja segmenteeritud ajalehetekstides (sh artiklid ja kuulutused), hiljem piirasin andmestiku ainult reklaamide ehk kuulutustega kuna antud töö uurimisobjektiks on tubakareklaamid. Välja toodud joonis hõlmab nii artikleid kui ka kuulutusi, et anda ülevaade DIGARi digiteeritud ajalehtede andmestiku ulatusest. Joonisel nimetatud nelja põhimärksõna on mainitud segmenteeritud ajalehtedes kokku 20 aasta jooksul 46 172 artiklis, illustratsioonis ja reklaamis. Siit graafikust on näha, et märksõnade kasutus oli üllatavalt stabiilne, kuid on märgata artiklite osakaalu kasvu. Reklaamide arv olid pärast 1920. aastate keskpaika langustrendis.[126] Graafiku puhul tuleb arvestada, et see ei esinda kõiki olemasolevaid ajalehti aastatel 1920 – 1940, sest segmenteeritud (s.t. ajalehtede tekst on jaotatud sisuliselt kooskõlalisteks üksusteks nagu artikliteks ja kuulutusteks) on vaid ligikaudu pooled olemasolevatest ajalehtedest.[127] Kahjuks oli segmenteerimata andmestikuga võimatu eristada nn traditsiooniliste andmepuhastusmeetoditega artikleid ja reklaame ilma pildituvastusalgoritme kasutamata. Masinõppega loodud algoritmidega saaks segmenteerimist kindlasti parandada, aga selline uurimissuund jääb tulevastele uurijatele.

[126] vt Joonis 4.
[127] Peeter Tinits, "Digiteeritud Eesti ajalehed uurimisallikana," (ilmumas).

Peale seda ühildasin ajalehetekstid metaandmetega toetudes digilabori "Ligipääs DEA tekstidele" juhendile[128] ning lõin filtri märksõna sõnale "suits". Filter osutus vajalikuks, kuna vaja oli lahti saada valeandmetest, nagu näiteks suitsukala ja suitsuvorsti kuulutused. Seega seadsin filtrile tingimuse, et tulemused sisaldaksid sõna "suits" vaid juhul, kui see esineb koos sõnadega "sigar", "pabeross" või "tubak". Teistel märksõnadel eritingimusi ei olnud. Analüüsi käigus selgus, et otsisõnale „suits" leidus kokku ligikaudu 7000 vastet. Filtri eritingimuse seadmisel langes otsisõna "suits" arv 2400 vasteni.

Piirdusin vaid nelja põhimärksõnaga. Põhjuseks oli asjaolu, et otsingule täiendavate tubakatööstusega seonduvate märksõnade lisamine, nagu näiteks liitsõnad "suitsuleht", "suitsumees", "suitsumärk", "tubakavabrik", "tubakatoodang", "sigaretivabrik" ja "piip", andis märkimisväärselt vähe lisavasteid. Ligi 17 061 esialgsele vastele leidsin täiendavalt vaid 500 vastet. See tulemus viitab selgelt otsingu kahanevale tootlikusele, mistõttu loobusin laiemast märksõnavalikust. Seetõttu võibki piirdumine nelja põhilise regulaaravaldisega olla mõistlik kompromiss. Minu andmestikku sattus ka tubakareklaamide väliseid kuulutusi. Näiteks tolliameti avaliku enampakkumise kuulutuses leitud märksõna "Tubak lehtedes".[129] Samuti ka "tubakatööstusmasinaid".[130]

Esialgse andmestikuga proovisin lemmatiseerimata reklaamtekste analüüsida, kuid see ei õnnestunud eriti. Näiteks sõnaanalüüside tegemisel tekkis tulemustesse sõna "pabeross" eri käändevorme (nt "paberosside"), mis takistasid tõelähedase sõnasageduse graafiku koostamist kuna sellest tekib eri käändevormidest omaette alajaotus. Seega otsustasin töös kasutada andmete töötlusel lemmatiseerimist, mida pakub digilabor.[131] Seetõttu jääb siin töös üks loomuliku keele töötluse osa puutumata. Lemmatiseerimine on üks loomuliku keele töötluse oluline etapp, mis hõlmab endas sõna eri vormide taandamist algvormile ehk sõna standardiseeritakse.[132] See tähendab, et kui reklaamides on näiteks mainitud sõna "paberossid" mitmuses, siis lemmatiseerimisega taandatakse sõna oma algkujule "pabeross". Lemma on sõnastikus esinev sõnavorm ehk algvorm. Lemmatiseerimine on tekstianalüüsis vajalik, kuna see võimaldab lugeda ühe sõna erinevad morfoloogilised vormid kõik ühe

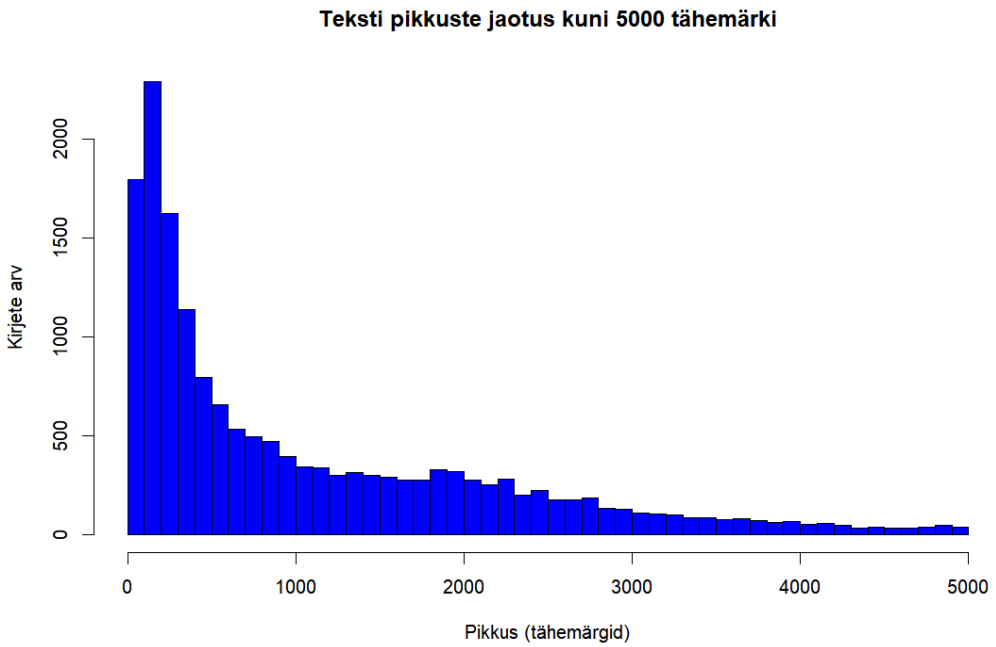
[128] Digilab – RaRa, "Ligipääs DEA tekstidele," https://digilab.rara.ee/tooriistad/ligipaas-dea-tekstidele/ (viidatud 14.05.2025).
[129] Postimees, 18.08.1927, lk 6.
[130] Päevaleht, 18.08.1927, lk 9.
[131] Funktsiooni kutsumisel (do_subset_search) määrasin käsurea parameetrina searchtype="lemmas" – "Ligipääs DEA tekstidele."
[132] Laura Katrin Leman, "Tehisnärvivõrgul põhinevate lemmatiseerijate võrdlev analüüs eesti keeles," bakalaureusetöö, juhendaja Kairit Sirts, Eesti ja üldkeeleteaduse instituut, Humanitaarteaduste ja kunstide valdkond, Tartu Ülikool, 2019, lk 5.

algvormi juurde kuuluvaks, vastupidiselt olukorrale, kus neid käsitletaks eraldiseisvate sõnadena. Tulenevalt digilabori pakutavast andmete lemmatiseerimise teenusest, kasutasin käesolevas töös juba eelnevalt lemmatiseeritud andmestikku. Seetõttu ei kuulunud spetsiifilise lemmatiseerimisprotsessi implementeerimine käesoleva töö ülesannete hulka.[133] Nagu ka digihumanitaaria, on loomuliku keele töötlus (LKT), inglise keeles Natural Language Processing (NLP), üheks arvutiteaduse, tehisintellekti ja arvutilingvistika teadusharuks, mis keskendub inimese ja arvuti vahelisele suhtlusele. Selle valdkonna peaaegu kõikides ülesannetes tuleb läbida mõned eeltöötluse etapid, et teisendada toortekst vormi, mis on mudeli ja masina jaoks loetav.[134]

2.3. Tekstituvastuse vead ja kuulutuste segmenteerimine

Järgmiseks oli peamiseks väljakutseks küsimus, kui palju on andmestikku sattunud selliseid sõnu, mis tegelikult tubakat ei puuduta. Selleks loendasin, kui palju oli eri suurusega tekste analüüsiks loodud tekstikogus, et aru saada, mitu sõna ühte digitaalsesse reklaamisegmenti keskmiselt kuulub.

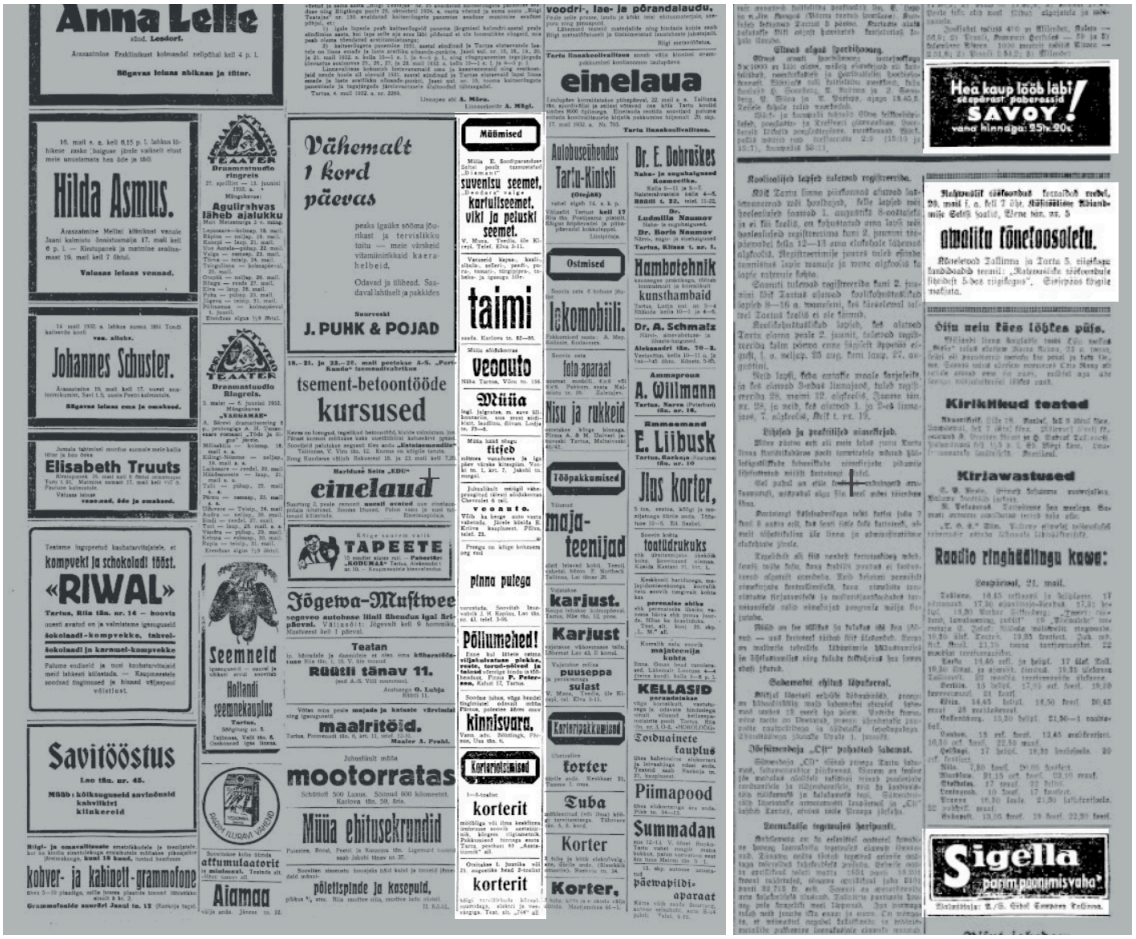
Joonis 3. Reklaamisegmendi teksti pikkused tähemärkide numbrite arvuna.



[133] Leman, "Tehisnärvivõrgul põhinevate lemmatiseerijate võrdlev analüüs eesti keeles," lk 5.
[134] Vladislav Šikirjavõi, "Temaatiliste mustrite kaevandamine seadustekstidest," magistritöö, juhendaja Ahti Lohk, Tallinna Tehnikaülikool, Infotehnoloogia teaduskond, Tarkvarateaduse instituut, 2018, lk 15.

Jooniselt on näha, et suur hulk reklaame olid lühikese tekstiga. Paljud reklaamidest (üle 1000 tähemärgi: 6875 vastet, alla 1000 tähemärgiga: 10186 vastet) ületas 1000 tähemärgi pikkuse, mis justkui tähendab, et need reklaamid peaksid olema keskmiselt 150 – 175 sõnalised. Üks probleem, mis tekkis andmestikuga töötamisel, oli kuulutuste segmenteerimise viis: kuulutused ehk reklaamid olid segmenteeritud veergudeks, kuhu võib kuuluda mitu reklaami.[135] Seega viisin läbi lähilugemise (ingl k *close reading*) analüüsi, mille eesmärk oli välja selgitada, mil määral reklaamtekstid segmendi piires kattuvad, ning otsida viise andmestiku filtreerimiseks, nii et see koosneks vaid tubakareklaamidest. Andmete lähilugemisel selgus, et DIGARi ajaleheandmestikus on segmenteeritud reklaamid veergudeks, kuhu võib kuuluda keskmiselt 1 – 15 reklaami.

Pilt 2 ja 3. Näited kuulutuste veergudeks segmenteerimisest (valgega on andmestikus kättesaadav tekst, mis on segmenteeritud "üheks" reklaamiks).[136]



[135] vt Pilt 2 ja 3, lk 24.
[136] Postimees nr 113, 18.05.1932; Postimees nr 115, 20.05.1932.

