

Opinnäytetyö (AMK)

Liiketalouden koulutus, kirjasto- ja tietopalvelu

2024

Olli Virtanen

Erikoismerkit Turun yliopiston kirjaston Volter-tietokannassa



Opinnäytetyö (AMK) | Tiivistelmä

Turun ammattikorkeakoulu

Liiketalouden koulutus, kirjasto- ja tietopalvelu

2024 | 34 sivua

Olli Virtanen

Erikoismerkit Turun yliopiston kirjaston Volter-tietokannassa

Opinnäytetyön toimeksiantaja on Turun yliopiston kirjasto. Tavoitteena oli selvittää, miten Turun yliopiston kirjaston Volter-tietokanta käsittelee latinalaisten aakkosten erikoismerkkejä sisältäviä sanoja ja miten tämä heijastuu hakutuloksiin.

Tutkittaviksi kieliksi valittiin suomi, ruotsi, saksa, ranska, italia ja espanja. Valinta tehtiin sillä perusteella, että nämä kielet käyttävät latinalaisia aakkosia, niiden aakkostot sisältävät erikoismerkkejä ja niitä voi opiskella Turun yliopistossa pääaineena. Selvitystyötä varten valikoitiin näiden kielten erikoismerkkejä sisältäviä sanoja, jotka tuottavat Volterissa tämän opinnäytetyön tarkoitukseen sopivan määrän hakutuloksia, noin 20–200 kappaletta. Näitä sanoja haettiin Volterissa niin erikoismerkkien kanssa kuin ilmankin, ja tarkasteltiin, millainen vaikutus kirjoitusasulla on hakutulospäämääriin.

Työssä pohdittiin myös digitaalisten palveluiden vaikutusta asiakaskokemukseen ja asiakkaiden tyytyväisyyteen. Volter-tietokanta on Turun yliopiston kirjaston digitaalisista palveluista käytetyin, joten sillä on asiakastyytyväisyyteen merkittävä vaikutus.

Asiasanat:

diakriittiset merkit, kirjastot, tietokantaohjelmat

Bachelor's Thesis | Abstract

Turku University of Applied Sciences

Degree programme in Business Administration,
Library and Information Services

2024 | 34 pages

Olli Virtanen

Diacritics in the Volter-database of Turku University Library

Keywords:

Database Programs, Diacritics, Libraries

Sisältö

1 Johdanto	5
2 Primo VE	6
3 Digitaalisen palvelun vaikutus asiakaskokemukseen	7
4 Työn rajaus ja erilaisia tarkemerkkejä	9
5 Volter-hakujen tulokset	11
6 Yhteenveto	28
Lähteet	31

1 Johdanto

Turun yliopiston kirjaston asiakastietokanta on nimeltään Volter (kirjaston ensimmäisen ylikirjastonhoitajan, kirjailija Volter Kilven mukaan). Volter-tietokannan käyttöliittymäohjelma on nimeltään Primo VE. Kyseisen ohjelman käytössä on ilmennyt ongelmia latinalaisen aakkoston erikoismerkkien suhteen. Useat latinalaisia aakkosia käyttävät kielet sisältävät erilaisia tarke- eli diakriittisiä merkkejä, kirjaimen ylä- tai joskus alapuolelle sijoitettavia pisteitä, aksenttimerkkejä ja muita lisukkeita; tällaisia ovat esimerkiksi suomen ä ja ö, ruotsin å, saksan ü ja ranskan é. Ohjelman logiikka on tarkemerkkien suhteen epäselvä, sillä haut tarkemerkin kanssa tai ilman sitä tuottavat usein näennäisen satunnaisesti toisistaan poikkeavia tulospäämiä. Ei ole selvää, mistä tulospäämien heitot johtuvat, eikä myöskään, missä määrin vähemmän osumia tuottavien hakujen tulokset sisältyvät enemmän osumia tuottavien hakujen tuloksiin. Turun yliopiston kirjasto on ottanut yhteyttä ohjelman luoneeseen yritykseen, mutta vastausta ei ole tullut.

Tämän opinnäytetyön tarkoituksena on yrittää selvittää ohjelman toimintalogiikkaa. Selvitystyötä varten on valittu sanoja, jotka sisältävät erikoismerkkejä, ja tutkailtu, millaisia hakutulospäämiä ne tuottavat eri tavoin kirjoitettuina. Sanat on pyritty valitsemaan niin, että ne tuottaisivat sopivan määrän hakutuloksia, mikä käytännössä tarkoittaa noin 20–200 tulosta. Liian suuren tulospäämän läpikäynti ei olisi tämän työn puitteissa mahdollista, toisaalta liian pienestä määrästä ohjelman logiikkaa olisi vaikea päätellä.

Työssä pohditaan myös sitä, miten digitaaliset palvelut vaikuttavat asiakaskokemukseen. Volter-tietokanta on suurimmalle osalle Turun yliopiston kirjaston asiakkaista se palvelu, jota he eniten käyttävät, joten sen merkitys asiakkaille kirjastosta muodostuvaan kuvaan on huomattava.

2 Primo VE

Primo VE -ohjelman on luonut israelilainen Ex Libris Group Ltd., joka puolestaan on nykyään osa brittiläis-yhdysvaltalaisista Clarivate-analytiikkayritystä. Primo VE:n kotisivuilla (ExLibris Knowledge Center) kerrotaan, että ohjelma tunnistaa automaattisesti seuraavat kielet:

- latinalaisia aakkosia käyttävistä englannin, saksan, ranskan, espanjan, italian ja tanskan
- "aasialaisista" kiinan, japanin ja korean
- muista kielistä arabian, heprean "ja niin edelleen"

Ohjelma tunnistaa kielen vertailemalla hakusanoja (tarkemmin määrittelemättömiin) sanakirjoihin: mikäli vastaavuus on vähintään 50 %, tunnistus onnistuu.

Kotisivuilla kerrotaan myös, että ohjelma osaa normalisoida ("normalize"), toisin sanoen muuttaa erikoismerkkejä sisältävät kirjaimet erikoismerkitöiksi kirjainyhdistelmiksi, seuraavien kielten osalta: saksa, islanti, liettua, norja, tanska, ruotsi, espanja, puola, korea, kiina ja japani. Sitä, millä perusteella juuri nämä kielet on valikoitu, ei sen tarkemmin perustella. Tämän työn kannalta olennaisin tieto on, että ohjelma tunnistaa ae-, oe- ja ue-yhdistelmät ä-, ö- ja ü-kirjaimiksi, toisin sanoen siis esimerkiksi haku "haen" tuottaa myös "hän"-haun tulokset. Toiseen suuntaan tämä ominaisuus ei kuitenkaan toimi, eli "hän"-haku ei tuota "haen"-haun tuloksia.

3 Digitaalisen palvelun vaikutus asiakaskokemukseen

Volter-tietokanta on Turun yliopiston kirjaston digitaalisista palveluista se, jonka kanssa keskivertoasiakas todennäköisesti on eniten tekemisissä. Näin ollen sen vaikutus asiakaskokemukseen on huomattava.

Tuotteen tai palvelun *käytettävyys* muodostuu kolmesta osa-alueesta: tuloksellisuudesta, tehokkuudesta ja miellyttävyydestä (de Mooij ym. 2005, 178). On selvää, että mikäli Volter-tietokannan käyttäjä ei löydä hakemaansa tietoa tai vähintään joutuu tekemään useita hakuja, kaikki kolme käytettävyiden osa-alueita kärsivät. Tulosten saaminen on epävarmaa tai ainakin aikaa vievää, eikä kokemus varmasti ole erityisen miellyttävä. Lyhyesti sanottuna siis käyttökokemuksesta ei muodostu niin myönteistä kuin olisi mahdollista.