Kuulutuste veergudeks segmenteerimise probleemi minimeerimiseks osutus mõistlikuks lahenduseks konkordantsi kasutamine. Konkordants on tekstianalüüsis kasutatav töövahend, mida defineeritakse kui etteantud sõna või sõnavormi kõigi esinemiste loendit digitaalses korpuses koos seda ümbritseva kontekstiga, mida tuntakse ka nimetuse all "keyword(s) in context" (KWIC).[137] Konkordantse kasutatakse mitmes rakenduslingvistika valdkonnas, sealhulgas leksikograafias ja keeleõppes. Konkordantsi tööriista olulisus seisneb selle võimes pakkuda ülisuurt hulka näiteid autentsest keelest ning võimalust teostada otsingut erinevate parameetrite alusel, mis on eriti väärtuslik teksti süvaanalüüsil.[138] Seeläbi võimaldab konkordants saada põhjaliku ülevaate sõnade kasutusest ja nende kontekstist analüüsitavas andmestikus, toetades oluliselt näiteks kindlate märksõnade tähendusvarjundite või kasutusmustrite uurimist.

Siinses töös on konkordantsile antud tavapärasest erinev kasutus: selle abil kuvati iga reklaamisegmendi (ideaalis üks tubakareklaam) kohta vaid üks tulemus. See võimaldas analüüsida sõnu, mis esinevad koos etteantud märksõnaga kindlas aknas ehk määratud sõnade arvu või tähemärkide ulatuses selle ümber. Kuigi sõnade koosesinemine samas reklaamisegmendis ei pruugi alati otsest semantilist seost näidata, on nende lähedus tekstis hea indikaator vastava seose olemasolule.[139] Konkreetse akna suuruseks piirdusin 100 tähemärgiga põhimärksõnadest vasakule ja paremale, mis on kokku umbes 10 – 15 sõna.

Kindla aknasuuruse rakendamise põhjenduseks oli digiteerimise käigus ilmnenud optilise tekstituvastuse (OCR ehk inglise keeles Optical Character Recognition) vead. Kuna digiteerimise protsessis ei jagatud iga kuulutust eraldi segmendiks, vaid ühte segmenti võis kuuluda 1 kuni 15 reklaami, mis digitaalarhiivis talletati ühe digitaalse segmendina, anti sellele ka vaid üks ID-kood. See tähendab, et ühele digitaalsele segmendi ID-le vastav tekstisisu võib pärineda mitmest kuulutusest. Selliste liitsegmentide põhjustatud analüüsimisvigade vältimiseks oli vajalik analüüsida sõnade koosesinemist vaid märksõna läheduses. Seetõttu piirdusin käesolevas töös määratletud kontekstiakna kasutamisega (200 tähemärki) põhimärksõnade ümber. Erinevalt standardsetest konkordantsipõhistest analüüsidest ei pidanud ma rohkem kui ühte tulemust (nt "Tubakatehases toodetakse

[137] "Text Mining Types," University of Oslo Libraries, https://www.ub.uio.no/english/libraries/dsc/research-methods/text-mining/text-mining-types.html (viidatud 14.05.2025).
[138] Jelena Kallas, Maria Tuulik ja Madis Jürviste, "Leksikograafilise tarkvara Sketch Engine eesti keele moodul," *Eesti ja soome-ugri keeleteaduse ajakiri* 3, nr. 2 (2012): lk 58.
[139] Wevers, "Consuming America," lk 141.

paberosse ja sigarette, mida saab suitsetada.") reklaamides oluliseks, kuna eesmärk oli saada ühele reklaamile külge üks nn ID-kood.

Kindla aknasuuruse rakendamise kitsaskoht oli, et ulatuslikumad reklaamid[140] ei mahu tervikuna määratletud kontekstiaknasse, mistõttu jääb nende sisu osaliselt hõlmamata. Võrreldes reklaamveergude segmenteerimisest tulenevate probleemidega, pidasin seda puudust aksepteeritavaks.

Andmestikus tuli esmalt lahendada tekstituvastuse vigadest ja reklaamide digitaalsest segmenteerimisest tingitud probleemid, mida loomuliku keele töötluses kutsutakse korpuse eelpuhastuseks ehk eeltööks. Selgus, et üks digitaalne "reklaamisegment" võib sisaldada teksti mitmest füüsiliselt eraldiseisvast reklaamist.[141] See põhjustas olulist teksti kattumist ja raskendas tubakakuulutuste isoleerimist võimalikult tõelähedase sõnade analüüsi teostamiseks. Selle probleemi minimeerimiseks piirasin reklaamide sõnahulga kontekstiakna 200 tähemärgi laiuses alas (100 tähemärki vasakule ja 100 tähemärki paremale) põhimärksõnade ümber. Kontekstiakna loomine aitas vähendada tubakaga mitteseotud kuulutusi, nii näiteks ei tulnud sagedusanalüüside ega lähivaatlusega nähtavale kuulutusi sõnadega "korter", "laps", "noormees", mis viitavad sootuks korteri või kaaslase otsimise kuulutustele.

Baasandmestiku loomisel kujunes probleemiks DIGARi kuulutuste segmenteerimise viis, mis luges mitmed (1 – 15) reklaamid üheks, andes tulemustesse palju mitteseotud kuulutusi. Kontekstiakna kasutusele võtmine, mille piiritlesin 200 tähemärgiga põhimärksõna ümber, aitas oluliselt korpuse täpsust parandada. Iteratiivsete sagedusanalüüside ja lähivaatluse kasutamine vähendas peale kontekstiakna loomist mitteseotud kuulutuste esinemist. Korpuse koostamisel tuleb täiendada ning kontrollida seejärel iga etappi lähilugemisega. Korpuse loomist saab ka tõhustada tekstituvastuse kvaliteedi ja segmenteerimisviiside parandamisel. Ajalehtede digiteeritud andmestik loob väärtusliku, kuid täiendavat andmepuhastust vajava aluse Eesti tubakareklaamide uurimiseks.

[140] vt Pilt 4, lk 27.
[141] vt Pilt 2 ja 3, lk 24.

Pilt 4. Laferme tubakareklaam Päevalehes 1938. aastal.[142] Illustreerijaks Paul Aleksander Pedersen.[143] Reklaamis jääb 200 tähemärgi sisse kaks põhimärksõna.

[142] Päevaleht, 05.03.1938, lk 3.
[143] A–S. Laferme, *Tubaka ajalugu: piltides*, illustreerija Paul Aleksander Pedersen, [Tallinn]: Laferme, [193-?]; elektrooniline reproduktsioon Tartu: Eesti Kirjandusmuuseum, 2022, https://kivike.kirmus.ee/AR-22066-64996-20343 (viidatud 16.05.2025).

3. Tubakakorpuse analüüs

3.1. Töövoo tõhususe hindamine

Selles peatükis hindan rakendatud andmestiku koostamise meetodeid. Eelnevalt uurisin, kas märksõnapäring ja kontekstiakna väljavõte parandab andmestiku täpsust ja katvust. Saadud tulemused aitavad paremini mõista rakendatud korpuse loomise piiranguid. Järgneva hindamise käigus analüüsin lähilugemise abil valimi töövoo veaprotsenti, eristades korrektselt tuvastatud tubakareklaamid muudest andmestikku sattunud reklaamidest. Samuti analüüsin lähilugemise kaudu tekstituvastuse (OCR) võimekust digiteeritud materjali lugemisel. Uurin põhimärksõnade "tubak", "sigar", "pabeross" ja "suits" kaudu esinevaid sõnasagedusi, et saada ülevaade, mis sorti reklaamid andmestikus on.

Otsingute täiustamisel ja korpuse loomisel tuleb juba eos teadvustada, et ideaalset otsingut pole võimalik saavutada. Otsingu täiustamist piirab keele rikkalikus ja keerukus, andmevormide varieeruvus ja digiteerimisvigadest tulenev müra, mistõttu iga valik nihutab tasakaalu täpsuse ja katvuse vahel.[144] Erinevalt kvantitatiivses andmeteaduses kasutatavatest rangetest valimi kujundamise protokollidest, kus kontrollitakse kõiki väliseid muutujaid, tuginevad digihumanitaaria alased uurimused paratamatult sellele, mis on juhuslikult säilinud ja digiteeritud.[145] Seetõttu on allpool kirjeldatud otsingu täiustamise strateegiate fookuses lisaks leidude suurendamisele ka korpuse ebatäiuslikkuse ja katvuse hindamine.

Töövoo kohasuses ja tõhususes veendumiseks kasutasin meetodina digihumanitaaria ja keeletöötluse kontekstis lähilugemist. Keeletöötluse (NLP) ja digihumanitaaria kontekstis nimetatakse lähilugemiseks (ingl k *close reading*) sellist kvalitatiivset meetodit, mille käigus vaadeldakse väikeses mahus teksti sõna, vormi või lause kaupa, et mõista semantilisi, stilistilisi või pragmaatilisi detaile. Lähilugemine on laiem sisu kui pelgalt andmete kontrollfunktsioon. See võib hõlmata erinevaid eesmärke ja lähenemisviise, näiteks: (1) keeruliste tekstide interpreteerimist (nt G.W.F. Hegeli teoste analüüs); (2) teose autori sõnumi ja taotluse mõistmist (nt luules või tsenseeritud kirjutistes); (3) keskendumist tekstis

[144] Sarah Oberbichler ja Eva Pfanzelter, "Topic-specific corpus building: A step towards a representative newspaper corpus on the topic of return migration using text mining methods," *Journal of Digital History* 1, no. 1 (September 1, 2021): lk 74 – 98, https://doi.org/10.1515/jdh-2021-1003.
[145] Mikko Tolonen, Eetu Mäkelä, Jani Marjanen ja Tuuli Tahko, "Arvutiteaduse kaasamine humanitaarharidusse," *Methis : studia humaniora Estonica* 21, nr. 26 (2020): lk 38, https://doi.org/10.7592/methis.v21i26.16909.

leiduvatele mustritele ja detailidele; (4) autoritööde analüüsimist autori loomingu arengu või elulooliste seoste paremaks hindamiseks. Lähilugemine on fundamentaalne meetod eelkõige kirjandusteaduses.[146]

Lähilugemise üheks eesmärgiks on kvaliteedikontroll kaugloetud andmestikule – automaatse korpuseanalüüsi tulemusi valideeritakse lähilugemisega, kontrollides, kas masin tuvastas tegelikult soovitud nähtuse.[147] Lähilugemise eesmärk siinses töös on lugeda kokku mitu reklaami juhuslikult valitud kuudes 1920 – 1940 perioodil esineb ja siis võrrelda, kas see vastab tekstianalüüsi tulemustele. Täpsemalt uurisin, kui suur osa korpusest on päriselt tubakaettevõtete reklaamid ning kui suur osa on tubakaga seotud kuulutused, näiteks tolliameti avalikud enampakkumiste teated, tubakatööstusmasinate müügi kuulutused või kaupluste reklaamid, mis mainivad oma tootekataloogis tubakatooteid. Seega minu töövoo tõhususe hindamisel on lähilugemisel kontrollimise roll, millega saan teada, kui suur osa minu andmestikust on tubakaettevõtete reklaamid.

Uuritud perioodil oli lähilugemisel kasutatud 1932. aasta maikuu andmeid, mil sain tüüpiliseks reklaamlause pikkuseks keskmiselt 15 sõna ning see ulatus ligikaudu 90 tähemärgini (koos tühikutega), sisaldades tihti ka lisainfot nagu aadress või hind.[148] Märkimisväärselt populaarne, eriti 1920. aastate keskpaigast alates, oli värssreklaam[149] ehk reklaam, mis oli sõnastatud üsna pika luuletusena ning sageli olid luuletuse kõrval ka koomiksilaadsed pildiveerud, mille puhul võis ühe kuulutuse maht ulatuda 40 – 80[150] sõnani.[151]

Andmestiku esialgne lähilugemine andis esmase ülevaate tubakareklaamide sisust ja kvaliteedist. Kvalitatiivse tähelepanekuna ilmnes, et ühe kuu jooksul avaldasid tubakareklaame vaid paar ettevõtet. 1932. aasta maikuu lähilugemise väljavõte, mis kajastas andmestikku enne põhimärksõnade otsingu parandusi, sisaldas suurel hulgal tubakaettevõtete väliseid kuulutusi ja selle täpsus oli ligikaudu 0,43 ehk 43% (TP=12, FP=16). Selles valimis tuvastati tekstituvastusega (OCR) reklaamitekstidest ligikaudu 82% sõnadest.

---

[146] "Close Reading," University of Wisconsin–Madison Writing Center, https://writing.wisc.edu/handbook/closereading/ (viidatud 14.05.2025).
[147] C. Aurnhammer, I. Cuppen, I. van de Ven ja M. van Zaanen, "Manual Annotation of Unsupervised Models: Close and Distant Reading of Politics on Reddit," *Digital Humanities Quarterly* 13, nr. 3 (2019), https://www.digitalhumanities.org/dhq/vol/13/3/000431/000431.html (viidatud 16.05.2025).
[148] Pealinna Teataja, 24.12.1938, lk 6.
[149] vt Lisa 4.
[150] Postimees, 08.11.1925, lk 3.
[151] Kaukvere, "Linnamuuseum avas vana reklaami näituse," 12.04.2014.

Korrektselt tuvastatud tubakaettevõtete reklaamide täpsust (ingl k *precision*) arvutasin kui tõsipositiivsete (TP - korrektselt tuvastatud reklaamid ehk tubakaettevõtetele kuuluvad reklaamid) ja valepositiivsete (FP - valesti tuvastatud ehk tubakaettevõttele mittekuuluvad reklaamid) suhtena (P = TP / (TP + FP)).

Pärast andmetöötluse parandusi (otsisõna "suits" eritingimuse filter, ja regulaaravaldiste metamärkide kasutamine) läbi viidud lähilugemine suuremal, 113 reklaamist koosneval valimil (koosnedes 1925. aasta novembri, 1927. aasta augusti ja 1935. aasta veebruari reklaamidest), näitas täpsuse olulist kasvu. Erinevate kuude valimite täpsused jäid vahemikku 56% kuni 71%. Kogutäpsus parandatud valimis oli ligikaudu 63%. Seega andmestiku väljavõtte kvaliteet paranes, kuid märkimisväärne veaprotsent jäi siiski püsima ka pärast parandusi, mida tuleb edasiste analüüside teostamisel arvestada. Lähilugemisel tuvastasin, et andmete ebatäpsust tekitasid eelkõige kaupluste tootereklaamid, tubakatööstusmasinate müügi reklaamid ja tolliameti teated.

Tabel 1. Lähilugemise valimi suurus ja täpsus.

	1925. november	**1927. august**	**1935. veebruar**	**Kokku**
Tõsipositiivsed	43	34	13	71
Väärnegatiivsed	19	14	9	42
Koguarv	62	48	22	113
Täpsus	55.8%	70.83%	59.09%	62.83%

Lisaks lähilugemisele analüüsisin valitud põhimärksõnu läbi sõnasageduste, mille kaudu hindasin, kas tegemist on tubakareklaamidega. Uurisin, missugused sõnad esinevad põhimärksõna otsinguga leitud reklaamides. Tulemustest selgus, et enne kui lisasin põhimärksõnale "suits" täiendava filtri, tuli sõnasageduse kaudu esile ka "suitsuvorsti" ja "suitsukala" kuulutusi.

Tabel 2. Põhimärksõnade sõnasageduse otsingu tulemused (märksõnadega koos esinenud sõnade esinemissagedused kõige enim nimetatus vähimani, võttes välja stoppsõnad)

Märksõna "tubak", vasteid kokku 8894	**Märksõna "sigar", vasteid kokku 1217**	**Märksõna "pabeross", vasteid kokku 10515**	**Märksõna "suits", vasteid kokku 2407**
tubakas	sigaret	pabeross	pabeross

pabeross	sigar	sent	suitsetaja
tubak	pabeross	headus	suitsetama
hind	tubakas	tükk	tubakas
müüma	müüma	hind	sent
tartu	kell	parem	suits
sent	portsigar	mark	eesti
parem	sigarett	tubak	parem
müük	hõbe	paberossi	headus
vabrik	hind	sort	tükk
tallinn	kuld	müük	tubak
sort	tükk	eesti	müük
nael	tallinn	maitse	hind
suur	inglise	suitsetaja	sort
tubakavabrik	müük	tallinn	nõudma
mark	suhkur	suitsetama	maitse
headus	suur	tartu	suitsetamine
kaup	kaup	ilmuma	uudis
maitse	maitse	nõudma	laskma
kell	sigareti	suur	mark
andma	seep	kõrge	tartu
eesti	paluma	vabrik	tallinn
tükk	valmistama	uudis	suitsumees
suitsetaja	uudis	karp	mikaado
vene	suurem	kest	vabrik
ilmuma	tubak	laskma	kõrge
tubakatehas	konfiskeeritud	astoria	tundma
tubakakauplus	vein	laferme	tähelepanu
uudis	ostma	kell	nael
soovitama	valmistatud	tundma	astoria
suurem	konfiskeerima	vatt	kord

tubakasaadus	tooma	kaup	teadma
türgi	riie	aroom	puhastama
tubaka	parem	vene	sigaret
kuld	sigari	nael	valmistama
suitsetama	sort	müüma	kaup
nõudma	soovitama	tuntud	suitsetavad
valmistatud	headus	kuld	suitsetava
vabriku	kõlblik	valmistatud	karp

Põhimärksõna "tubak" otsinguga tulid esile eelkõige liitsõnad: "tubakavabrik", "tubakatehas", "tubakakauplus" ja "tubakasaadus". Sõnad "vene" ning "türgi" kirjeldasid arvatavasti riike, kust tubakas oli sisse toodud. Põhimärksõna "sigar" eripäraks olid "sigaretid", "portsigarid". Samas tähistab sõna "konfiskeeritud" tolliameti müügikuulutusi, mis on samuti andmestikku sattunud.

Põhimärksõna "pabeross" andis andmestikus kõige suurema hulga vasteid. Selle märksõna eripäradeks oli "sort" ehk erinevad paberossisordid, samuti tulid esile sõnad "maitse", "suitsetaja" ja "suitsetama", "nõudma". Koos märksõna otsinguga "pabeross" tulid välja ka ettevõtted Astoria ja Laferme.

Kui vaadata põhimärksõna "suits" sisaldavat sõnasageduste otsingut, oli esimesel kohal sõna "pabeross" ehk paberossireklaamidel esines sõna "suits" tihti koos erinevate sõnavormidega, nagu "suitsetaja" või "suitsetama". Samuti tulid selle otsinguga välja sõnad "eesti", "headus", "müük", "hind", "sort", "maitse", "mark", "suitsumees", "uudis" ning "tähelepanu". See oli ainukene põhimärksõna, kus tuli esile ka bränd "Mikaado", mis oli H. Anton ja Ko[152] ning Astoria[153] suitsumark. Esinesid ka sõnad "puhastama" ja "sigaret". Kõikide põhimärksõnade ühisomadusteks olid "headus", "parem", "müüma" ning ühikud "sent", "mark", "tükk".

Põhisõnade sagedustabeleid analüüsides jääb mulje, et tubakaettevõtete reklaame on justkui palju rohkem kui lähiloetud veaprotsent näitas, kuna enamus kasutatavaid sõnu näivad pärinevat tubakareklaamidest, välja arvatud "konfiskeerima" (regulaaravaldisega "konfisk" leidub tekstikorpuses 126 vastet 17 061 vastest), mis tuleneb tolliameti müügipakkumistest. Sõnad "suhkur", "seep" ja "vein" tulid sagedustabelites esile, kuna Eesti Tarvitajate

[152] Postimees, 17.06.1934, lk 3.
[153] Postimees, 22.08.1934, lk 8.

Keskühisuse (ETK) puhul reklaamiti oma tooteid tootekataloogina.[154] Samamoodi reklaamisid oma tooteid järjestades erinevad toidukauplused[155] ja laod.[156]

Lähilugemise 1932. aasta maikuu reklaamide põhjal ilmnes, et kuigi üldine masinloetud (OCR) teksti kvaliteet oli hea, osutus ettevõtete nimede tuvastamine siiski problemaatiliseks kuna need olid tihti dekoratiivselt kaunistatud. Näiteks tuvastati AS "A. Reier & Ko" hoopis kirjapildina "n/5. B. Reter & Ko".[157]

Kokkuvõttes saab öelda, et andmeanalüüs on kasutatud meetoditega võimalik. Kuigi andmestiku lähilugemine andis täpsuseks 63% (enne parandusi 43%), siis põhimärksõnade sagedusanalüüsil ei tõusnud esile märkmimisväärsel hulgal sõnu, mis tubakaettevõtete reklaamidega ei seostunud. Näiteks märksõna "pabeross" esines ligi 10 000 reklaamis, märksõna "tubak" 9000 reklaamis. Tubakaga mitteseonduvad sõnad nagu "konfiskeeritud", "vein", "suhkur" ja "seep" oli madalama sagedusega (iga sagedus ligikaudu 25 – 100 korda). Lähilugemisel esinesid need näiteks ETK tootekataloogidena esitatud reklaamides või tolliameti teadetes. Kasutatud kontekstiaken (100 märki paremale, 100 vasakule põhimärksõnast) oli põhjendatud, kuna laiema kontekstiaknaga, näiteks 300 või 400 tähemärgiga, vähenes põhimärksõnade sagedustabelite kvaliteet, sagedamini ilmusid esile tubakareklaamidega mitteseotud sõnad. Tekstituvastuse puhul tekkisid probleemid eelkõige illustreeritud kirjapildiga ettevõttenimede lugemisel. Seega näitab sagedusanalüüs, et koostatud andmestiku üldine puhtus ja spetsiifilisus on piisavalt hea hoolimata OCRi ja märksõnapõhise otsingu piirangutest.

3.2. Tubakareklaamide sõnakasutuse analüüs

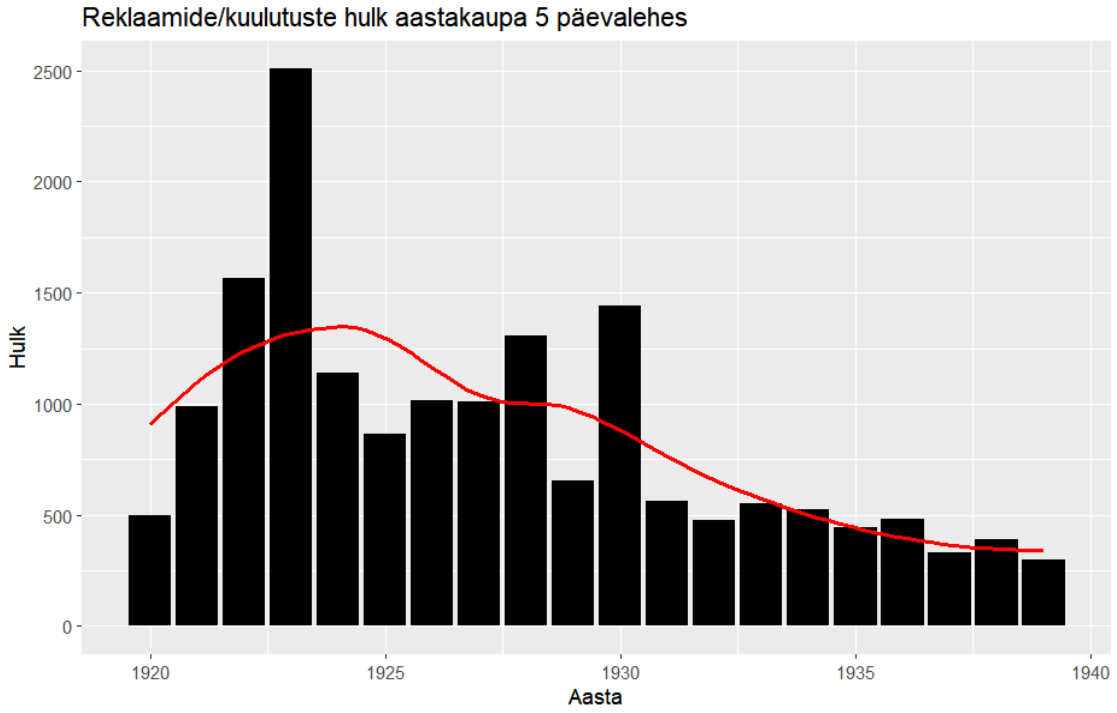
Käesolev andmestik sisaldab 17 061 reklaami, mis on võetud 1920. – 1940. aastatel järjepidevalt ilmunud viiest päevalehest, et tagada kvaliteet ühtlasema andmestiku toel. Andmestiku piiramine viie peamise päevalehega oli vajalik, et vältida statistilisi ebatäpsusi, mida võivad põhjustada ebaregulaarselt ilmuvad ajalehed, kasutatavatel ajalehtedel peab olema piisavalt suur hulk ühiseid omadusi. Niimoodi sai tagada, et reklaamide koguhulka ei mõjutaks ühe ajalehe teke, sulgemine või väljaannete ilmumissageduse muutumine. Samuti

[154] Uus Eesti, 18.12.1938, lk 1.
[155] Päevaleht, 04.11.1925, lk 8.
[156] Päevaleht, 01.11.1925, lk 12.
[157] Kaja, 08.11.1925, lk 5. OCR-tekst kättesaadav: https://dea.digar.ee/article/kaja/1925/11/08/1/54.1 (viidatud 15.05.2025.)

on oluline, et igast ajalehest oleks kaasatud piisavalt suur valim reklaame. Seetõttu on valitud nendeks viie päevalehe Päevaleht 8029, Postimees 6641, Kaja 3086 ning selle järglase Uus Eesti[158] 771 ning Sakala 2217 tubakareklaami.

Iga järgmise väiksema ilmumissagedusega ajalehe lisamine ei tõstaks statistilist piirkasulikkust (ingl k *diminishing returns*), ei pakuks palju infot ning vähendaks andmestiku ühtlast kvaliteeti. Seetõttu keskendutaksegi suuremate, regulaarselt ilmunud päevalehtede reklaamidele. Ka Eesti Reklaami Agentuur tõdes 1937. aastal, et "Loomulikult tuleb kuulutada seal, kus on rohkem lugejaid ja mille lugejaskonnast võib loota endale rohkem ostjaid. Juhuslikes ja väheloetavais väljaandeis on reklaamil väga väike mõju või pole seda üldse."[159]

Joonis 4. Tubakareklaamide kuulutuste hulk korpuses aastate lõikes viies päevalehes. (Päevaleht, Postimees, Sakala, Kaja (1919 kuni 17.09.1935) ja selle järglane Uus Eesti (alates 18.09.1935)[160])



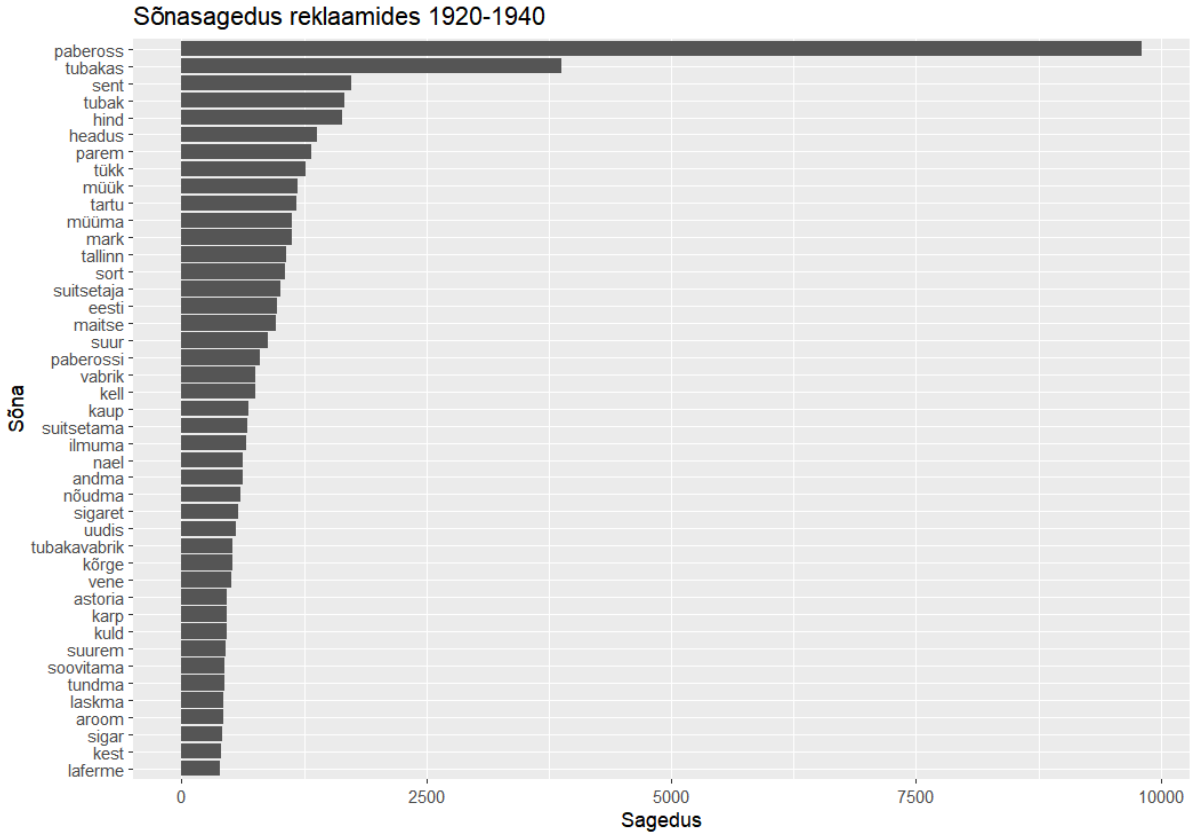
Mõtteeksperimendina saagise hindamiseks võtsin aluseks hüpoteetilise arvutuse, et kui näiteks kolm peamist ettevõtet (ETK, Laferme ja Astoria) reklaamivad oma tubakatooteid

[158] Lauk, *Peatükke Eesti ajakirjanduse ajaloost*, lk 14 – 15
[159] Eesti Reklaami Agentuur, *ERA Reklaam kataloog käsiraamat*, lk 6.
[160] Epp Lauk, *Peatükke Eesti ajakirjanduse ajaloost*, lk 14 – 15.

igapäevaselt kõigis viies päevalehes (Uus Eesti asendas ajaleht Kaja aastal 1935),[161] teeb see kokku 12 reklaami päevas, mis on ligikaudu 360 reklaami kuus ning uuritava ajaperioodi vältel ehk 20 aasta jooksul tähendaks see 86 400 reklaami. Leitud reklaamide koguhulk on aga palju väiksem, ulatudes 17 000 juurde. Kui see number jagada 20 aasta päevadega, siis ilmus nendes viies päevalehes kokku ligikaudu 2,3 tubakareklaami päevas. Tulemustest selgus, et enim reklaame ilmus 1923. aastal, kokku 2500 reklaami, mis teeb 20 aasta peale kõikidest reklaamidest kokku 14,7%. Keskmiselt teeb 2500 reklaami kokku esinemissageduselt 6,8 reklaami päevas.