Käyttökokemus tarkoittaa tuotteen tai palvelun käyttöön liittyviä tunnuseikkoja. Käyttökokemuksella on ratkaiseva merkitys siinä, millaisen käsityksen käyttäjä saa yrityksestä tai organisaatiosta. (de Mooij ym. 2005, 180) Mikäli käyttökokemus on miellyttävä, käyttäjäkin on tyytyväinen ja suosittelee tuotetta tai palvelua usein myös tuttavilleen. Mikäli kokemus puolestaan on epämiellyttävä, sana lähtee kiertämään vähintään yhtä helposti. On tärkeää muistaa, että asiakaskokemus on aina täysin yksilöllinen, eikä asiakkaan kokemusta voi ruveta kiistämään (Filenius 2015, 25.) Tässä asiassa asiakkaan voi siis sanoa olevan aina oikeassa.

Asiakaskokemukseksi ymmärretään usein pelkkä henkilötason kohtaaminen joko kasvokkain tai puhelimesta. Tämä on toki olennaisen tärkeä osa asiakaskokemusta, mutta se sisältää muitakin osa-alueita. Nykyaikana digitaalinen asiakaskokemus on tietotekniikan myötä muuttumassa yhä tärkeämmäksi, ja lisäksi asiakaskokemukseen kuuluu niin sanottu tiedostamaton tunnekokemus. Nämä yhdessä synnyttävät *kokonaisvaltaisen* asiakaskokemuksen. (Ahvenainen, Gylling & Leino 2017, 34)

Digitaalinen asiakaskokemus syntyy silloin, kun asiakas käyttää digitaalisesti tuotettua palvelua. Yksittäisellä sivulla tai ohjelmalla, oli se miten hyvä tahansa, ei vielä voida taata onnistunutta asiakaskokemusta, mutta yksittäinenkin epäonnistuminen voi sen kyllä pilata. (Filenius 2015, 30.) Digitaalinen maailma on siis tässä suhteessa varsin armoton. Sivustojen ja ohjelmistojen toimivuuteen onkin syytä kiinnittää huomiota, jotta yksi viallinen osa ei pilaa kokonaisuutta.

Käytettävyyden parantaminen on parhaita keinoja tehdä digitaalisesta käyttökokemuksesta entistä myönteisempi. On selvää, että mitä suurempaa hyötyä asiakkaat kokevat tuotteesta tai palvelusta saavansa, sitä tyytyväisempiä he ovat (de Mooij ym. 2005, 19). Se, että tieto löytyy helposti ja nopeasti, on puolestaan omiaan lisäämään asiakkaan luottamusta yrityksen palveluihin (Komulainen 2018, 23).

Onkin siis huomattava, että digitaalisissa kanavissa pelkkä tiedon saanti ei ole ratkaisevaa, vaan myös asiakaskokemuksella on erittäin suuri merkitys (Komulainen 2018, 24). Tässä nimenomaisessa tapauksessa ne tosin lienevät pitkälti sama asia, koska tieto on se tuote, jota tieteellinen kirjasto "myy". Juuri tämän vuoksi kirjastojen on tärkeää pitää huolta siitä, että tieto on paitsi tarjolla, myös mahdollisimman helposti ja vaivattomasti löydettävissä.

4 Työn rajaus ja erilaisia tarkemerkkejä

Tässä kappaleessa määritellään tutkittavat kielet ja esitellään yleisimmät eurooppalaisperäisten kielten aakkostoissaan käyttämät tarkemerkit.

Työn rajaus

Tämä työ keskittyy pääasiassa niihin kieliin, jotka ovat suomalaisille opiskelijoille tutuimpia ja joilla siis voi olettaa tehtävän Volter-tietokannassa eniten hakuja. Näitä kieliä ovat tekijän arvion mukaan suomi, ruotsi, saksa, ranska, espanja ja italia, joista kaikkia voi opiskella Turun yliopistossa pääaineena. Luonnollisesti myös englanti kuuluu tähän ryhmään, mutta englannin perusaakkosto ei sisällä tarkemerkkein varustettuja kirjaimia, joten se rajautuu tämän työn aihepiirin ulkopuolelle. Lisäksi Turun yliopistossa voi opiskella pääaineena venäjää ja kiinaa, mutta nämä kielet eivät tunnetusti käytä latinalaista aakkostoa, joten niiden tutkiminen ei tule tässä työssä kyseeseen.

Treema

Treema merkitsee kirjaimen yläpuolista pisteparia. Treemallisia kirjaimia esiintyy suomalaisille tutuista kielistä muun muassa suomessa (ä, ö), ruotsissa

(ä, ö), saksassa (ä, ö, ü) (GermanVeryEasy.com), virossa (ä, ö, ü) (Omniglot) ja ranskassa (ë, î, ù) (Langster), minkä lisäksi niitä on esimerkiksi islannissa (ö) (Omniglot), slovakissa (ä) (Slovak.EU), albaniaassa (ë) (Omniglot), unkarissa (ö, ü) (Omniglot) ja turkissa (ö, ü) (Middle East and North African Languages Program).

Akuutti

Akuutti on oikealle kallistuva aksenttimerkki. Se esiintyy yleisenä esimerkiksi ranskassa (é) sekä portugalissa (PortugueseLanguage), tšekissä (Omniglot) ja unkarissa (á, é, í, ó, ú, tšekissä myös ý, unkarissa ő ja ű). Lisäksi akuuttia aksenttia käytetään jonkin verran myös italiassa ja espanjassa ja harvakseltaan esimerkiksi hollannissa ja skandinaavisissa kielissä (esim. ruotsin idé, armé). Puolan kielessä akuuttia aksenttia käytetään tiettyjen konsonanttien yhteydessä (ń, ś, ź) merkitsemään liudennusta eli konsonantin ääntymistä j-mäisenä (Mówić Po Polsku).

Gravis

Gravis on vasemmalle kallistuva aksenttimerkki. Sitä käytetään suomalaisille tutuista kielistä lähinnä ranskassa (è sekä tietyissä sanoissa à ja ù) ja italiassa (à, è, ì, ò, ù) (ItalianPod101.com).

Sirkumfleksi

Sirkumfleksi muistuttaa ylöspäin osoittavaa nuolta. Sitä käytetään esimerkiksi ranskassa (â, ê, î, ô, û), portugalissa (â, ê, ô) ja romaniassa (â, î) (Omniglot) sekä suomalaisille hieman vieraammista kielistä muun muassa kymrissä (Wales Guidebook).

Tilde

Tilde eli aaltoviiva on vokaalien yhteydessä käytössä virossa (õ) ja portugalissa (ã, õ). Aaltoviivan merkitys on näissä kielissä eri: viron õ ääntyy takaisena e-äänteenä, portugalin ã ja õ puolestaan ovat nasaalisia eli nenän kautta ääntyviä

vokaaleja. Espanjan kielessä (Spanish411.net) aaltoviiva on puolestaan käytössä kirjaimessa ñ, joka ääntyy liudentuneena n-äänteenä.

Muita tarke- ja erikoismerkkejä

Edellä esiteltyjen merkkien lisäksi latinalaisissa aakkostoissa käytettyjä tarkkeita ovat muun muassa hattu kirjaimissa ě, ň, š, ž ja isoissa Ď- ja Ť-kirjaimissa (tšekki, slovakki), yläpuolinen piste (puolan ż, liettuan é [Academic Accelerator]), yläpuolinen ympyrä (ruotsin, norjan [Life in Norway] ja tanskan [Danishnet.com] å, tšekin ů), lyhyysmerkki eli breve (romanian ă, turkin ğ), pituusmerkki eli makron (latvian [Wikiwand] ā, ē, ī, ū, liettuan ū) ja sedilji (ranskan, portugalin ja turkin ç). Omia kirjaimiaan ovat esimerkiksi saksan ß ja norjan ja tanskan æ ja ø.