Kõige lihtsam viis tekstikorpusest esmase ülevaate saamiseks on analüüsida selles esinevate sõnade sagedust.[162] Sõnasageduste analüüs annab võimaluse mõista, milliseid sõnu kasutati reklaamides kõige sagedamini ning millised teemad ja rõhuasetused joonistuvad välja uuritava perioodi reklaamides.

Joonis 5. Tubakareklaamide sõnasagedus aastatel 1920 – 1940 (17 061 vastet).



[161] Lauk, *Peatükke Eesti ajakirjanduse ajaloost*, lk 14 – 15.
[162] Wevers, "Consuming America," lk 57.

Sõnasageduse põhjal reklaame analüüsides saab infot reklaamide temaatilise sisu ja enim kasutatud reklaamisõnavara kohta. Teksti sisule keskendumiseks eemaldatakse analüüsil tihti stoppsõnad. Need on valitud sõnad, mis meid analüüsil ei huvita. Analüüsil kasutamise eelnevalt loodud stoppsõnade nimekirja, mis eemaldab tekstidest levinud sidesõnad, asesõnad ja muud sagedasti eesti keeles esinevad sõnad.[163] Kõigepealt tegelesin sõnestamisega (ingl k *tokenization*) – protsess, mida defineeritakse kui teksti jagamist väiksematekst üksusteks ehk sõnedeks (ingl k *token*), milles lasin sagedustabeli koostamisel lugeda kõik sõnad eraldi ükstusteks. Sõned ei pea ainult sõnu sisaldama, vaid võivad hõlmata endaks ka lauseid või muid segmente.[164] Segmenteerides näiteks numbrid sõnedeks saab tekstikaeve abil uurida paberosside hindade muutumist ajas.

Sagedustabeli järgi oli kõige populaarsem sõna "pabeross", mida esines ligi 10 000 korda, järgnev sõna "tubakas" avaldus ligi 4000 korral, mis oli ootuspärane kuna tegu on toote nimetusega. Selle ajastu reklaamides esines ka hinna ("sent", "mark") ja kogusega ("tükk", "nael") seotud sõnu. Huvitavateks leidudeks olid sellised sõnad nagu "parem", "headus", "maitse", "nõudma", "kõrge", "uudis", "aroom", "suitsetaja".

Lähilugemisel leidsin, et sõna "uudis" kasutati kutsuva pealkirjana, näiteks "Uudis suitsetajatele! Uus paberossisort." Samuti sõna "nõudma", mida kirjutati näiteks pealkirjana "Nõudke igal pool!"[165] Samuti kasutati sõna "nõudma" reklaami pealkirjas "Uued nõuded suitsetamisel!",[166] "millede peale omal ajal suur nõudmine oli"[167] ning "seda tõendab suur nõudmine"[168] See ei olnud omane vaid ühele ettevõttele, seda kasutasid AS "A. Reier & Ko", OÜ "Havanna", AS "Laferme" ja Eesti Tarvitajate Keskühisus (ETK). Iseloomulik oli ka reklaamides "suitsetaja" poole pöördumine.[169] Sõnad "parem", "kõrge", "headus", "aroom", "maitse" tunduvad olevat antud ajastu kõnepruuk, millega kirjeldati tubakatoodete omadusi. Samas eksisteeris ka H. Anton ja Ko piibutubakas kaubamärgiga "Aroom".[170]

Sagedustabelis tõusid esile ettevõtted AS "Astoria" ja AS "Laferme", just need kaks ettevõtet olid tugevamini seotud reklaamindusega. Tubakaettevõttel "Astoria" oli oma

[163] Šikirjavõi, "Temaatiliste mustrite kaevandamine seadustekstidest," lk 15.
[164] Šikirjavõi, "Temaatiliste mustrite kaevandamine seadustekstidest," lk 22, 34.
[165] Postimees, 13.12.1921, lk 1; Kaja, 16.08.1922, lk 1.
[166] Uus Eesti, 24.09.1938, lk 5.
[167] Postimees, 15.04.1922, lk 1.
[168] Päevaleht, 16.10.1930, lk 3.
[169] Päevaleht, 30.09.1931, lk 3.
[170] Uus Eesti 13.12.1936, lk 2.

reklaamiagentuur ning AS "Laferme" juhatuse liige Adam Grünbaum kuulus Eesti Reklaam-Klubi juhtivate liikmete sekka.

Lisaks sõnasagedustele võimaldab kollokatsioonianalüüs ehk naabersõnade uurimine paremini mõista tubakareklaamide tähenduslikke seoseid ja keelelisi mustreid. Kollokatsioonid (ehk statistiliselt silmapaistvad naabersõnad) on sõnad või sõnaühendid, mis esinevad tekstikorpuses märkimisväärselt sagedamini koos kui juhuslik esinemine. Kvantitatiivse kollokatsioonianalüüsi eesmärgiks on tuvastada ühe sõna jaoks selle kasutuskonteksti iseloomustavad teised sõnad.[171] Kvantitatiivne kollokatsioonianalüüs võimaldab seega (1) identifitseerida tüüpilisi diskursiivseid mustreid (nt tarvitamise sidumine naudingu ja aroomidega), (2) suunata lähilugemist täpselt nendesse tekstikohtadesse, kus huvipakkuvad seosed esile kerkivad, ning (3) jälgida ajas muutusi (nt 1923. a. "keemiliselt puhastatud" ja 1930. a. "pneumaatiliselt puhastatud tubakas"). Nii toetab kollokatsioonitabel ka sõnasagedusloendeid, pakkudes kvaliteetseid „sõlmpunkte", mille kaudu kirjeldada tubakareklaamide retoorilisi võtteid ja nende arengut.[172]

Tabel 3. Kõige levinumad kollokatsioonid ehk naabersõnad.

kollokatsioon	**hulk**	**lambda**	**z**
pabeross paberossi	710	6.07700	36.1905
tükk sent	416	4.39612	66.1579
parem pabeross	363	2.26163	35.7946
tükk mark	361	4.77702	65.4884
sort pabeross	332	2.51892	36.6886
wabrik wabriku	307	10.08417	39.7593
kõrge headus	262	5.41133	56.3016
ilmuma müük	237	4.86097	55.3614
müük laskma	211	5.39420	52.4173
maitse aroom	157	5.11536	47.6605
nael mark	156	4.41339	44.4507

[171] Anatol Stefanowitsch, *Corpus Linguistics: A Guide to the Methodology* (Berlin: Language Science Press, 2020), lk 218, https://doi.org/10.5281/zenodo.3735822.
[172] Wevers, "Consuming America," lk 58 – 59.

o_ü tubak	151	5.40608	40.7917
kõrgem sort	134	5.58255	43.5777
a_s astoria	100	5.14863	41.0324

Lambda (λ) mõõdab sõnadevahelise seose tugevust kollokatsioonis, ning z-statistik (z) näitab, kui statistiliselt oluline ehk usaldusväärne see leitud seos on.[173] Mõlemad näitajad arvutab automaatselt R-paketi quanteda.textstats funktsioon textstat_collocations.[174] Tegemist on statistiliste testidega, mis aitavad määrata, kui usaldusväärne ja märkimisväärne on uuritav keeleline nähtus.

Kollokatsioonanalüüsis tulevad eelkõige välja kogus (“tükk”, “nael”) ja hind (“sent”, “mark”) ning ka sõnakasutus “kõrge headus”. Sõnapaar “kõrge headus” ilmnes just Laferme[175] ja H. Anton ja Ko[176] reklaamides. Samuti kasutas A. Le Coq seda sõnastust[177] ning leidus ka reklaame, kus ettevõtte nimi puudus.[178] Tundub, et sõna “headus” oli oma ajastu tüüpiline kõnepruuk.[179]

Kui vaadata mõningaid välja nopitud näiteid, siis reklaamijate sõnul olid paberossid kõikvõimsad, näiteks AS “Astoria” paberossid “Kabaree” suitsetamine pidi lahendama kõik tülid,[180] lahutama meelt, andma uut jõudu, ülendama tuju ja tegema elu mugavaks.[181] AS “Sirena” paberossid olid ainukesed Eestis, millel puudus nikotiinimürk.[182]

3.3. Tubakaettevõtete võrdlus korpuses

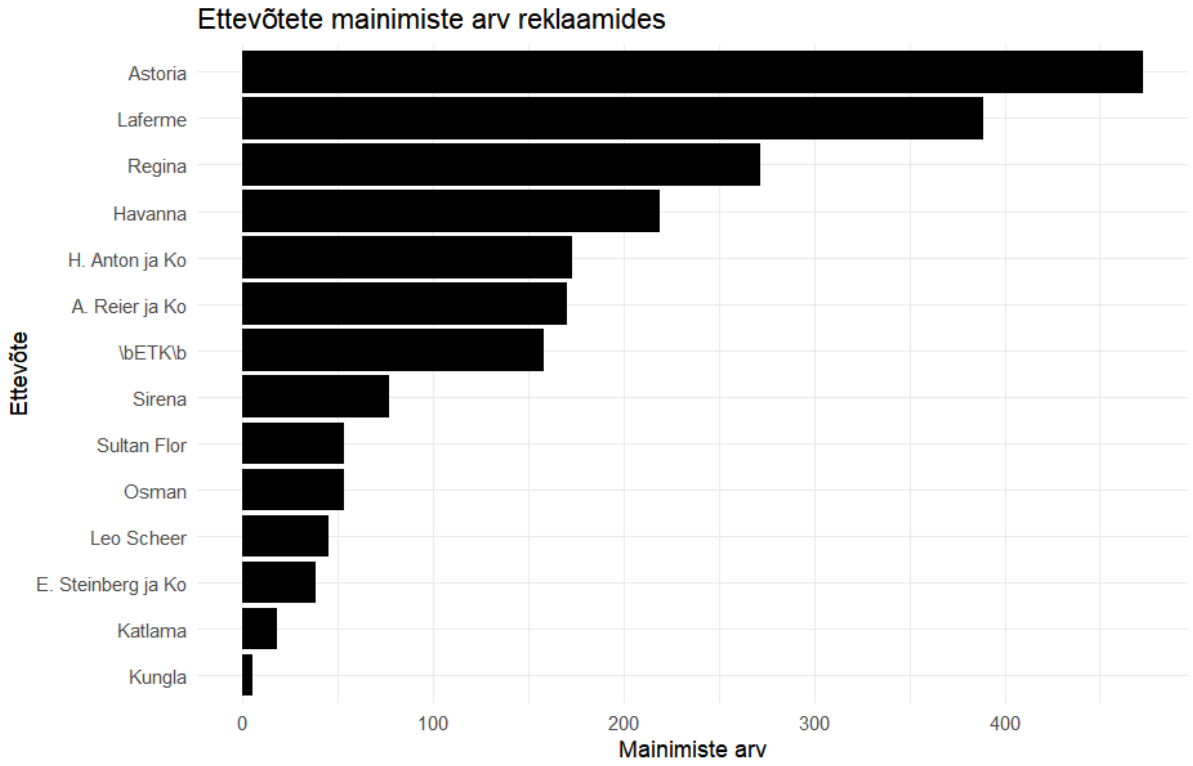
Oluliseks metoodiliseks põhimõtteks antud töös oli tööriistade mitmekesisus. Töö digitaalsete tööriistadega ei ole lineaarne protsess, vaid iteratiivne, mis hõlmab korduvat edasi-tagasi

[173] Don Blaheta ja Mark Johnson, “Unsupervised Learning of Multi-Word Verbs,” *Proceedings of the ACL 2001 Workshop on Collocation: Computational Extraction, Analysis and Exploitation*, Toulouse, France, 2001, Association for Computational Linguistics, lk 54 – 60.
[174] Valem paketi lähtekoodis.
[175] Kaja, 10.07.1926, lk 5.
[176] Päevaleht, 15.06.1929, lk 4.
[177] Kaja, 30.10.1924, lk 5.
[178] Päevaleht, 19.08.1931, lk 3.
[179] Nool, 11.10.1930, lk 5; Päevaleht, 01.10.1931, lk 3; Päevaleht, 13.03.1932, lk 5; 1932. a. Valter Kõrveri tehtud reklaamis on näha Hollywood’ilikku tarbija kujutamist.
[180] Päevaleht, 30.10.1925, lk 6.
[181] Postimees 06.10.1925, lk 3.
[182] Päevaleht, 16.12.1930, lk 32.

liikumist erinevate tööriistade väljundi genereerimise, nende interpreteerimise ja allikate lähilugemise (ing k *close reading*) vahel.[183]

Otsisin ettevõtete kaupa korpusest reklaame, et siduda iga reklaam konkreetse turuosalisega, mõistmaks kuidas reklaamid ettevõtete kaupa erinevad. Esialgu koostasin nimekirja firmanimedest "Laferme", "Sirena", "Katlama", "Regina", "Havanna", "Sultan Flor", "Reier", "Astoria", "H. Anton", "Kungla", "Steinberg", "Leo Scheer", "Osman" ning "ETK". Seejärel analüüsisin kui palju kordi ettevõttenimed olid tubakareklaamides mainitud.

Joonis 6. Ettevõtete mainimiste arv reklaamides.



Statistiliselt oli ettevõtete mainimiste koguarv 17 061-st reklaamist üle 2000 vaste, mis on analüüsitud reklaamide koguarvust 11,7%. Lähilugemisel selgus, et praktiliselt olid kõik tõsipositiivsed reklaamid ettevõtte nimedega.

Kuna DIGARi tekstituvastuse (OCR) algoritmid on treenitud peamiselt standardsete ajalehetrüki fontide põhjal (eesmärgiga artikleid digiteerida), siis ei taga tekstituvastus (OCR) täielikku täpsust reklaamitekstide tuvastamisel, kus kasutatakse sageli standardsetest erinevaid kirjatüüpe, nagu dekoratiivsed ja kalligraafilised ettevõtete logod.

[183] Wevers, "Consuming America," lk 47.

Tabel 4. Mõningaid OCR veanäiteid ettevõtete nimede tuvastamisel ajalehereklaamides.

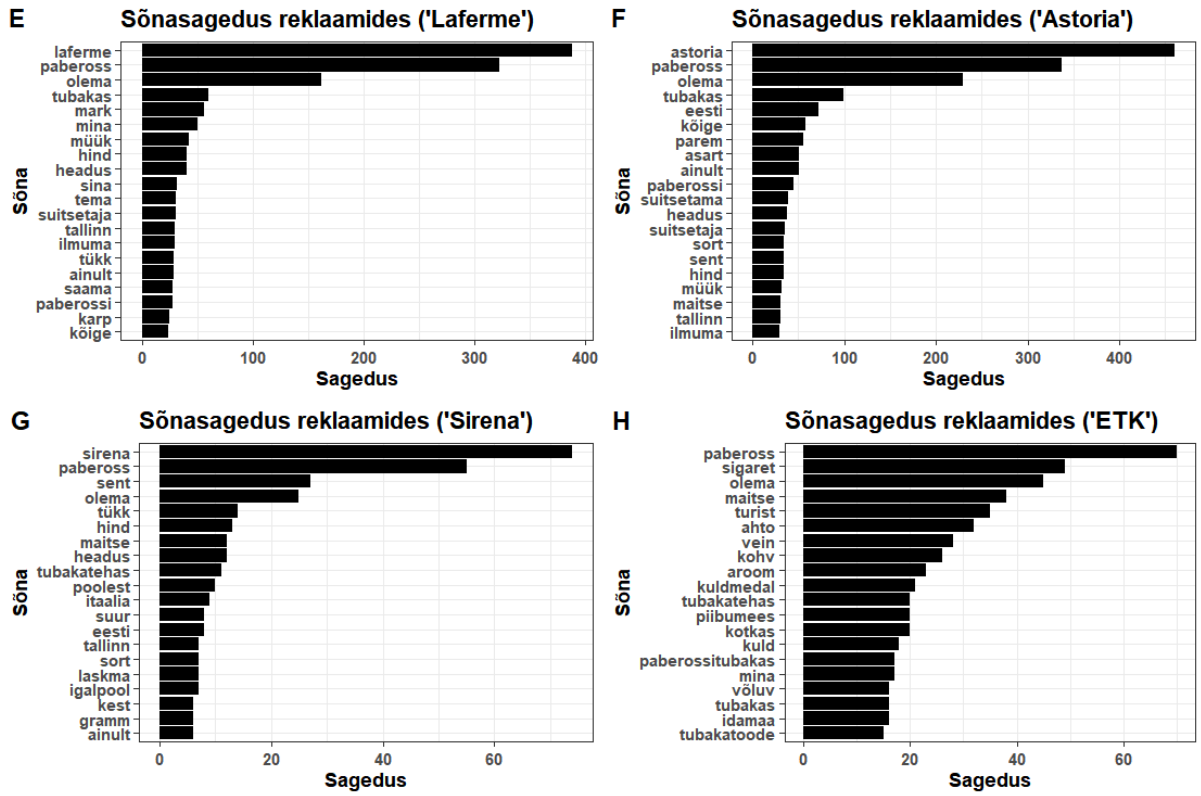
Ettevõtte nimi	**OCR vead**
Havanna	Havann, Kavanna, havänna, kaualla, havaima, Jfauonna
H. Anton ja Ko	nton ja Ko, Dnfon ja Ke
Laferme	£aferme, £afcritie, JBafermo, mferme
Sirena	Sirema, SIgSena, Sirina, sirerr
B. Katlama	AA. Jvadtema, £O. Jiaflama, J Katta 100 “Priima”
A. Reier ja Ko	Reier ja Co, n/5. B. Reter & Ko
Sultan Flor	9ultan Flor

Vigadest tulenevalt proovisin leida meetodi, et ettevõtte nimesid paremini masinlugeda. Proovisin laiendada regulaaravaldiste metamärkidega katvust, kuid sellega tuli kaasa hulk mittekuuluvaid reklaame. Seejärel parandasin metamärkide katvust iteratiivselt nimemustrite otsingu protsessi täpsustamiseks. Proovisin ka *fuzzy matching* lähenemisviisi ettevõtete nimekirja koostamisel, mida nimetatakse ka ligikaudseks sõnede sobitamiseks, et leida selliseid tekstilõike, mis sarnanevad otsitava mustriga, kuid ei pea täpselt kattuma. Uurija saab ise määrata, kui suur on tähemärkide erinevus (*edit distance*). Näiteks päring “america^5” leiab kõik variandid, mis erinevad Sirenast kuni viie tähemärgi võrra, mille tulemusena leitakse ka “Ameerika”.[184] Antud analüüs andis eelkõige teadmise tekstituvastuse (OCR) võimaluste kohta edaspidisteks uurimusteks. Tekstide kalligraafilisi vorme tuleb analüüsida teistsuguste meetodite ja algoritmidega.

Otsingumeetodite piiranguid arvesse võttes analüüsisin sõnasagedusi valitud tubakaettevõtete reklaamides, tuues esile erinevate firmade keelelised eripärad ja rõhuasetused.

[184] Wevers, “Consuming America,” lk 53.

Joonis 7. Sõnasagedused nelja tubakaettevõtte reklaamides.



Ettevõtete kaupa sõnasagedusi vaadates võtsin vaatluse alla neli enama reklaamide arvuga ettevõtet (vähemalt 200 reklaami), ülejäänud ettevõtete tabeleid saab uurida lähemalt GitHub'i repositooriumist.[185]

Laferme, Astoria ja Sirena reklaamides tuleb esile sõna "headus", mis tundus sel ajal olevat põhiliseks tubakatooteid iseloomustavaks sõnaks. ETK kasutas seevastu omadussõna "võluv". ETK kaubamärkidele viitavad ka "Turist", "Kuld", "Kotkas" ning "Kuldmedal". 1936. aastal sooviti võtta kasutusele ka kaubamärk „Palusalu", kuid rahvusvaheline spordimäärus ei lubanud seda teha ning selle asemel hakati kasutama „Kuldmedal"it.[186] Eesti Tarvitajate Keskühisus tootis ka paberosse nimega "Ahto", mille nimi oli pandud tollase rahvuskangelase, purjetaja Ahto Valteri järgi, kes oli esimene Eesti lipu all ümbermaailma purjetaja.[187] Vaatamata paberossi markide paljususele, moodustas „Ahto" poole turu kogutarbimisest.[188] Laferme puhul näiteks ei tule välja ühtegi domineerivat kaubamärki, Astoriat esindas vaid üks kaubamärk "Asart".[189]

[185] Lust, "Tubakakorpus," https://github.com/tormil/tubakakorpus/
[186] Samas, lk 52.
[187] Kaukvere, "Linnamuuseum avas vana reklaami näituse."
[188] Juurmaa, lk 53.
[189] Kaja, 26.09.1923, lk 4.

Astoria ja Laferme reklaamides on kasutatud "suitsetajat", tundub et reklaamides pöörduti lugejate poole neid "suitsetajateks" kutsudes. ETK reklaamijad kutsusid suitsetajaid seevastu "piibumeesteks". ETK reklaamide "vein" ja "kohv" viitavad sellele, et nemad reklaamisid oma tooteid tihti tootekataloogina.[190]

Sõnapaar "kõrge headus" ilmnes just Laferme ja H. Anton ja Ko reklaamides. Samuti kasutas A. Le Coq seda sõnastust ning leidus ka reklaame, kus ettevõtte nimi puudus.

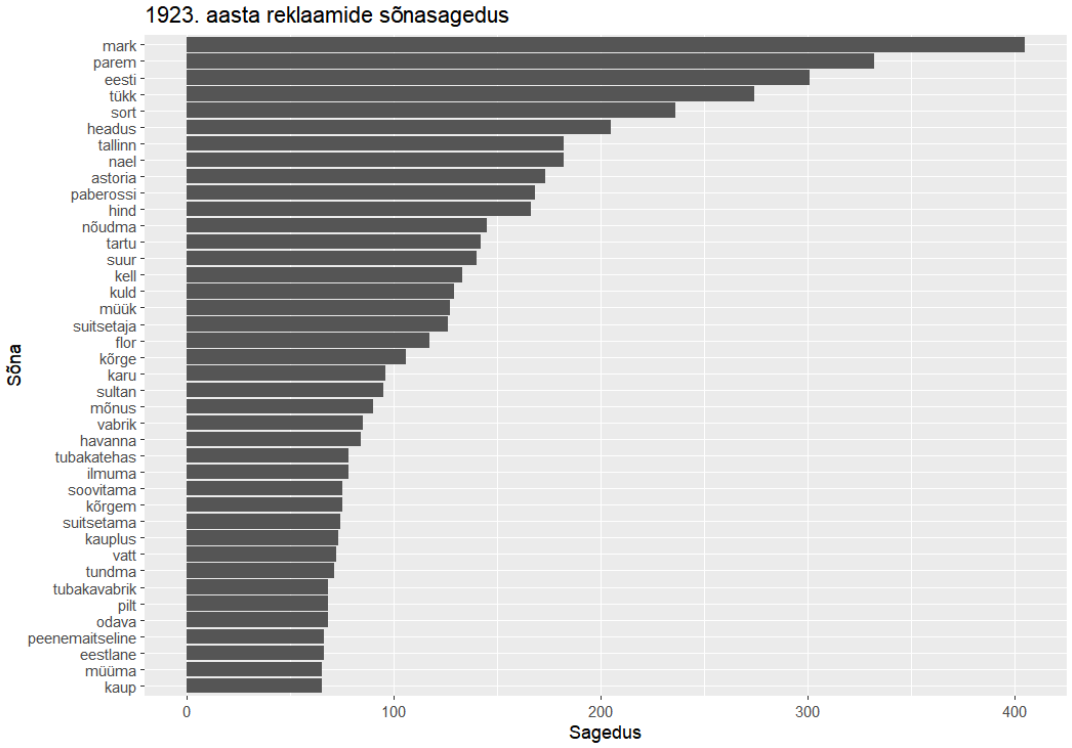
Otsides vastust küsimusele, kuidas erinesid ettevõtete reklaamid sõnastuselt, rakendasin ettevõtete tuvastamiseks ja reklaamidega sidumiseks regulaaravaldisi. Ettevõtete tuvastamise tegi keeruliseks tekstituvastuse (OCR) ebatäpsus, mis suutis tuvastada 17 061 reklaamist vaid 2000-s ettevõtte nime, eriti ilmnes probleem kunstipäraselt kujundatud nimede puhul. Ent lähilugemisel selgus, et ettevõtete nimed olid reklaamides praktiliselt alati esindatud. Selgus, et masin suutis oluliselt paremini lugeda lauseid, kui ettevõtete nimesid, mis oli tihti kalligraafilises vormis. (Näiteks Laferme tähistas oma reklaame L-kujulise logoga.)[191] Sõnasageduste analüüsist ettevõtete kaupa sai selgeks, et brändinimedes kasutati sageli kuulsate inimeste nimesid,[192] näiteks ETK "Ahto", Laferme "Kuldmedal". Samuti oli kasutusel AS "Sirena" kaubamärk "Male",[193] viimane oli tõenäoliselt nimetatud Eesti male suurmeistri Paul Kerese auks. Saadud tulemused näitavad vajadust arvestada OCR-i piiranguid ajalooliste reklaamide digitaalsel töötlemisel, mistõttu tuleks edaspidistes uuringutes sarnaseid kalligraafilisi tekste käsitleda alternatiivsete tekstitöötluse meetoditega.

3.4. 1923. ja 1930. aastate eripärad

Reklaamide hulga analüüs aastate lõikes[194] näitas, et 20 aastase perioodi kestel jäävad silma kaks aastat – 1923 ja 1930, mil reklaamide hulk oli ülejäänud aastatest märgatavalt suurem. 1923. aastal saavutas reklaamide arv uuritud perioodi kõrgeima taseme (2509 reklaami) ning 1930. aastal oli see ligi 1500, samas kui aastate keskmiseks oli 853 reklaami. Küsimuseks on – mida näitab aasta sõnasageduste analüüs reklaamide temaatika ja sisu kohta, ja kas või kuidas reklaamitegevus peegeldab konkreetseid ajaloolisi, näiteks majanduslikke või kultuurilisi tegureid.

[190] Uus Eesti, 18.12.1938, lk 1; Päevaleht, 01.04.1939, lk 9; Päevaleht, 14.12.1937, lk 1.
[191] vt Pilt 4.
[192] Samuti müüs Tartu Tubakavabrik maadleja Jaan Jaago nimelisi paberosse ning AS "Laferme" tõi Paul Pinna juubeliaasta auks turule "Pinna" paberossid. – Päevaleht, 21.05.1933, lk 5; Päevaleht, 21.08.1926, lk 5 – 6.
[193] Päevaleht 12.05.1938, lk 7.
[194] vt 3.2. alapeatükki

Joonis 8. 1923. aasta reklaamide sõnasagedus

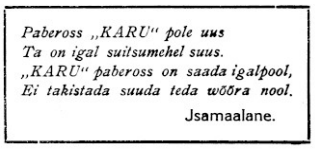


1923. aasta sagedusloendi põhjal tulevad esile sel ajal tegutsenud suurematest ettevõtetest – Astoria (asutatud 1922),[195] Sultan Flor (varaseim mainimine 1922)[196] ja Havanna (asutatud 1921).[197] Kaubamärkidest ilmnevad OÜ "Tubak" kaubamärgi "Karu" reklaamid. Samas ei tulnud esimese kolmekümne hulgas suurimatest ettevõtetest välja kolme nime – AS "Laferme", "H. Anton ja Ko" ning "A. Reier ja Ko", millest võib järeldada, et ettevõtted, mis otsinguga välja ei tulnud, panustasid sel aastal reklaamidesse vähem.