5 Volter-hakujen tulokset

Jokaisesta tutkittavana olevasta kielestä on valikoitu tarkastelun kohteeksi joukko sanoja, jotka sisältävät kyseisessä kielessä käytettyjä erikoismerkkejä. Sanojen valikointiperustana on käytetty erikoismerkkien lisäksi niiden tuottamien hakutulosten määrää: käytettävissä olevilla menetelmillä pystyy analysoimaan tarkemmin enintään noin 200 hakutulosta antavia sanoja. Mikäli hakutuloksia on tätä enemmän, tulosten tarkan erittelyn sijaan pitää tyytyä yksinkertaisesti tarkastelemaan eri hakusanojen tuottamien osumamäärien eroja.

Kaikkien hakusanojen perään on merkitty hakupäivä, sillä Volter päivittyy jatkuvasti ja tuloksissa tapahtuu muutoksia.

Suomenkieliset hakusanat

Suomenkielisiksi hakusanoiksi on valikoitu ä:n ja ö:n sisältäviä sanoja. Sanoja on etsitty kolmella tavalla: normaalilla oikeinkirjoituksella, ilman yläpuolisia pisteitä (ä → a, ö → o) sekä e:llisinä yhdistelminä ae ja oe. Suomen kielessä ä:tä ja ö:tä ei ole tapana korvata ae- ja oe-yhdistelmillä (ainakin yhdessä sanaparissa, hän vs. haen, ä:n ja ae-yhdistelmän välinen valinta aiheuttaa

merkityseron), mutta tämä tapa on käytössä esimerkiksi saksassa ja joskus englannissakin, joten sen sisällyttäminen hakuihin on perusteltua.

Hälytys

Hakusana *hälytys* tuottaa Volterissa 82 osumaa (27.10.2023). *Halytys* tuottaa osumia 30 ja *haelytys* 52 (27.10.2023). On mielenkiintoista, että *halytys*- ja *haelytys*-haut eivät tuota yhtäkään samaa tulosta, toisin sanoen siis *hälytys*-haun tulokset jakautuvat jäännöksettömästi noiden kahden hakusanan kesken. *Halytys* tuottaa selvästi enemmän englanninkielisiä tuloksia, 18/30, kuin *haelytys*, 4/52. Lisäksi *haelytys* tuottaa neljä ruotsinkielistä tulosta, jollaisia *halytys* ei tuota yhtään.

Neljännes

Hakusana *neljännes* tuottaa Volterissa 95 osumaa (27.10.2023). *Neljannes* antaa 16 tulosta, *neljaennes* puolestaan 79. Näiden hakusanojen kohdalla toistuu sama ilmiö kuin *hälytys*-sanana variaatioiden kohdalla: *neljannes* ja *neljaennes* eivät tuota yhtäkään samaa hakutulosta. *Neljannes*-hakusanan kaikki osumat ovat suomenkielisiä, kun taas *neljaennes*-sanana tuloksista 18 on englanninkielisiä. Tulokset ovat siis päinvastaisia kuin *hälytys*-sanalla, jossa a:llinen kirjoitusasu antoi enemmän englanninkielisiä tuloksia kuin ae:llinen.

Kaikissa *neljaennes*-hakusanan osumissa Volter ilmaisee myös teoksen julkaisuvuoden. *Neljannes*-haun tuloksista julkaisuvuosi sen sijaan ilmoitetaan vain kahdessa.

Äkillinen

Hakusana *äkillinen* tuottaa Volterissa 135 osumaa (27.10.2023). *Akillinen* tuottaa 75 osumaa ja *aekillinen* 61 osumaa (27.10.2023), yhteensä siis 136 tulosta. Nämä tulokset jakautuvat niin, että yksi ainoa tulos on ”päällekkäinen”, toisin sanoen sekä *akillinen* että *aekillinen* tuottavat sen. Tämä kyseinen tulos on G. F. Aminoffin ”Äkillinen kihlaus: Leikki-näytelmä ; yhdessä näytöksessä”

vuodelta 1861. (Kirjoitusasu virheineen on kopioitu suoraan Volterista. Oikeasti näytelmän nimi on ”Äkillinen kihlaus”, ei suinkaan ”Akillinen”). Tämä hakutulos on huomattavasti vanhempi kuin kaikki muut hakutulokset, joiden julkaisuvuosi on merkitty Volteriin; toiseksi vanhimmat tulokset ovat vuodelta 1974.

Teosten julkaisuvuoteen liittyy toinenkin mielenkiintoinen havainto, joka on linjassa *neljannes*-hakusanan tulosten kanssa. Kaikissa *aekillinen*-haun 61 osumassa on mainittu teoksen julkaisuvuosi, kun taas *akillinen*-haun 75 osumasta julkaisuvuosi mainitaan ainoastaan viidessä. Ei ole suoraan selvää, miten tämä ero liittyy erikoismerkkeihin, mutta se on joka tapauksessa silmäänpistävä. Esimerkiksi edellä mainitusta kihlajaisnäytelmästä on toinenkin hakutulos, mutta siinä ilmestymisvuotta ei ole mainittu. Se löytyy *akillinen*-haulla, muttei *aekillinen*-haulla.

Akillinen-haun tuloksista englanninkielisiä on 35, *aekillinen*-haun tuloksista 20. A:llinen muoto tuottaa siis enemmän englanninkielisiä tuloksia kuin ae:llinen, mutta ero ei ole yhtä selkeä kuin kahdella aikaisemmalla hakusanalla.

Miehistö

Hakusana *miehistö* tuottaa Volterissa 67 osumaa (26.10.2023). *Miehisto* tuottaa osumia vain kaksi, *miehistoe* 65 (26.10.2023); samoja tuloksia nämä hakusanat eivät tuota.

Mökki

Hakusana *mökki* tuottaa Volterissa 72 osumaa (29.9.2023). Mikäli ö:n pisteet jättää pois ja hakee sanalla *mokki*, tuloksia tulee vain kymmenen (29.9.2023). Mielenkiintoista on, että hakusana *moekki* tuottaa 67 tulosta (29.9.2023).

Kaikki *moekki*-hakusanan 67 tulosta mahtuvat *mökki*-haun 72 tulokseen. *Mokki*-sanon suhteen tilanne on erilainen: kymmenestä hakutuloksesta vain viisi sisältyy *mökki*-hakuun, toiset viisi sen sijaan eivät. Mielenkiintoista on, että ne *mokki*-haun viisi tulosta, jotka sisältyvät *mökki*-hakuun, ovat juuri ne samat viisi, jotka *moekki*-hausta jäävät puuttumaan; toisin sanoen siis *mokki*- ja *moekki*-haut eivät tuota yhtäkään samaa tulosta.

Tarkastellaan lähemmin niitä viittä *mokki*-hakusanan tulosta, jotka sisältyvät myös *mökki*-hakuun. Niistä kolme on vieraskielisiä artikkeleita, kaksi englanninkielisiä ja yksi espanjankielinen. Sen paremmin englannin kuin espanjankaan aakkosto ei sisällä ö:tä, mikä lienee osaltaan selityksenä siihen, että Volter käsittelee näiden artikkelien sisältämää ”mökki”-substantiivia samanmuotoisena kuin o:llista muotoa ”mokki”. Loput kaksi hakutulosta sisältävät suomenkielisen runon tai romaanin englanninkielisellä alustalla, eli jälleen kerran alustan kielellä vaikuttaa olevan merkitystä siinä, mitä tuloksia Volter millekin hakusanalle antaa.

Tätä olettamusta vahvistavat myös ne viisi *mokki*-haun tulosta, jotka eivät sisällä *mökki*-hakuun. Ne kuuluvat kaikki helsinkiläisen Kehittämiskeskus Opinkirjon Mokki-kirjasarjaan, ja ovat siis suomenkielisiä. Hakutulokset voivat toisin sanoen antaa myös sellaisia osumia, jotka eivät lähtökohtaisestikaan sisällä erikoismerkkejä.