Kirjeldavad omadussõnadest olid iseloomulikud "mõnus", "peenemaitseline", "headus", "parem", "iseäraline", neist oli uueks sõnaks "peenemaitseline". AS Astoria kasutas oma reklaamides tarbija poole pöördudes sõnaühendit "peenemaisteline suitsetaja",[198] millega sihiti justkui peenemate kommetega sihtrühma.

[195] Michelson, "Tubakatööstus," lk 606.
[196] Vaba Maa, 06.12.1922, lk 1.
[197] Kannel, "Tubakatööstus Eesti Vabariigis 1920–1940," lk 16.
[198] Kaja, 10.10.1923, lk 3.

Pildid 5 ja 6. Paberossid “Karu” reklaamid 1923. aastast.[199]



Vaatasin ka 1923. aasta reklaamide (2509 reklaami) kollokatsioone ehk naabersõnu. Seal tulid suurimatest ettevõtetest samuti välja Tubak, Sultan Flor, Astoria ja Havanna, mis kutsus ennast “Esimeseks Eesti tubakatehaseks”.[200]

Tabel 5. 1923. aasta reklaamide kollokatsioonid.

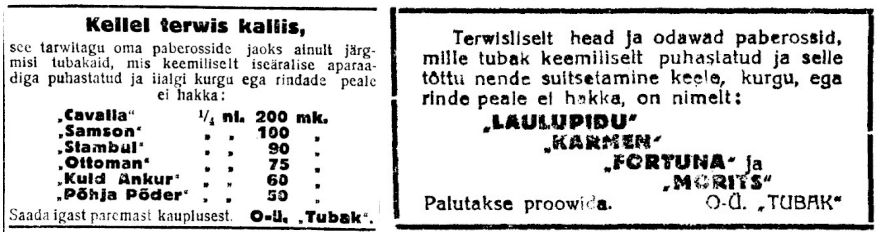
kollokatsioon	hulk	lambda	z
tükk mark	149	5.044446	36.15871
sort pabeross	129	3.302912	23.58565
o-ü tubak	99	4.877909	28.11052
nael mark	80	4.512175	27.10792
kõrge headus	74	6.423103	27.41262
eesti parem	58	3.356873	21.06696
kõrgem sort	54	6.377795	23.62751
tubak keemiliselt	35	4.867705	18.97251
hügieeniline wati	33	7.740752	21.40612
sultan flor	32	7.312791	21.92829
eesti tubakatehas	32	4.661074	18.47399
kuld ankur	32	6.984989	18.22587
keemiliselt iseäraline	30	6.742624	21.92292
kotkas kuld	30	6.912157	18.48055
keemiliselt puhastama	28	6.96843	21.10379
a-s astoria	28	5.03754	19.67746
suitsetamine keel	28	9.243804	18.19759
warustatud hügieeniline	25	7.345725	20.32781
müük laskma	25	5.920342	18.88422
paberossi asart	25	4.970701	18.37357

1923. aasta kollokatsioone uurides selgus, et reklaamidel kajastati tüki hinda ehk mitu marki teatud kogus paberosse maksis (149 vastet), nagu ka paberossi markide nimesid Kotkas, Kuld, Ankur (mis vastavad markidele Kuld Kotkas, Kuld Ankur). Huvitav leid oli 28 korda

[199] Sakala, 26.01.1923, lk 2; Sakala, 09.05.1923, lk 2.
[200] Sakala, 19.09.1923, lk 2.

esinenud sõnaühend “keemiliselt puhastama” ehk keemiliselt puhastatud ja “hügieeniline wati” (33 vastet) ehk tegemist on hügieenilise vatiga (kuna otsingul oli ära võetud käändevorm, siis olid tulemused sõna algvormis).

Pildid 7 ja 8. “Keemiliselt puhastatud väljendite kasutus tubakareklaamides.[201]



Keilel terwis kaliis,

see tarwitagu oma paberosside jaoks ainult järgmisi tubakaid, mis keemiliselt iseäralise aparaadiga puhastatud ja iialgi kurgu ega rindade peale ei hakka:

„Cavalla“	¼ nl.	200 mk.
„Samson“	„ „	100 „
„Stambul“	„ „	90 „
„Ottoman“	„ „	75 „
„Kuld Ankur“	„ „	60 „
„Põhja Põder“	„ „	50 „

Saada igast paremast kauplusest. **O-ü. „Tubak“.**

Terwisliselt head ja odawad paberossid, mille tubak keemiliselt puhastatud ja selle tõttu nende suitsetamine keele, kurgu, ega rinde peale ei hakka, on nimelt:

„LAULUPIDU“
„KARMEN“
„FORTUNA“ ja
„MORITS“

Palutakse proowida. O-Ü. „TUBAK“

Järelduste tegemiseks, miks just 1923. aastal reklaame kõige rohkem esines, tuleks arvestada majanduslike teguritega. Toetudes M. Kannela tööle, näeme, et kui 1919. aastal tegutses Eestis vaid 1 tubakavabrik, siis 1922. aastal oli neid juba 11. Kõige paremini annab olukorra edasi tubakatööstuse juhi, Hans Antoni sõnastus: “nii hakkas Eestis tubakavabrikuid tekkima nagu sügisel seeni: 1923. aastal tegutsesid Eestis juba ligi 15 masinatega töötavat tubakavabrikut.”[202] Samuti kasvas toodang.[203] Kui 1920. aastal toodeti 12 miljonit paberossi aastas, siis 1921. aastaks kasvas see 13 korda kuni 156 miljonile paberossini. 1923.aastal toodeti 665 miljonit paberossi. 1925. aastaks jõuti 1 miljardi paberossi tootmiseni, kus see püsis stabiilsena kuni majanduskriisini.

Analüüsides 1930. aastat tuli sõnasageduse otsinguga välja ainsa ettevõttenimena Sirena (asutatud 1929). Huvitavateks leidudeks olid sõnad “tolm” ja “pneumaatiline”. Konteksti mõistmiseks teostasin kollokatsioonide analüüsi.

Tabel 6. 1930. aasta reklaamide kollokatsioonid.

kollokatsioon	hulk	lambda	z
tükk sent	178	5.12815	30.95423
puhastama tubakas	129	5.02211	29.47322
tubakas pneumaatiline	113	5.664051	25.76296
vabrik puhastama	85	5.264447	27.87927
pneumaatiline viis	85	7.963728	22.25528
viis tolm	78	7.496819	21.42603

[201] Postimees, 16.11.1923, lk 3; Esmaspäev, 24.09.1923, lk 4.
[202] Eesti Nädal, 31.08.1934, lk 5.
[203] vt Lisa 3.

tolm kahjulik	75	8.311986	20.11613
kahjulik lisandus	67	8.62278	21.12253
wabriku puhastama	52	5.935619	20.14321
müük laskma	50	6.887142	23.22036
maitse aroom	48	5.847336	23.4862
kollane kest	44	5.350068	22.2954
peen maitse	37	7.046893	17.624

Kui 1923. aasta reklaamides rõhutati keemiliselt puhastatud tubakat ja hügieenilist vatti, siis nüüd hakati esile tooma tubaka puhastamist pneumaatilise meetodiga. 1930. aastal tõusis esile tehnoloogia ja tööstusliku innovatsiooni rõhutamine reklaamide sisus, minu korpuses leidus pneumaatiliselt puhastatud tubaka mainimist ligi 200 reklaamis. Kui uurida laiemalt sõna „pneumaat“ levikut kõikides selle ajavahemiku DIGARi ajalehtedes ja artiklites, siis selgub, et 1930. aastal kasutati seda sõnavormi erakordselt palju – ligikaudu 750 korda, samas kui teistel aastatel piirdus esinemine vaid kuni 10 korda aastas.

AS “Laferme” reklaamides võis sel ajal kohata tüüpteksti “Tähelepanu. Meie vabrikus puhastatakse tubakas pneumaatilisel viisil tolmust ja teistest kahjulikest lisandustest.”[204] Sõna “keemiliselt puhastatud” asendumine sõnaga “pneumaatiline” võib seletada üldise trendiga rõhutada tööstuse tehnoloogilisi uuendusi, mis aitas ettevõtetel turul silma paista ning näidata oma kaasaegsust. Näiteks Kalevi šokolaadivabriku eelkäija Kawe kasutas reklaamides nimetust „aurušokolaadivabrik“, et rõhutada aurumasinate abil valmistatud toodete kvaliteeti. Lisaks sellele kasutati tööstuse arengule viitamiseks sageli korstnatega reklaame, sümboliseerides masintootmist ja modernset tootmistehnoloogiat.[205] Samal ajal hakati reklaamidesse lisama erinevaid tervisealaseid märkusi, sest 1930. aastatel tajuti ja teadvustati tubakasuitsu kahjulikku mõju inimorganismile.[206]

[204] Nool, 13.09.1930, lk 2; Eesti Spordileht, 23.05.1930, lk 3, 5, 7.
[205] Kaukvere, “Linnamuuseum avas vana reklaami näituse.”
[206] Spiridis, “Tubakasuitsu mõju organismi ja närvisüsteemi.”; Edasi, 22.03.1925, lk 6.

Pilt 9 ja 10. Üks Laferme reklaamidest, kus mainitakse pneumaatilisi tubakapuhastamise sisseseadmeid[207] ning artiklist "Millist tubakat võib suitsetada sportlane" illustratsioon pneumaatilisest aparaadist.[208]

Kaitske oma kopse!

Meie saaduste edu on tingitud mitte ainult **kõrgewäärtuslikkude tubakate** tarwitamisest, waid esi oones ka sellest, et me l on kasu ada kõik mcodsa tubakatööstuse tehnika saawutused.

Nende tehniliste uuenduste hulgas omab pneumaatiline tubakapuhastamise sissesead suure tähtsuse just **terwishoidlikust seisukohast.** Selle sisseseade abil puhastatakse kõik meie poolt ümber töötatawad tubakad kõigest tolmust, kotiebemetest ja teistest **kopsudele kahjul. mõjuwatest** lisandustest.

Seepärast wõime täie õigusega öelda:

Kaitske oma kopse!

A-s. „Laferme"

Meie esiksaadusi:

MANON ✿ LIA ✿ ORIENT

Aparat, mis korjab tuhakast tolmu.

Uurides 1923. ja 1930. aasta reklaamide eripärasid, sai keskseks küsimus – miks oli neil aastatel reklaame märkimisväärselt rohkem. Majanduslikult seletab reklaamimahu kasvu 1923. aastal see, et Eestis tubakatööstuse jõudis kiire kasvu haripunkti, kus vabrikute arv kasvas vaid nelja aastaga ühest ligi üheteiskümneni ning toodetud paberosside hulk tõusis märkimisväärselt võrreldes varasemate aastatega.[209]

1930. aasta reklaamide analüüsimisel tuli välja, et siis hakkasid reklaamid rõhutama tehnoloogilist innovatsiooni. Kui 1923. aastal puhastati tubakat keemiliselt ning tutvustati tarbijatele hügieenilise vati lisamist paberossi, siis 1930. aasta reklaamides rõhutati tubaka pneumaatilist puhastamist, millega püüti eristuda konkurentidest kaasaegsete tööstuslike meetoditega.

Peatükis analüüsisin loodud tubakareklaamide korpuse töövoo tõhusust, andmekvaliteeti ja sõnakasutust tubakareklaamides. Hoolimata tekstituvastuse OCR väljakutsetest (veaprotsent 37%), osutus koostatud andmestik pärast täpsustusi piisavalt asjakohaseks. Sõnasageduste ja kollokatsioonide analüüs andis esmase ülevaate reklaamides kasutatud terminitest, müügiargumentidest ning keelelistest eripäradest. Sagedus- ja kollokatsioonianalüüsid näitasid, et valitud põhimärksõnad iseloomustasid hästi uurimisobjekti, samuti tuli esile iseloomulik ajastuomane keelekasutus ("kõrge headus", "nõudma", "peenemaitseline"), rõhutati ka tervist ning tehnoloogilisi uuendusi ("pneumaatiline", "hügieeniline vatt").

[207] Päevaleht, 15.03.1931, lk 3.
[208] Sport ja auto, 06.06.1930, lk 3.
[209] vt Lisa 3.

Kokkuvõte

Bakalaureusetöö eesmärgiks oli luua DIGARi digiteeritud materjalide ja Rahvusraamatukogu digilaboriga andmestiku loomise ja analüüsi töövoog, millega astuda samm edasi traditsioonilistest ajalooalaste uurimustest, näidates kuidas digitaalsed tööriistad ja avatud teaduse põhimõtted võivad laiendada ajaloolaste, sotsiaalteadlaste, andmeteadlaste ja teiste uurijate metoodilisi võimalusi. Loodud töövoog ja korpus on GitHub'is vabalt kättesaadavad, mis pakub teistele uurijatele hüppelaua: neid saab kohandada nii reklaamide, ajaperioodide kui ka artiklite uurimiseks, et seeläbi laiendada digihumanitaaria tööriistakasti.

Töö uurimisobjektiks olid Eesti tubakavabrikute reklaamid aastatel 1920 – 1940. Konkurents tubakavabrikute vahel oli tihe ja tootmine suur, näiteks paberosside tootmine kasvas 12 miljonist (1920. a) kuni 1 miljard paberossini aastas (1927. a). Järk-järgult läks turg suurtootjate kontrolli alla. Reklaamipraktika arenes läbi aastate lihtsamatest teadaannetest keerukamate ning visuaalsemate reklaamideni. Suuremad ettevõtted palkasid reklaamide tegemiseks kunstnikke ja AS "Astoria" asutas lausa oma reklaamiagentuuri. Tööstusharu areng katkes 1940. aastal kui Nõukogude okupatsioon tõi kaasa kõigi tubakaettevõtete natsionaliseerimise.

Alusmaterjaliks valiti DIGARi arhiividest digiteeritud viis suuremat Eesti päevalehte (Päevaleht, Postimees, Sakala, Kaja ja selle järglane Uus Eesti). Ajalehtede andmetele pääses ligi tänu Rahvusraamatukogu digilabori tööriistadele. Tubakaettevõtete reklaamidest moodustati regulaaravaldistega nelja põhimärksõna ("tubak", "pabeross", "sigar" ja "suits") põhjal korpus, millel rakendati JupyterLabi ja RStudio vahendeid, et eraldada tubakareklaamid teistest reklaamidest.

Korpuse eeltöötlusel kasutati loomuliku keele töötluse meetodeid, kuid korpuse loomisel ja analüüsimusel kerkisid esile mitmed metodoloogilised väljakutsed. Peamiseks probleemiks osutus tekstituvastuse (OCR) kvaliteet, mis oli dekoratiivsete reklaamide puhul ebatäpne. Teiseks takistuseks kujunes DIGARi ajalehtede segmenteerimisviis, kus üks digitaalne tekstisegment võis sisaldada mitut erinevat reklaami. Selle probleemi vähendamiseks kasutati 200-tähemärgilist kontekstiakent põhimärksõnade ümber. Loodud märksõnapõhine korpus oli piisav, et sõnasageduste ja kollokatsioonianalüüse teostada. Tulevased uurijad, kes töövoogu kasutada soovivad, peavad korpuste loomisel otsima lahendusi nende probleemide lihvimiseks, parandades näiteks tehisintellektiga DIGARi ajalehtede segmenteerimist ja

tekstituvastust. Töövoo tõhususe hindamiseks viidi läbi ka lähilugemine juhuslikult valitud kuudel. Esialgne täpsus tubakaettevõtete reklaamide eristamisel teistest oli madal (43%), kuid pärast korpuse täiustamist tõusis see valimis 63%ni. Töö demonstreerib erinevate digihumanitaaria meetodite kasutamise võimalikkust, mille käigus selekteeriti välja millised andmeanalüüsi meetodid olid uurimuse eesmärkide saavutamiseks kõige sobivamad.

Leidmaks vastuseid püstitatud uurimisküsimustele, kasutati sõnasagedusi ja kollokatsioone ehk naabersõnu. Viie päevalehega loodud andmestikule põhinedes oli korpuses kokku 17 061 reklaami. Uurides milline sõnakasutus joonistub välja tubakareklaamide massanalüüsimisel, selgus et kõige sagedamini esinesid neis hinna ja kogusega seotud sõnad (sent, mark, tükk, nael), omadusi kirjeldavad sõnad (kõrge headus, maitse, aroom, mõnus, võluv, peenemaitseline) ning erinevad kaubamärgid (näiteks "Turist" ja "Kotkas"). Esile tulid ka tuntud isikute järgi nimetatud kaubamärgid, nagu "Ahto", mis viitas purjetaja Ahto Valterile või "Kuldmedal", osundades maadleja Kristjan Palusalule. Huvitavateks leidudeks olid sõnad "nõudma" ja "uudis". Lähilugemisel selgus, et nende sõnadega pöörduti tarbija poole soovitades neil tooteid "nõuda" või anti mõista, et neil toodetel on "kõrge nõudmine". Sõna "uudis" kasutati uue toote reklaamimisel. Uurides ettevõtteid võrdlevalt, jäi silma erinev sõnakasutuses – AS Astoria ja AS Laferme reklaamides tuli välja sõna "suitsetaja", ETK reklaamides aga "piibumees". ETK reklaamides esinenud sõnad "vein" ja "kohv" viitasid tootekataloogile. Enamuses reklaamidest oli ettevõtte nimi alati välja toodud.

Kõige enam jäid silma kaks aastat, 1923 ja 1930, mis võeti lähema vaatluse alla. 14.7% kogu korpuse reklaamidest (2500 reklaami) pärines aastast 1923, mida võib siduda tubakatööstuse kiire laienemisega. Nii 1923. kui 1930. aastate eripäradeks oli tehnoloogilise uuenduslikkuse toonitamine, tubakatooteid reklaamiti kui "keemiliselt" ja "pneumaatiliselt" puhastatud tooteid.

Antud töö eesmärgiks oli luua töövoog, mille abil saaksid ka teised uurijad digilabori võimalusi kasutada. Ühtlasi võimaldab see lisaks ajaleheartiklitele analüüsida ka reklaame. Loodud töövooga edasi minnes annaks tekstianalüüsile teha meelsusanalüüsi ja vastavaid teemasid sõnakasutusanalüüsi alusel mudeldada. Samuti oleks võimalik laiendada uurimist teistele reklaamikanalitele, hinnata artiklite ja reklaamide seoseid või ehitada tehisintellektiga pildituvastustööriist. Numbrite tokeniseerimisega ehk sõnestamisega saaks uurida paberosside hindade muutumist läbi aja.

Kasutatud allikad ja kirjandus

Publitseeritud allikad

Laferme, A–S. *Tubaka ajalugu: piltides*. Illustreerija Paul Aleksander Pedersen. [Tallinn]: Laferme, [193-?]. Elektrooniline reproduktsioon, Tartu: Eesti Kirjandusmuuseum, 2022. https://kivike.kirmus.ee/AR-22066-64996-20343.

Seadus napside ja likööride valmistamise ja müügi kohta, 03.09.1920. – *Riigi Teataja* 1920, nr 145 – 146, lk 1153 – 1154.

Tubakamaksu seadus, 10.04.1920. – *Riigi Teataja* 1920, nr 65 – 66, lk 517 – 522.

Tubakatööstuse aktsiaseltsi „Tubak" põhikiri – *Riigi Teataja* 1920, nr 69 – 70, lk 651 – 657.

Tubakatööstuse osaühisuse „Tubak" põhikiri – *Riigi Teataja* 1919, nr. 82 – 83, lk 553 – 555.

Käsikirjalised uurimused

Juurmaa, Ronald. "Eesti tubakatööstus 1919–1940 tubakavabrik 'Laferme' näitel." Magistritöö, Tartu Ülikool, Filosoofiateaduskond, Ajaloo ja arheoloogia instituut, juhendajad Jaak Valge ja Maie Pihlamägi, 2012.

Kannel, Marge. "Tubakatööstus Eesti Vabariigis 1920–1940." Lõputöö, Tartu Ülikool, Filosoofiateaduskond, Ajaloo osakond, juhendaja Jaak Valge, 2005.

Leman, Laura Katrin. "Tehisnärvivõrgul põhinevate lemmatiseerijate võrdlev analüüs eesti keeles." Bakalaureusetöö, Eesti ja üldkeeleteaduse instituut, Humanitaarteaduste ja kunstide valdkond, Tartu Ülikool, Juhendaja Kairit Sirts, 2019.

Šikirjavõi, Vladislav. "Temaatiliste mustrite kaevandamine seadustekstidest." Magistritöö, Tallinna Tehnikaülikool, Infotehnoloogia teaduskond, Tarkvarateaduse instituut, juhendaja Ahti Lohk, 2018.

Kirjandus

Aurnhammer, C., I. Cuppen, I. van de Ven ja M. van Zaanen. "Manual Annotation of Unsupervised Models: Close and Distant Reading of Politics on Reddit." *Digital Humanities*

Quarterly 13, nr. 3 (2019). https://www.digitalhumanities.org/dhq/vol/13/3/000431/000431.html (viidatud 17.05.2025).

Bendixen, Mike T. "Advertising Effects and Effectiveness." *European Journal of Marketing* 27, no. 10 (November 1, 1993): 19–32. https://doi.org/10.1108/03090569310045861

Bernays, Edward L. *Crystallizing Public Opinion*. New York: Boni and Liveright, 1923.

Blaheta, Don, ja Mark Johnson. "Unsupervised Learning of Multi-Word Verbs." In *Proceedings of the ACL Workshop on Collocations: Computational Extraction, Analysis and Exploitation*, Toulouse, France, 2001, lk 54 – 60.

Calder, Bobby J., and Edward C. Malthouse. "Media Engagement and Advertising Effectiveness." *Kellogg on Advertising & Media: The Kellogg School of Management*, edited by Bobby J. Calder, 1 – 36. Hoboken, NJ: Wiley, 2015. https://doi.org/10.1002/9781119198154.ch1

Eesti Reklaami Agentuur. *ERA Reklaam kataloog käsiraamat*. Tallinn: ERA, 1937. https://www.digar.ee/arhiiv/et/raamatud/15578

Eesti Reklaam-klubi. *Eesti Reklaam-klubi põhikiri: asutatud 24. veebr. 1929. a.* Tallinn: Eesti Reklaam-klubi, 1937. https://www.digar.ee/arhiiv/et/raamatud/95494

Grzybowski, Andrzej. "The History of Antitobacco Actions in the Last 500 Years. Part. II. Medical Actions." *Przeglad Lekarski* 63, nr. 10 (2006): 1131–1134. Kättesaadav: https://pubmed.ncbi.nlm.nih.gov/17288236/

Hartmann, Wesley R., and Daniel Klapper. "Super Bowl Ads." Stanford University Graduate School of Business Research Paper no. 15 – 16 (August 1, 2016). SSRN. https://ssrn.com/abstract=2385058

Kallas, Jelena, Maria Tuulik ja Madis Jürviste. "Leksikograafilise tarkvara Sketch Engine eesti keele moodul." *Eesti ja soome-ugri keeleteaduse ajakiri* 3, nr. 2 (2012): lk 57 – 77.

Karma, Otto. *Tööstuslikult revolutsioonilt sotsialistlikule revolutsioonile Eestis*. Tallinn: Eesti NSV Teaduste Akadeemia, 1963.

Kaukvere, Tiina. "Linnamuuseum avas vana reklaami näituse." *Postimees*, 12.04.2014. https://www.postimees.ee/2759506/linnamuuseum-avas-vana-reklaami-naituse

Kodumaa Saaduste Propaganda Keskkorraldus. *Eesti tööstus ja kaubandus.* Tallinn, 1934. Addressraamat. Kättesaadav: https://www.digar.ee/arhiiv/et/kollektsioonid/78753

Lauk, Epp. *Peatükke Eesti ajakirjanduse ajaloost*. Tartu: TÜ Kirjastus, 2000.

Michelson, J. "Tubakatööstus," *Eesti: maa. Rahvas. Kultuur*, toim Hans Kruus, lk 603–606. Tartu, 1926.

Nõulik, A., ja M. Ets, toim. *Tubakaraamat*. Tallinn: Eesti Entsüklopeediakirjastus, 2002.

Oberbichler, Sarah, ja Eva Pfanzelter. "Topic-specific corpus building: A step towards a representative newspaper corpus on the topic of return migration using text mining methods." *Journal of Digital History 1*, no. 1 (September 1, 2021): lk 74 – 98. https://doi.org/10.1515/jdh-2021-1003.

Paulus, Karin. *Eesti disaini ja reklaami 100 aastat*. Tallinn: Post Factum, 2018.

Samji, Hussein A., and Robert K. Jackler. "'Not one single case of throat irritation': misuse of the image of the otolaryngologist in cigarette advertising." *The Laryngoscope* 118, no. 3 (March 2008): lk 415 – 27. https://doi.org/10.1097/MLG.0b013e31815ad5c6.

Spiridis, Apollon. "Tubakasuitsu mõju organismi ja närvisüsteemile." Uurimustöö, Tartu Ülikool, 1930.

Stefanowitsch, Anatol. *Corpus Linguistics: A Guide to the Methodology*. Textbooks in Language Sciences 7. Berlin: Language Science Press, 2020. https://doi.org/10.5281/zenodo.3735822.

Talvik, Merle. "Ajakirjagraafika 1930. aastate Eestis: stereotüübid ja ideoloogia." Doktoritöö, Tallinna Ülikool, Kunstide Instituut, 2010.

Talvik, Merle. "Eesti kunstnikud ajakirjandusgraafikas 1930. aastail." *Mäetagused* 33 (2006): lk 7 – 40. https://doi.org/10.7592/mt2006.33.talvik

Tellis, Gerard J. *Effective Advertising: Understanding When, How, and Why Advertising Works*. Thousand Oaks, CA: Sage Publications, 2004.

Tinits, Peeter. "Digiteeritud Eesti ajalehed uurimisallikana." *Acta Historica Tallinnensia,* 2025 (ilmumas).

Tolonen, Mikko, Eetu Mäkelä, Jani Marjanen ja Tuuli Tahko. "Arvutiteaduse kaasamine humanitaarharidusse." *Methis : studia humaniora Estonica* 21, nr. 26 (2020): lk 35 – 50. https://doi.org/10.7592/methis.v21i26.16909.

Wevers, M. J. H. F. 2017. *Consuming America: A Data-Driven Analysis of the United States as a Reference Culture in Dutch Public Discourse on Consumer Goods, 1890–1990*. Doktoritöö, Utrecht University, 15.09.2017.

Wevers, Melvin, ja Jesper Verhoef. "Coca-Cola: An Icon of the American Way of Life. An Iterative Text Mining Workflow for Analyzing Advertisements in Dutch Twentieth-Century Newspapers." *Digital Humanities Quarterly* 11, nr. 4 (2017). https://www.proquest.com/scholarly-journals/coca-cola-icon-american-way-life-iterative-text/docview/2555183804/se-2 (viidatud 14.05.2025).

Wevers, Melvin, Jianbo Gao ja Kristoffer Laigaard Nielbo. "Tracking the Consumption Junction: Temporal Dependencies between Articles and Advertisements in Dutch Newspapers." *Digital Humanities Quarterly* 14 (2019). https://digitalhumanities.org/dhq/vol/14/2/000445/000445.html (viidatud 27.11.2024).

Wevers, Melvin. "Mining Historical Advertisements in Digitised Newspapers." *De Gruyter eBooks*, 2022, lk 227 – 52. https://doi.org/10.1515/9783110729214-011.

Perioodika

Edasi 1925.

Eesti Nädal 1934.

Eesti Postimees 1868, 1886, 1892, 1907, 1919, 1921 – 1923 1925, 1927, 1930 – 1934, 1937.

Eesti Spordileht 1930.

Esmaspäev 1923.

Kaja 1922 – 1926.

Nool 1930, 1939.

Päevaleht 1925 – 1931, 1933, 1937 – 1940.

Pealinna Teataja 1938.

Sakala 1897, 1923, 1936.

Sport ja auto 1930.

Uus Eesti 1936, 1938.

Loengud

Vilo, Jaak. Loeng „Tekst, infootsingud, masintõlge – ekskurss“ (kursus Digitaalne maailmapilt, LTAT.00.020), 7. loeng, Tartu Ülikool, kevadsemester 2025. Kättesaadav: https://courses.cs.ut.ee/2025/digit/spring/Main/L07

Veebileheküljed

DIGAR – Eesti Rahvusraamatukogu digiarhiiv. “DIGARist.” https://www.digar.ee/arhiiv/et/info/digarist (viidatud 14.05.2025).

Digilab – RaRa. https://digilab.rara.ee/ (viidatud 14.05.2025).

Digilab – RaRa. “Ligipääs DEA tekstidele.” https://digilab.rara.ee/tooriistad/ligipaas-dea-tekstidele/ (viidatud 14.05.2025).