Öljynporaus

Öljynporaus tuottaa 19 hakutulosta, *oljynporaus* kaksi ja *oeljynporaus* 17 (26.10.2023). Jälleen kerran huomaamme, etteivät pisteetön variantti ja e-yhdistelmä tuota lainkaan samoja tuloksia.

Ruotsinkieliset hakusanat

Ruotsinkielisiksi hakusanoiksi on valikoitu å:n, ä:n ja ö:n sisältäviä sanoja. Sanoja on etsitty kolmella tavalla: normaalilla oikeinkirjoituksella, ilman yläpuolisia merkkejä (å,ä → a, ö → o) sekä ä:n ja ö:n tapauksissa e:llisinä yhdistelminä *ae* ja *oe*.

Fångenskap

Hakusana *fångenskap* (vankeus) tuottaa Volterissa 238 osumaa, *fangenskap* 117 osumaa (23.10.2023). Kaikki *fangenskap*-haun tulokset sisältyvät *fångenskap*-hakuun. Vertailtaessa hakusanojen tuottamia osumia nousee esiin eräs osumien kieleen liittyvä seikka: *fångenskap*-haku tuottaa 84 suomenkielistä osumaa, *fangenskap*-haku sen sijaan ei yhtäkään. Å:n

puuttuminen näyttää siis tämän hakusanan kohdalla suodattavan suomenkieliset tulokset kokonaan pois.

Minderårig

Hakusana *minderårig* (alaikäinen) tuottaa 115 hakutulosta, *minderarig* 28 tulosta (23.10.2023). Havaitsemme saman kuin edellisenkin hakusanan kohdalla: *minderårig* tuottaa 74 suomenkielistä hakutulosta, *minderarig* ei yhtäkään.

Årsdag

Hakusana *årsdag* (vuosipäivä) tuottaa Volterissa 142 tulosta, *arsdag* 46 tulosta (23.10.2023). Jälleen kerran å:n muuttaminen a:ksi tiputtaa suomenkieliset tuokset pois: *årsdag* tuottaa 13 suomenkielistä tulosta, *arsdag* ei yhtäkään.

Lärrinna

Hakusana *lärrinna* (opettajatar) tuottaa Volterissa 20 hakutulosta, *lararina* 14 tulosta ja *laerarina* kuusi tulosta (23.10.2023). A:llinen ja ae:llinen muoto eivät, kuten suomenkielistenkään sanojen kohdalla, tuota lainkaan samoja tuloksia. Volter ilmoittaa kaikkien *laerarina*-haun tulosten julkaisuvuosiluvut, mutta *lararina*-haun tuloksista vuosiluku on mainittu vain kahdessa, joissa se on molemmissa 1913.

Ryggmarg

Hakusana *ryggmarg* (selkäydin) tuottaa Volterissa 104 osumaa, *ryggmarg* 68 ja *ryggmaerg* 39 tulosta (24.10.2023). Sekä a:llisen että ae:llisen haun kaikki tulokset sisältyvät ä:lliseen hakuun, joten huomaamme, että tällä kertaa nämä kaksi hakua tuottavat kolme kappaletta samoja tuloksia. Kyseiset teokset on

julkaistu vuosina 1840, 1847 ja 1869, eli kyse on, samoin kuin *äkillinen-* sanankin kohdalla, varsin vanhoista julkaisuista.

Äventyrare

Hakusana *äventyrare* (seikkailija) tuottaa Volterissa 193 hakutulosta, *aventyrare* 152 tulosta ja *aeventyrare* 41 tulosta (24.10.2023). Jälleen kerran a:llisen ja ae:llisen haun tulokset jakautuvat niin, ettei tuloksissa ole lainkaan samoja teoksia. Mielenkiintoista on, että *aventyrare*-haku ei tuota lainkaan suomenkielisiä tuloksia, mutta *aeventyrare* tuottaa niitä peräti 33 kappaletta; tämä muistuttaa hieman sitä jakaumaa, jonka å:n sisältävät hakusanat tuottivat.

Mörkret

Hakusana *mörkret* (pimeys, määräinen muoto) tuottaa Volterissa 101 hakutulosta, *morkret* 55 tulosta ja *moerkret* 49 tulosta (24.10.2023). Tämän hakusanan kohdalla o:llinen ja oe:llinen muoto tuottavat kolme samaa tulosta, mikä vaikuttaa olevan jossakin määrin poikkeuksellista ö:n sisältävien hakusanojen kohdalla. Kaikki nämä kolme tulosta ovat ruotsinkielisiä, eikä yhteenkään niistä valitettavasti ole merkitty julkaisuvuotta.

Nödfall

Hakusana *nödfall* (hätätapaus) tuottaa Volterissa yhdeksän hakutulosta, *nodfall* kahdeksan ja *noedfall* yhden (24.10.2023). Kaikki *nodfall*-haun tulokset ovat ruotsinkielisiä, *noedfall*-haun ainoa tulos on sitä vastoin suomenkielinen.

Ögonvittne

Hakusana *ögonvittne* (silminnäkijä) tuottaa Volterissa 30 hakutulosta, *ogonvittne* kuusi ja *oegonvittne* 24 tulosta (24.10.2023). Taaskaan eivät o:llinen ja oe:llinen haku tuota samoja tuloksia. O:llinen ei myöskään tuota suomenkielisiä hakutuloksia, kun taas oe:llinen tuottaa niitä peräti 21 kappaletta; on jossain määrin erikoista, että selkeä enemmistö tämän ruotsinkielisen hakusanan tuloksista on suomenkielisiä.

Saksankieliset hakusanat

Saksankielisiksi hakusanoiksi on valikoitu ä:n, ö:n ja ü:n sisältäviä sanoja. Sanoja on etsitty kolmella tavalla: normaalilla oikeinkirjoituksella, ilman yläpuolisia merkkejä (ä → a, ö → o, ü → u) sekä e:llisinä yhdistelminä ae, oe ja ue.

Maschinenführung

Saksan kielen substantiivin *Maschinenführung* voi kääntää suomeksi sanoilla ”koneen käyttö” tai ”koneen ohjaaminen”. Volter-haku kyseisellä sanalla tuottaa 22 osumaa (26.9.2023). Substantiivit kirjoitetaan saksan kielessä aina isolla, ja tätä tapaa käytetään myös tässä työssä.

Saksan kielessä on tapana tietyissä yhteyksissä korvata umlaut-kirjaimet ä, ö ja ü kirjainyhdistelmillä ae, oe ja ue. Näin tehdään esimerkiksi sanaristikoissa ja usein sellaisissa yhteyksissä, joissa pisteellisiä vokaaleja ei syystä tai toisesta ole saatavilla. Näin ollen siis kirjoitusmuoto ”Maschinenfuehrung” olisi, jos ei nyt varsinaisesti oikein, niin joka tapauksessa vakiintuneiden normien mukainen ja saksalaiselle oitis ymmärrettävä. Sen sijaan u:llinen kirjoitusasu ”Maschinenfuhrung” olisi yksiselitteisesti virheelinen.

Volterin hakutoiminto ei kuitenkaan näytä kunnioittavan tätä käytäntöä. Haulla *Maschinenfuhrung* saa täsmälleen samat 22 osumaa (26.9.2023) kuin haulla *Maschinenführung*, kun taas *Maschinenfuehrung* tuottaa vain 17 osumaa (26.9.2023).