Eesti Rahvusraamatukogu. “DIGARi Eesti artiklid.” *DIGARi Eesti artiklite portaal*. https://dea.digar.ee/?a=p&p=about (viidatud 14.05.2025).

HPC Public Documentation. “Jupyter.” 24.11.2020; muudetud 15.10.2024. https://docs.hpc.ut.ee/public/services/jupyter.hpc.ut.ee/

Lust, Tormi. “Tubakakorpus.” GitHub repositoorium. Viimati muudetud 17.mai 2025. https://github.com/tormil/tubakakorpus/

RStudio kasutusjuhend. Avaldatud 05.05.2025. https://docs.posit.co/ide/user/

RStudio Team. GitHub-repositoorium. https://github.com/rstudio/rstudio (viidatud 14.05.2025).

Stanford Research into the Impact of Tobacco Advertising. https://tobacco.stanford.edu/ (viidatud 14.05.2025).

The Comprehensive R Archive Network (CRAN). https://cran.r-project.org/ (viidatud 14.05.2025)

Tinits, Peeter. “Elekter, aur ja hobujõud 20. saj. vahetusel.” 16.11.2020. https://data.digar.ee/samples/elekter_aur_hobu.html (viidatud 14.05.2025).

Tinits, Peeter. “Estonian National Library Overviews.” OSF-projekt. https://doi.org/10.17605/OSF.IO/3GZXE (viidatud 14.05.2025).

University of Oslo Libraries. “Text Mining Types.” https://www.ub.uio.no/english/libraries/dsc/research-methods/text-mining/text-mining-types.html (viidatud 14.05.2025).

University of Wisconsin–Madison Writing Center. “Close Reading.” https://writing.wisc.edu/handbook/closereading/ (viidatud 14.05.2025).

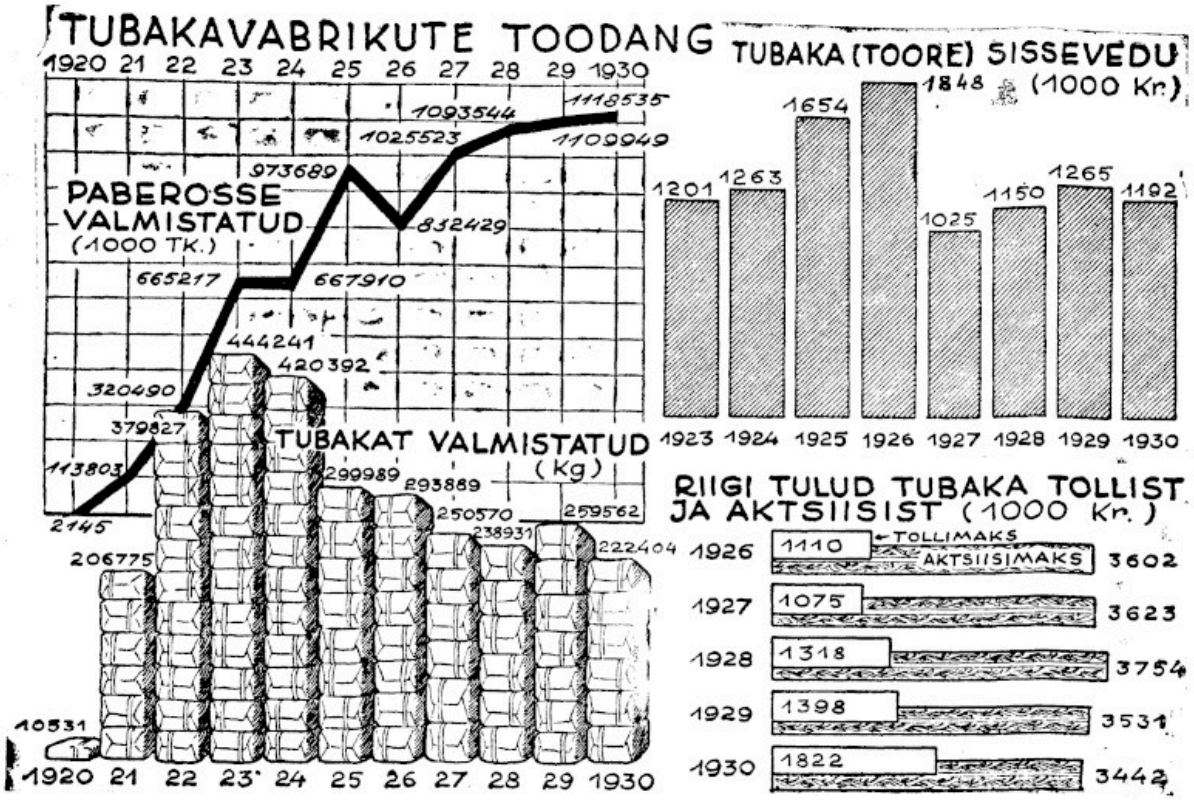
Lisa 1. Tubakatööstuse ettevõtete tööliste arv ja toodang ajavahemikus 1922 – 1938.[210]

Tabel 1. Tubakatööstuse ettevõtete töölistearv ja toodang.*

	Suurettevõtete arv Eesti tubaka-tööstuses**	Töötajate arv (aasta keskmine)	Tubakas (tonnides)	Paberossid (miljonites)	Sigarid (tuhandetes)	Toodangu netoväärtus (tuhandetes kroonides)
1922	6	582	379,8	323	132	3910,2
1923	7	771	444,2	665,2	12	2857,4
1924	7	658	420,4	667,9	-	4472
1925	7	706	300	973,7	-	5140,3
1926	7	746	293,9	832,4	-	5017,4
1927	7	722	250,6	1025,5	-	4825,9
1928	8	782	238,9	1093,5	228	4735,1
1929	8	767	259,6	1110	145	4842,4
1930	5	752	222,4	1118,5	75	5195,3
1931	4	732	198,9	1077	121	5620,5
1932	6	689	139,4	642,6	12	4687,6
1933	7	686	138	633,1	37	4567,3
1934	5	611	144,5	619,7	-	4226,9
1935	4	-	216,2	672,8	42	4325
1936	4	620	309,2	778	-	5055
1937	4	562	231,2	872	-	6680
1938	-	512	236,7	951	-	-

[210] Juurmaa, "Eesti tubakatööstus 1919–1940 tubakavabrik 'Laferme' näitel," lk 17. Tabel 1.

Lisa 2. Tubakavabrikute toodang, sissevedu ja riigi tulud tubaka tollist ja aktsiisist ajavahemikus 1920 – 1930.[211]



[211] Postimees 29.08.1931, lk 5.

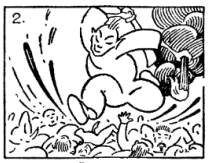
Lisa 3. Tubakatööstuse kapital, tööliste arv, sissevedu ja toodang ajavahemikus 1919 – 1924.[212]

	1919.a.	1920.a.	1921.a.	1922.a.	1923.a.	1924.a.
Tubakavabrikuid	1	5	8	11	11	11
Põhikapital (kogusummas Emk)	3.000.000	3.000.000	38.000.000	80.000.000	80.000.000	80.000.000
Tööliste arv	63	134	480	540	580	585
Toodang:						
Paberosse (tk)	5.000.000	12.000.000	156.196.400	341.296.800	665.200.000	667.900.000
Suitsutubakat (nl)[40]	-	-	410.314	816.628	1.084.800	1.108.160
Toore tubaka tarvitus:						
(sissevedu pd)	-[41]	1.748	17.541	39.217	50.700	48.829
Tubaka valmissaadus						
(sissevedu pd)	-	1411	601	268pd 21nl	135	121

[212] Kannel, "Tubakatööstus Eesti Vabariigis 1920–1940," lk 14. Tabel 1.

Lisa 4. 1925. a. AS “Astoria” värsireklaam, kunstnikuks Karl Jürgens.[213]

MIKS KAKLEB JÄLLE BALKAN ?..

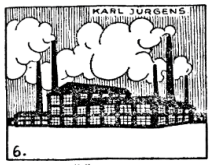


MIKS MÄRATSEB ? HIINA ?..

MIKS MÜRAB ? MAROKKO ?..

MIKS KOLLITAB ? KOMMUNIST ?..

MIKS ON RAHULIKLIL „REINU“-VADERID..?



MIKS TÖÖTAB „ASTORIA“ ÖÖL JA PÄEVAL ?..

SEST ET CHAMBERLAIN EI TEE — PABEROSSE „KABAREE“ TEEME TÖÖD KÜLL PÄEV JA ÖÖ, ETTE AGA SEE EI LÖÖ, — ÜKSINDA MAAILMA RAHUKS „ASTORIA“ EI JÄTKU KAHJUKS.

[213] Postimees 08.11.1925, lk 3.

Summary

Tobacco Advertisements in Estonian Newspapers 1920–1940: A Corpus-Based Analysis

The aim of the Bachelor's thesis is to create a workflow for building and analyzing datasets based on DIGAR's digitized materials and the National Library's digilab, thereby advancing beyond traditional historical research and demonstrating how digital tools and the principles of open science can expand the methodological possibilities for historians, social scientists, data scientists, and other researchers. The developed workflow and corpus are freely available on GitHub, providing a springboard for other researchers: they can be adapted for studying advertisements, time periods, or articles, thereby expanding the digital humanities toolkit.

The object of this study was the advertisements of Estonian tobacco factories from 1920 to 1940. Competition in the field was fierce and production was high; for example, cigarette production grew from 12 million to one billion cigarettes per year. Gradually, the market came under the control of large manufacturers. Advertising practices evolved over the years from simpler announcements to more complex and visual advertisements. Larger companies hired artists for their advertising, and AS "Astoria" even established its own advertising agency. The development of the industry was cut short in 1940, when the Soviet occupation led to the nationalization of all tobacco companies.

As source material, five major Estonian daily newspapers digitized in DIGAR's archives were chosen: Päevaleht, Postimees, Sakala, Kaja, and its successor Uus Eesti. Access to the newspaper data was provided by the National Library of Estonia's digilab. A corpus of tobacco company advertisements was formed using regular expressions with four main keywords ("tubak" [tobacco], "pabeross" [cigarette], "sigar" [cigar], and "suits" [smoke]), and JupyterLab and RStudio tools were applied to distinguish tobacco advertisements from other ads.

Natural language processing methods were used for preprocessing the corpus, but several methodological challenges arose during the creation and analysis of the corpus. The main problem was the quality of text recognition (OCR), which proved inaccurate for decorative ads. Another obstacle was DIGAR's method of segmenting newspaper ads, where a single digital text segment could contain several different ads. In order to mitigate this issue, a

200-character context window around each keyword was used. The resulting keyword-based corpus was sufficient to carry out word frequency and collocation analyses. Future researchers wishing to use this workflow should seek solutions for refining these problems, for example, by improving DIGAR's newspaper segmentation and text recognition with artificial intelligence. With the aim of evaluating the effectiveness of the workflow, close reading was also conducted on randomly selected months over the 20-year period. The initial accuracy in distinguishing tobacco company ads from others was low (43%), but after improving the corpus, this rose to 63% in the sample. The work demonstrates the feasibility of using various digital humanities methods and shows which data analysis techniques were most suitable for achieving the research objectives.

To solve the research questions posed, word frequencies and collocations (neighboring words) were analyzed. Based on the dataset constructed from the five newspapers, there were a total of 17,061 advertisements in the corpus. Examining the patterns of word usage in a mass analysis of tobacco ads, the most common words were related to price and quantity (sent [cent], mark [mark], tükk [piece], nael [pound]), descriptive adjectives (kõrge headus [high quality], maitse [taste], aroom [aroma], mõnus [pleasant], võluv [charming], peenemaitseline [refined]), and mentions of brand names ("Turist" and "Kotkas" [Eagle]). Also highlighted were brand names referring to well-known figures, such as "Ahto" (referring to the sailor Ahto Valter), "Kuldmedal" [Gold Medal] (referring to the wrestler Kristjan Palusalu). Interesting findings included words such as "nõudma" (to demand/request) and "uudis" (news). Close reading revealed that these words were used to address consumers, urging them to "demand" products or implying that the products were in "high demand." The word "uudis" was used to promote new products. Differences in word usage between companies also emerged—for example, "suitsetaja" ("smoker") appeared in the ads of AS Astoria and AS Laferme, while "piibumees" ("pipe smoker") was typical for ETK (Estonian Consumers' Cooperatives Union) ads. The words "vein" ("wine") and "kohv" ("coffee") found in ETK advertisements indicate that their products were often promoted via product catalogues. In most companies' advertisements, the company name was always stated.

Two years, 1923 and 1930, stood out and were examined more closely. 14.7% of all advertisements in the corpus (2,500 ads) came from 1923, which can be linked to the rapid expansion of the tobacco industry. Both 1923 and the 1930s were characterized by an

emphasis on technological innovation, with tobacco products advertised as “chemically” and “pneumatically” purified.

The purpose of this work was to create a workflow that enables other researchers to also utilize the possibilities of the digital lab. In addition to newspaper articles, it also allows for the analysis of advertisements. Moving forward with the created workflow, in addition to text analysis, sentiment analysis and topic modeling can be applied. It would also be possible to expand the study to other advertising channels, assess the relationships between articles and advertisements, or develop an AI-based image recognition tool. By tokenizing numbers, one could study changes in cigarette prices over time.

Lihtlitsents lõputöö reprodutseerimiseks ja üldsusele kättesaadavaks tegemiseks

Mina, **Tormi Lust**

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) minu loodud teose **Tubakareklaamid Eesti ajalehtedes aastatel 1920-1940: korpusepõhine analüüs**,

mille juhendajad on Peeter Tinits ja Aigi Rahi-Tamm, reprodutseerimiseks eesmärgiga seda säilitada, sealhulgas lisada digitaalarhiivi DSpace kuni autoriõiguse kehtivuse lõppemiseni.

2. Annan Tartu Ülikoolile loa teha punktis 1 nimetatud teos üldsusele kättesaadavaks Tartu Ülikooli veebikeskkonna, sealhulgas digitaalarhiivi DSpace kaudu Creative Commonsi litsentsiga CC BY NC ND 3.0, mis lubab autorile viidates teost reprodutseerida, levitada ja üldsusele suunata ning keelab luua tuletatud teost ja kasutada teost ärieesmärgil, kuni autoriõiguse kehtivuse lõppemiseni.

3. Olen teadlik, et punktides 1 ja 2 nimetatud õigused jäävad alles ka autorile.

4. Kinnitan, et lihtlitsentsi andmisega ei riku ma teiste isikute intellektuaalomandi ega isikuandmete kaitse õigusaktidest tulenevaid õigusi.

Tormi Lust
19.05.2025