Varsin mielenkiintoinen havainto on, että hakutulospäättäjän kannalta ei ole merkitystä, minkä vokaalin päälle pisteet kirjoittaa. Haut *Mäschinenfuhrung* ja *Maschinenfuhrüng* tuottavat nekin 22 osumaa, ja niin tuottavat myös *Maschïnenfuhrung* ja *Maschinënfuhrung* – vaikka saksan aakkosto ei edes tunne ë- ja ï-kirjaimia. Pisteet voi myös kirjoittaa vaikka kaikkien vokaalien päälle, hakutulosten määrä ei siitä muutu. Lyhyesti sanottuna Volter ei siis ainakaan tämän nimenomaisen sanan kohdalla vaikuta ottavan lainkaan huomioon sitä, minkä vokaalin tai vokaalien yläpuolella pisteet sijaitsevat.

E:llisten ”yhdistelmävokaalien” suhteen tilanne on hieman erilainen. Haut *Maeschinenfuhrung* ja *Maschinenfuhrung* tuottavat nekin 17 tulosta, eli samat kuin *Maschinenfuehrung*. Mikäli lisä-e:n kirjoittaa sanassa jo olevien i:n tai e:n perään, siis *Maschiennenfuhrung* tai *Maschineenfuhrung*, tuloksia ei sen sijaan tule lainkaan. Volter siis nähtävästi osaa erottaa ae-, oe- ja ue-yhdistelmät, joita käytetään saksassa ä:n, ö:n ja ü:n korvikkeina, ee- ja ie-yhdistelmistä, jotka eivät tavallisesti korvaa mitään kirjaimia. Epäloogista kyllä, sillä ei nähtävästi ole merkitystä, minkä vokaalin e-yhdistelmällä korvaa.

Näistä viidestä ”puuttuvasta” osumasta neljä on englanninkielisiä, mikä lienee ainakin osaselitys sille, miksi *Maschinenfuehrung*-haku ei tavoita näitä tuloksia. Aukoton tämä selitys ei kuitenkaan ole, sillä kyseinen haku tuottaa useammankin englanninkielisen tuloksen.

Absatzmöglichkeit

Saksan kielen sana *Absatzmöglichkeit*, suomeksi suurin piirtein ”markkinatilanne”, tuottaa Volterissa 31 osumaa (6.10.2023). Haku *Absatzmoglichkeit* tuottaa saman verran osumia, *Absatzmoeglichkeit* sen sijaan vain 24 osumaa (molemmat 6.10.2023). Tilanne on siis samanlainen kuin edellisen hakusanan suhteen: ”pisteellinen” ja ”pisteetön” muoto tuottavat samat osumat, e-yhdistelmä sen sijaan vähemmän.

Absatzmöglichkeit-haun tuottamista 31 tuloksesta kymmenen on englanninkielisiä, *Absatzmoeglichkeit*-haun tuloksista englanninkielisiä on vain kolme. Kaikki seitsemän jälkimmäiseen hakuun sisältymätöntä tulosta ovat siis englanninkielisiä, mikä antaa osviittaa siitä, että englannin kieli toimii erotusta selittäväenä tekijänä.

Abenddämmerung

Saksan kielen substantiivi *Abenddämmerung* (iltahämärä) antaa Volterissa 36 osumaa (6.10.2023). *Abenddammerung* tuottaa 35 osumaa, *Abenddaemmerung* puolestaan 24 osumaa (molemmat 6.10.2023). Tämän

hakusanan osalta tilanne on siis siinä suhteessa erilainen kuin kahden aikaisemman kohdalla, että ”pisteellinen” ja ”pisteetön” muoto eivät tuota samaa osumamäärää. Sen sijaan e-yhdistelmä tuottaa tämänkin hakusanan kohdalla vähemmän osumia kuin kaksi muuta muotoa.

Pisteellisen ja pisteettömän muodon toisistaan poikkeava tulospäättely lienee tämän hakusanan mielenkiintoisin piirre, joten tarkastellaan aluksi sitä. Hakusana *Abenddämmerung* tuottaa kolme tulosta, joiden kaikkien niminä on ”Die Eule der Minerva in Hegels Rechtsphilosophie” (”Minervan pöllö Hegelin oikeusfilosofiassa”), hakusana *Abenddammerung* vain kaksi. Huomaamme ”puuttuvan” tuloksen kohdalla jo tutuksi tulleen kaavan: sen kieleksi on merkattu vain saksa, kun taas kahdesta muusta tuloksesta toisen kieliksi on merkattu saksan lisäksi englanti. Toisaalta toinen on sekin merkitty vain saksankieliseksi, eli englannin kielen selitysvaikutus ei ole aukoton.

Hakusana *Abenddaemmerung* tuottaa 24 tulosta, 12 vähemmän kuin ä:llinen muoto. Hakusana *Abenddämmerung* tuottaa yksitoista englanninkielistä tulosta, hakusana *Abenddaemmerung* vain kolme; näyttää siis siltä, että englannin kieli on jälleen merkittävä selittävä tekijä ainakin kahdeksalla näistä 12 tuloksesta.

Lopuista neljästä tuloksesta kolme on kuvia. Mielenkiintoista on, että vaikka kuvien nimet ovat saksaa, niiden ”kieleksi” on merkitty englanti. On jossain määrin epäselvää, mistä tämä johtuu, sillä kuvissahan ei ole kieltä. Joka tapauksessa vaikuttaa siltä, että englannin kieli toimii jälleen selittävä tekijänä.

Muita saksankielisiä sanoja

Edellisen kolmen sanan perusteella vaikuttaa siltä, että saksankielisten sanojen suhteen kaava olisi seuraavanlainen: ”pisteettömät” vokaalit (a, o, u) tuottavat yhtä paljon tai lähes yhtä paljon tuloksia kuin umlaut-vokaalit ä, ö ja ü, kun taas e-yhdistelmät ae, oe ja ue tuottavat selvästi vähemmän hakutuloksia. Tämän hypoteesin testaamiseksi seuraavassa tarkastellaan sellaisia umlautin sisältäviä saksan kielen sanoja, jotka tuottavat niin paljon hakutuloksia, että niiden yksityiskohtainen tutkailu ei ole käytettävissä olevan ajan ja välineiden puolesta

mahdollista. Hakusanoiksi on tarkoituksella valittu pitkiä sanoja, sillä lyhyiden sanojen kanssa olisi vaarana, että jossakin toisessa kielessä esiintyisi sattumalta sama sana ilman umlautia (esimerkiksi *Käse* eli juusto tai *Tür* eli ovi olisivat huonoja sanoja testaamaan tätä hypoteesia, sillä jossakin toisessa kielessä saattavat hyvinkin esiintyä sanat "kase" ja "tur"). Tämä puolestaan vääristäisi tuloksia.

Oberfläche (pinta)	213 730	Oberflache	213 688	Oberflaeche	156 413
Qualitätssicherung (laadunvarmistus)	8 232	Qualitäts- sicherung	8 157	Qualitaets- sicherung	4 343
Wörterbuch (sanakirja)	85 338	Worterbuch	84 185	Woerterbuch	14 799
Zauberflöte (taikahuilu)	1 069	Zauberflote	1 049	Zauberfloete	291
Übersetzung (käännös)	46 219	Übersetzun g	46 199	Uebersetzung	37 059
Dünndarm (ohutsuoli)	1 384	Dunndarm	1 384	Duenndarm	843

(Kaikki hakusanat 18.10.2023)

Nämä tulokset vaikuttavat tukevan sitä teoriaa, että ilman pisteitä kirjoitetut vokaalit tuottavat pääsääntöisesti lähes yhtä paljon hakutuloksia kuin umlaut-vokaalitkin, kun taas e-yhdistelmät tuottavat hakutuloksia huomattavasti vähemmän. On tietysti selvää, että poikkeuksia saattaa olla, sillä umlaut-vokaalin sisältäviä saksan sanoja on tuhansittain, eikä niistä voida tämän työn puitteissa tutkia kuin pienen murto-osan satunnaisotantaa. Se, että kaikki tähän satunnaisotantaan mahtuneet hakusanat käyttäytyivät samalla tavalla, tuottivat ne hakutuloksia muutamia kymmeniä tai satoja tuhansia, antaa kuitenkin olettaa, että tämä hypoteesi on yleensä pätevä.

Ranskankieliset hakusanat

Ranskan kieli käyttää pääasiassa viittä erilaista erikoismerkkiä: nousevaa (é) ja laskevaa (à, è, ù) aksenttia, sirkumfleksiä (â, ê, î, ô, û), treemaa eli vokaalin yläpuolista pisteparia (ë, ï, ü) ja sediljiä eli kirjaimen alapuolista koukkua (ç). Ranskankieliset hakusanat tuottivat uudenlaisen ongelman, sillä mikään näitä merkkejä sisältävistä hakusanoista ei tuottanut merkittävästi eroavia tulospäämiä, kirjoittipa ne sitten erikoismerkkien kera – eli oikein – tai ilman niitä. Seuraavassa tarkempaa selostusta eri erikoismerkkejä sisältävistä hakusanoista.

égout (viemäri)	5 962	egout	5 962
sécheresse (kuivuus)	9 132	secheresse	9 132
prévenir (varoittaa)	110 934	prevenir	110 934
événement (tapaus)	185 925	evenement	185 925

15.10.2023

voilà (kas tässä)	17 209	voila	17 209
nanomètre (nanometri)	20 338	nanometre	20 340
obsolète (vanhentunut)	42 173	obsolete	42 174
déjà (jo)	114 715	déja/dejà/deja	114 671
où (jossa)	3 843 929	ou (tai)	3 641 885

16.10.2023

hawaïen (havaijin kieli)	116	hawaien	116
exigüe (pikkuruinen)	1 007	exigüe	1 007
gagüere (haaste)	1 019	gagüere	1 019

laïcité (sekularismi)	7 057	laicite	7 057
naïf (naiivi)	7 328	naif	7 332
coïncider (sattua samaa aikaan)	10 060	coincider	10 237
taïga (taiga)	17 042	taiga	17 042
canoë (kanootti)	30 450	canoe	30 454
haïe (muuan vihata-verbin muoto)	34 676	haie	34 676
Citroën (automerkki)	54 299	Citroen	54 270
Noël (joulu)	180 929	Noel	180 945

12.10.2023

embêter (häiritä)	163	embeter	163
verdâtre (vihertävä)	220	verdatre	220
croûton (krutonki)	604	crouton	607
surcoût (lisämaksu)	1 213	surcout	1 213
bientôt (pian)	5 532	bientot	5 532
fantôme (haamu)	8 449	fantome	8 449
mâitre (isäntä)	66 928	maitre	67 139
faïte (laki, harja)	429 409	faite	429 409
être (olla)	2 404 375	etre	2 404 465

14.10.2023

glaçure (lasite)	860	glacure	860
garçon (poika)	15 731	garcon	15 785
façade (julkisivu)	109 781	facade	109 940

16.10.2023

Kuten näemme, oli hakutuloksia sitten vähän, paljon tai jotakin siltä väliltä, hakutulospäämäärät ovat lähes aina hyvin samanlaisia riippumatta siitä, kirjoittaanko sanan oikein eli erikoismerkin kera vai ilman erikoismerkkiä. Erot ovat tavallisesti pieniä, käytännössä korkeintaan muutamia kymmeniä tuloksia, ja esiintyvät, mikäli niitä ylipäättäen on, vain sellaisissa hakusanoissa, jotka tuottavat vähintään useita satoja, useimmiten useita tuhansia hakutuloksia. Tämä merkitsee sitä, että tämän työn puitteissa näiden erojen selvittäminen ei tule kyseeseen. Myös siinä yhdessä tapauksessa, jossa aksentti tai sen puuttuminen tuottaa huomattavan eron, eli sanaparissa *ou* ja *ou̇*, hakutuloksia on niin paljon, ettei tarkempi tutkimus valitettavasti ole mahdollinen. Sen voimme kuitenkin sanoa, että aksentillinen haku tuottaa noin 5,5 % enemmän tuloksia kuin aksentiton. Myös hakusanassa *coïncider* pisteparin poisjätö tuottaa noin 1,7 % enemmän tuloksia. Muilla hakusanoilla erot lasketaan promilleissa tai niidenkin osissa.

Toisaalta voi ajatella niinkin, ettei erojen syiden selvittäminen ole välttämätöntäkään; mikäli sanan kirjoitusasu ei käytännössä vaikuta hakutulosten määrään juuri mitenkään, ei juuri ole vaarana, että hakutuloksia jäisi väärän kirjoitusasun vuoksi uupumaan. Hankaluuden muodostavat tosin sellaiset sanat, joiden ranskankielinen ja toisaalta jonkin muun kielinen kirjoitusasu eroavat toisistaan ainoastaan erikoismerkin suhteen (esimerkiksi ranskan *façade* vs. englannin *facade*). Tähän auttane se, että käyttää Volterin tarkennettua hakua ja suodattaa tuloksista vain ranskankieliset hakutulokset.

Espanjankieliset hakusanat

Espanjankielisiksi hakusanoiksi on valikoitu ñ-kirjaimen sisältäviä sanoja. Tämän kirjaimen lisäksi espanjassa käytetään myös ü:tä sekä aksentillisiä vokaaleja á, é, í, ó ja ú, mutta näiden kirjainten käyttö on niin vähäistä, ettei niitä sisältäviä sanoja ole sisällytetty tähän työhön.

Apañar

Hakusana *apañar* (korjata, valmistaa) tuottaa Volterissa 17 hakutulosta (30.10.2023). Hakusana *apanar* (leivittää) tuottaa tuloksia kymmenen

(30.10.2023). Tämä sana poikkeaa siis muista tässä työssä käytetyistä erikoismerkkejä sisältävistä sanoista siinä, että erikoismerkitön kirjoitusasu ei ole virheellinen, vaan oma, erillinen sanansa.

Kaikki *apanar*-haun kymmenen tulosta sisältyvät *apañar*-hakuun, eli Volter ei osaa erottaa sanoja toisistaan mainittavan hyvin. *Apañar*-haun 17 tuloksesta espanjankielisiä on vain seitsemän. Englanninkielisiä tuloksia on kolme, ja loppuista seitsemästä yksi on norjan- ja kuusi ruotsinkielisiä. Ruotsin- ja norjankieliset osumat johtuvat siitä, että Volter sisällyttää tuloksiin myös *apa*- eli *apina*-sanan sisältäviä osumia. *Apanar*-haku antaa tuloksiksi samat seitsemän espanjankielistä osumaa kuin *apañar*, minkä lisäksi se antaa yhden norjankielisen ja kaksi ruotsinkielistä tulosta. Mielenkiintoinen huomio on, että seitsemään *apañar*-haun osumaan on Volterissa merkitty julkaisuvuosi, ja nämä ovat juuri ne seitsemän osumaa, jotka *apanar*-haun tuloksista puuttuvat.

Entrañablemente

Hakusana *entrañablemente* (rakastavasti) tuottaa Volterissa 34 osumaa, *entranablemente* 22 osumaa (30.10.2023). Nämä hakusanat tuottavat lähes pelkästään espanjankielisiä tuloksia. *Entrañablemente* tuottaa kaksi englanninkielistä tulosta ja *entranablemente* yhden, muutoin kaikki osumat ovat espanjankielisiä. Jälleen kerran on huomionarvoista, että ne 12 *entrañablemente*-haun tulosta, joihin on Volterissa merkitty teoksen ilmestymisvuosi, ovat ne samat 12, jotka *entranablemente*-hausta puuttuvat.

Tacañamente

Hakusana *tacañamente* (ilkeästi) tuottaa Volterissa kahdeksan tulosta, *tacanamente* kuusi (30.10.2023). *Tacañamente*-haun kahdeksasta tuloksesta englanninkielisiä on kaksi, *tacanamente*-haun kuudesta tuloksesta yksi. Tällä kertaa myös *tacanamente*-haku tuottaa yhden osuman, johon on Volterissa merkitty ilmestymisvuosi. *Tacañamente* tuottaa tällaisia osumia kolme.

Muita espanjankielisiä hakusanoja

Suurin osa ñ-kirjaimen sisältävistä hakusanoista tuottaa saman tai lähes saman määrän tuloksia, vaikka ne kirjoittaisikin ilman tildeä eli yläpuolista aaltoviivaa. Seuraavassa esitellään muutamia esimerkkejä tällaisista sanoista. Kuten saksankielistenkin sanojen kohdalla, tällaisiksi hakusanoiksi on pyritty löytämään pitkiä sanoja, jotta muissa kielissä sattumalta esiintyvien samanmuotoisten sanojen riski saadaan minimoitua. Sanat on luetteloitu sen mukaan, miten paljon hakutuloksia ne Volterissa tuottavat. Kaikki haut on tehty 15.11.2023.

mozambiqueño (mosambikilainen)	76	mozambiqueño	76
Matarraña (paikkakunta Koillis-Espanjassa)	82	Matarrana	82
escudriñar (tarkastella)	242	escudrinar	242
señorita (neiti)	942	senorita	942
montañoso (vuoristoinen)	1 453	montanoso	1 430
mañana (huomenna)	6 464	manana	6 450
ordeñador (lypsäjä)	26 659	ordenador	26 659
pestaña (silmäripsi)	29 730	pestana	29 729
pequeño (pieni)	137 016	pequeno	137 016

Kuten taulukko osoittaa, riippumatta siitä, onko hakusanalla osumia vähän, keskiverto määrä vai huomattavankin paljon, hakutulosten määrät pysyvät samoina tai lähes samoina riippumatta siitä, kirjoitetaanko sana oikeaoppisesti ñ-kirjaimella vai pelkästään n:llä. Hakutulosten määrissä on eroja ainoastaan sanoissa *montañoso* (23), *mañana* (6) ja *pestaña* (1), eikä hakusanan tuottamien osumien määrällä näytä olevan yhteyttä siihen, vaikuttaako kirjoitusasu osumamäärään vai ei. Esimerkiksi hakusana *pequeño* tuottaa lähes sata kertaa enemmän osumia kuin *montañoso*, mutta silti kirjoitusasu vaikuttaa

ainoastaan jälkimmäisen hakusanan osumamäärään. Ei myöskään ole selvää, kumpaan suuntaan kirjoitusasu vaikuttaa: hakusanoilla *montañoso* ja *pestaña* aaltoviivallinen kirjoitusmuoto tuottaa enemmän osumia kuin aaltoviivaton, hakusanalla *mañana* asia on sen sijaan päinvastoin.

Italiankieliset hakusanat

Italia käyttää pääasiassa kuutta aksentillista vokaalia: laskeva aksentti on vokaaleissa à, è, ì, ò ja ù, nouseva aksentti vokaalissa é.

Kaikki haut on tehty 15.–16.11.2023.

dàrsena (telakka)	319	darsena	319
longevità (pitkäikäisyys)	2 260	longevita	2 260
àncora (ankkuri)	13 186	ancora (vielä)	13 186
città (kaupunki)	17 870	citta	17 870
libertà (vapaus)	52 907	liberta	52 870

trentatré (kolmekymmentäkolme)	25	trentatre	25
cosicché (siksi)	303	cosicche	303
poiché (koska)	1 623	poiche	1 623
séguito (jatko-osa)	7 370	seguito	7 373
perché (miksi, koska)	13 973	perche	13 921

vabbè (jaaha, vai niin)	387	vabbe	387
ahimè (voi kauhistus)	398	ahime	398
sèttembre (syyskuu)	9 047	settembre	9 047
pèsca (persikka)	21 231	pesca	21 271

bebè (vauva)	27 313	bebe	27 313
--------------	--------	------	--------

lunedì (maanantai)	198	lunedì	198
vivono (he elävät)	440	vivono	611
seguito (minä seurasin)	7 370	seguito	7 373
utensile (väline)	11 026	utensile	11 022
principi (periaatteet)	194 111	principi	194 111

hòrrido (hirveä)	27	horrido	27
uditòri (yleisöt)	65	uditori	65
nòcciolo (ydin)	300	nocciolo	300
perciò (siksi)	5 070	percio	5 034
però (mutta)	238 827	pero	238 827

laggiù (alhaalla)	34	laggiu	34
frústino (ratsupiiska)	45	frustino	45
autogrù (nosturi)	123	autogru	123
dùtile (käytännöllinen)	571	dutile	571
virtù (hyve)	18 752	virtu	18 752

Italian suhteen toistuu siis sama kuin ranskan aksentillisten hakusanojen kohdalla: muutamia poikkeuksia lukuun ottamatta aksenttimerkeillä tai niiden puuttumisella ei ole juuri minkäänlaista vaikutusta siihen, montako hakutulosta sana tuottaa. Suhteellisesti – ja itse asiassa määrällisestikin – huomattavin poikkeus tähän sääntöön on hakusana *vivono*, joka tuottaa aksentillisena 611

osumaa, aksentittomana vain 440. Aksentti lisää siis hakutuloksia noin 39 %. Myös hakusanoissa *perché* (42), *pèsca* (40), *perciò* (36), *libertà* (27), *utensile* (4) ja *séguito/seguito* (3) aksentin poisjättäminen saa osumissa aikaan suluissa mainitun määräeron, mutta prosentuaalinen muutos on huomattavasti pienempi, koska mainituilla hakusanoilla on paljon enemmän osumia. On myös mielenkiintoinen havainto, että se, paljonko osumia hakusana tuottaa, ei vaikuta olevan yhteydessä siihen, muuttavatko aksenttimerkit osumamääriä. Esimerkiksi hakusanat *però* (238 827) ja *principi* (194 111) tuottavat yli kolmesataakertaisen määrän hakutuloksia *vivono*-sanaan verrattuna, mutta silti aksenttimerkin poisjätto ei lainkaan vaikuta niiden osumamääriin.

6 Yhteenveto

Ä, Ö, Ü

Näitä kirjaimia sisältäviä sanoja haettiin kolmella eri kirjoitusasulla: oikealla sekä korvaamalla ä, ö ja ü yhtäältä kirjaimilla a, o ja u, toisaalta yhdistelmillä ae, oe ja ue. Jakauma näiden hakujen kesken oli hyvin selkeä: lähes kaikissa tapauksissa pisteettömät ja yhdistelmälliset haut tuottivat yhdessä kaikki ne tulokset, jotka oikeinkirjoitettu hakukin, mutta samoja tuloksia ne eivät muutamia poikkeuksia lukuun ottamatta antaneet. Volter siis tulkitsee nämä kirjaimet tietyissä tapauksissa ae-, oe- ja ue-yhdistelmiksi, tietyissä taas ei.

Å

Å:n kohdalla havaittiin erikoinen ilmiö: oikein kirjoitettuina hakusanat tuottivat runsaasti suomenkielisiä tuloksia, mutta kun å:n korvasi a:lla, kaikki suomenkieliset tulokset tippuivat pois. A:lliset hakusanat tuottivat ylipäätään selvästi vähemmän tuloksia kuin å:lliset.

AKSENTIT

Työssä tutkittuja aksenttimerkkejä sisältäviä kirjaimia olivat à, â, é, è, ê, ì, î, ò, ô, ù ja û. Kaikki aksenttimerkit käyttäytyivät Volterissa samalla tavalla: suurimmalla osalla aksenttimerkkejä sisältävistä sanoista aksenttimerkin poisjättäminen ei

vaikuttanut osumamäärään mitenkään, ja mikäli eroja oli, ne olivat tavallisesti erittäin pieniä, muutamia tuloksia tuhansista. Tietyissä tapauksissa erot olivat huomattavasti suurempia, mutta valitettavasti tällaisia eroja tuottivat ainoastaan sellaiset hakusanat, joiden kokonaistulosmäärät olivat niin suuria, ettei niiden tarkempi tutkiminen ole tässä työssä mahdollista. Siinä, minkälaiset sanat tuottivat hakutuloksissa määräeroja, ei ollut havaittavissa mitään erityistä säännönmukaisuutta: vain muutamia satoja osumia tuottavat sanat saattoivat antaa eri määrän tuloksia aksentillisina ja aksentittomina, ja toisaalta satoja tuhansia tuloksia tuottavat sanat saattoivat antaa täsmälleen saman osumamäärän riippumatta siitä, hakiko niitä aksentin kera vai ilman sitä.

Ç ja Ñ

Näihin erikoismerkkeihin pätee pitkälti sama kuin aksenttimerkkeihin: useimmissa tapauksissa niiden korvaaminen c:llä ja n:llä ei tuottanut hakutuloksien osumamääriin eroja, ja jos niitä oli, ne vaikuttivat pääosin sattumanvaraisilta. Esimerkiksi osumamäärän suuruudesta sinänsä ei voinut päätellä, millä todennäköisyydellä erikoismerkkien poisjätö vaikuttaisi osumamäärään.

Aakkosjärjestys

Kiintoisa huomio on, että hakusanojen kirjoitusmuodolla on vaikutusta myös osumien aakkosjärjestykseen. Esimerkiksi haettaessa sanalla *äkillinen* Volter järjestää osumat tavalliseen suomen aakkosjärjestykseen, eli ä:llä alkavat tulokset ovat viimeisenä (hakusana ei tuota ö:llä alkavia tuloksia). Sen sijaan haettaessa sanalla *akillinen* ä:llä alkavat tulokset järjestyvätkin a:lla alkavien tulosten joukkoon, eli ilman ääkkösiä haettaessa Volter ei nähtävästi erota pisteellisiä ja pisteettömiä vokaaleja toisistaan.

Teoksen ikä

Tietyissä tapauksissa teoksen iällä näytti olevan vaikutusta siihen, mitkä haut antavat sen tulokseksi. Tavallisesti ä-, ö- ja ü-kirjaimia sisältävien sanojen

yhtäältä pisteettömät ja toisaalta e-yhdistelmälliset haut eivät tuottaneet lainkaan samoja tuloksia. Tähän sääntöön poikkeuksen tekeviä osumia oli muutamia, ja ne olivat kaikki yli sata vuotta vanhoja teoksia. Epäselväksi jäi, minkä tähden julkaisuvuosi sai aikaan tällaisen ilmiön, mutta se vaikutti joka tapauksessa jokseenkin johdonmukaiselta.

Lopuksi

Tämän opinnäytetyön tekeminen osoittautui opettavaiseksi kokemukseksi. Minulla on ohjelmoinnista ja tietokoneohjelmista yleensäkin ainoastaan maallikkotason taidot, joten olin työhön ryhtyessäni jossakin määrin huolissani siitä, että olen haukannut liian suuren palan. Kirjoitustyön edetessä ohjelmointitaitojen puute ei kuitenkaan osoittautunut sellaiseksi esteeksi, ettei työn saattaminen loppuun onnistuisi.

Tämän työn kirjoittaminen sai minut tiedostamaan sen, miten suuri merkitys hakusanan kirjoitusasulla on hakutuloksiin. Lisäksi olen oppinut huomaamaan, miten erilaisia – ja erikielisiä – tuloksia eri hakusanat antavat. Tällainen tietotaito tulee jatkossa olemaan hyödyllistä etenkin työelämän asiakaspalvelutilanteissa.

Lähteet

Academic Accelerator. Lithuanian Orthography. Viitattu 18.12.2023.

<https://academic-accelerator.com/encyclopedia/lithuanian-orthography>

Ahvenainen, P., Gylling, J. & Leino, S. 2017. Viiden tähden asiakaskokemus. Tee asiakkaistasi faneja. 1. painos. Viro: Meedia Zone OÜ.

Danishnet.com. The Danish Alphabet. Viitattu 18.12.2023.

<https://www.danishnet.com/culture/danish-alphabet/>

de Mooij, M., Kortesmäki, T., Lammi, M., Lautamäki, S., Pekkala, J. & Sinkkonen, I. 2005. Kompassina asiakas. Tampere: Tammer-paino Oy.

ExLibris Knowledge Center. Linguistic Features for Primo VE. Viitattu 19.12.2023.

[https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_\(English\)/160Linguistic_Features_for_Primo_VE](https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/020Primo_VE/Primo_VE_(English)/160Linguistic_Features_for_Primo_VE)

Filenius, M. 2015. Digitaalinen asiakaskokemus. Menesty monikanavaisessa liiketoiminnassa. Saarijärven Offset Oy.

GermanVeryEasy.com. The German Alphabet. Viitattu 18.12.2023.

<https://www.germanveryeasy.com/german-alphabet>

ItalianPod101.com. È with an Accent or E? All about Italian Accent Marks.

Viitattu 18.12.2023. <https://www.italianpod101.com/blog/2021/01/18/italian-written-accents/>

Komulainen, M. 2018. Menesty digimarkkinoinnilla. 1. painos. Viro: Meedia Zone OÜ.

Langster. French Accent Marks. Viitattu 18.12.2023.

<https://langster.org/en/blog/french-accent-marks-use-and-type/>

Life in Norway. The Norwegian Alphabet. Viitattu 18.12.2023.

<https://www.lifeinnorway.net/norwegian-alphabet/>

Middle East and North African Languages Program. The Turkish Alphabet.

Viitattu 18.12.2023 <https://mena-languages.northwestern.edu/language-learning/languages-with-different-scripts/turkish-alphabet.html>

Mówić Po Polsku. Polish Alphabet & Pronunciation. Viitattu 18.12.2023.

<https://mowicpopolsku.com/polish-alphabet-pronunciation/>

Omniglot. Albanian. Viitattu 18.12.2023

<https://www.omniglot.com/writing/albanian.htm>

Omniglot. Czech. Viitattu 18.12.2023.

<https://www.omniglot.com/writing/czech.htm>

Omniglot. Estonian. Viitattu 18.12.2023.

<https://www.omniglot.com/writing/estonian.htm>

Omniglot. Hungarian. Viitattu 18.12.2023.

<https://www.omniglot.com/writing/hungarian.htm>

Omniglot. Icelandic. Viitattu 18.12.2023.

<https://www.omniglot.com/writing/icelandic.htm>

Omniglot. Romanian. Viitattu 18.12.2023.

<https://www.omniglot.com/writing/romanian.htm>

PortugueseLanguage. Diacritics. Viitattu 18.12.2023.

<http://www.portugueselanguageguide.com/pronunciation/vowels/diacritics.asp>

Slovak.EU. Slovak Alphabet. Viitattu 18.12.2023.

<https://slovak.eu/learning/grammar/pronunciation%2Fabc>

Spanish411.net. The Spanish Alphabet. Viitattu 18.12.2023.

<https://spanish411.net/Spanish-Alphabet.asp>

Wales Guidebook. Welsh Alphabet [+ How To Pronounce the Letters]. Viitattu

18.12.2023. <https://walesguidebook.com/language/welsh-alphabet/>

Wikiwand. Latvian Orthography. Viitattu 18.12.2023.

https://www.wikiwand.com/en/Latvian_orthography